ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 121/26

Trends and Structure of Educational Assortative Mating

Edoardo Ciscato

www.rfberlin.com APRIL 2026

Trends and Structure of Educational Assortative Mating

Authors

Edoardo Ciscato

Reference

JEL Codes: D13, J11, J12

Keywords: marriage markets, educational assortative mating, matching models

Recommended Citation: Edoardo Ciscato (2026): Trends and Structure of Educational Assortative Mating. RFBerlin Discussion Paper No. 121/26

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Trends and Structure of Educational Assortative Mating

Edoardo Ciscato*

March 16, 2026

Abstract

We study the structure and evolution of educational assortative mating in the U.S. using 1962-2025 Current Population Survey data. From observed marital patterns, we recover the full matrix of marital gains across educational groups within the equilibrium matching framework of Choo and Siow (2006) and provide a novel characterization of its complementarity structure using Singular Value Decomposition. This approach allows us to infer the number of relevant sorting dimensions, to describe how different educational attainments are valued along each dimension, and to develop aggregate measures of assortative mating based on local odds ratios. We show that the first dimension explains more than 90% of total sorting in all periods and ranks individuals monotonically by educational attainment, consistent with the presence of human capital complementarities à la Becker. On the other hand, higher-order dimensions capture homogamous preferences that help explain the prevalence of same-education couples in the data. We document evidence of a long-run increase in educational assortative mating, with a marked slowdown since the 1990s, although the magnitude of the long-run increase is sensitive to how educational categories are defined, partly due to a coding change in the CPS educational variable in 1992. We also show that high-school dropouts have become increasingly isolated, college and postgraduate groups have converged, and gender asymmetries in returns from sorting have declined.

Keywords: marriage markets, educational assortative mating, matching models.

JEL Classification: D13, J11, J12.

*KU Leuven. Email: edoardo.ciscato@kuleuven.be. Website: https://sites.google.com/site/ciscatoedoardo.

1. Introduction

Assortative mating on education contributes to inequality both within and across generations (Fernández and Rogerson, 2001; Eika et al., 2019). However, the preferences and constraints that drive observed patterns of marriage market sorting remain less well understood. Educational assortative mating may arise because partners' human capital is complementary in consumption, leisure, and home production (Becker, 1973, 1974). In this case, individuals agree on a shared ranking of potential partners by educational attainment. Empirical evidence, however, suggests that both men and women tend to prefer partners with similar educational levels and to avoid partners with higher levels of education, with men being particularly reluctant to marry a more educated spouse (Hitsch et al., 2010). In addition, search frictions reinforce similarity in matches, with educational institutions playing a substantial role in generating educational homogamy (Shimer and Smith, 2000; Nielsen and Svarer, 2009; Kirkebøen et al., 2021).

These considerations have important implications for our understanding of marriage markets. First, a one-dimensional model of sorting is not necessarily a good characterization of the marriage market if we want to study educational assortative mating.[1] Individuals with different educational levels may rank potential partners differently, particularly if they have a preference for similar partners, if they are averse to more educated partners, or if search frictions reinforce similarities. Hence, whether different diplomas and degrees can actually be ranked along a single dimension and whether their ranking in the marriage market reflects their natural order in the education system are open empirical questions.

Second, educational complementarities are inherently local, i.e., they may differ in strength and sign for different diplomas and degrees, and they may be gender asymmetric. If the structure of educational complementarities is inconsistent with a single-index model, it is harder to provide a comprehensive measure of assortative mating and track its changes over time. Unless we resort to a binary classification (e.g., college vs non-college), it may be challenging to derive a sufficient statistic to capture the overall strength of assortative mating, with different aggregate measures possibly yielding conflicting conclusions (Gihleb and Lang, 2020; Chiappori et al., 2025).

In this paper, we study the evolution of educational assortative mating in the U.S. using Current Population Survey (CPS) data from 1962 to 2025 within the equilibrium matching framework of Choo and Siow (2006). We first obtain the matrix of marital

[1] This is also an obvious concern when researchers are interested in studying mating patterns with respect to multiple observed variables. For instance, Chiappori et al. (2012) estimate a single-index model to explain sorting patterns with respect to income, education, and Body Mass Index (BMI). Other papers estimate models of multidimensional matching to understand patterns of assortative mating with respect to other traits, such as personality or health-related behavior (Dupuy and Galichon, 2014; Chiappori et al., 2018, 2024).

gains from cross-sectional marital patterns, where each entry corresponds to the average match surplus between two educational groups. In this model, assortative mating arises locally due to complementarities in the match surplus (Graham, 2011; Siow, 2015). Local complementarities are identified from odds ratios,[2] which are widely used measures of assortative mating in social sciences (Edwards, 1963; Mare, 1991; Chiappori et al., 2025).

The contribution of the paper lies in a convenient characterization of the complementarity structure that allows us to uncover the orthogonal dimensions that can rationalize the aggregate patterns of educational sorting. Starting from the match surplus matrix, we show how to infer the number of relevant matching dimensions, the strength of assortative mating on each dimension, their relative importance and composition. In this way, we can investigate whether a one-dimensional characterization of the marriage market is a good approximation or if a high-dimensional model is needed, we can provide an exhaustive description of each dimension, and we can derive appropriate sufficient statistics for overall assortative mating. The approach can be readily applied to study the structure of assortative matching based on (ordered or unordered) categorical variables in other markets.

In practice, drawing from Dupuy and Galichon (2014), we perform a Singular Value Decomposition of the surplus matrix and show that we can interpret the singular values as measures of assortative mating on each dimension.[3] If the surplus matrix has rank one, then the first singular value is a comprehensive measure of assortative mating. Moreover, the left and right singular vectors can be interpreted as 'indices of attractiveness', which quantify how different educational attainments are valued on the marriage market without imposing any ex-ante restriction on their relative importance and allowing us to test what margins are the most salient. If instead the surplus matrix has full rank, then we need to look at the entire set of singular values to understand how each dimension contributes to assortative mating. In this case, the Frobenius norm, a measure of total 'energy' of the matrix, can be used to quantify overall assortative mating.

Our findings show that, while we formally reject that there exists only one dimension of educational sorting, a single-index model is actually an excellent approximation. The first sorting dimension accounts for more than 90% of the overall assortative mating in all periods. Along the first dimension, individuals are ranked based on their educational attainment, from the least to the most educated. We show that, in our baseline five-group

[2]The odds ratio for, say, college degrees and high-school diplomas is defined as the ratio of college degrees and high-school diplomas odds to match with the same type and with a different type. We formally define it in equation (2.5).

[3]Dupuy and Galichon (2014) show how to infer the number of sorting dimensions and the indices of attractiveness in a setting where individuals sort on continuous variables with a biquadratic parameterization of the match surplus. The SVD is applied to the so-called 'affinity matrix'. In our paper, we do not impose any parametric restriction on the match surplus and directly apply SVD to the full double-centered complementarity matrix $\tilde{P}$ in the discrete-type setting of Choo and Siow (2006).

categorization, a higher educational attainment is always associated with positive returns from sorting, which validates the use of five distinct categories. Yet, we also show that sorting on higher-order dimensions reflects homogamous preferences and helps explain the prevalence of same-education couples in the data. This suggests that, while human capital complementarities à la Becker can by and large explain the marital patterns observed in the data, they coexist with an inclination to form homogamous matches, possibly due to a pure taste for similarity or search frictions.

When we look at changes in assortative mating over time, we find evidence of a long-run increase in educational assortative mating, although its rise has substantially slowed down in the last three decades and its exact magnitude depends on how educational categories are defined. This confirms previous findings obtained with different methods but similar data (Eika et al., 2019; Hirschl et al., 2024). Moreover, when we look at changes in the indices of attractiveness, we show that high-school dropouts are increasingly isolated on the marriage market, whereas college graduates and postgraduates have grown closer, confirming previous findings for sorting at the top and bottom of the educational distribution (Eika et al., 2019; Chiappori et al., 2025). We also find that returns from educational sorting have become more and more similar across genders, which suggests that social norms or hurdles preventing highly educated women from marrying down may have weakened.

On the other hand, it is important to note that, as already noticed by Schwartz and Mare (2005) and Gihleb and Lang (2020), changes in coding of the CPS educational variable in 1992 make the comparison between earlier and later survey waves problematic. In particular, we note a spike in our measures of assortative mating exactly in 1992, suggesting that earlier measures might be downward bias and casting doubt on whether there has really been a long-run increase. Hence, we explore coarser categorizations to address these concerns and show that merging high-school dropouts and graduates (two categories potentially affected by misclassification before 1992) partly solves the comparability issue, but results in a smaller long-run increase in assortative mating. However, as argued by Chiappori et al. (2025, Theorem 4), our measures of assortative mating are not robust to categorization, which means that neglecting certain margins can lead to different conclusions. We conclude that, while it is possible that changes in coding of the educational variable lead to upward biased estimates of the long-run increase, results obtained when aggregating high-school dropouts and graduates represent a lower bound. More broadly, we reiterate the recommendations of Gihleb and Lang (2020) and Chiappori et al. (2025) that exploring different categorizations remains a necessary exercise.

From a methodological perspective, our paper is closest to Siow (2015) and Chiappori et al. (2017), who estimate different parametric versions of the additively separable matching models by Choo and Siow (2006) and Galichon and Salanié (2022) in order to assess which specification of the match surplus can best explain changes in educational assortative mating over time. Our approach is not based on a comparison between different parameterized models, but consists in providing a concise and intuitive representation of the saturated model of Choo and Siow (2006).

Our paper also provides a methodological alternative to a large literature that has explored different statistical methods to measure educational assortative mating and track its changes over time. Previous works have used Poisson regression models with two-way fixed effects (Mare, 1991; Schwartz and Mare, 2005; Hirschl et al., 2024), linear or rank correlation rates (Greenwood et al., 2014; Gihleb and Lang, 2020), aggregate likelihood ratios (Eika et al., 2019; Almar and Schulz, 2024), or the normalized trace (Cheremukhin et al., 2024) as overall measures of assortative mating.

How does our new measure of assortative mating differ from existing ones? Chiappori et al. (2025) review and compare the methodologies used in the above-mentioned papers. They show that odds ratios enjoy particularly desirable properties, being the only metric satisfying marginal independence, but they also stress their local nature. In other words, odds ratios lack a natural extension to measure overall assortative mating with more than two categories. In this paper, starting from local odds ratios, we derive aggregate measures of assortative mating that preserve their desirable properties and that, contrary to previously proposed aggregate measures (Altham, 1970), also admit a structural interpretation.[4]

The structure of the paper is as follows. In Section 2, we introduce the theoretical framework from which we derive the metrics of assortative mating used in our analysis. In Section 3, we present CPS data and detail our sample restrictions and measures of educational attainment. In Section 4, we present the results obtained with our baseline sample, while Section 5 contains a number of robustness checks with alternative samples and educational categorizations. Section 6 concludes.

2. Conceptual framework

2.1. Setup

We study sorting on education in a large population of size normalized to one. Every man is characterized by a level of education $x \in \{1, 2, ..., E\}$. Similarly, every woman is characterized by a level of education $y \in \{1, 2, ..., E\}$. We let $m(x)$ and $n(y)$ be two probability mass functions respectively corresponding to the marginal frequencies of men and women. They are scaled so that $\sum_x m(x) + \sum_y n(y) = 1$.

[4] The structural interpretation of our approach still rests on further parametric assumptions, and in particular on the additively separable extreme value type I distributed taste shocks, as noted by Galichon and Salanié (2022) and Gualdani and Sinha (2023).

The equilibrium assignment is characterized by the probability mass function $\mu(x, y)$, which gives the joint frequency of couples with type (x, y) in equilibrium. The functions $\mu(x, 0)$ and $\mu(0, y)$ respectively denote the frequency of single men of type x and women of type y in equilibrium. We assume that the matching function is the same as in Choo and Siow (2006):

$$\mu(x, y) = \exp\left(\frac{\Phi(x, y)}{2}\right)\sqrt{\mu(x, 0)\mu(0, y)} \tag{2.1}$$

where $\Phi(x, y)$, a model primitive, is the average match surplus for couples of type (x, y).[5] We know from Choo and Siow (2006) and Galichon and Salanié (2022) that, for given (Φ, m, n), there exists a unique function μ so that (2.1) is respected and feasible, i.e., the following constraints hold:

$$m(x) = \mu(x, 0) + \sum_y \mu(x, y) \quad \forall x \tag{2.2}$$

$$n(y) = \mu(0, y) + \sum_x \mu(x, y) \quad \forall y. \tag{2.3}$$

Complementarities in $\Phi(x, y)$ lead to assortative mating (Graham, 2011; Siow, 2015). Hence, whether Φ is supermodular is informative about the direction and strength of assortative mating. For given (x, y) and (x', y'), Φ is locally supermodular if the following double difference is positive:

$$\Delta(x, x', y, y') \equiv \Phi(x', y') - \Phi(x', y) - \Phi(x, y') + \Phi(x, y) > 0. \tag{2.4}$$

We can invert the matching function (2.1) and retrieve $\Phi(x, y)$ as a function of the equilibrium frequencies in order to show that the complementarity measure $\Delta(x, x', y, y')$ corresponds to the logarithm of the odds ratio:

$$\Delta(x, x', y, y') = \log \frac{\mu(x', y')\mu(x, y)}{\mu(x', y)\mu(x, y')}. \tag{2.5}$$

This establishes a link with a large literature in demography and sociology, which recognizes that odds ratios enjoy particularly desirable properties (Edwards, 1963; Altham, 1970). As discussed by Chiappori et al. (2025), they are the only measure of assortative mating that, aside from satisfying other basic axioms, satisfies marginal independence.[6]

[5] For a derivation of the matching function (2.1) starting from individual utilities and programs under the assumption of additively separable, extreme value type I preference shocks, see Choo and Siow (2006), Chiappori et al. (2017), or Galichon and Salanié (2022).

[6] In other words, odds ratios are invariant to changes in the marginal distributions that do not affect the conditional choice probabilities of marrying with different types. For any male type x, when $m(x)$ is multiplied by a factor $\lambda > 0$ and all $\mu(x, y)$ are also multiplied by λ, the odds ratio stays unchanged. The same holds for any female type y. See Chiappori et al. (2025) for a comparison of the odds ratio with other measures of assortative mating.

Moreover, two-way fixed effect Poisson regression models for contingency tables are used to measure assortative mating by fitting local odds ratios (Mare, 1991; Schwartz and Mare, 2005; Hirschl et al., 2024).

2.2. Factorization of the match surplus

We let P be a square matrix of size $E \times E$, where each cell $p_{x,y}$ corresponds to $\Phi(x, y)$. The singular value decomposition (SVD) of P yields:

$$P = U \Lambda V^{\intercal} \tag{2.6}$$

where Λ is a diagonal matrix with non-negative diagonal values $\{\lambda_k\}_{k=1}^{E}$ representing the singular values of P, while U and V are orthogonal square matrices of size $E \times E$, and their k-th columns $u_{\cdot,k}$ and $v_{\cdot,k}$ respectively represent the k-th left and right singular vectors of P. We let $K \leq E$ be the rank of P, so that only the first K singular values $\{\lambda_k\}_{k=1}^{K}$ are strictly positive. Hence, K indicates the number of orthogonal matching dimensions.

The SVD representation implies that we can express the match surplus $\Phi(x, y)$ as the following weighted sum:

$$\Phi(x, y) = \sum_{k=1}^{K} \lambda_k u_{x,k} v_{y,k} \tag{2.7}$$

where $u_{x,k}$ and $v_{y,k}$ respectively represent the value of educational attainment x and y along the k-th matching dimension, whereas λ_k represents the weight assigned to the k-th dimension.

On each dimension, the singular vectors $u_{\cdot,k}$ and $v_{\cdot,k}$ can be interpreted as indices of mutual attractiveness, in the same vein as Dupuy and Galichon (2014). The structure of $u_{\cdot,k}$ and $v_{\cdot,k}$ is fully flexible. The difference $u_{x',k} - u_{x,k}$ represents the additional value of educational attainment x' over x along dimension k. This difference does not need to be positive; if x' and x are equally valued, then $u_{x',k} - u_{x,k}$ is zero. Moreover, the index structure can be gender asymmetric, i.e., the order of the elements of $u_{\cdot,k}$ can be different (and even reverse) with respect to the order of the elements of $v_{\cdot,k}$. Hence, by recovering the index structure, we can identify the relevant margins in the marriage market without imposing ex-ante restrictions.

2.3. Numerical examples

In Becker's original model, a single index of attractiveness subsumes all individual traits (Chiappori et al., 2012). How individual characteristics map into such an index depends on the nature of the marital gains. For instance, when parents' educational levels are complementary inputs in the production of children's human capital, the index reflects a common preference order of one gender over the other gender's educational attainment (Chiappori et al., 2017).

To provide a numerical example that satisfies these basic assumptions, assume the match surplus is defined as follows:

$$\Phi_1(x, y) = \alpha \log x \log y \tag{2.8}$$

with $\alpha > 0$. We can show that, when the match surplus is given by (2.8), individuals sort on a single dimension. In fact, P is symmetric and has rank one, with $\lambda_1 = \alpha\sigma^2$, while the first singular vectors $u_{x,1} = v_{x,1} = \log x/\sigma$, with $\sigma = [\sum_{k=1}^{E}(\log k)^2]^{1/2}$, represent the (gender-symmetric) indices of attractiveness. Table 1 shows the index structure when $\alpha = 0.4$ and $E = 5$. As noted by Siow (2015), in this context, we have global positive assortative mating, since $\Delta(x, x', y, y') > 0$ for any (x, x', y, y') with $x \neq x'$ and $y \neq y'$.

Table 1: SVD representation of a single-index model

	Dim. 1
Type 1	0.00
Type 2	0.28
Type 3	0.44
Type 4	0.56
Type 5	0.65
λ_k	2.48

Notes. The table reports the first singular vector $u_{\cdot,1}$ and singular value λ_1 for the match surplus function $\Phi_1(x, y)$ with $\alpha = 0.4$ and $E = 5$. Since the match surplus is symmetric with respect to x and y, P is symmetric and $U = V$.

In contrast, we consider a market where the match surplus is given by the following function:

$$\Phi_2(x, y) = \gamma_x \mathbb{1}\{x = y\}, \tag{2.9}$$

i.e., individuals have a taste for homogamy captured by the parameters $\{\gamma_x\}_{x=1}^{E}$. In this case, Siow (2015) note that we should expect positive assortative mating only locally, since $\Delta(x, x', y, y') > 0$ for any $x = y$ and $x' \neq y'$, but equals zero elsewhere. Under this scenario, P is symmetric and has full rank. In the simplest case, when $\gamma_x = \gamma$ for any x, the singular values correspond to $\lambda_k = \gamma$ for any k, while the singular vectors are unit vectors, with $U = V = I$. Table 2 shows the index structure when the taste for similarity γ_x is allowed to differ across educational levels. The first singular value λ_1 is equal to γ_4, which represents the taste for similarity of type 4, the highest among all types. Along the first dimension, individuals of type 4 distance themselves from the rest of the market and form a cluster. On each further dimension, the k-th singular value reflects the taste for homogamy of the other types, which results in the marriage market being divided into homogeneous clusters.

Next, we consider an intermediate case where the match surplus is given by the

Table 2: SVD representation of a homogamy model

	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
Type 1	0.00	0.00	1.00	0.00	0.00
Type 2	0.00	0.00	0.00	0.00	1.00
Type 3	0.00	0.00	0.00	1.00	0.00
Type 4	1.00	0.00	0.00	0.00	0.00
Type 5	0.00	1.00	0.00	0.00	0.00
λ_k	0.99	0.81	0.76	0.26	0.20

Notes. The table reports the singular vectors $u_{\cdot,k}$ and singular values λ_k for the match surplus function $\Phi_2(x, y)$ with $\{\lambda_k\}_{k=1}^{E} = \{0.76, 0.20, 0.26, 0.99, 0.81\}$. Since the match surplus is symmetric with respect to x and y, P is symmetric and $U = V$.

mixture $\Phi_3(x, y) = 0.5\Phi_1(x, y) + 0.5\Phi_2(x, y)$. In this case, P still has full rank, but Table 3 shows that its first singular value plays a much more important role relative to the pure homogamy model of Table 2. It is now harder to establish a correspondence between the match surplus parameters and the singular values and singular vectors, but the intuition behind the first two examples carries over. The values of the first singular vector are increasing in x, similarly to the single-index case of Table 1 and reflecting the complementarities introduced by the function Φ_1. However, while the first dimension suggests that types 4 and 5 are fairly similar and should often match, the second dimension accounts for the additional segmentation introduced by the homogamy parameters γ_4 and γ_5, which set types 4 and 5 apart in the market. The largest singular vector component along each following dimension respectively captures the taste for similarity of types 1, 3, and 2.

Table 3: SVD representation of a mixed model

	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
Type 1	0.00	0.00	1.00	0.00	0.00
Type 2	0.23	-0.04	0.00	0.36	0.90
Type 3	0.37	-0.07	0.00	0.82	-0.42
Type 4	0.62	0.76	0.00	-0.22	-0.03
Type 5	0.66	-0.65	0.00	-0.37	-0.05
λ_k	1.61	0.46	0.38	0.19	0.11

Notes. The table reports the singular vectors $u_{\cdot,k}$ and singular values λ_k for the match surplus $\Phi_3(x, y) = 0.5\Phi_1(x, y) + 0.5\Phi_2(x, y)$ and $E = 5$. Since the match surplus is symmetric with respect to x and y, P is symmetric and $U = V$.

So far, we only considered symmetric match surplus functions. However, what happens if, for instance, couples where the wife is more educated than the husband experience a penalty? In this case, the male and female indices would differ from each other, as shown in Table A1 in the Online Appendix. In particular, it is worth noting that, while the first male index increases in education, more educated women experience small or even negative returns from sorting on the marriage market.

2.4. Double-centered surplus matrix

Since our goal is to analyze assortative mating, we now focus on the structure of complementarities given by $\Delta(x, x', y, y')$ for any (x, y) and (x', y'). First, we note that changes in male or female type fixed effects in the match surplus leave $\Delta(x, x', y, y')$ unchanged, since the latter is obtained from the double difference (2.4). Hence, we will work with a normalized, double-centered version of P, which we name $\tilde{P} = \tilde{U}\tilde{\Lambda}\tilde{V}^\intercal$, with elements:

$$\tilde{p}_{x,y} = \Phi(x,y) - \underbrace{\frac{1}{E}\sum_{y}\Phi(x,y)}_{\equiv\bar{\Phi}(x)} - \underbrace{\frac{1}{E}\sum_{x}\Phi(x,y)}_{\equiv\bar{\Phi}(y)} + \underbrace{\frac{1}{E^2}\sum_{x,y}\Phi(x,y)}_{\equiv\bar{\Phi}} \tag{2.10}$$

where we note that the rank of $\tilde{P}$ cannot be larger than $E - 1$.

By construction, the elements of the double-centered matrix $\tilde{P}$ only depend on the complementarities $\Delta(x, x', y, y')$ and, contrary to the elements of the match surplus matrix P, they are invariant to changes in mean surplus $\bar{\Phi}(x)$ and $\bar{\Phi}(y)$. In fact, we have:

$$\tilde{p}_{x,y} = \sum_{\substack{x'\neq x\\ y'\neq y}} \frac{\Delta(x,x',y,y')}{E^2}, \tag{2.11}$$

Differences in mean surplus $\bar{\Phi}(x)$ and $\bar{\Phi}(y)$ across types and over time are interesting per se, in that they document changes in the extensive margin, i.e., whether individuals get married or not. However, when we work with the double-centered matrix $\tilde{P}$, we can abstract from changes in the extensive margin and focus on the intensive margin, i.e., assortative mating. Moreover, since $\tilde{P}$ is identified exclusively from data on couples, this also implies that the empirical analysis of assortative mating is robust to different definitions of singlehood, which potentially imply different values for the empirical frequencies $\mu(x, 0)$ and $\mu(0, y)$.

2.5. Measuring assortative mating

We can now rewrite $\Delta(x, x', y, y')$ as follows:[7]

$$\Delta(x,x',y,y') = \sum_{k=1}^{K} \tilde{\lambda}_k(\tilde{u}_{x',k} - \tilde{u}_{x,k})(\tilde{v}_{y',k} - \tilde{v}_{y,k}), \tag{2.12}$$

which suggests that $\tilde{\lambda}_k$ is an overall measure of the strength of assortative mating along dimension k, while differences in complementarities across educational levels depend on the contrast patterns of the male and female indices $\tilde{u}_{\cdot,k}$ and $\tilde{v}_{\cdot,k}$. While other normalizations for P are also possible, double-centering is advantageous since not only $\sum_x \tilde{u}^2_{x,k} = 1$

[7]This is simply due to $\Delta(x, x', y, y') = p_{x,y} + p_{x',y'} - p_{x,y'} - p_{x',y} = \tilde{p}_{x,y} + \tilde{p}_{x',y'} - \tilde{p}_{x,y'} - \tilde{p}_{x',y}$.

and $\sum_y \tilde{v}_{y,k}^2 = 1$,[8] but also $\sum_x \tilde{u}_{x,k} = 0$ and $\sum_y \tilde{v}_{y,k} = 0$ for any k. This facilitates the comparison of the indices $\tilde{u}_{\cdot,k}$ and $\tilde{v}_{\cdot,k}$ over time.

In order to analyze aggregate assortative mating patterns, we look at the singular values $\{\tilde{\lambda}_k\}_{k=1}^K$ and how they changed over time. First, we can estimate $\tilde{P}$ and test its rank K as in Kleibergen and Paap (2006) and Dupuy and Galichon (2014). Second, we look at two key summary statistics: the spectral norm and the Frobenius norm. Their formulas are respectively given by:

$$\|\tilde{P}\|_2 = \max_k \tilde{\lambda}_k = \tilde{\lambda}_1 \tag{2.13}$$

$$\|\tilde{P}\|_F = \sqrt{\sum_{x,y} \tilde{p}_{x,y}^2} = \sqrt{\sum_k \tilde{\lambda}_k^2}. \tag{2.14}$$

Since the elements of $\tilde{P}$ exclusively depend on complementarities $\Delta(x, x', y, y')$, the matrix norms $\|\tilde{P}\|_2$ and $\|\tilde{P}\|_F$ result from the aggregation of local odds ratios. In particular, $\|\tilde{P}\|_F$ is proportional to the metric proposed by Altham (1970), which is given by the root sum of squared log odds ratios.[9] Moreover, when $E = 2$, it is easy to show that $\|\tilde{P}\|_2 = \|\tilde{P}\|_F = \Delta(1, 2, 1, 2)/2$, i.e., the two norms correspond to the (only) log odds ratio divided by two.

What do we learn from each of the two metrics? The spectral norm measures the strength of sorting on the first and most important dimension, but it disregards other dimensions. The Frobenius norm measures the total energy of the matrix and informs us on the overall strength of sorting, but does not say anything about how it is distributed across dimensions. However, the two measures correspond in single-index models, when the rank of $\tilde{P}$ is equal to one, while they are the farthest away when all dimensions play an equal role. We can generalize this idea and introduce the following ratio:

$$\rho_k = \frac{\tilde{\lambda}_k^2}{\sum_l \tilde{\lambda}_l^2}, \tag{2.15}$$

which captures the fraction of overall assortative mating explained by dimension k and helps us assess the relative importance of each dimension, with $\sum_k \rho_k = 1$.[10]

2.6. Changes in the index structure

The decomposition of $\Delta(x, x', y, y')$ in equation (2.12) shows that local changes in assortative mating do not only stem from changes in singular values, but can also reflect

[8] For any orthogonal matrix Q, we have $Q^\intercal Q = I$.

[9] The link between the two measures is also visible in our empirical application. The results obtained with the Frobenius norm in panel (a) of Figure 1 are identical to those obtained with the Altham's index in panel (b) of Figure A5, in the Online Appendix. The only difference between the two is the scale.

[10] A related statistic is the (normalized) Herfindahl-Hirschman concentration index suggested by Cheremukhin et al. (2024), which measures to what extent different sorting dimensions contribute equally.

changes in the singular vectors of $\tilde{P}$, i.e., in the composition of the indices of attractiveness. Hence, we document how different singular vectors $\tilde{u}_{\cdot,k}$ and $\tilde{v}_{\cdot,k}$ have changed over time in order to analyze local changes in assortative mating and better understand what types of preferences can rationalize the observed marital patterns.

When comparing different sets of singular vectors over time, two additional remarks are in order. First, singular vectors are uniquely determined up to a sign. Hence, it is convenient to flip the signs of certain singular vectors to ease their interpretation. Second, the relative importance of different sorting dimensions, as measured by the corresponding singular values, may change over time. Hence, it is convenient to permute the singular values $\tilde{\Lambda}$ and the columns of matrices $\tilde{U}$ and $\tilde{V}$ in order to preserve the interpretation of each sorting dimension over time. This reordering is merely a relabeling procedure, which may not be necessary if the singular values remain well distinct over time.[11]

Finally, additional care is needed if we compare different sets of singular vectors across samples with a different education variable, particularly if the number of categories E changes. In fact, orthogonality of $\tilde{U}$ and $\tilde{V}$ implies $\text{Var}[\tilde{u}_{\cdot,k}|k] = \text{Var}[\tilde{v}_{\cdot,k}|k] = E^{-1}$, which means that singular values are artificially inflated when E increases. Consider two samples s and t with E_s and E_t. In this case, it is convenient to compare $\sqrt{E_s}\tilde{U}_s$ with $\sqrt{E_t}\tilde{U}_t$ (same for right singular vectors). It also follows that, if we want to compare singular values $\tilde{\Lambda}_s$ and $\tilde{\Lambda}_t$ (or the resulting matrix norms), we should first divide them by E_s and E_t respectively.

3. Data and estimation

3.1. Current Population Survey

We use data from the Annual Social and Economic Supplements (ASEC) of the Current Population Survey (CPS) for the period 1962–2025.[12] Following Eika et al. (2019), we focus on a sample of couples with at least one partner aged between 26 and 60 upon the survey date.[13] We look at both legally married and cohabiting couples in primary and secondary families, although we also run a robustness check with only legally married household heads and their spouses, due to changes in the coding of intrahousehold relationships in the CPS.[14] We also use data on singles aged between 26 and 60 to provide some auxiliary results on changes in marriage rates over time, although our analysis is

[11] Note that reordering explains why $\tilde{\lambda}_k$ is not necessarily larger than $\tilde{\lambda}_{k+1}$ in our findings.

[12] CPS data do not contain information about education for 1963. The year is excluded from the sample.

[13] We explore alternative age restrictions and look at couples with at least one partner aged between 35 and 44, as in Chiappori et al. (2025), but our conclusions remain unchanged. Results are available upon request.

[14] Since 1962, we can easily identify the household head and their spouse. Since 1988, the CPS has more detailed information on the relationships between members other than the head and their spouse (e.g., we can identify the head's siblings and parents). From 1995 only, we can safely identify unmarried partners.

robust to alternative definitions of singlehood, as anticipated in Section 2.5.

In our benchmark specification, we divide individuals into five educational groups, as previously done by Greenwood et al. (2014), Siow (2015), Chiappori et al. (2017), and Chiappori et al. (2025) to study assortative mating, but also by Acemoglu and Autor (2011) to document changes in the wage distribution. The educational groups are: high-school dropouts (<HS), high-school diplomas (HS), some college (SC), college degrees (C), and postgraduate degrees (C+). However, we note that the coding of educational attainment in the CPS is not entirely consistent over time. Until 1991, education is essentially measured as years of schooling, but it is not always clear whether the respondent obtained a certain degree.[15] Since 1992, educational categories correspond to obtained diplomas and degrees. We also explore both less and more detailed educational categorizations to understand how this affects our findings. For instance, following Eika et al. (2019), we group college and postgraduate degrees and replicate our findings with four categories. Moreover, for data until 1991, we divide individuals into nine groups: until 11th grade (<HS), 12th grade but diploma is unclear (HS?), 12th grade with high-school diploma (HS), one year of college (C1), two years (C2), three years (C3), four years (C4), five years (C5), six or more years (C6+). Finally, for data since 1992, we use an alternative categorization with eight groups: high-school dropouts (<HS), high-school diplomas (HS), some college but no degree (SC), associate degrees (AD), college degrees (C), master's degrees (M), professional degrees (PD), and doctoral degrees (PhD).

Our main sample contains 1,655,531 couples, 690,089 single women, and 544,322 single men for the period 1962–2025. The descriptive statistics, contained in the Online Appendix, replicate some well-known facts. Figure A1 shows that men were more educated than women in the 1960s, but their education did not grow as fast. Since the early 2000s, women have been more educated than men. Figure A2 shows that the share of singles among men aged 26 to 60 was very similar across educational groups until the early 1980s, but a strong educational gradient emerged since then, with less educated men much more likely to be single in recent years. Similarly, a large gap in the probability of being single emerged between women with and without a college degree. Finally, the singlehood rate of women with a postgraduate degree used to be the highest among all educational groups in the 1960s, whereas it is now the lowest.

[15]This issue is clearly discussed by Schwartz and Mare (2005) and Gihleb and Lang (2020) among others. Similar issues are present also in the data from the U.S. Census Bureau, with similar changes in coding starting from the 1990 Census. For instance, it is not always clear if individuals reaching 12th grade obtained a high-school diploma, while tertiary education is measured in number of years, so that we cannot clearly identify individuals with a master or PhD degree.

3.2. Estimation

We estimate empirical frequencies $\mu_{t,t+4}(x,y)$ separately for every five-year interval $[t, t+4]$, starting from 1962-1966 and ending with 2021-2025.[16] Empirical frequencies are calculated using statistical weights. On average, each five-year interval contains about 133,000 couples and 98,000 singles. Then, we compute the double-centered surplus matrix $\tilde{P}_{t,t+4}$ using the matching function (2.1) and the double-centering formula (2.10). Binning the data into overlapping intervals has two advantages. First, it reduces noise and results in a smoother sequence of estimates for the matrices $\{\tilde{P}_{t,t+4}\}_{t=1962}^{2021}$. Second, it ensures strictly positive values for empirical frequencies $\mu(x,y)$, even for infrequent combinations of types (x,y), which is needed for the estimation of $\Phi(x,y)$.

Starting from the sample estimate $\tilde{P}_{t,t+4}$, we obtain the singular values $\tilde{\Lambda}_{t,t+4}$ and singular vectors $\tilde{U}_{t,t+4}$ and $\tilde{V}_{t,t+4}$. As anticipated in Section 2.6, since singular vectors are uniquely determined up to a sign, we adjust the signs of $\tilde{U}_{t,t+4}$ and $\tilde{V}_{t,t+4}$ to maximize their similarity with estimates from the previous interval, $\tilde{U}_{t-1,t+3}$ and $\tilde{V}_{t-1,t+3}$. Moreover, since the relative importance of the different sorting dimensions may change over time, we reorder the singular values $\tilde{\Lambda}_{t,t+4}$ and the columns of $\tilde{U}_{t,t+4}$ and $\tilde{V}_{t,t+4}$, always to maximize their similarity with $\tilde{U}_{t-1,t+3}$ and $\tilde{V}_{t-1,t+3}$.

We obtain confidence intervals for the elements of $\tilde{P}_{t,t+4}$ with 1,000 bootstrap samples, using Bayesian bootstrapping to ensure bootstrap estimates of the empirical frequencies $\mu(x,y)$ remain strictly positive. In order to obtain confidence intervals for the singular values and vectors of the estimated $\tilde{P}_{t,t+4}$, we adjust the signs of the singular vectors and reorder the sorting dimensions in order to minimize the distance between the bootstrap and sample estimates (Milan and Whittaker, 1995; Chiappori et al., 2024). In practice, when the singular values of $\tilde{P}_{t,t+4}$ are well distinct, inference is straightforward, whereas confidence intervals for singular vectors are larger when the singular values of $\tilde{P}_{t,t+4}$ are close.[17] Finally, for every t, we infer the rank of the estimated matrix $\tilde{P}_{t,t+4}$ as in Kleibergen and Paap (2006) and Dupuy and Galichon (2014).

4. Main results

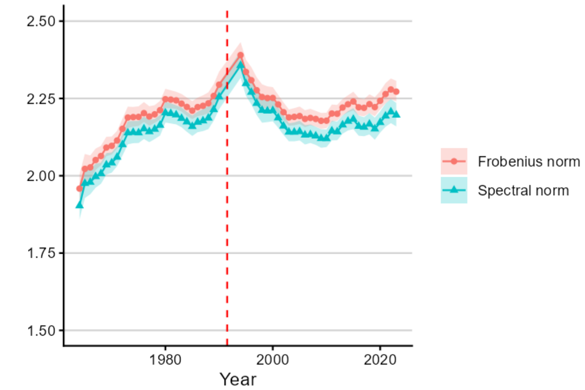
4.1. Number of sorting dimensions

Panels (a) and (b) in Figure 1 show that the spectral and Frobenius norms not only exhibit similar trends over time, but also that they are very close in levels in all periods. Moreover, panel (c) shows that the first singular value is always four to five times larger than the second, while panel (d) shows that the first sorting dimension accounts for 93%

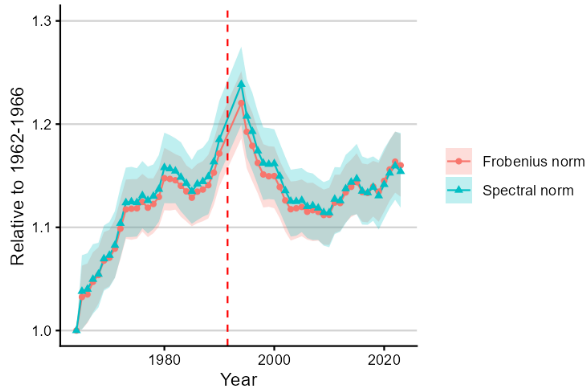
[16] We do not pool data from years with different specifications of the educational variable. Namely, we do not pool data collected before and after 1992. This means our last interval before the break is 1987-1991 and the first after the break is 1992-1996.

[17] See also Section 7.2 in Tibshirani and Efron (1993) and Sections 3.6 and 3.7 in Jolliffe (2002) for a discussion of this problem.

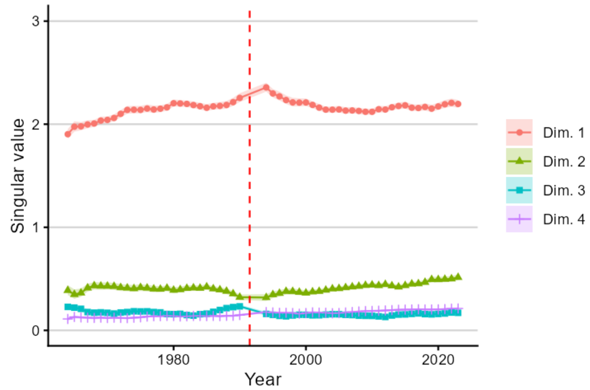
Figure 1: Changes in assortative mating over time - singular values and matrix norms



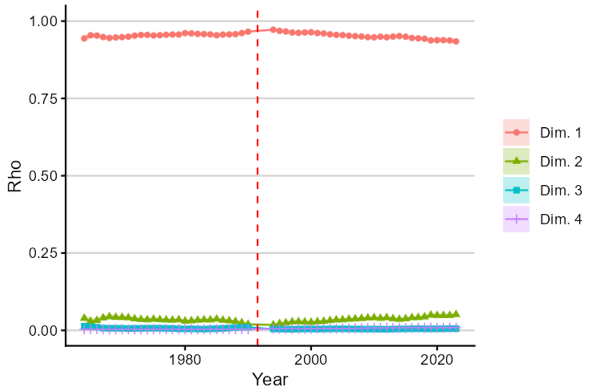
(a) Norms: $\|\tilde{P}\|_2$ and $\|\tilde{P}\|_F$



(b) Norms relative to 1962-1966



(c) Singular values: $\tilde{\lambda}_k$



(d) Relative contributions: ρ_k

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Five educational groups. Panel (a) shows the spectral and Frobenius norm of the double-centered surplus matrix $\{\tilde{P}_{t,t+4}\}_{t=1962}^{2021}$. Panel (b) shows how these norms changed relative to their values in the 1962-1966 period, in which they are normalized to one. Panel (c) shows the singular values $\tilde{\lambda}_k$ with $k = 1, 2, 3, 4$. Panel (d) shows the relative contribution of these sorting dimensions to assortative mating, as measured by ρ_k, defined in equation (2.15). The shaded areas represent 95% two-tailed confidence intervals.

to 98% of overall assortative mating in all periods. Hence, a one-dimensional model of sorting can, by and large, replicate the aggregate educational sorting patterns in the data, even though, due to our large sample size, we formally reject that the matrix $\tilde{P}_{t,t+4}$ has rank lower than $E - 1$ in all periods.[18]

Concretely, how well does a one-dimensional representation of the marriage market replicate the observed marital patterns? To answer this question in a more intuitive way, we calculate the predicted patterns if individuals sorted exclusively on the first dimension, i.e., if $\tilde{\lambda}_k$ were zero for any $k \geq 2$ and the normalized match surplus were equal to $\tilde{\Phi}(x, y) =$

[18] We present the estimated rank in Figure A4 in the Online Appendix. We iteratively test the hypothesis $\text{rank}(\tilde{P}_{t,t'}) = k$ with $k = 1, 2, ...E - 1$, with $E - 1$ being the largest possible after the double centering. We conclude that the rank of $\tilde{P}_{t,t'}$ is equal to k if the corresponding p-value is higher than 0.01. When we consider a higher number of educational categories, we occasionally find that the rank is lower than $E - 1$.

Table 4: Predicted sorting with a one-dimensional model

	<HS	HS	SC	C	C+
<HS	0.110	0.062	0.013	0.003	0.001
HS	0.042	0.185	0.059	0.022	0.006
SC	0.011	0.072	0.087	0.034	0.011
C	0.003	0.033	0.045	0.075	0.023
C+	0.001	0.010	0.019	0.038	0.035

(a) Observed (1962-2025)

	<HS	HS	SC	C	C+
<HS	0.086	0.084	0.016	0.003	0.001
HS	0.061	0.151	0.069	0.026	0.008
SC	0.015	0.079	0.063	0.041	0.015
C	0.003	0.036	0.050	0.060	0.029
C+	0.001	0.012	0.024	0.043	0.024

(b) Counterfactual (1962-2025)

	<HS	HS	SC	C	C+
<HS	0.044	0.076	0.035	0.025	0.009
HS	0.053	0.115	0.070	0.054	0.023
SC	0.031	0.075	0.051	0.040	0.018
C	0.025	0.061	0.042	0.034	0.016
C+	0.014	0.035	0.025	0.020	0.010

(c) Random matching (1962-2025)

	<HS	HS	SC	C	C+
<HS	-0.024	0.021	0.003	-0.000	-0.000
HS	0.019	-0.034	0.010	0.004	0.001
SC	0.004	0.008	-0.024	0.007	0.005
C	0.001	0.003	0.005	-0.015	0.006
C+	-0.000	0.002	0.006	0.005	-0.012

(d) Difference (b) - (a)

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Panel (a) shows the observed relative frequencies for the entire 1962-2025 period. Panel (b) shows the relative frequencies if individuals only sorted along the first dimension. The counterfactual frequencies are obtained for every five-year interval $[t, t+4]$, using the estimated $\tilde{u}_{\cdot,1}$, $\tilde{v}_{\cdot,1}$, and $\tilde{\lambda}_1$ to calculate the counterfactual assignment, and then averaged over the entire period. Similarly, panel (c) shows the frequencies if individuals were randomly paired for every five-year interval and then averaged over the entire period.

$\tilde{\lambda}_1 \tilde{u}_{x,1} \tilde{v}_{y,1}$.[19] Table 4 compares observed and counterfactual relative frequencies, obtained by averaging contingency tables over the entire 1962-2025 period. When we disregard the higher-order dimensions, we still obtain strong educational sorting; if compared to the random assignment, we see that couples with a large educational gap are rare, while there is substantially more mass on and around the diagonal. Yet, our one-dimensional model underpredicts the share of couples exactly on the diagonal, which suggests that higher-order dimensions help explain the prevalence of same-education couples by capturing homogamous preferences. As we will show in Section 4.4, the structure of the indices of attractiveness supports this interpretation.

4.2. Overall assortative mating

In this section, we present trends in overall assortative mating. Figure 1 is based on five educational categories, and panel (a) shows that assortative mating, as measured by either the spectral or the Frobenius norm, increased over time. Panel (b) shows that these measures are about 16% higher in 2021-2025 relative to the starting period 1962-1966. While assortative mating exhibits a long-run upward trend, it peaked in the 1992-1996 period, when both the spectral and Frobenius norms were 22% larger than in the 1962-1966 period. Between the mid-1990s and the mid-2000s, assortative mating decreased,

[19] The counterfactual assignment can be obtained using the iterative procedure introduced by Mosteller (1968) to calculate counterfactual contingency tables that preserve the original margins. Dupuy and Galichon (2014) and Galichon and Salanié (2022) review iterative methods to solve for the equilibrium matching in models with additively separable extreme value type I taste shocks.

then stabilized and possibly increased again until 2021-2025, but at a much slower pace than in previous decades.

How do these findings compare to the literature? Using U.S. Census or CPS data, existing studies have reached conflicting conclusions depending on the statistic used to measure assortative mating and the definition of the educational groups (Gihleb and Lang, 2020; Chiappori et al., 2025). From a methodological perspective, our study is closest to Siow (2015) and Chiappori et al. (2017), who both build on Choo and Siow (2006) and show that assortative mating has increased over time. Using different methodologies, Greenwood et al. (2014), Eika et al. (2019), and Chiappori et al. (2025) also show that overall assortative mating has increased in the long run. However, Gihleb and Lang (2020) and Chiappori et al. (2025) cast doubt on the robustness of these findings.[20] Some of their concerns are methodological: here we provide an alternative metric to those used in the above-mentioned papers and, using a comparable sample and variable definition, we bring additional evidence of a rise in educational assortative mating over time. Yet, part of the concerns are due to sensitivity to categorization and changes in the coding of the educational variable: we extensively discuss these issues in Section 5.

Finally, in spite of the long-run increase, previous works have also documented periods of faster growth and others of stagnation or even decline. Gihleb and Lang (2020) argue that assortative mating has likely increased and then decreased since the 1960s, while Eika et al. (2019) find that assortative mating has stabilized since the 1990s. Similarly, using Poisson regression models with two-way fixed effects, Mare (1991) and Schwartz and Mare (2005) find that assortative mating has increased in the second part of the 20th century, while Hirschl et al. (2024) show that this trend has stopped and possibly reverted over the last three decades. Our findings confirm that the rise of educational assortative mating has substantially slowed down and stabilized in the last three decades.

4.3. First dimension: complementarities à la Becker

To have a complete picture of how educational sorting works, we look at the composition of the relevant indices of attractiveness, corresponding to the singular vectors of $\tilde{P}_{t,t+4}$, and at their evolution over time. Panels (a) and (b) in Figure 2 show that, along the first dimension, individuals are ranked based on their educational attainment, from the least to the most educated. Any progression from category k to $k+1$ results in an upward movement in the marriage market ranking, for both men and women. Hence, sorting along the first dimension is consistent with the presence of human capital complementarities in consumption, leisure, and home production (Chiappori et al., 2017).

[20] Eika et al. (2019) and Chiappori et al. (2025) measure overall assortative mating by the weighted likelihood ratio, while Greenwood et al. (2014) use the Kendall's rank correlation. Chiappori et al. (2025) also show that the results are ambiguous when using the normalized trace, while Gihleb and Lang (2020) argue that results obtained with linear and rank correlation rates are not robust. In the Online Appendix, Figure A5 shows trends for selected statistics as a means of comparison with these papers.

The first index structure is remarkably stable over time and validates the use of five distinct categories in the literature, since the differences $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$ and $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$ are always positive and of comparable magnitude across levels. Yet, we do observe some changes in the value of different educational attainments over time. For men, the penalty for dropping out of high school (<HS) relative to obtaining a high-school diploma (HS) is larger today than in the past, while this is not the case for women. The value of obtaining a college degree (C) relative to dropping out of college (SC) has slightly increased for women, but it has remained stable for men. In contrast, the distance between college and postgraduate degrees (C+) has decreased for men, while it has slightly increased for women. In Figure A7 in the Online Appendix, we plot differences $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$ and $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$ over time in order to provide a clear representation of their evolution over time.

Panel (a) also suggests that, overall, the first index has become more gender symmetric over time. We test elementwise symmetry in Figure A8 by plotting the difference $\tilde{u}_{k,1} - \tilde{v}_{k,1}$ for every educational level k and the respective confidence interval. Gender differences in the index composition have vanished over time, which implies that educational returns from sorting along the first dimension have become more similar for men and women in more recent decades. This strong symmetry also suggests that, if there exist norms dictating that husbands should be more educated than wives, as documented by Hitsch et al. (2010), they have substantially weakened over time and are no longer noticeable from aggregate marital patterns.

4.4. Higher-order dimensions: homogamous preferences

The higher-order sorting dimensions play a less important, but still non-negligible role. As shown by Table 4, they capture what we loosely refer to as homogamous preferences, by which we mean a pure taste for similarity rather than complementarities à la Becker. We note, however, that search frictions can generate observationally equivalent patterns (Shimer and Smith, 2000), so that the two interpretations cannot be formally distinguished within our reduced-form framework. Panels (c) and (d) of Figure 2 show that, along the second sorting dimension, men and women without a high-school diploma and with a postgraduate degree rank high, while others rank low. By construction, individuals with similar index values tend to match with each other. Hence, sorting along the second dimension results in clusters at both ends of the educational distribution, i.e., postgraduate degrees (C+) and high-school dropouts (<HS) distance themselves from the rest of the market and tend to match within their educational group. Note that, while postgraduate degrees and high-school dropouts can potentially maximize their returns from sorting on the second dimension by matching with each other, this will typically not happen. In fact, sorting is multidimensional, and postgraduate degrees and high-school dropouts are already distant along the first dimension. This can easily be understood by

looking at panels (e) and (f), where we provide a two-dimensional map of the marriage market in order to visually assess the distance between educational groups.

In the Online Appendix, Figure A9 shows that the third and fourth sorting dimension further reflect the homogamous preferences of other educational groups.[21] The indices have once again a strongly gender-symmetric composition and contribute to the creation of homogeneous clusters in the marriage market. In the Online Appendix, we provide further guidance to understand how each sorting dimension contributes to educational homogamy.[22] For every $k = 2, ..., E-1$, Table A2 shows changes in marital patterns if we suppressed the k-th sorting dimension by setting $\tilde{\lambda}_k$ to zero. For instance, panels (e) and (f) show that the fourth sorting dimension captures a divide between high-school diplomas (HS) and individuals with some college but no degree (SC). This clearly reflects the index structure shown in panels (c) and (d) of Figure A9, where these two categories respectively have very low and high values along this dimension.

4.5. Sorting at the top and bottom of the distribution

Panels (c) and (d) in Figure 2 also show changes in the index structure along the second dimension. For both men and women, the gap between college (C) and postgraduate degrees (C+) has closed over time, while the gap between high-school diplomas (HS) and dropouts (<HS) has widened. Panels (e) and (f) allow us to graphically assess changes in the distance between educational groups in the marriage market in a two-dimensional space. Due to changes along the first and second dimension, male and female high-school dropouts are more isolated in the 2021-2025 period than in the 1962-1966 period, as their distance from high-school diplomas increased. Eika et al. (2019) also report that high-school dropouts have become less and less likely to marry outside of their group.

At the other end of the distribution, college graduates have pulled away from individuals who attended college without a degree, with the distance between the two groups increasing between 1962-1966 and 2021-2025. Chiappori et al. (2025) also find a local increase in assortative mating for the two groups, while Hirschl et al. (2024) emphasize that obtaining a college degree still remains a very important dividing line in the marriage market. In our framework, this increasing divide is entirely explained by changes in the second sorting dimension. In contrast, postgraduate degrees have become closer to college degrees. Once again, this result is in line with Chiappori et al. (2025), who document a decrease in sorting between college graduates and postgraduates.

[21] Panel (c) of Figure 1 shows that the third and fourth singular values are close, at least in some years. As anticipated in Section 3.2, this sometimes results in larger confidence intervals for some components of the third and fourth singular vectors.

[22] We do this exercise separately for the 1962-1991 and 1992-2025 periods since the composition of the second and higher-order indices is seemingly sensitive to changes in coding of the educational variable starting from 1992.

5. Robustness checks

5.1. Changes in coding and coarser categorizations

While our measures of assortative mating based on odds ratios conserve their appealing theoretical properties, our aggregate measures remain sensitive to categorization, an issue discussed by Gihleb and Lang (2020) and formalized by Chiappori et al. (2025). In particular, it is important to note that the peak in assortative mating observed in 1992-1996 coincides with a discontinuity in the educational variable in the CPS, corresponding to the vertical red line in Figure 1.[23] The upward spike suggests that assortative mating is stronger when individuals are grouped following the newer CPS classification, and thus that individuals tend to sort based on their attainments (i.e., diplomas and degrees) rather than on years spent in school. Hence, assortative mating measures before 1992 may be downward biased, a problem already highlighted by Gihleb and Lang (2020). This casts doubt on whether there has really been a long-run increase in assortative mating, since the rise during the first part of our time frame may have been completely eroded by a decline in the second.

In order to better understand how changes in coding affects our conclusions, we explore coarser educational categorizations. If the downward bias were due to individuals being misclassified with the pre-1992 coding, then aggregating two or more categories might alleviate the problem. Figure A6 in the Online Appendix shows that the only alternative categorization that substantially reduces the hump around 1992 is the one where high-school dropouts and diplomas are pulled together (panel (b)).[24] When we aggregate these two groups, our metrics suggest that the rise of assortative mating is much more modest, with the Frobenius and spectral norm respectively increasing by 3.6% and 5%.[25] However, adopting this coarser categorization also implies that we disregard an increasingly important margin in the marriage market. In fact, Eika et al. (2019) find that assortative mating has strengthened among the low educated, a conclusion that we also reach when we look at changes in the structure of the indices of attractiveness in Sections 4.5 and 5.3. Hence, while it is possible that changes in coding of the educational variable lead to upward biased estimates of the long-run increase in assortative mating, results obtained when aggregating high-school dropouts and graduates likely constitute

[23] Another potential issue is the change in how intrahousehold relationships are coded in the CPS, which mainly occurred between 1988 and 1995. However, Figure A3 shows that the results are very similar when we exclusively look at household heads and their spouse, removing unmarried couples and secondary families from the sample.

[24] This is likely due to the impossibility of distinguishing individuals who completed 12th grade without obtaining a diploma and high-school graduates before 1992.

[25] We reach a similar conclusion when we aggregate high-school dropouts and diplomas into one group and college graduates and postgraduates in another group (panel (d)). With this three-group categorization, assortative mating is unchanged in the long run, although we see a more pronounced increase in earlier and later waves, eroded by a substantial decline in the middle of our time frame, between 1980 and 2000.

a lower bound. We therefore conclude that the true long-run increase in assortative mating lies between this lower bound and the baseline estimate, with the exact magnitude depending on the degree of misclassification before 1992.

5.2. More refined categorizations

We present additional findings obtained with a more detailed categorization of education. Panels (a) and (b) in Figure 3 show that the spectral and Frobenius norms are close in all periods, while panels (c) and (d) confirm that the first sorting dimension accounts for about 85%–91% of assortative mating across the two different specifications. Hence, even with a more detailed categorization of education, a one-dimensional model remains a good representation of the marriage market.

Panel (a) in Figure 3 also confirms that educational assortative mating increased over the 1962-1991 period, with both norms increasing by about 13%. The growth was much stronger in the period until 1975, while it slowed down over time. Instead, panel (b) shows a sharp decline in the 1990s and a mild increase thereafter, which results in a 5% decrease in the Frobenius norm in three decades. Comparing the measures of assortative mating across the two samples,[26] the values are higher in the 1992-2025 period, confirming that individuals tend to sort on attainments rather than years in school.

These results broadly confirm those of Section 4, with a period of rapid growth in the earlier decades, followed by a decline in the 1990s and a period of modest growth in the last two decades. Hence, adopting a more refined categorization does not radically alter the conclusions about temporal changes in educational assortative mating. However, due to the coding changes in 1992, it remains difficult to say whether the increase in educational assortative mating during the 1962-1991 period was large relative to the subsequent decline.

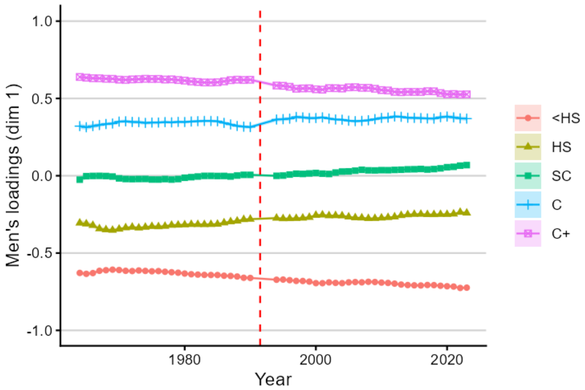
5.3. Which margins matter?

Figure 4 shows the elements of the first singular vectors for men and women with a larger number of educational groups. Individuals with higher education are ranked higher along the first dimension in the marriage market. Once again, this result is remarkably stable over time and across the two specifications. However, this approach is advantageous since, starting from a large number of categories, it allows us to identify the most important margins on the marriage market.[27] Hence, it can help us justify the choice of a particular categorization, i.e., whether two categories should be treated as distinct or

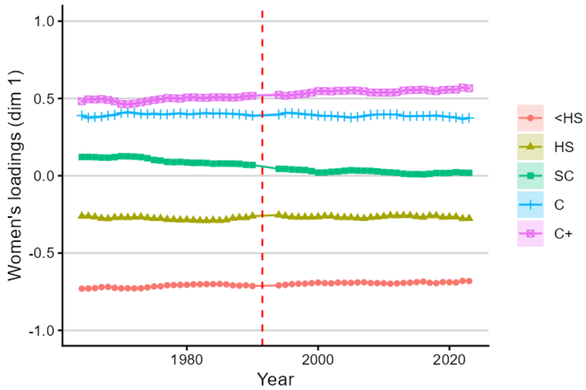
[26] The comparison is meaningful if the norms are appropriately adjusted. As explained in Section 2.6, the adjustment simply consists in dividing the singular values and norms by the number of educational categories E.

[27] Figure 4 is also useful to better understand how to treat residual or ambiguous categories. For instance, panels (a) and (b) show that individuals who have done twelve years of high school but may not have obtained a diploma (category HS?) rank somewhere in between high-school dropouts (<HS) and high-school diplomas (HS).

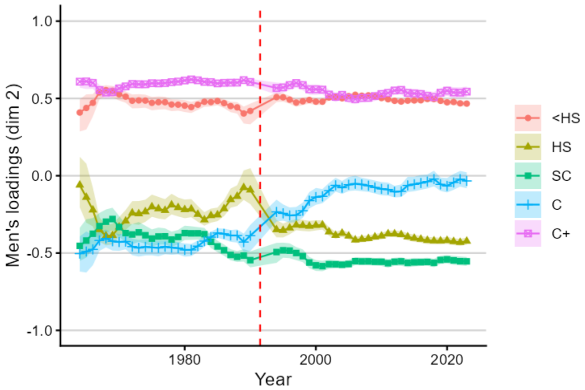
Figure 2: Changes in index composition over time, first and second dimension



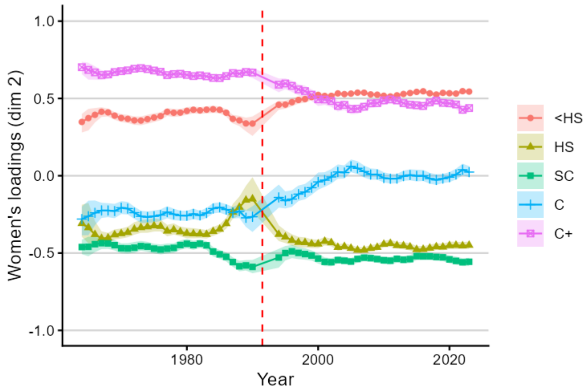
(a) Men, dim. 1: $\tilde{u}_{\cdot,1}$



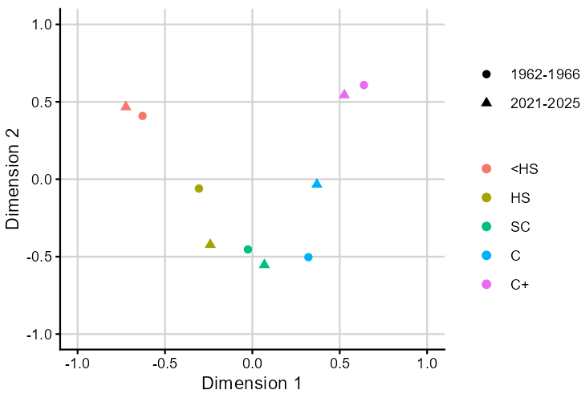
(b) Women, dim. 1: $\tilde{v}_{\cdot,1}$



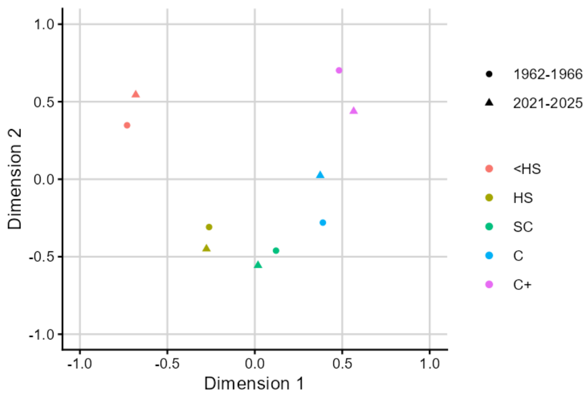
(c) Men, dim. 2: $\tilde{u}_{\cdot,2}$



(d) Women, dim. 2: $\tilde{v}_{\cdot,2}$



(e) Men, dim. 1 and dim. 2



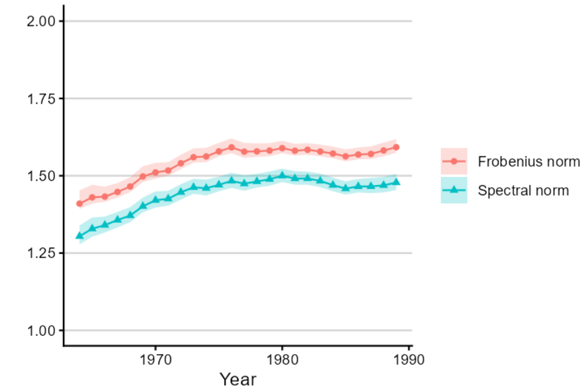
(f) Women, dim. 1 and dim. 2

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Five educational groups. Panels (a) to (d) show how the elements of the first two singular vectors changed over time, for men and women separately. Panel (e) shows the elements of the first two singular vectors for men for two selected periods; every dot corresponds to $(\tilde{u}_{k,1}, \tilde{u}_{k,2})$ for $k = 1, ..., E$. Panel (f) does the same for women. The educational groups are: high-school dropouts (<HS), high-school diplomas (HS), some college (SC), college degrees (C), and postgraduate degrees (C+). The shaded areas represent 95% two-tailed confidence intervals.

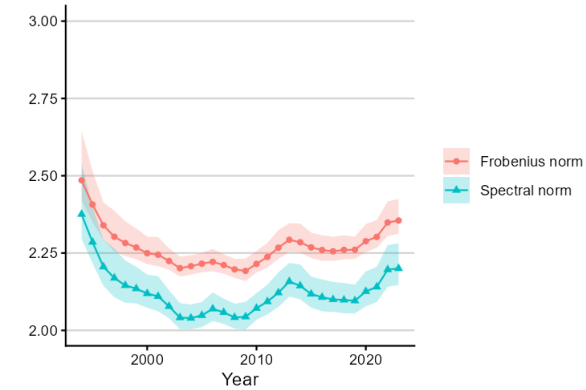
not.

Completing only one year of college education (C1) results in small returns from

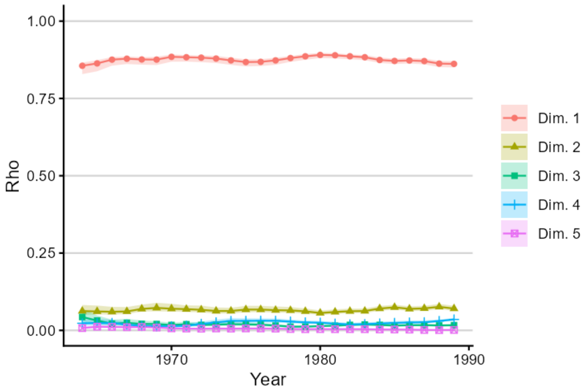
Figure 3: Changes in assortative mating over time - detailed categorization



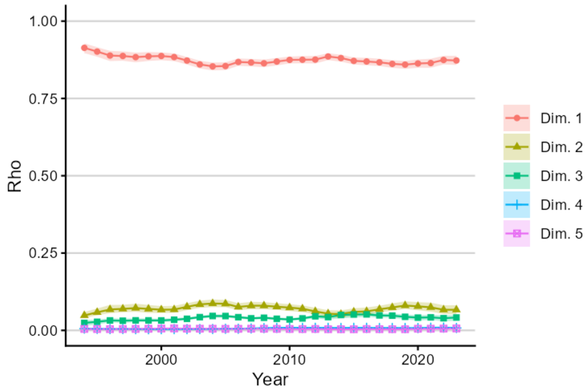
(a) Norms (1962-1991)



(b) Norms (1992-2025)



(c) Relative contributions (1962-1991)



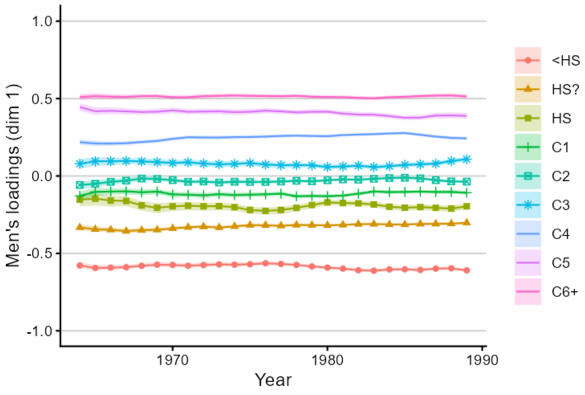
(d) Relative contributions (1992-2025)

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Nine educational groups for 1962-1991, eight for 1992-2025 (see Section 3.1 for details). Panels (a) and (b) show changes in the spectral and Frobenius norm of the double-centered surplus matrix, respectively until 1991 and starting from 1992. Panels (c) and (d) show the relative contribution of the first five sorting dimensions to assortative mating, as measured by ρ_k, defined in equation (2.15). The shaded areas represent 95% two-tailed confidence intervals.

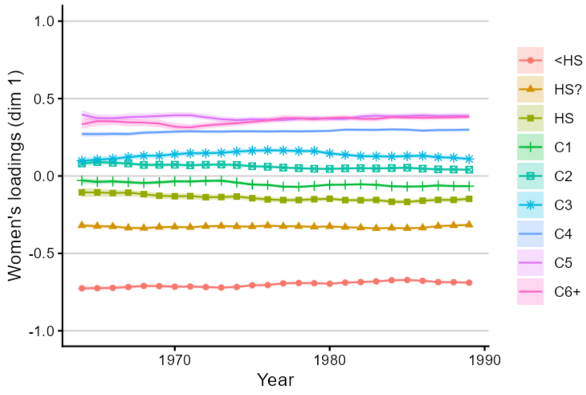
sorting along the main dimension relative to obtaining a high-school diploma (HS). On the other hand, obtaining a high-school diploma results in substantial returns from assortative mating if compared with dropping out (<HS). Similarly, obtaining a college degree (C4 before 1992, C after) also represents an important margin if compared with dropping out (C1, C2, or C3) or with obtaining an associate degree (AD). There also exist important returns from assortative mating for postgraduate degrees relative to a college degree, with larger returns for professional (PD) and doctoral degrees (PhD) and smaller but significant returns for master's degrees (M). Figure A10 conveniently shows that obtaining a high-school diploma (compared with dropping out) and obtaining a college degree (compared with obtaining an associate degree) are the most important margins along the first dimension.

Finally, we detect the presence of homogamous preferences also with this more detailed categorization. Interestingly, Figure A11 in the Online Appendix shows that high-school dropouts on one end and doctoral and professional degrees on the other have the strongest homogamous preferences of all groups and tend to form separate marriage market clusters.

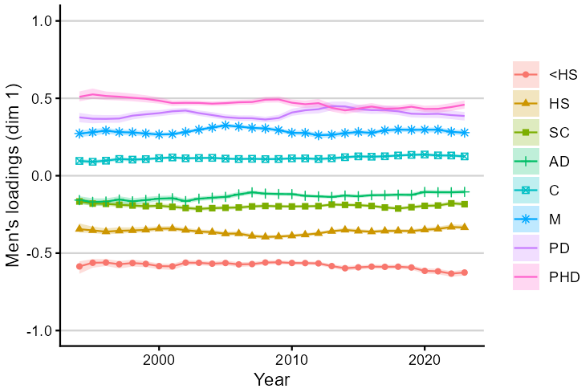
Figure 4: Changes in index composition over time - detailed categorization, first dimension



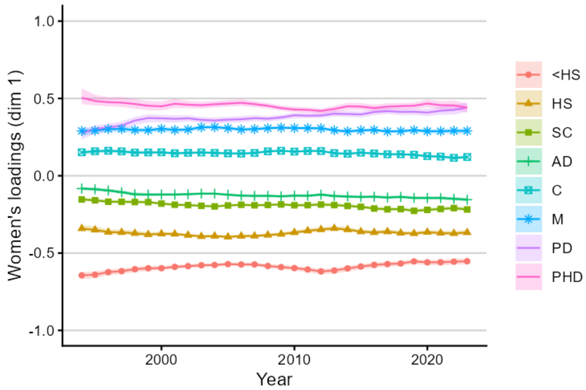
(a) Men, dim. 1: $\tilde{u}_{\cdot,1}$



(b) Women, dim. 1: $\tilde{v}_{\cdot,1}$



(c) Men, dim. 1: $\tilde{u}_{\cdot,1}$



(d) Women, dim. 1: $\tilde{v}_{\cdot,1}$

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Nine educational groups for 1962-1991, eight for 1992-2025 (see Section 3.1 for details). Panels (a) to (b) show how the elements of the first singular vectors changed over time, for men and women separately. The shaded areas represent 95% two-tailed confidence intervals.

6. Conclusion

In this paper, we provide new methods to analyze patterns of educational assortative mating and answer questions such as: are sorting patterns consistent with a single-index matching model? If not, what is the number of sorting dimensions that can rationalize the data patterns? How do we measure overall assortative mating under these two distinct scenarios? How do individuals with different diplomas and degrees rank along each dimension? Is there evidence that individuals display some taste for homogamous matches? These methods can be readily applied to study assortative mating in other dimensions and are particularly suited to analyze sorting patterns between (ordered or unordered) categorical traits also in other markets.

The paper contributes to a methodological literature recently reviewed by Chiappori et al. (2025) that explores and compares different metrics for assortative mating. Our approach differs from others in that we build aggregate measures of assortative mating starting from local odds ratios, while providing a structural interpretation for our measures based on the standard model of Choo and Siow (2006). In short, we recover the matrix of match surplus complementarities starting from local odds ratios, then we perform a Singular Value Decomposition in order to recover its singular values, interpretable as measures of assortative mating along different orthogonal dimensions, and singular vectors, interpretable as indices of attractiveness as in Dupuy and Galichon (2014). In practice, the rank of the match surplus complementarities is revealing of the number of relevant sorting dimensions, while different matrix norms can be used as aggregate measures of assortative mating.

Using CPS data for the 1962-2025 period, we show that educational sorting patterns are broadly consistent with a single-index model where individuals are ranked based on their educational attainment, from the least to the most educated. However, sorting on higher-order dimensions reveals the existence of homogamous preferences, which are practically important to explain the high prevalence of same-education couples in the data. Our findings show that educational assortative mating has increased in the long run, but its rise has substantially slowed down over the last three decades, confirming previous findings by Eika et al. (2019). Moreover, as previously noted by Schwartz and Mare (2005) and Gihleb and Lang (2020), changes in the coding of the educational variable in 1992 complicate the comparison of earlier and later CPS waves. Yet, our findings also reveal more nuanced patterns, such as the increased isolation of high-school dropouts on the marriage market and the increasing affinity between college degrees and postgraduates, while gender differences in the value of different educational attainments have progressively vanished.

References

Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier, 2011.

Frederik Almar and Bastian Schulz. Optimal weights for marital sorting measures. *Economics Letters*, 234:111497, 2024.

Patricia ME Altham. The measurement of association of rows and columns for an r× s contingency table. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(1):63–73, 1970.

Gary S. Becker. A theory of marriage: Part i. *Journal of Political Economy*, 81(4): 813–846, 1973.

Gary S. Becker. A theory of marriage: Part ii. *Journal of Political Economy*, 82(2): S11–S26, 1974.

Anton Cheremukhin, Paulina Restrepo-Echavarria, and Antonella Tutino. Marriage market sorting in the us. 2024.

Pierre-André Chiappori, Sonia Oreffice, and Climent Quintana-Domeque. Fatter attraction: Anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy*, 120(4):659–695, 2012.

Pierre-André Chiappori, Bernard Salanié, and Yoram Weiss. Partner choice, investment in children, and the marital college premium. *American Economic Review*, 107(8): 2109–67, 2017.

Pierre-André Chiappori, Sonia Oreffice, and Climent Quintana-Domeque. Bidimensional matching with heterogeneous preferences: education and smoking in the marriage market. *Journal of the European Economic Association*, 16(1):161–198, 2018.

Pierre-André Chiappori, Edoardo Ciscato, and Carla Guerriero. Analyzing matching patterns in marriage: Theory and application to italian data. *Quantitative Economics*, 15(3):737–781, 2024.

Pierre-André Chiappori, Monica Costa Dias, Costas Meghir, and Hanzhe Zhang. Changes in marital sorting: Theory and evidence from the us. *Journal of Political Economy*, 2025.

Eugene Choo and Aloysius Siow. Who marries whom and why. *Journal of Political Economy*, 114(1):175–201, 2006.

Arnaud Dupuy and Alfred Galichon. Personality traits and the marriage market. *Journal of Political Economy*, 122(6):1271–1319, 2014.

Anthony WF Edwards. The measure of association in a 2× 2 table. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 126(1):109–114, 1963.

Lasse Eika, Magne Mogstad, and Basit Zafar. Educational assortative mating and household income inequality. *Journal of Political Economy*, 127(6):2795–2835, 2019.

Raquel Fernández and Richard Rogerson. Sorting and long-run inequality. *Quarterly Journal of Economics*, 116(4):1305–1341, 2001.

Alfred Galichon and Bernard Salanié. Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629, 2022.

Rania Gihleb and Kevin Lang. Educational homogamy and assortative mating have not increased. In *Change at home, in the labor market, and on the job*, volume 48, pages 1–26. Emerald Publishing Limited, 2020.

Bryan S Graham. Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers. *Handbook of social economics*, 1: 965–1052, 2011.

Jeremy Greenwood, Nezih Guner, Georgi Kocharkov, and Cezar Santos. Marry your like: Assortative mating and income inequality. *American Economic Review*, 104(5): 348–353, 2014.

Cristina Gualdani and Shruti Sinha. Partial identification in matching models for the marriage market. *Journal of Political Economy*, 131(5), 2023.

Noah Hirschl, Christine R Schwartz, and Elia Boschetti. Eight decades of educational assortative mating: A research note. *Demography*, 61(5):1293–1307, 2024.

Gunter J Hitsch, Ali Hortaçsu, and Dan Ariely. Matching and sorting in online dating. *American Economic Review*, 100(1):130–63, 2010.

Ian T Jolliffe. *Principal Component Analysis, Second Edition.* Springer Science & Business Media, 2002.

Lars Kirkebøen, Edwin Leuven, and Magne Mogstad. College as a marriage market. Working paper 28688, National Bureau of Economic Research, 2021.

Frank Kleibergen and Richard Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126, 2006.

Robert D Mare. Five decades of educational assortative mating. *American Sociological Review*, pages 15–32, 1991.

Luis Milan and Joe Whittaker. Application of the parametric bootstrap to models that incorporate a singular value decomposition. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(1):31–49, 1995.

Frederick Mosteller. Association and estimation in contingency tables. *Journal of the American Statistical Association*, 63(321):1–28, 1968.

Helena Skyt Nielsen and Michael Svarer. Educational homogamy how much is opportunities? *Journal of Human Resources*, 44(4):1066–1086, 2009.

Christine R Schwartz and Robert D Mare. Trends in educational assortative marriage from 1940 to 2003. *Demography*, 42(4):621–646, 2005.

Robert Shimer and Lones Smith. Assortative matching and search. *Econometrica*, 68(2): 343–369, 2000.

Aloysius Siow. Testing becker's theory of positive assortative matching. *Journal of Labor Economics*, 33(2):409–441, 2015.

Robert J Tibshirani and Bradley Efron. An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57(1):1–436, 1993.

A. Supplementary tables

Table A1: SVD representation of a gender-asymmetric mixed model

	Dim. 1		Dim. 2		Dim. 3		Dim. 4		Dim. 5	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Type 1	-0.44	0.00	-0.34	-0.00	-0.28	-0.00	0.17	0.00	0.76	1.00
Type 2	-0.06	0.35	0.27	0.66	0.66	0.51	-0.54	-0.42	0.45	0.00
Type 3	0.24	0.51	0.71	0.45	-0.01	-0.47	0.58	0.56	0.32	0.00
Type 4	0.49	0.57	0.16	-0.36	-0.63	-0.43	-0.53	-0.60	0.25	0.00
Type 5	0.71	0.54	-0.53	-0.48	0.32	0.57	0.24	0.38	0.24	0.00
λ_k	1.12	1.12	0.20	0.20	0.16	0.16	0.13	0.13	0.00	0.00

Notes. The table reports the singular vectors $u_{\cdot,k}$ and singular values λ_k for the gender-asymmetric match surplus function $\Phi_4(x, y) = 0.5\Phi_1(x, y) + 0.5\delta\mathbb{1}\{y > x\}$ with $\delta = -0.5$ and $E = 5$, while Φ_1 is defined in (2.8). The penalty δ for couples where the wife is more educated than the husband introduces an asymmetry in the match surplus.

Table A2: Changes in educational sorting if dimension k were suppressed

	<HS	HS	SC	C	C+
<HS	-0.020	0.016	0.004	0.000	-0.000
HS	0.013	-0.012	-0.002	-0.000	0.000
SC	0.006	-0.003	-0.004	-0.000	0.002
C	0.002	-0.002	-0.003	-0.000	0.003
C+	-0.000	0.001	0.004	0.001	-0.005

(a) $\tilde{\lambda}_2 = 0$ (1962-1991)

	<HS	HS	SC	C	C+
<HS	-0.012	0.008	0.004	-0.000	-0.000
HS	0.008	-0.008	-0.005	0.001	0.003
SC	0.004	-0.004	-0.011	0.004	0.008
C	0.000	0.001	0.002	-0.004	0.001
C+	-0.000	0.003	0.010	-0.000	-0.012

(b) $\tilde{\lambda}_2 = 0$ (1992-2025)

	<HS	HS	SC	C	C+
<HS	-0.011	0.011	-0.000	-0.000	0.000
HS	0.010	-0.015	0.002	0.004	0.000
SC	0.001	-0.002	-0.001	0.002	-0.000
C	-0.000	0.007	0.000	-0.007	0.000
C+	0.000	-0.001	-0.001	0.002	-0.001

(c) $\tilde{\lambda}_3 = 0$ (1962-1991)

	<HS	HS	SC	C	C+
<HS	-0.001	0.001	0.000	-0.000	0.000
HS	0.001	-0.003	0.000	0.003	-0.000
SC	-0.000	0.001	-0.005	0.005	-0.001
C	0.000	0.002	0.005	-0.015	0.007
C+	-0.000	-0.000	-0.001	0.007	-0.006

(d) $\tilde{\lambda}_3 = 0$ (1992-2025)

	<HS	HS	SC	C	C+
<HS	-0.002	0.003	-0.000	0.000	0.000
HS	0.003	-0.010	0.007	-0.000	0.000
SC	-0.001	0.009	-0.010	0.002	-0.000
C	-0.000	-0.002	0.003	-0.002	0.000
C+	-0.000	-0.000	-0.000	0.000	-0.000

(e) $\tilde{\lambda}_4 = 0$ (1962-1991)

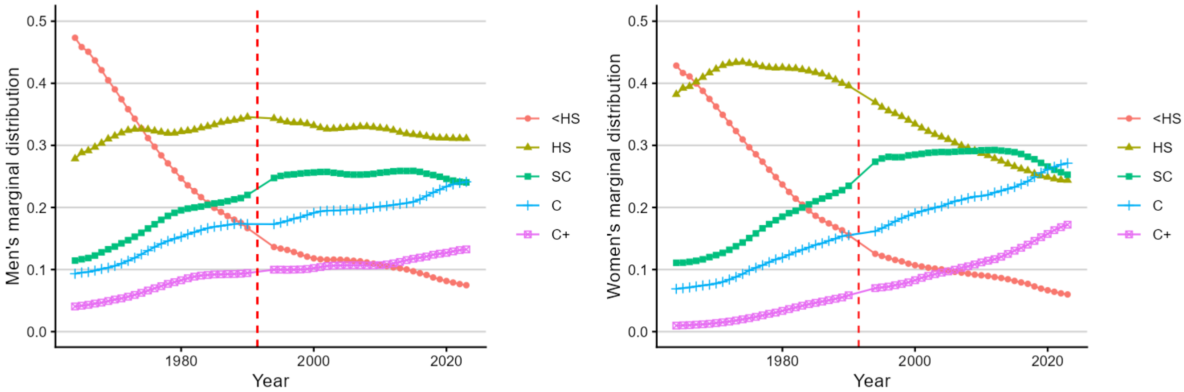
	<HS	HS	SC	C	C+
<HS	-0.002	0.003	-0.001	-0.000	-0.000
HS	0.002	-0.021	0.019	0.001	-0.000
SC	-0.000	0.018	-0.020	0.001	0.001
C	0.000	0.001	0.001	-0.002	0.000
C+	0.000	0.000	0.000	0.000	-0.001

(f) $\tilde{\lambda}_4 = 0$ (1992-2025)

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Each panel shows changes in relative frequencies if we calculated a counterfactual assignment after suppressing sorting along a particular dimension k by setting $\tilde{\lambda}_k = 0$. The counterfactual frequencies are obtained for every five-year interval $[t, t+4]$, using the estimated $\tilde{U}$, $\tilde{V}$, and $\tilde{\Lambda}$ to calculate the counterfactual assignment, and then averaged over the periods 1962-1991 and 1992-2025 separately. We divide the results into two periods since Figures 2 and A9 suggest that changes in coding in the educational variable lead to changes in the index composition.

B. Supplementary figures

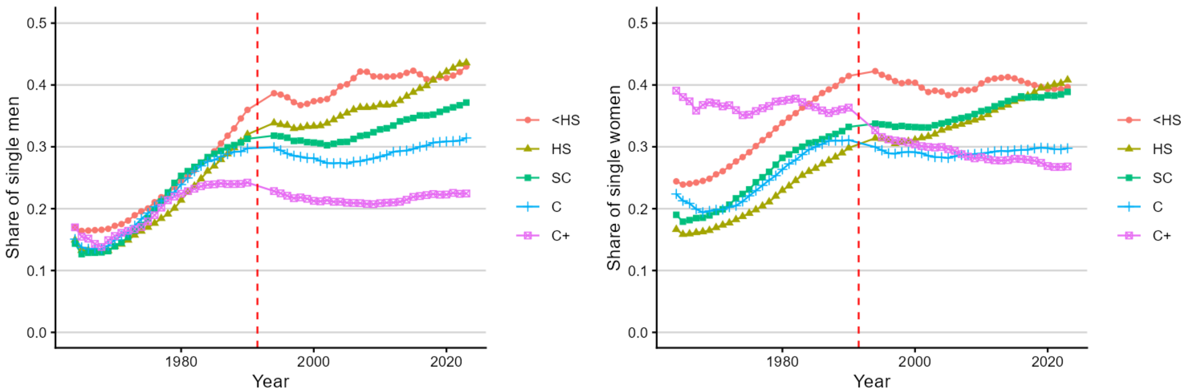
Figure A1: Marginal distribution of education by gender



(a) Men (b) Women

Notes. 1962-2025 CPS data, individuals aged between 26 and 60. The educational groups are: high-school dropouts (<HS), high-school diplomas (HS), some college (SC), college degrees (C), and postgraduate degrees (C+).

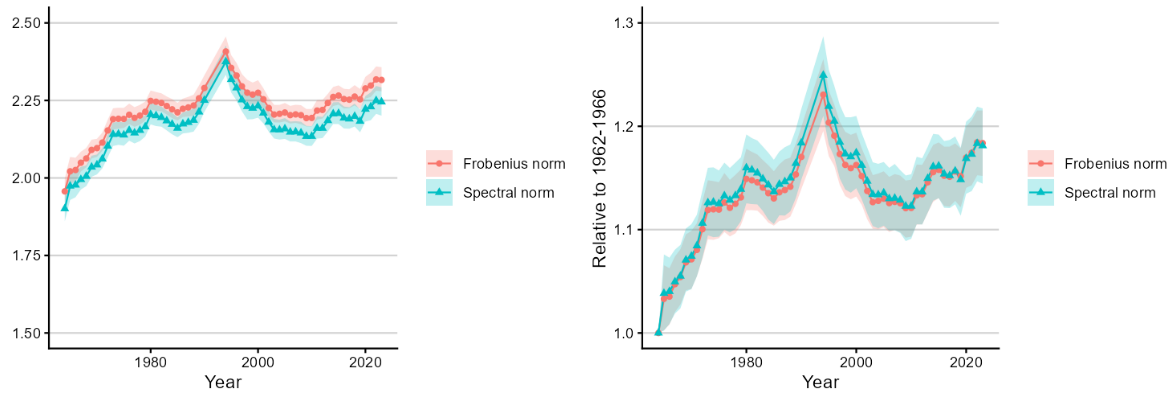
Figure A2: Share of singles by gender and educational group



(a) Men (b) Women

Notes. 1962-2025 CPS data, individuals aged between 26 and 60. Singles are defined as individuals living without a partner, regardless of their marital status.

Figure A3: Changes in assortative mating over time - head & spouse only

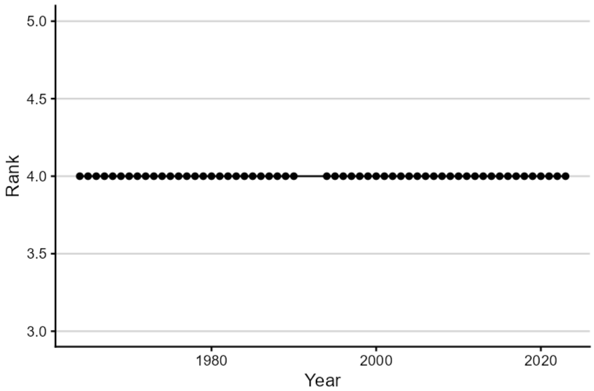


(a) Norms: $\|\tilde{P}\|_2$ and $\|\tilde{P}\|_F$

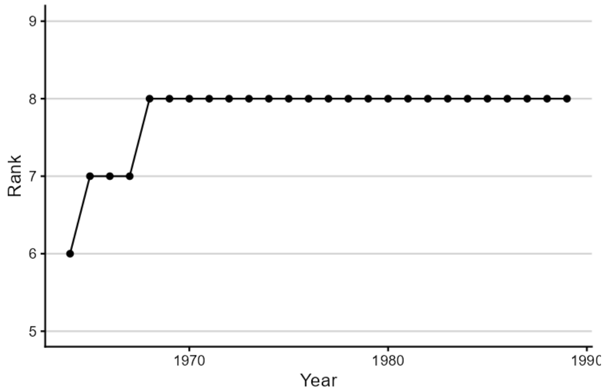
(b) Norms relative to 1962-1966

Notes. 1962-2025 CPS data, legally married couples with at least one partner aged between 26 and 60. One of the two partners must be the household head. Five educational groups. Panel (a) shows the spectral and Frobenius norm of the double-centered surplus matrix $\{\tilde{P}_{t,t+4}\}_{t=1962}^{2021}$. Panel (b) shows how these norms changed relative to their values in the 1962-1966 period, in which they are normalized to one. The shaded areas represent 95% two-tailed confidence intervals.

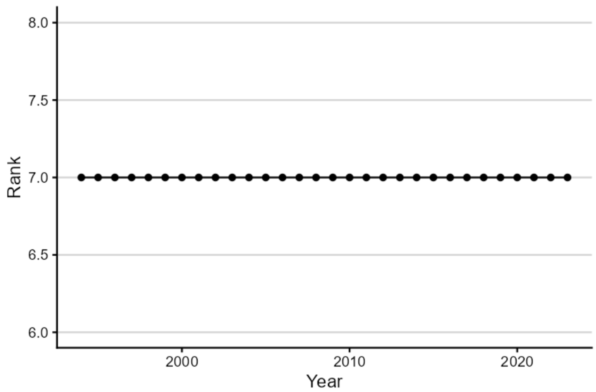
Figure A4: Rank of $\tilde{P}_{t,t'}$



(a) $E = 5$



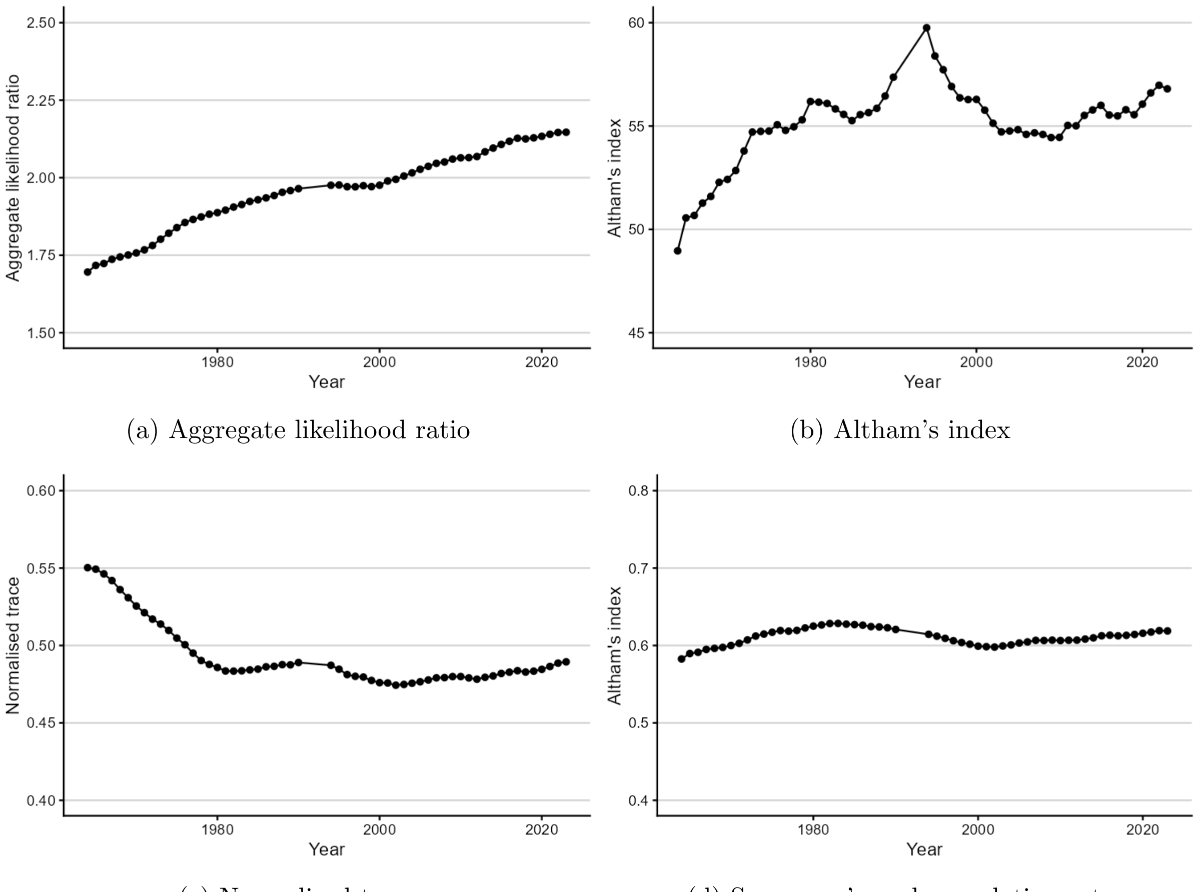
(b) $E = 9$



(c) $E = 8$

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. For different educational categorizations, we present the estimated rank of the match surplus matrix $\tilde{P}_{t,t'}$. In practice, we iteratively test the hypothesis $\text{rank}(\tilde{P}_{t,t'}) = k$ with $k = 1, 2, ...E - 1$. We conclude that the rank of $\tilde{P}_{t,t'}$ is equal to k if the corresponding p-value is higher than 0.01.

Figure A5: Alternative measures of assortative mating



(a) Aggregate likelihood ratio

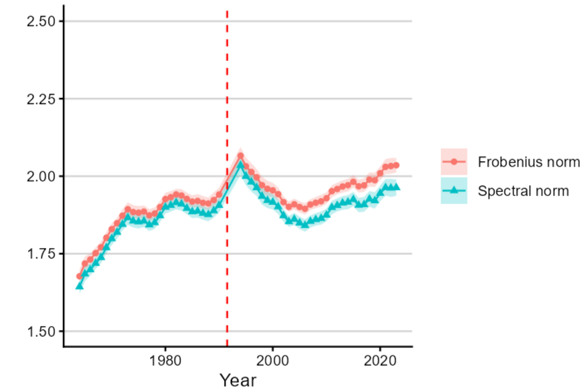
(b) Altham's index

(c) Normalized trace

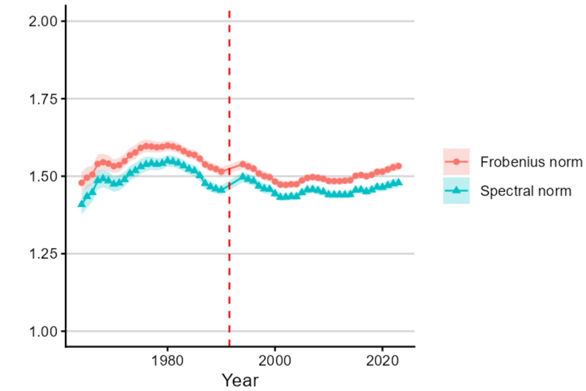
(d) Spearman's rank correlation rate

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Five educational groups. The shaded areas represent 95% two-tailed confidence intervals.

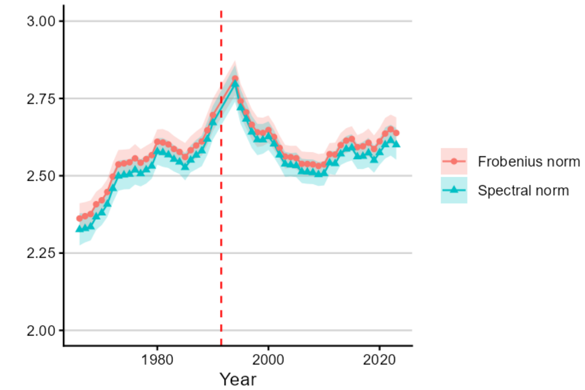
Figure A6: Changes in assortative mating over time - coarse categorizations



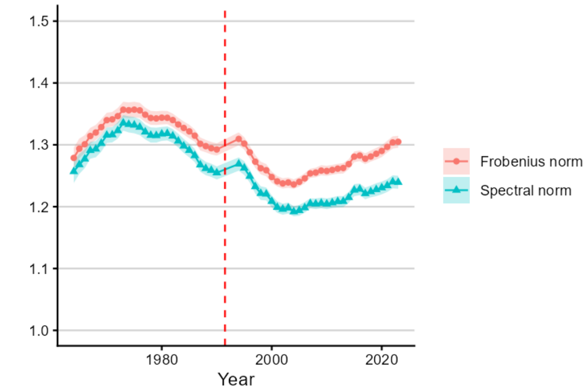
(a) C and C+ merged ($E = 4$)



(b) <HS and HS merged ($E = 4$)



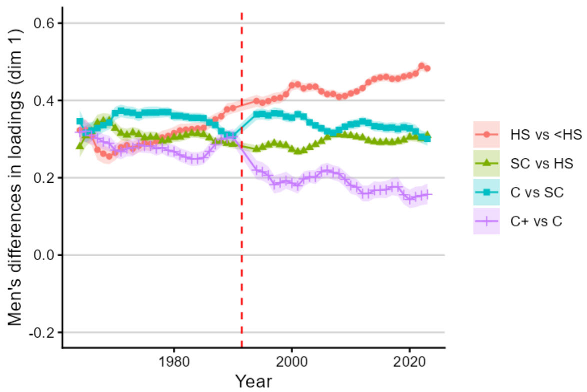
(c) HS and SC merged ($E = 4$)



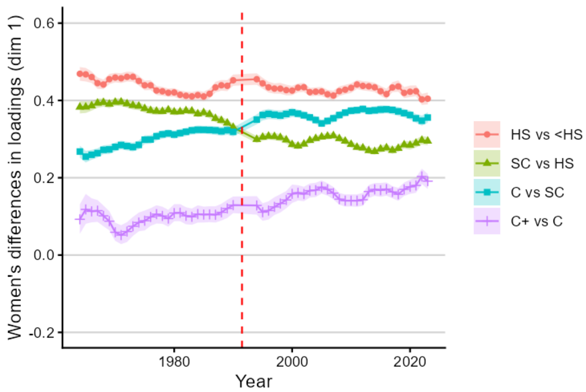
(d) <HS and HS merged,
C and C+ merged ($E = 3$)

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Three or four educational groups. The shaded areas represent 95% two-tailed confidence intervals.

Figure A7: Differences $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$ and $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$ over time



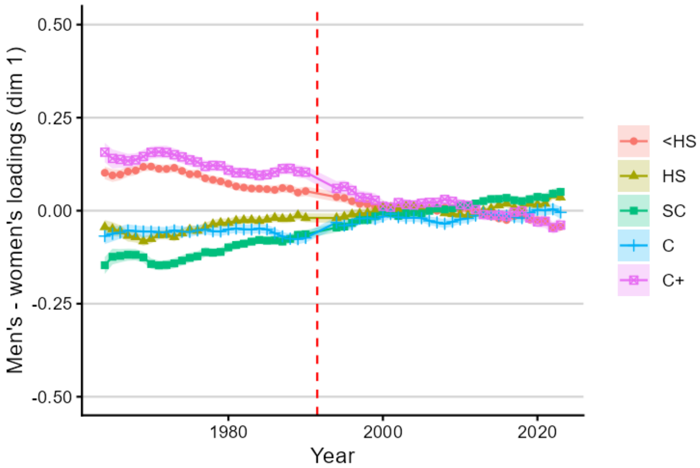
(a) Men, dim. 1: $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$



(b) Women, dim. 1: $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$

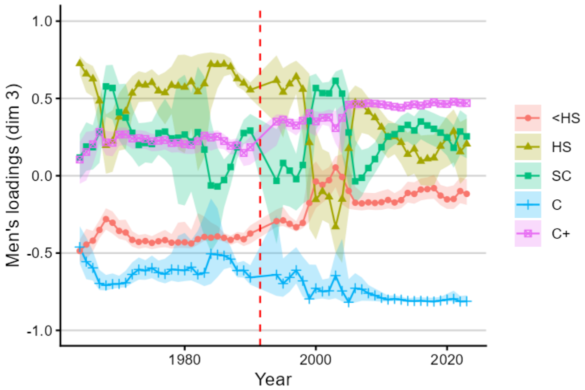
Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Five educational groups. The figure shows the differences $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$ and $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$ for the first male and female indices, which can also be seen in panel (a) and (b) of Figure 2. These differences are always positive, which means that higher attainments are associated with positive returns. The educational groups are: high-school dropouts (<HS), high-school diplomas (HS), some college (SC), college degrees (C), and postgraduate degrees (C+). The shaded areas represent 95% two-tailed confidence intervals.

Figure A8: Differences $\tilde{u}_{\cdot,1} - \tilde{v}_{\cdot,1}$ over time

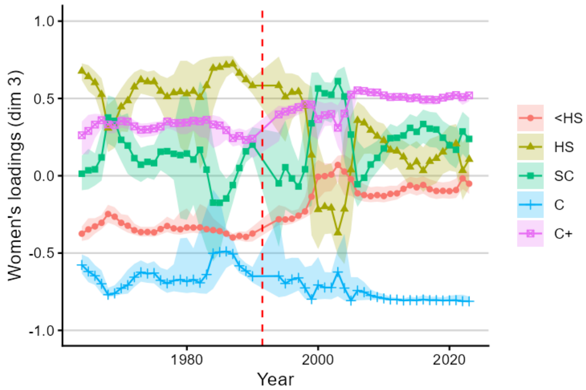


Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Five educational groups. The figure shows the difference between the first male and female indices, which can also be seen in panel (a) and (b) of Figure 2. The figure shows that differences in the first indices have vanished over time. This implies that, along the first dimension, the gains from sorting have become gender symmetric. The educational groups are: high-school dropouts (<HS), high-school diplomas (HS), some college (SC), college degrees (C), and postgraduate degrees (C+). The shaded areas represent 95% two-tailed confidence intervals.

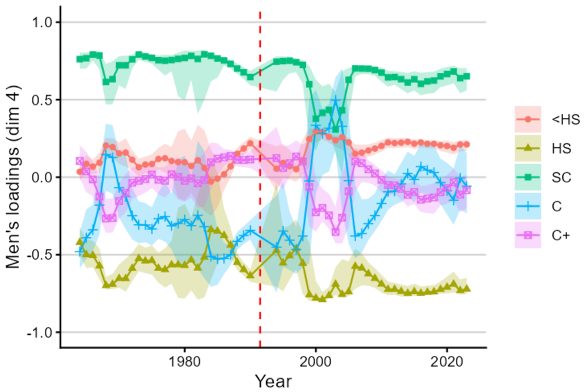
Figure A9: Changes in index composition over time, third and fourth dimensions



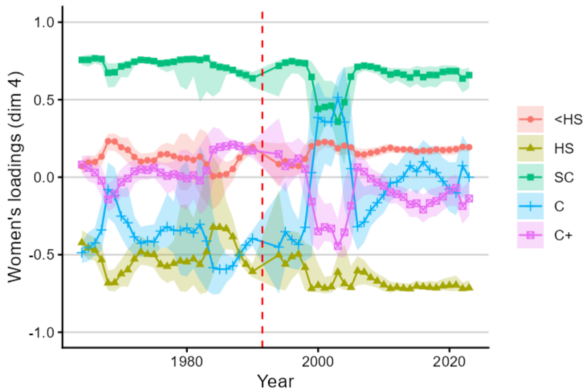
(a) Men, dim. 3: $\tilde{u}_{\cdot,3}$



(b) Women, dim. 3: $\tilde{v}_{\cdot,3}$



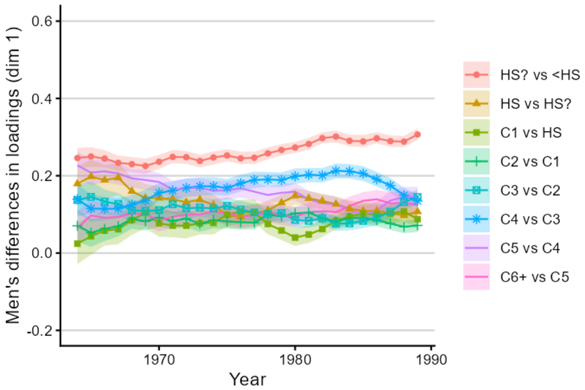
(c) Men, dim. 4: $\tilde{u}_{\cdot,4}$



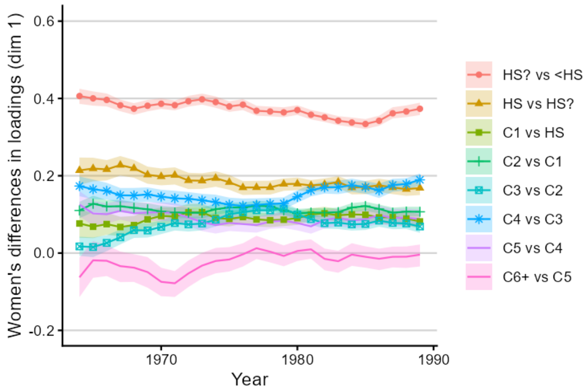
(d) Women, dim. 4: $\tilde{v}_{\cdot,4}$

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Five educational groups. The different figures show how the elements of the third and fourth singular vectors changed over time, for men and women separately. Figure 2 shows the first two singular vectors. The educational groups are: high-school dropouts (<HS), high-school diplomas (HS), some college (SC), college degrees (C), and postgraduate degrees (C+). The shaded areas represent 95% two-tailed confidence intervals.

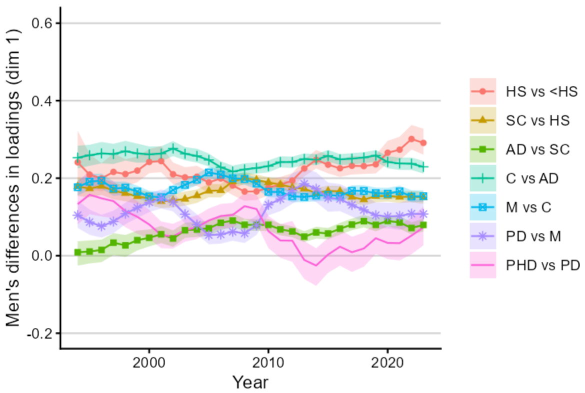
Figure A10: Differences $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$ and $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$ over time



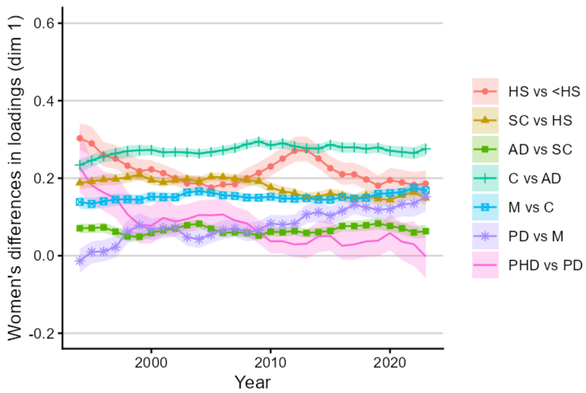
(a) Men, dim. 1: $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$



(b) Women, dim. 1: $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$



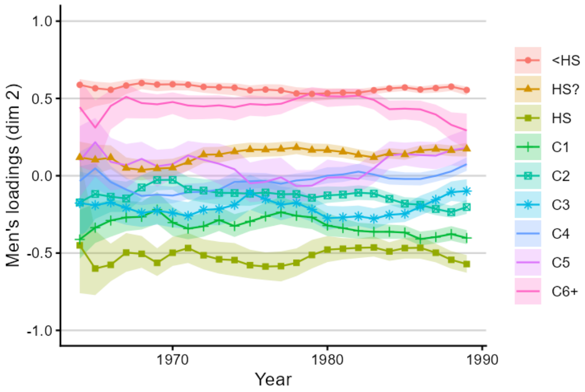
(c) Men, dim. 1: $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$



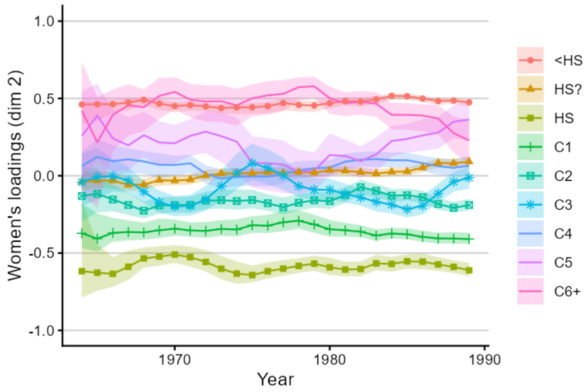
(d) Women, dim. 1: $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Nine educational groups for 1962-1991, eight for 1992-2025 (see Section 3.1 for details). The figure shows the differences $\tilde{u}_{k+1,1} - \tilde{u}_{k,1}$ and $\tilde{v}_{k+1,1} - \tilde{v}_{k,1}$ for the first male and female indices, which can also be seen in panel (a) and (b) of Figure 2. The larger these differences, the greater the returns associated with obtaining an additional diploma or degree. The shaded areas represent 95% two-tailed confidence intervals.

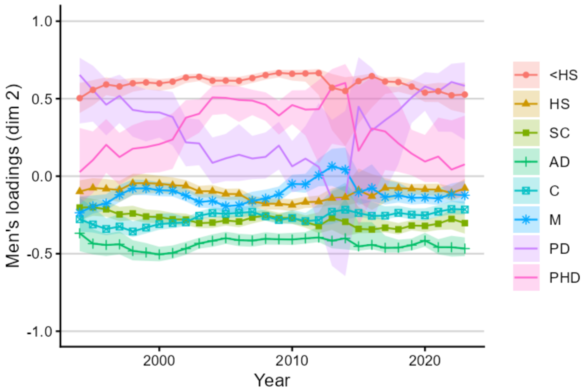
Figure A11: Changes in index composition over time - detailed categorization, second dimension



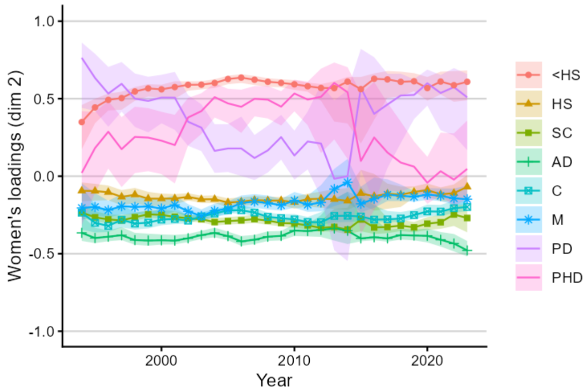
(a) Men, dim. 2: $\tilde{u}_{\cdot,2}$



(b) Women, dim. 2: $\tilde{v}_{\cdot,2}$



(c) Men, dim. 2: $\tilde{u}_{\cdot,2}$



(d) Women, dim. 2: $\tilde{v}_{\cdot,2}$

Notes. 1962-2025 CPS data, couples with at least one partner aged between 26 and 60. Nine educational groups for 1962-1991, eight for 1992-2025 (see Section 3.1 for details). Panels (a) to (b) show how the elements of the second singular vectors changed over time, for men and women separately, while Figure 4 shows the elements of the first singular vectors. The shaded areas represent 95% two-tailed confidence intervals.