Federal Reserve Bank of Dallas

Marriage Market Sorting in the U.S.

Anton Cheremukhin, Paulina Restrepo-Echavarria and Antonella Tutino

Working Paper 2406 **September 2024**

Research Department

https://doi.org/10.24149/wp2406

Marriage Market Sorting in the U.S.*

Anton Cheremukhin†, Paulina Restrepo-Echavarria‡ and Antonella Tutino§

September 16, 2024

Abstract

We examine shifts in the U.S. marriage market, assessing how online dating, demographic changes, and evolving societal norms influence mate choice and broader sorting trends. Using a targeted search model, we analyze mate selection based on factors such as education, age, race, income, and skill. Intriguingly, despite the rise of online dating, preferences, mate choice, and overall sorting patterns showed negligible change from 2008 to 2021. However, a longer historical view from 1960 to 2020 reveals a trend toward preferences for similarity, particularly concerning income, education, and skills. Our findings refute two out of three potential explanations – reduced search costs and growing spatial segregation – as potential causes of these long-term shifts. In particular, we conclude that people's capacity to process and evaluate information hasn't improved despite technological advancements. Among the remaining demographic factors, we identify enhanced workforce participation and college attainment among women as the primary drivers of the U.S. marriage market transformation. Furthermore, we find that the corresponding changes in mate preferences and increased assortativeness by skill and education over this timeframe account for about half of the increased income inequality among households.

Keywords: Marriage, Sorting, Online dating

JEL Classification: J12, D83, C78

* This work was supported by computational resources provided by the Big Tex High Performance Computing Group at the Federal Reserve Bank of Dallas. The views expressed in this paper are our own and do not necessarily reflect the views of the Federal Reserve Banks of Dallas or St. Louis or the Federal Reserve System.

†Anton Cheremukhin, Federal Reserve Bank of Dallas, 2200 N. Pearl St., Dallas, TX 75201, chertosha@gmail.com.

‡Paulina Restrepo-Echavarria, Federal Reserve Bank of St. Louis, One Federal Reserve Bank Plaza, Broadway and Locust St., St. Louis, MO 63166, paulinares@me.com.

§Antonella Tutino, Haverford College, tutino.antonella@gmail.com.

1 Introduction

To understand who married whom and why, we need to understand the dating process; how people search for love and potential marriage partners. Do people prefer someone like them or different? How much effort is placed into finding their better half? How strong are people's preferences for certain characteristics? Where do people search for a partner?

As chronicled by Rosenfeld et al. (2019), the percentage of married couples who met online went from only 2% in 1998, to 20% in 2008, to nearly 50% in 2017, becoming the dominant form of initial contact between married couples. Nothing has transformed the dating process more than the proliferation of online dating. Furthermore, research shows that a prevalent feature of online dating is that people target potential partners who occupy higher positions on the perceived desirability scale. In other words, people would ideally like a partner who is "above their league."[1]

This finding represents a complication for the field of search theory as existing models of the marriage market fail to incorporate the dating process with its nuances brought about by online dating. Existing models posit that individuals seek partners of similar rank or with similar characteristics as them, and they fail to replicate the observed "shooting above their league" behavior that is observed in online dating in equilibrium. Furthermore, most existing theoretical models can't explain how mate selection choices affect aggregate sorting patterns. This limitation extends to both models of individual marital choice and those addressing aggregate marriage market sorting, which often do not specify individual selections.[2]

In Cheremukhin et al. (2020), we developed a model of targeted search that endogenizes the selection of marriage partners and is consistent with (1) patterns observed in online data and (2) implications of mate selection choices on aggregate marriage outcomes, which allows us to uncover the patterns of mate selection and understand the underlying causes of observed sorting using only aggregate matching data. In this paper, we use the above-mentioned targeted search model to analyze the U.S. mar-

[1] See Bruch and Newman (2018) for a particularly neat measurement of this pattern and Hitsch et al. (2010a), Finkel et al. (2012), Bruch and Newman (2018), and Dinh et al. (2022) for an overview of online dating behavior.

[2] For an overview of models of search and matching models see Chade et al. (2017). For an overview of models used to study marriage markets see Chiappori (2020).

riage market and its evolution over time. We do this by estimating people's preferences and developing new tools to assess their degree of horizontality (to what extent people prefer someone like themselves), the degree of assortativeness of matching outcomes, and how selective people are when searching for a potential partner. We also explore the degree of complementarity or substitutability of different attributes (which we refer to as affinity), and analyze how income inequality has been affected by the degree of selectivity of people. We ask if search costs have gone down over time due to the innovation in meeting technologies brought about by the increased popularity of online dating websites.[3]

We use the American Community Survey for the 2008-2021 period, and for longer-term comparisons we employ data from the U.S. Census for the years 1960 and 1980. Using the yearly data for the 2008-2021 period allows us to study the effects of the proliferation of online dating.

In the model, selectivity (lopsidedness of the probability distribution toward a target) and the resulting aggregate sorting are dictated by the configuration of preferences. If preferences are horizontal, implying an inclination toward similarity,[4] then partners face limited competition and they can be very selective, which results in strongly assortative aggregate matching outcomes. In contrast, if preferences are vertical, signifying a commonly held hierarchy of potential partners, then the majority of market participants find themselves vying for the top-tier mate types, obligating all participants to hedge their bets. Vertical preferences make individuals less selective (because they have to be more strategic) and yield mixed sorting outcomes wherein the most desirable types experience an elevated matching rate from all potential partners, and the least desirable types scarcely secure any matches.

We find pronounced horizontal preferences with respect to race, moderate horizontal tendencies for education and age, and considerable vertical preferences with respect to income and skill. Furthermore, we find minimal change in these results over the 2008-2021 period, but when we include 1960 and 1980, our findings indicate a trend toward an increased degree of horizontality in preferences, predominantly with respect to income

[3]We restrict our attention to traditional marriages. We also focus on preferences, selection, and sorting by education, age, race, income and skill, as physical attributes reflecting looks and healthiness are unavailable in our data set.

[4]More generally, preferences are horizontal if types of partners on two sides can be combined into pairs of mutual best matches.

and skill.

The estimated preference structure for the 2008-2021 period produces considerable assortativeness by race, education, and age while evoking more attenuated sorting by income and skill (consistent with the literature). The increase in horizontality that we observe in the long-run is accompanied by an increase in aggregate assortativeness, with the largest increases observed for assortativeness by income and skill.

In terms of selection we find that people are most selective on race followed by a large difference by education and age. Also, people are more selective on the interaction of income and skill than on the two characteristics individually. During the 2008-2021 period both males and females became slightly more selective on age. When we include the results for 1960 and 1980, it's clear that both males and females became less selective on race and more selective on income and skill.

Our model is particularly well suited to study how marriage market outcomes affect inequality because, unlike existing models, it allows us to distinguish interactions of preferences for income, education, and skill from their correlational structure in the population. In other words, the model allows us to fully jointly identify preferences toward attributes while taking into account the frequencies and correlations of attributes in the population. This has been a long-standing issue in the literature for which we provide comprehensive resolution.

We find that the observed increase in assortativeness accounts for approximately half of the increase in household income inequality between 1980 and 2020.[5] The remainder of this disparity is attributable to changes in the socioeconomic composition of prospective brides and grooms and would have occurred even if mates were chosen randomly. The most important factors contributing to household income inequality through mate selection are selection on education (35%) and skill (30%), with selection on income (15%) and age (15%) trailing significantly, while selection by race (5%) plays a relatively inconsequential role. The collective influence of partner selection on household income inequality is substantial, leading to a 14-point increase in the coefficient of variation, or a 3-point increase in the Gini coefficient.

Finally, the fact that we find minimal change in preferences, selectivity, and aggregate sorting over the 2008-2021 period is surprising because given the proliferation of online dating we would expect to observe a substantial improvement in the ability to

[5] Among households that contain a married couple.

find and meet potential partners, reflecting a reduction in search costs brought about by this improved search technology. Assuming a reduction in search costs in the model while maintaining preferences would instigate a higher selectivity of partner choice, and a larger increase in the assortativeness of overall matching. We show that the data contradicts these predictions and indicates that search costs have not changed over time despite the wide use of online dating sites.

In addressing the drivers of these medium-to long-term changes, we evaluate and reject two potential explanations: diminished search costs and increased spatial segregation. First, we find that the reduction in search costs cannot be the explanation of these medium-to long-term changes. The recent literature[6] documents a similar lack of improvement in matching efficiency in labor and product markets despite substantial apparent enhancement in search technologies and associated reduction in physical search costs. Martellini and Menzio (2020) propose that increasing selectivity in search could compensate for reduced search costs thereby resolving this conundrum. However, within the context of the marriage market we observe no decrease in search costs and no discernible increase in selectivity since the 1960s. We offer an alternative resolution for this puzzle: effective search costs reflect individuals' capabilities to process and evaluate information, which subsequently determines their proficiency in mate selection within the marriage market. Despite technological progress, individual capabilities to process information remain unaltered, accounting for the apparent lack of enhancements in matching efficiency and selectivity.

The second potential driver of long-term changes, based on the growing literature on spatial segregation,[7] encompasses a broad range of theories capturing the fact that individuals today face a biased set of possibilities skewed toward encountering similar partners, stemming from shared educational experiences, professional environments, or residential proximity.[8] To encapsulate this biased capacity to locate analogous partners,

[6]Martellini and Menzio (2020) documented that the unemployment and vacancy rates in the labor markets have not declined much over the past century, while Kaplan and Menzio (2015) have shown that dispersion of prices for consumer products has not declined over the past half a century. An explanation proposed by Menzio (2023) is that the decline in search frictions has been undone by the endogenous rise in selectivity of workers and firms, consumers and producers respectively.

[7]For a good overview of trends in spatial segregation in the U.S. see Moretti (2011), Moretti (2012), and Bischoff and Reardon (2014). Also see more recent work on sorting across neighborhoods by Diamond and Gaubert (2022), and Couture and Handbury (2022). On the effect of spatial segregation on local marriage markets see Fan and Zou (2021).

[8]This explanation includes the mechanism described by Kalmijn (1998) and Hitsch et al. (2010a)

we incorporate skewed prior knowledge into our model. Our analyses indicate that this inclusion results in negligible modifications in our estimates, with the principal outcome being a reduction in estimated horizontality – a direct contradiction to our long-term estimates of preferences.

The remaining set of explanations are the demographic factors that have not only influenced the composition of available partners – an aspect accounted for in our model—but also potentially altered the preference structure. To understand the influence of this broad factor on our estimates and point out specific mechanisms at play, we conduct a separate examination of the selection patterns of females and males of different levels of income, skill, and education, noting significant initial disparities between the two, which are slowly diminished over time. We point out a significant demographic trend that aligns with many of our observations and is capable of explaining the observed changes—the rise in female labor force participation and education attainment.

We observe that females typically pursue males with a combination of higher income and skill levels, while predominantly opting for partners within their own educational bracket.[9] In contrast, we find that a substantial proportion of males select partners with the lowest education, skill, and income levels, implying a preference for spouses who assume domestic roles—a pattern much more prevalent in the years 1960 and 1980.

Thus, our findings demonstrate a temporal decrease in the number of males opting for partners inclined toward domestic roles. This shift is explained by the increased availability and preference for females with higher education, income, and skill. Additionally, we find that while females in the past selected partners with the highest education, they have in recent years increasingly chosen partners with similar education levels. As both of these alterations in selection behaviors align nicely with the increased horizontality of preferences that we document, we conclude that the rise in female labor force participation and college attainment is the likely culprit.

The paper proceeds as follows. Section 2 distills the related literature, Section 3 summarizes the theory used in the empirical part. Section 4 describes data and Section

that the improved ability to find partners while in school or college affects observed sorting in the marriage market.

[9]Females strongly factor in potential partners' current income and skill levels (measured as the average wage in the current occupation, and serving as a proxy for anticipated future income) into their vertical ranking of potential partners. This behavior aligns with microeconomic models of marital choice, which postulate a preference for higher expected lifetime income, see Ermisch (2016), Chiappori et al. (2020), and Altonji et al. (2023) among others.

5 the methodology, and Section 6 documents the empirical results. Section 7 discusses the effects of sorting on income inequality and welfare. Section 8 concludes.

2 Related Literature

Our paper contributes to the research on multidimensional matching in the marriage market, building upon previous work as described in Chiappori (2020). Initial efforts in this field, such as Choo and Siow (2006), studied frictionless matching with unobserved characteristics, but their model could not be tested due to the assumption of preference separability and restrictions on the distributions of preferences. To address this, Dupuy and Galichon (2014) developed a framework incorporating a quadratic surplus and normal distribution. They used a Dutch survey of households, assessing variables such as education, height, BMI, health, attitude toward risk, and personality traits of the spouses to estimate an affinity matrix that defines the quadratic surplus. Their findings underscore the multidimensionality of spouse selection with trade-offs between attributes and the differing preferences among men and women. While our work confirms these observations, we do not impose restrictions on the surplus shape and distribution, allowing for a more comprehensive estimation of preferences.

Building on this framework, Ciscato and Weber (2020) studied the changing gains from marriage in the United States from 1964 to 2017 using Current Population Survey data. They find that the significance of education has grown while that of age has declined, confirming the findings of Chiappori et al. (2017). They also noted a slight increase in racial segregation in the 1960s to the 1970s followed by a reversal in recent years. Chiappori et al. (2020) observed a similar rise in assortative matching based on education. Our study aligns with these findings, and analysis confirms the finding on education mostly for women and strong horiziontal preferences on race. However, in our sample, preferences and search effort across attributes have not significantly changed. Taking into account vertical preferences across all the characteristics, even if individual characteristics reflect a relatively high degree of assortativeness, considering them jointly results in low assortativeness in matching patterns.

This paper also relates to the literature investigating the decline in search frictions on economic outcomes over time. Ellison and Ellison (2018) show that the reduction in trading frictions brought about by the Internet has led to better matching between products and consumers and, in doing so, to an increase in consumer surplus. Focusing also on matching in product markets, Menzio (2023) finds that the growth rate of the surplus depends on the rate at which search frictions decline and on the elasticity of buyers' utility with respect to the degree of specialization in attributes. For labor markets, Martellini and Menzio (2020) report that the decline in search costs has not been matched by improvements in unemployment, labor productivity growth, vacancies, or transition rates. They attribute this finding to an increase in selectivity canceling out the abetting of search frictions. Fleshing out the trade-off between selectivity and declining search costs in marriage markets is the purpose of a recent paper by Antler et al. (2022). They show that learning and search frictions have ambiguous effects on sorting patterns as more informative dating, due to technological improvements, leads to an endogenous increase in effort to find the best match. Different from these contributions, our paper provides a direct measure of selectivity overall and across attributes and allows us to quantify their contribution over time.

Rosenfeld (2008) documents the prevalence of endogamy (i.e., horizontal preferences) with respect to race and religion in the U.S. over the 20th century. He finds less evidence of endogamy in income and education over time than we do, but our estimated preferences on race are horizontal, so both males and females prefer partners of the same race, consistent with the findings in Rosenfeld (2008). To our knowledge, no paper has studied the joint incidence of several attributes on the sorting patterns and documented how much effort is allocated to the search of individual as well as joint characteristics of a potential match.

We find that taking into account multidimensionality in search matters. Both males and females spend about 40% of their search effort on the joint characteristics of a partner. That is, all attributes (skills, education, income, race) must be simultaneously taken into account in targeting a potential partner. By looking at individual characteristics, we find that while preferences appear horizontal along some dimensions, taking multidimensionality into account results in a very low level of assortativeness in matches. These results show the importance of taking into account complexity in the decision of finding a match.

Race together with age and education are the most sought – after characteristics in a partner. The search effort that both men and women put on these attributes is the highest among all attributes. Recent literature[10] found evidence of more interracial marriages in online dating that partially overturns the long-lasting horizontal preferences in race. Our results show that many cognitive resources are spent by both men and women in finding a partner of the same race.

With respect to education, the evidence on assortative matching in education has been mixed.[11] We find that both men and women value education highly, devoting the most effort among the attributes (after race) to this particular category. This focus on education has important implications for inequality in the U.S. and abroad.[12]

We contribute in a methodological way to the new and growing literature on measurement of sorting. We use the concentration index of the singular value distribution of the preference and strategy matrices to construct and benchmark indexes of horizontality of preferences and assortativeness of matching. Our assortativeness measure, which can be directly compared with the frameworks of Eika et al. (2019)), Chiappori et al. (2021), and Zhang (2024), represents a monotone transformation of the normalized trace and exhibits behavior akin to the aggregate likelihood ratio across various scenarios highlighted in existing studies. Distinct from previous approaches, our methodology leverages the equilibrium search strategies of agents instead of relying on axiomatic assumptions regarding behavior. It identifies rather than presupposes matching patterns, seamlessly integrates the probabilistic nature of search and matching, accounts for the number of searchers, and allows for a natural decomposition by attribute in cases of multidimensional matching.

3 Model of targeted search

In this section we briefly discuss the model we use for the analysis, which we borrow from Cheremukhin et al. (2020). The economy contains a large finite number of females and males. Females and males are characterized by a multidimensional set of attributes, such as (but not restricted to) income, age, education, and race. We define sets of types of females and males, which contain all the combinations of attributes available. We assume F types of females indexed by x, and μ_x identical females of each type x. There

[10] See, e.g., Thomas (2019), Smith et al. (2014) and Lin and Lundquist (2013))

[11] See, e.g., Mare (2016), Eika et al. (2019), Gihleb and Lang (2020)).

[12] See, e.g., Skopek et al. (2010), Greenwood et al. (2014), Lee (2016), Eika et al. (2019), Ciscato and Weber (2020), Chiappori et al. (2017), Guiso et al. (2022).

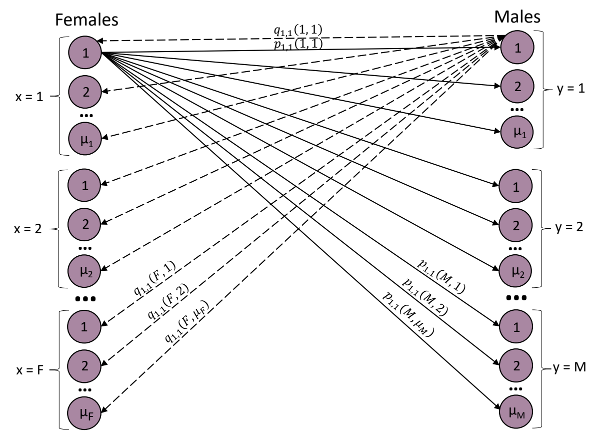

Figure 3.1: Strategies of Males and Females

are also M types of males indexed by y, and μ_y identical males of each type y. Types x and y are in general unranked indices that aggregate all attributes.

Males and females are heterogeneous in their type and simultaneously search for a match. Both males and females know the distribution and their preferences over types on the other side of the market, but there is noise—agents cannot locate potential partners with certainty. However, they can pay a search cost to help locate them more accurately. We model this by assuming that each agent chooses a discrete probability distribution over types. Each element of this distribution reflects the likelihood of contacting a particular agent on the other side. Let $p_x(y)$ be the probability that a female of type x targets a male of type y, and $q_y(x)$ be the probability that a male of type y targets a female of type x. Figure 3.1 illustrates the strategies of males and females. Once these are selected, both males and females make one draw from their respective distributions to determine which individual they will contact.

A match between any female of type x and any male of type y generates a non-negative payoff (surplus) Φ_{xy}. If a male and a female match, the payoff is split between them so that the payoff appropriated by the female ε_{xy} and the payoff appropriated by the male η_{xy} sum up to the total surplus $\Phi_{xy} = \varepsilon_{xy} + \eta_{xy}$.

A more targeted search, or a probability distribution that is more concentrated on a particular group of agents (or agent), is associated with a higher cost, as the agent needs to exert more effort to locate a particular person more accurately. As a baseline we might assume that agents enter the search process with a uniform prior of whom to target, $\tilde{p}_x(y) = 1/\sum_{y=1}^{M} \mu_y$ and $\tilde{q}_y(x) = 1/\sum_{x=1}^{F} \mu_x$. Choosing a more targeted strategy implies a larger distance between the chosen strategy and the uniform prior and is associated with a higher search effort. A natural way to introduce this feature into our model is the Kullback-Leibler divergence (relative entropy), which provides a convenient way of quantifying the distance between any two distributions, including discrete distributions as in our model. We assume that the search effort of a female of type x is defined as follows:

$$\kappa_x = \sum_{y=1}^{M} \mu_y p_x(y) \ln \frac{p_x(y)}{\tilde{p}_x(y)}. \tag{3.1}$$

Likewise, a male's search effort is defined as

$$\kappa_y = \sum_{x=1}^{F} \mu_x q_y(x) \ln \frac{q_y(x)}{\tilde{q}_y(x)}. \tag{3.2}$$

We assume that the search costs $c_x(\kappa_x) = \theta_x \kappa_x$ and $c_y(\kappa_y) = \theta_y \kappa_y$ are linear functions of the search effort.

To capture congestion in meetings among identical agents we introduce a congestion function $g_{xy} = g(p_x(y), q_y(x))$, which depends in some general way on the strategies of the agents as well as the number of agents of each type. Given this, the total number of matches formed between females of type x and males of type y is given by

$$M_{x,y} = \mu_x \mu_y p_x(y) q_y(x) g_{xy}.$$

Both males and females maximize the expected value of their payoffs net of the search costs, assuming that they take the meeting rates as given. For a female of type x, the problem is

$$Y_x = \max_{p_x(y)} \sum_{y=1}^{M} \mu_y \varepsilon_{xy} q_y(x) g_{xy} p_x(y) - \theta_x \sum_{y=1}^{M} \mu_y p_x(y) \ln \frac{p_x(y)}{\tilde{p}_x(y)} \tag{3.3}$$

Likewise, a male of type y solves

$$Y_y = \max_{q_{y,j}(x,i)\in S_{y,j}} \sum_{x=1}^{F} \mu_x \eta_{xy} p_x(y)\, g_{xy} q_y(x) - \theta_y \sum_{x=1}^{F} \mu_x q_y(x) \ln \frac{q_y(x)}{\tilde{q}_y(x)} \tag{3.4}$$

A *matching equilibrium* is then a Nash equilibrium in the admissible strategies for females and males, $p_x(y)$ and $q_y(x)$, which solve the problems in (3.3) and (3.4) for each individual male and female. We utilize the results describing the properties of the matching equilibrium, which we briefly summarize below.

First, a matching equilibrium must satisfy the necessary conditions, which facilitate computation of equilibria:

$$p_x^*(y) = \tilde{p}_x(y)\, exp\left(\frac{\varepsilon_{xy} q_y^*(x)\, g_{xy}^*}{\theta_x}\right) / \sum_{y'=1}^{M} \mu_{y'} \tilde{p}_x(y)\, exp\left(\frac{\varepsilon_{xy'} q_{y'}^*(x)\, g_{xy'}^*}{\theta_x}\right), \tag{3.5}$$

$$q_y^*(x) = \tilde{q}_y(x)\, exp\left(\frac{\eta_{xy} p_x^*(y)\, g_{xy}^*}{\theta_y}\right) / \sum_{x'=1}^{F} \mu_{x'} \tilde{q}_y(x)\, exp\left(\frac{\eta_{x'y} p_{x'}^*(y)\, g_{x'y}^*}{\theta_y}\right). \tag{3.6}$$

Second, if the congestion function takes the form $g_{xy} = p_x^{-\alpha} q_y^{-(1-\alpha)}$, $0 < \alpha < 1$, and search costs θ_x and θ_y are positive, then the matching equilibrium exists, is unique, and the aggregate matching function exhibits constant returns to scale. In addition, if the the surplus is split proportionally as $\frac{\varepsilon_{xy}}{\Phi_{xy}} = 1 - \alpha$, and the parameter α is the same for all pairs of types (x, y), then the competitive equilibrium is constrained efficient.

4 Data and Estimation

In this section we describe the data we use and the estimation procedure.

To study the U.S. marriage market, we use data from the Integrated Public Use Microdata Series (IPUMS), more specifically the American Community Survey (ACS), a 1 in 100 representative sample of the U.S. population, for 14 consecutive years from 2008 to 2021. The variables we use include sex, age, race, hispanic origin, total current income, educational attainment, occupation, marital status, and whether a person got married in the past 12 months. The last two variables allow us to identify singles and newly married couples. We assign unmarried males and females and members of newly married couples to bins corresponding to types in the model and therefore compute the number of searchers and couples formed for each bin.

We consider discretizations in several important dimensions: income, education, age, race, and skill. We split the income distribution into tertiles, quintiles or deciles based on total personal income in the current year (two bottom deciles are merged together since they both represent zero income). We split the education distribution into 3 unequal bins (school, in college, and post-college) or 2 bins (school, some college). The 21-40 age range is broken into 3 or 9 similarly sized bins. As the marriage rates are low for people below 21 and above 40, we consider only people in this age range. The implicit assumption is that everybody in this age range that is single is actively searching for a match. The data allow us to distinguish by race, forming 4 bins (White, Asian, Hispanic, Black) or 2 bins (combining White with Asian, and Hispanic with Black due to revealed similarity of preferences).

We utilize the data on occupations to obtain an estimate of "skilled" and "unskilled" occupations and split persons by their rank in the occupational skill distribution. Our skill variable represents an estimate of the average income a person might earn in the future if he or she stays in that occupation, and thus augments the current income variable described earlier. To this end, we sort occupations by their population average wages and obtain a mapping from occupation to skill level. We then assign each person to one of 3 or 6 equal occupational skill bins. Our skill bins roughly correspond to white-collar workers (top bins), blue-collar workers (bottom bins) and services (middle bins).

In the estimation section, we consider both uni- and multi-dimensional combinations of attributes. For instance, if we want to split the population into 3 bins by income, 3 bins by skill and 3 bins by education and evaluate the multi-dimensional sorting along these attributes, the model will have a total of $F = M = 27$ types spanning all possible combinations of these 3 attributes. Based on these criteria, we then compute the data split to obtain two 27-by-1 vectors containing the numbers of female and male singles of each type (corresponding to μ_x and μ_y in the model), and a 27-by-27 matrix containing the number of pairwise marriages for each combination of types ($M_{x,y}$).

To look at longer-term changes in excess of our 14-year consecutive sample, we have been able to construct a comparable dataset on demographic characteristics and simultaneously identify the recently married couples in the 1960 and 1980 census data. This is made possible by a variable reporting age at first marriage available in these datasets. Combining this variable with the current age and controlling for the number of marriages, we can identify couples that got married in the preceding 12 to 24 months. As the census data available to us are a 5-in-100 representative sample of the population, we obtain overall more precise measures from the census than from the 1-in-100 annual surveys. Otherwise, the American Community Survey is modeled based on the census and therefore most of the variables are defined in the same way. One important difference between these two samples and the more-recent period is the occupational coding and therefore the set of occupations that are considered high- or low-skill. While the procedure applied to the data remains the same, the specific occupations that are considered high-skill in 1960, 1980, and the more recent years may be somewhat different. For the counterfactual exercises looking at long-term changes in inequality, we construct a mapping between the occupational codes in the 1960, 1980 and more recent samples. Apart from that, the earlier samples are aggregated using the same methodology.

We observe the number of searchers, μ_x and μ_y, and the number of matches, $M_{x,y}$, between each pair of types x and y in a particular year. The goal of estimation is to use the model to recover the unknowns in the model: preferences $\Phi_{xy}, \eta_{xy}, \varepsilon_{xy}$, costs θ_x, θ_y, and priors $\tilde{p}_x(y), \tilde{q}_y(x)$. As we discuss in Cheremukhin et al. (2020), identifying all of these simultaneously is not possible, but under certain assumptions some of the fundamentals are possible to recover. As described in the model section, the equilibrium is unique and exhibits constant returns to scale if the congestion function takes the Cobb-Douglas form $g_{xy} = p_x^{-\alpha} q_y^{-(1-\alpha)}$. We maintain this as our first identifying assumption for all estimation exercises.

Second, we assume that $\alpha = 0.5$, which implies symmetric congestion. Third, we assume an equal split of the surplus $\frac{\varepsilon_{xy}}{\Phi_{xy}} = 0.5$, which therefore maintains that the equilibrium is not only unique but also constrained efficient. The assumption of equal split of the surplus is supported by micro estimates. In particular, Bruze et al. (2015) find that the surpluses of husbands and wives vary in the range from 0.35 to 0.65 depending on education levels, with the population average close to equal. Fourth, we assume that all males and females face identical marginal costs of search $\theta_x = \theta_y = \theta$. When making these assumptions, we deliberately impose as much symmetry as possible so that any asymmetry we observe in the estimated model cannot be attributed to our assumptions and must therefore be reflecting some asymmetric features of the data.

Under these assumptions, we provide a computational algorithm that allows us to use the model to identify the ratios of preferences to costs Φ_{xy}/θ, which in combination reproduce the empirically observed matching rates as an equilibrium of the model. The mapping from the matrix of suplus-to-cost ratios to matching rates has an additional degree of freedom that determines the matrix's scale. To fix this degree of freedom, we impose a technical assumption that the smallest element of the matrix $\frac{\Phi_{xy}}{\theta}$ is normalized to 1. Finally, under our identifying assumptions we can safely set the cost parameter to 1 for estimation purposes and work with the surplus matrix Φ_{xy}.

Our computational algorithm starts with an initial guess for the unknown surplus matrix Φ_{xy}, computes the unique equilibrium strategies $p_x^*(y), q_y^*(x)$ and matching rates M_{xy}^* that correspond to the proposed surplus, and then computes the likelihood that the empirically observed matching rates $\mathcal{M}_{x,y}^{Data}$ are an equilibrium outcome of the proposed surplus. We define the likelihood as follows:

$$\mathscr{L}\left(\mathcal{M}_{x,y}^{Data}|\Phi_{xy},\mu_x,\mu_y\right)=\sum_{x=1}^{F}\sum_{y=1}^{M}\mathcal{M}_{x,y}^{Data}\left(\ln M_{x,y}^*-\ln\mathcal{M}_{x,y}^{Data}\right).$$

We combine several standard likelihood maximization and gradient descent algorithms commonly used in the literature to converge to a local maximum from the initial guess. The procedure is then repeated from a large number of random initial guesses to obtain the global maximum. For lower dimensional estimations (below 20) the algorithm is very efficient in that convergence from 1000 initial guesses can be obtained in reasonable time and most of the local maximums are close to the global maximum. Moreover, the model fits the data well, as all the estimated surpluses reproduce the empirical matching rates very closely.

However, we face several natural computational challenges as the dimensionality of the model increases. The first limitation is that the amount of computations (and memory) grows proportionally to $(M*F)^2$. Despite multiple computational shortcuts that we came up with to significantly reduce the constant of proportionality,[13] there are

[13] Representing Hessian matrixes for the gradient descent as sparce matrices and forcing a diagonal structure upon them substantially improved the speed of computing the equilibrium, of overall likelihood maximization, and simultaneously reduced the memory requirements.

limits to how large the dimension can be. The BigTex supercomputer that we employ for estimation runs into memory limitations for surplus matrices exceeding 54 by 54. At this dimensionality, parallel optimization from 100 starting points often approaches the runtime limit of 48 hours. Thus, 54 by 54 is the natural computational limit that we face.

The second limitation is related to data sparsity. Ideally we would like to estimate an intersection of the maximum number of bins for all attributes simultaneously, e.g. 9 income bins x 6 skill bins x 3 education bins x 9 age bins x 4 race bins. However, estimating a 5832 by 5832 matrix of preferences not only is infeasible, but makes little sense since the matrix distributing a few million people into 34 million boxes would be extremely sparse. Even for a minimal estimation intersecting all five attributes, i.e. 3 income x 3 skill x 3 education x 3 age x 2 race, the 162x162 matrix of marriage rates has multiple columns and multiple rows that contain only zeros. Therefore, we have to not only cut the number of bins along each dimension, but also omit at least one attribute in each estimation. To circumnavigate this problem, we intersect different combinations of four attributes at a time, estimate the surplus for each such combination, and then combine the estimates. In our baseline, we omit income, education and age, one attribute at a time. Thus, we estimate the surplus for three setups: 3 income x 3 education x 3 skill x 2 race, 3 age x 3 education x 3 skill x 2 race, 3 income x 3 age x 3 skill x 2 race, and then combine the results of these three estimation sets to obtain the overall multidimensional properties of preferences.[14]

5 Measuring Assortativeness, Horizontality, Selectivity, and Affinity

To describe the results of the estimation for each breakdown of the data into a combination of attributes, we develop new measures and employ concepts and computational techniques proposed in the literature. First, we import and improve upon the measures of horizontality of preferences and of assortativeness of the equilibrium matching

[14] We have conducted a large number of different other combinations of attributes and bin splits of the population and found that the main properties of estimated results are broadly similar and consistent across different estimation setups. We thus report only the most salient systematic estimates in the paper and omit the rest.

presented in Cheremukhin et al. (2020). Then, using a decomposition of multivariate information, we measure selectivity with respect to each attribute and combination of attributes. Finally, we compute a discrete affinity matrix with respect to different attributes to capture the strength of preferences and interactions between attributes.

5.1 Horizontality

The measure of horizontality that we proposed in our theoretical paper counted the number of different best matches present in the population. In particular, let $\omega_x = \left|\{\arg\max_y (\varepsilon_{xy})\}_{x\in\{1,\ldots,F\}}\right| \in \{1,\ldots,M\}$ be the number of different types of males who are best matches for at least one type of female and let $\omega_y = \left|\{\arg\max_x (\eta_{xy})\}_{y\in\{1,\ldots,M\}}\right| \in \{1,\ldots,F\}$ be the number of different types of females who are best matches for at least one type of male. Then we can define a measure of horizontality as $(\omega_x + \omega_y - 2) / (M + F - 2)$, representing the number of different best matches scaled by the total number of types. Preferences are fully vertical if every type's best match is the same type, and preferences are fully horizontal if every type's best match is a different type.

This measure captures the main idea of horizontal and vertical preferences, but it has many problems. First, it can take only a small number of discrete values and therefore is very coarse when the number of types is small. Second, it does not properly span the whole range of preferences when taking the value of 1 for a substantial subset of possible preferences. As a consequence, this measure is not comparable across estimation setups with different dimensionalities.

To fix these problems we develop a more elaborate measure of horizontality of preferences. Instead of counting the number of maximums, we count the number of different patterns in rows (or columns) of the matrix of preferences. The measure we use is based on spectral properties of the matrix of preferences. It is well known that the distribution of singular values of a matrix (absolute values of its eigenvalues) describes the effective dimensionality of a matrix, with the sum of squared singular values capturing its "energy" and the relative energies of each eigenvalue representing its contribution to that effective dimensionality. To capture this effective dimensionality by a single number, we apply the Herfindahl-Hirschman concentration index measure to the distribution of singular values. Thus, denote the singular values of Φ_{xy} by $\{\lambda_j\}_{j\in\{1,M\}}$. Compute the concentration index as $HHI = \sum_j \frac{\left(\sum_j \lambda_j\right)^2}{\lambda_j^2}$. Then the inverse of HHI captures the effective number of different patterns of preferences in Φ_{xy}, which can take values between 1 and $\min(M, F)$. Properly rescaling this measure gives a better version of the horizontality index $H(\Phi_{xy}) = (1/HHI - 1)/(\min(M, F) - 1)$. To reduce this measure's dimensionality dependence, we further apply a Monte-Carlo technique to weighted sums of a perfectly vertical and a perfectly horizontal matrix of preferences to benchmark the values of the H-measure for different dimensionalities of the matrix. This technique allows us to compute a complete mapping between preference weights and the H-measure for each dimensionality. We then use this mapping as a lookup table to infer the underlying preference weight for each estimated matrix Φ_{xy} and treat the underlying weight as our estimated horizontality index.[15] This new horizontality index preserves the property that when preferences are vertical, the horizontality index equals 0, and when preferences are horizontal, the horizontality index equals 1. It spans the whole set of possible preference matrices in a monotone way and is comparable for matrices of different dimensions.

5.2 Assortativeness

We measure assortativeness in a similar way. Let us denote by $P_{xy} = [p_x(y)\,\mu_y]$ the matrix of all female strategies and by $Q_{yx} = [q_y(x)\,\mu_x]$ the matrix of male strategies. For each of these matrices we compute its singular values, compute a concentration index and rescale. This allows us to compute the assortativeness index of females' strategies as $A(P_{xy}) = (1/HHI_P - 1)/(M - 1)$ and of males' strategies as $A(Q_{xy}) = (1/HHI_Q - 1)/(F - 1)$. We use the same benchmarking procedure as for the horizontality index to look up the corresponding weights that are comparable across dimensionalities. Instead of representing the number of different targets of search, the assortativeness index now represents the number of different patterns of search relative to the maximum possible number of patterns. For an assortative equilibrium where each type has a different search pattern, the assortativeness index equals 1, while for a mixing equilibrium, where all females/males have similar patterns of search, the assortativeness index equals 0. We can then define an aggregate measure of assortativeness as the (weighted) average of the two measures for males and females.

For comparison with existing measures of assortativeness, we have applied our proce-

[15] For more details on this procedure see appendix.

dure, along with alternative procedures in the literature, to a large number of matching patterns spanning the range of equilibria from perfect assortativeness to random matching. We find that our aggregate measure of assortativeness is a monotone transformation of the normalized trace and behaves similarly to the aggregate likelihood ratio[16] for all the interesting cases. Our measure has multiple advantages over existing measures. First, our measure properly accounts for the probabilistic nature of search. The agents that searched but did not find a match are a natural outcome of the model, rather than having to assume that agents chose to match with themselves or, alternatively, that everybody got matched. Second, the number of searchers (including those that did not find a match) is properly taken into account when comparing relative matching rates, as more searchers would imply more matches, but that is not taken into account in existing measures. Third, our measure of assortativeness does not impose a pattern of who should be matched with whom for matching to be considered assortative, but instead identifies the sorting pattern. Perfect positive and perfect negative assortative matching both would be evaluated as highly assortative. Fourth, characterizations of sorting must ultimately be based on properties of individual choices of whom to look for, rather than on some abstract axioms that may or may not be features of individual choices. Our measure achieves that goal.

For multi-dimensional types, we naturally extend these definitions to compute assortativeness and horizontality indexes with respect to each dimension separately. For instance, when the estimation is for an intersection of income, skill and education bins, we can add up the matrices Φ_{xy}, P_{xy}, and Q_{xy} to a lower dimensionality corresponding to only, for example, the income dimension to obtain estimates of assortativeness and horizontality of preferences with respect to income alone.

5.3 Selectivity

Another concept we introduce into the search and matching literature is a measure of selectivity by agents with respect to attributes. Recall that the amount of search effort

[16]For definitions of these concepts for multi-dimensional matching see Zhang (2024).

that each agent exerts in equilibrium, defined as

$$\kappa_x^0 = \sum_{y=1}^{M} \mu_y p_x(y) \ln \frac{p_x(y)}{\tilde{p}_x(y)},$$

measures the amount of information processed by the agents in equilibrium and reflects how selective agents' strategies are. This will be our overall measure of selectivity. Using recent research on decomposition of multivariate information (see, e.g., Williams and Beer, 2010), we can decompose total selectivity into selectivity with respect to each attribute and to combinations of attributes.

Let the set of types $\{1, ..., M\}$ be broken into subsets $a = \{a_1, ..., a_A\}$ corresponding to different values an attribute can take. For instance, in the example of 3 income x 3 education x 3 skill with 27 types, a_1 would contain low-income, a_2 medium-income, and a_3 high-income types. If we aggregate the probability distribution $p_x(y)$ (and the priors) based on subsets of attribute a as $p_x(a_i) = \sum_{y \in a_i} \mu_y p_x(y)$, then we can define the amount of information (selectivity) with respect to attribute a as follows:

$$\kappa_x(a) = \sum_{i=1}^{A} p_x(a_i) \ln \frac{p_x(a_i)}{\tilde{p}_x(a_i)}.$$

Now consider two attributes a and b each breaking down the original distribution of types into overlapping subsets. Define the intersection of subsets $a \cap b$ as consisting of all possible intersections of subsets in a and b. We can then define the first-order interaction between attributes a and b as follows:

$$\kappa_x(a, b) = \kappa_x(a \cap b) - \kappa_x(a) - \kappa_x(b)$$

Williams and Beer (2010) show that this first-order interaction term is non-negative and that continuing this logic defines all first-order interactions, second-order interactions and so forth, which together add up to the total information (selectivity) κ_x^0 making up a multivariate information decomposition. In our estimates we have computed the full decomposition and found that all the terms beyond the first-order interactions are negligible and can be altogether omitted in practice. Moreover, we can re-attribute all first-order interactions by splitting them equally between the pair of attributes involved, which would then give a decomposition of selectivity with respect to all attributes z of the form $\kappa_x^0 = \sum_z \kappa_x(z)$.

These measures of selectivity represent how picky an agent is with respect to an attribute or combination of attributes. Selectivity is measured in bits of effectively processed information, reflecting the skewness of probabilistic strategies chosen by agents in equilibrium.

5.4 Affinity

For an estimated surplus shape, Dupuy and Galichon (2014) define an affinity matrix as a quadratic form approximation of preferences with respect to a continuum of attributes. Here we show how to accommodate this definition and compute the affinity matrix in the discrete case for a surplus with multi-dimensional attributes. As in the original definition, first consider the case when types x and y represent sorted values of one attribute, such as income or age. The affinity matrix A approximates the estimated surplus matrix Φ_{xy} as a quadratic form with respect to attributes of the form: $\Phi_{xy} = x'Ay$. In the continuous case, the matrix $A = \partial^2\Phi_{xy}/\partial x\partial y$ measures the partial second derivatives with respect to the attributes and measures the strength of mutual attractiveness between males and females on a pair of attributes.

To compute the discrete case version of this matrix with respect to multiple attributes, we first construct a K by F matrix $\mathcal{X}$, which contains in its rows attribute values or indexes of attribute values that preserve ordering of attributes from lowest to highest for each of K attributes of females. Similarly, the L by M matrix $\mathcal{Y}$ will contain in its rows attribute values for each of L attributes of males (in our estimated cases $\mathcal{X} = \mathcal{Y}$ due to symmetry). We standardize the scale of each row of $\mathcal{X}$ and $\mathcal{Y}$ by subtracting respective means and dividing by standard deviations. Then we compute the affinity matrix:

$$A_{xy} = \left(\mathcal{X}^T\right)^+ \Phi_{xy}\mathcal{Y}^+$$

where $\mathcal{Y}^+$ denotes the Moore-Penrose pseudo-inverse of a non-square matrix $\mathcal{Y}$.

The L by K affinity matrix A_{xy} contains all the partial cross derivatives measuring the curvature of preferences, with diagonal elements capturing strength of mutual attractiveness based on one attribute, and off-diagonal elements capturing intensity of complementarity/substitutability between attributes of women and men.

6 Empirical Results

6.1 Uni-dimensional estimates

For comparison with the literature and as a sanity check, we start by estimating preferences and equilibrium strategies for each attribute of interest separately. In each case we break down an attribute into the largest reasonable number of distinguishable bins, as shown in the first two columns of summary Table 1. For each attribute we apply the measures described in the previous section to estimate horizontality and strength of preferences (affinity), as well as assortativeness and selectivity of search strategies for each attribute. We show these in the following columns of Table 1. The last column shows the findings in the literature for assortativeness and horizontality for each attribute, for a quick comparison. The Table presents estimates averaged over the 2008-2019 period.

We find that mutual attractiveness is strongest based on race and education, and a lot weaker based on age, skill and income. We find that preferences are strongly horizontal for race and education (values above 0.2 typically indicate that every type is somebody's best match), somewhat less for age and close to vertical for income and skill. Consistent with the idea that horizontal preferences lead to assortativeness, while vertical preferences lead to looking up the desirability ladder and a mixed sorting equilibrium, we find high levels of assortativeness by race and education, intermediate level of assortativeness by age, and low assortativeness by skill and income. Naturally, the strength and horizontality of preferences are reflected in selectivity of individual strategies, which show that people are very selective on race, somewhat selective on education and age, and much less selective on income and skill.

Attribute	Bins	Strength (Affinity)	Horizon-tality	Assorta-tiveness	Selec-tivity	Literature
Income	9	0.06	0.04	0.07	0.05	PAM, vertical
Skill	6	0.13	0.08	0.12	0.06	mixed, vertical
Age	9	0.24	0.14	0.26	0.19	PAM, horizontal
Education	3	0.85	0.30	0.57	0.20	PAM/mixed, mixed
Race	4	0.82	0.52	0.90	0.88	PAM/mixed, mixed

Table 1: Uni-dimensional estimates of preferences and sorting

It is instructive to compare our results with the existing literature summarized in the last column of Table 1.[17] The literature has found mixed or horizontal preferences for race, education and age, and vertical preferences for skill and income.[18]

Our findings are mostly consistent with the literature on preferences. However, the evidence on the degree of assortativeness varies by study and by method of analysis. The literature is largely split, arguing about the degree of assortativeness in race, education and skill, and finds some assortativeness in income and age.[19] Keeping in mind the differences in measures and definitions, in contrast, we document a high degree of assortativeness in race, education and age, and non-assortativeness in income and skill. These results also provide a uni-dimensional benchmark against which multi-dimensional estimates are evaluated.

6.2 Multi-dimensional estimates

As discussed in detail in section 4, we estimate multi-dimensional preferences and sorting for all five attributes of interest jointly for each year. We split the population into 3 income bins, 3 education bins, 3 skill bins, 3 age bins, and 2 race bins.[20] For this section, we average the results across the 2008-2021 period and show the results in Table 2.[21]

We find that multidimensional results are in general agreement with uni-dimensional results along all five attributes under consideration. The strength of preferences represented by the diagonal elements of the affinity matrix is generally consistent with uni-dimension results presented in Table 1: preferences are strongest for race, with less importance placed on age, education, income and skill, in that order. Curvature of preferences with respect to education is apparently stronger when considered separately than when evaluated in combination with other characteristics, which suggests that ed-

[17] We have a longer discussion relegated to the appendix, in particular Table B.1.

[18] See Rosenfeld (2008), Hitsch et al. (2010a), Hitsch et al. (2010b), Skopek et al. (2010), Lin and Lundquist (2013), Lee (2016), Lewis (2016)), Bruch and Newman (2018), Thomas (2019).

[19] See Kalmijn (1994), Jepsen and Jepsen (2002), Choo and Siow (2006), Schwartz and Graf (2009), Greenwood et al. (2014), Smith et al. (2014), Bertrand et al. (2015), Mare (2016), Qian (2017a), Chiappori et al. (2017), Fiorio and Verzillo (2018), Mansour and McKinnish (2018), Ciscato and Weber (2020), Eika et al. (2019), Ciscato et al. (2020), Gihleb and Lang (2020), Chiappori et al. (2020)), Guiso et al. (2022).

[20] We estimate three 54 by 54 models each omitting one attribute at a time, and then combine the estimates through averaging.

[21] The full affinity matrix is shown in the appendix in Table C.1.

ucation often serves as a proxy for other attributes or combinations of attributes, such as current income and skill (future income).

Consistent with the unit-dimensional results, preferences are horizontal in race, vertical in skill and income. The ordering of horizontality for education and age is reversed — another indication that one or both of these attributes may interact with others in the search strategies.

Our cross-sectional results illustrate that attributes with more horizontal preferences tend to have more curvature and produce stronger equilibrium assortativeness and selectivity, while attributes with vertical preferences tend to have less curvature and generate more mixed-sorting patterns.

The selectivity measures in Table 2 reveal a much more elaborate pattern compared with the uni-dimensional case. It is still true that people are more selective on horizontal attributes themselves. However, the multi-dimensional estimates allow us to estimate selectivity with respect to an attribute, but also selectivity with respect to interactions of attributes. Strikingly, we observe the strongest interaction terms among vertical attributes, which exceed their direct selectivity measures. For instance, for men the interaction term for selectivity on income and skill of 0.16 exceeds their direct selectivity terms, which sum up to 0.13. Altogether, the interaction terms account for about half of overall selectivity. The largest interaction terms, selectively reported in Table 2, are between vertical attributes, such as income and skill, as well as their interactions with less vertical attributes, such as education and age. These results demonstrate that strategies targeting combinations of attributes, such as high income, high skill and high education, play a profound role in sorting that cannot be accounted for with uni-dimensional studies of sorting. This result indicates that it is necessary to use a method like ours, jointly estimating preferences with respect to multi-dimensional attributes to accurately uncover the properties of preferences, sorting and selectivity.

6.3 Changes over time

6.3.1 Medium-term Changes

An important question often discussed in the literature is that of changes in preferences and sorting over time. Our estimates are uniquely tailored to answer this question. In the previous subsection, we averaged measured properties across the recent period.

Attribute	Strength (Affinity)	Horizon-tality	Assorta-tiveness	Selectivity Men	Selectivity Women
Overall		0.50	0.67	1.12	1.18
Income	0.09	0.09	0.33	0.06	0.09
Skill	0.08	0.08	0.29	0.07	0.10
Age	0.28	0.22	0.56	0.16	0.18
Education	0.22	0.16	0.47	0.13	0.14
Race	0.35	0.26	0.60	0.15	0.17
Interactions				0.55	0.51
Income		Skill		0.16	0.14
Income		Education		0.08	0.05
Skill		Education		0.11	0.08
Income		Age		0.05	0.06

Table 2: Multi-dimensional estimates of preferences and sorting

In this subsection, we look more closely at the time dimension and document the changes over a sequence of annual estimates for the period 2008-2021. The time series for horizontality by attribute, assortativeness by attribute, strength by attribute, and selectivity by attribute and by gender are shown in Figure 6.1. Note here that for selectivity measures, we split the interaction terms equally between attributes, so that total selectivity equals the sum of contributions of five attributes shown in the bottom panels of Figure 6.1.

The first, most remarkable result is that there is no identifiable trend in the majority of the series. The degree of horizontality of preferences is stable overall and for each category. The degree of assortativeness is stable overall and for each category. The strength of preferences remained unchanged for income and skill, increased slightly for age, but reduced somewhat for race and education. There may be some mild related trends where the decline in curvature and horizontality of preferences for race led to a decrease in assortativeness on race. Similarly, stronger curvature and horizontality on age in the last few years likely contributed to increased sorting on age. Related to that, the degree to which both males and females select on age increased slightly in the most recent periods, and the degree to which males select on race and education marginally declined, but other components of selectivity saw little change.

Although overall selectivity of females is slightly higher than that of males, both

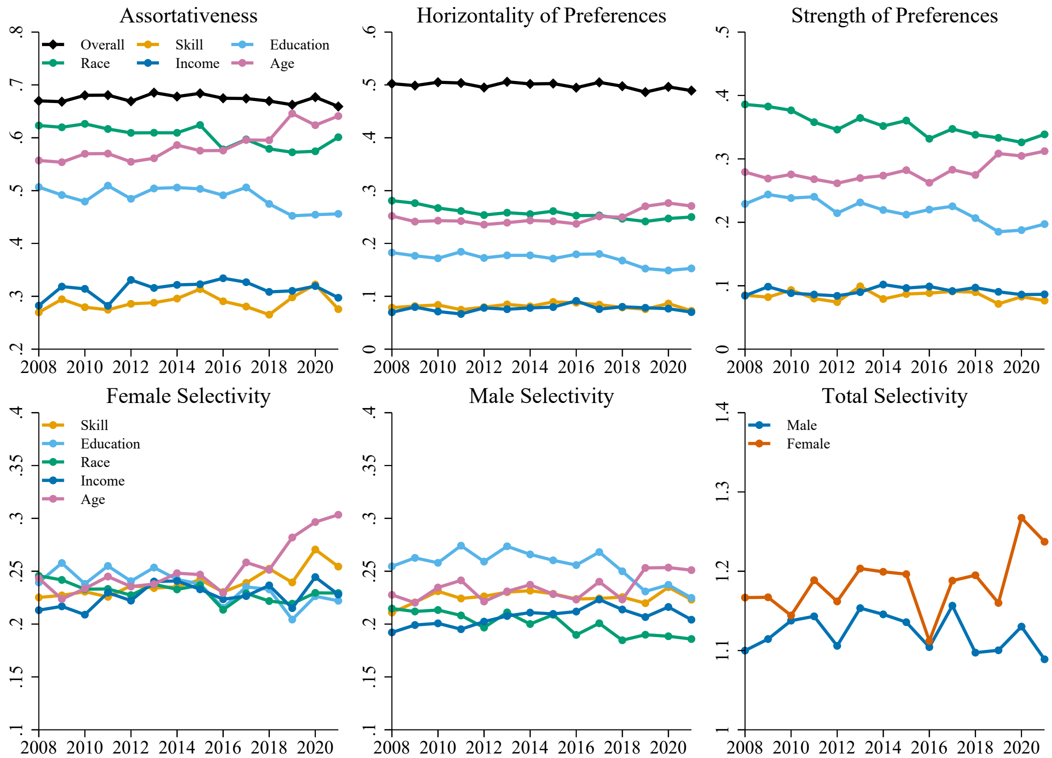

Figure 6.1: Properties of preferences and sorting over time

remained unchanged throughout the period we consider. This is especially striking taking into account the fact that the methods of finding a mate changed dramatically between 2008 and 2021. As documented by Rosenfeld et al. (2019), in 2008 less than one in ten marriages were conceived online, while by 2018 more than half of the marriages originated online. One would expect a profound effect of such a change in the search method on overall selectivity and search patterns, but we find essentially no change in how people search and who marries whom over the 2008-2021 period.

6.3.2 Long-term Changes

To compare the earlier and later periods and show the results in a transparent way, we average the two halves of the more recent sub-period into two years, 2010 and 2020. Thus, the Figures that follow show four observations: 1960, 1980, 2010 and 2020. Figure 6.2 illustrates the changes in assortativeness, horizontality, affinity, and

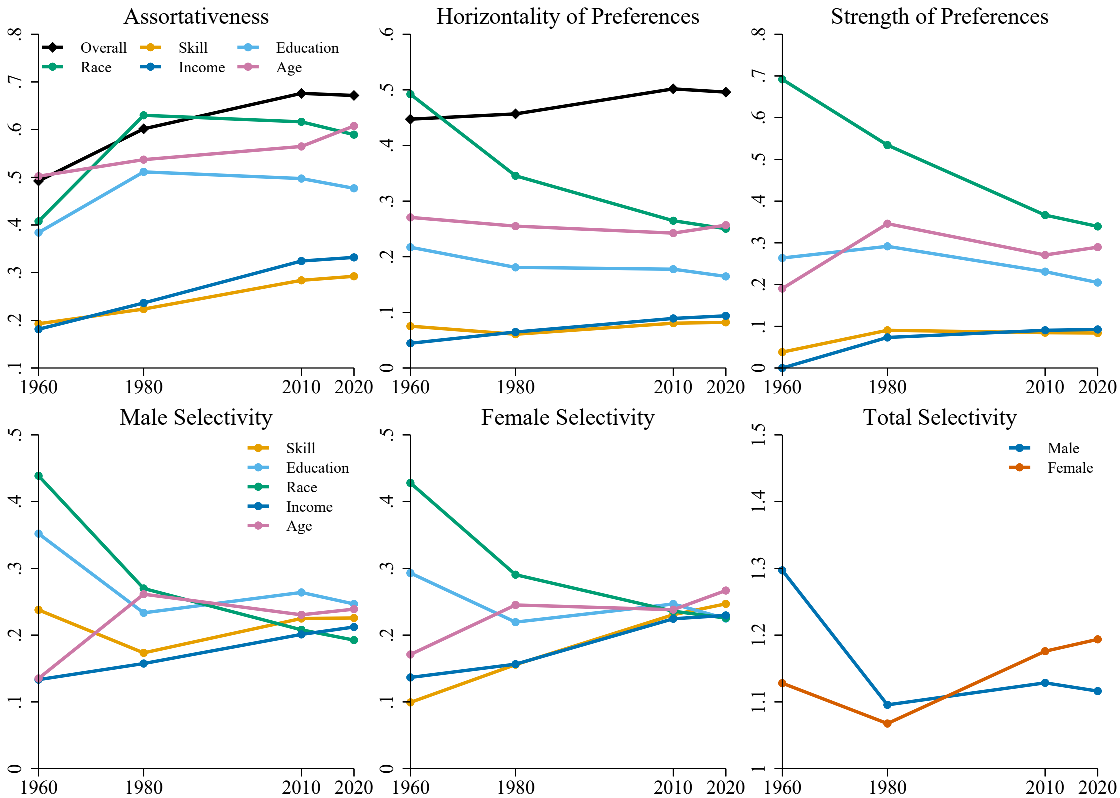

Figure 6.2: Long-term changes in preferences and sorting

selectivity together with their components, as described earlier in Section 5.

When looking at the long-term changes, a number of observations stand out regarding both preferences and sorting. First, we observe a trend decline in the strength of preferences for race and education with a simultaneous trend increase in preferences for income and skill. Second, we observe a trend decline in horizontality of preferences with respect to race and education with a simultaneously trend increase in horizontality of preferences with respect to income and skill. However, overall preferences gradually became *more horizontal* over time.

Third, consistent with the long-term changes in preferences, we observe an increase in assortativeness both overall and with respect to most categories, with the exception of a slight decline in assortativeness with respect to race and education in the most recent period. Fourth, females, and to a large extent males, became more selective on income, skill and age, but less selective on race and education. An interesting observation is the

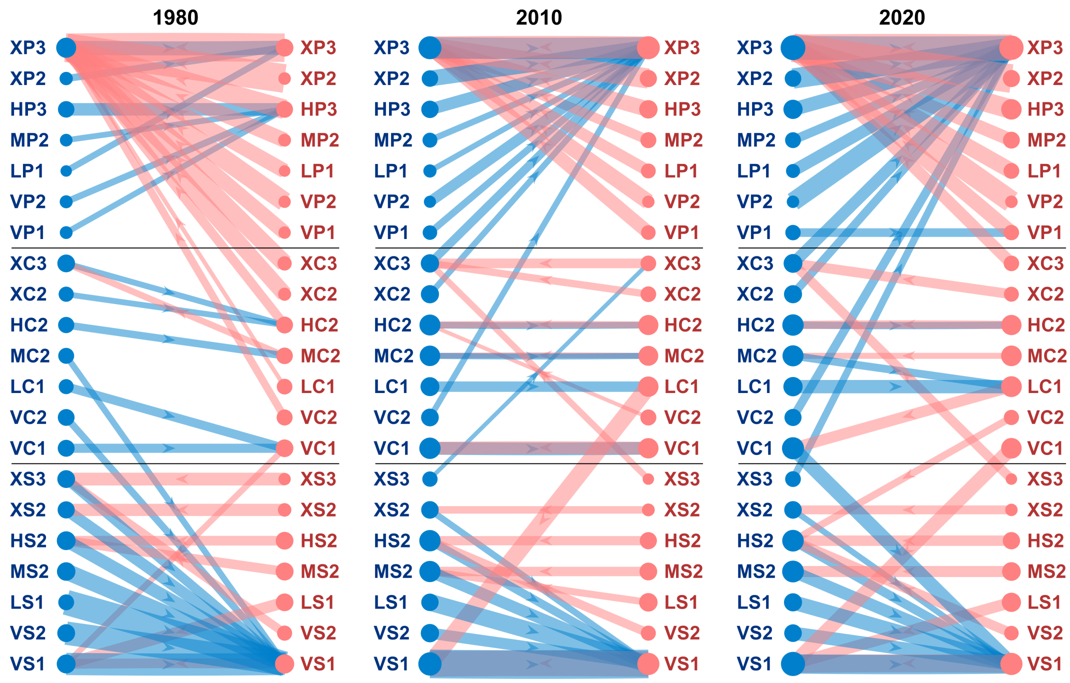

Figure 6.3: Change in sorting patterns by educational attainment: 1980-2020

near equalization of search effort devoted to sorting on different attributes, which was largely focused on race and education at the expense of income and skill in the earlier periods. However, we see *no trend in total selectivity*. Females overall became slightly more selective, surpassing males who reduced selectivity between 1960 and 1980, and then it remained flat since 1980.

An interesting observation related to the increase in horizontality of preferences is the substantial overlap in matching across education levels in 1960 and 1980. Targeting across education levels becomes less frequent in the subsequent periods. In particular, Figure 6.3 compares targeting strategies of females (red) and males (blue) in 1980, 2010, and 2020 sorted through education from post-college (P, upper bin), to college (C, medium bin), and to school (S, lower bin) for different incomes (X, H, M, L, V) and skill (3, 2, 1) levels. The figure uncovers two important patterns in the evolution of targeting strategies with respect to education.

First, females show an increasing interest in partners who share the same education level, from 1980 to the most recent data in 2020. For females, horizontality in education is also paired with strong vertical preferences for potential partners' incomes and skill levels, with the minor exception of less-educated females over time. Second, males' evolution of targeting strategies reflects a preference toward potential partners who are equally educated, and away from potential partners fulfilling domestic roles, as they were seeking in 1980. Both patterns are consistent with the rise in female labor force participation and the gender difference in gains in the market for college-educated participants favoring women over men in recent periods.[22]

6.4 Potential Drivers of Long-Term Trends

In the previous subsection we documented minimal changes in preferences, selectivity and sorting over the 2008-2021 period despite the proliferation of on-line dating platforms has made it easier for people to meet and form a match. Moreover, over the longer-period 1960-2021 we uncovered an upward trend in horizontality of females' preferences for partner's education coupled with a downward trend in males' preferences for partners assuming domestic roles.

In this subsection we consider other potential drivers of these patterns suggested by the literature that differ from our rationale based on females' labor force participation and educational attainment. In particular, we use our model to evaluate two potential explanations prominent in the literature: increasing selectivity and spatial segregation. We also consider the mechanical effects of changes in the proportions of searchers with different characteristics over time. Once these three forces are accounted for, the remaining changes are attributable to changes in preferences, that is, the changes in how men and women evaluate potential partners.

6.4.1 Reduced Search Costs and Increased Selectivity

Martellini and Menzio (2020) suggested that, while the introduction and proliferation of online matching platforms have contributed to a reduction in search costs, they have also increased selectivity of the participants in the search process. They posit that the reduction in costs has been undone by the increase in selectivity, resulting in minimal

[22] For instance, Cortes et al. (2018) use wage data from the US Census and the American Community Survey (1980-2014) to document a change in gender trends: they show a rise of women in the labor market for workers with at least a college degree as educational attainment. A similar finding for Europe is documented in Grow and Van Bavel (2020). Goldin (2023) traces out the evolution of women's labor force participation and educational attainments in the US since 1905.

change in quantity and quality of matches before and after the occurrence of these technological improvements in different search and matching markets.

In the context of our model applied to the marriage market, the diffusion of online dating may imply an overall effective decrease in the cost of search θ. Given that the sorting patterns and the shape of preferences overall do not seem to have changed over the 2008-2021 period, as a first check, we assess whether the average values of the elements of the matrix Φ_{xy}/θ have increased over time.

Figure 6.4 displays quantiles of (log) estimated surplus-to-cost ratios for which we have data.[23] From this Figure, there seems to be no evidence of a significant decline in the value of costs, neither more recently nor in the longer term. As we discussed in the previous section, we also find no evidence of an increase in selectivity, i.e. the precision with which agents are able to identify their best matches, reflecting how selective or targeted their strategies are.

We delve deeper into the analysis of whether and how we would be able to detect a decline in search costs by running the following counterfactual experiment. Suppose that, as a result of technological improvements in the dating market, the effective cost of search decreased to a third of the original value for all market participants. To model this change, we take preferences from the 1980 period with respect to income, education and race, and divide them by a decreasing factor reflecting the change in costs. We then compute the equilibrium and generate artificial matching rates for each counterfactually scaled matrix of preferences and apply our estimation procedure to these artificial matching rates to recover the parameters. We perform this counterfactual using a Monte-Carlo technique and report our findings in Figure 6.5.

We find that a reduction in costs by a factor of 3 would have increased substantially the 95th percentile of estimated elements of surplus-to-cost ratios.[24] Consistent with the intuition of Martellini and Menzio (2020), this reduction in costs would be compensated by a substantial increase in selectivity. We would also expect a uniform increase in our preferred measures of horizontality and assortativeness with respect to all attributes. Had such a cost reduction actually occurred, we would have expected to see an increase in the top quantiles of Φ_{xy}/θ as well as a substantial rise in our measure of selectivity.

[23] We have tried various approaches to computing this value and employed different statistical approaches. The resulting measures all tell the same story.

[24] The 5th percentile does not change because of our normalization of the minimum element of the estimated matrix to be equal to 1, as described in Section 4.

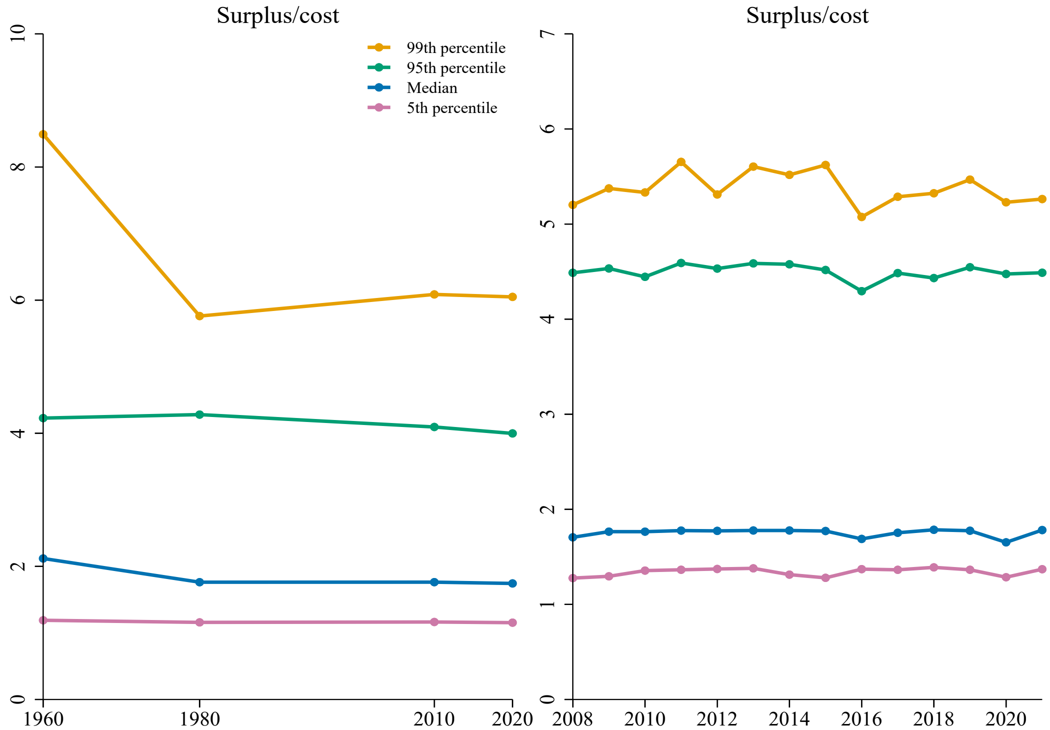

Figure 6.4: Inverse costs over time

As we find no evidence of an increase in the top quantiles of the estimated surplus-to-cost ratios, nor any evidence of increased selectivity, we conclude there has been no decline in search costs over our period of observation.

We think this finding is compelling. By showing that improvements in the technology of matching are not paired with significant cost reduction, the results suggest that the nature of the cost of search θ is cognitive rather than technical. In simple terms, the cognitive ability of the human brain facing a complex choice of whom to pick from the myriad of alternatives is the real bottleneck that determines the effective search costs, how selective people are, and who matches whom in the end. The technology improvements over time have mostly increased the number of available options, but have not improved our cognitive ability to make informed choices.

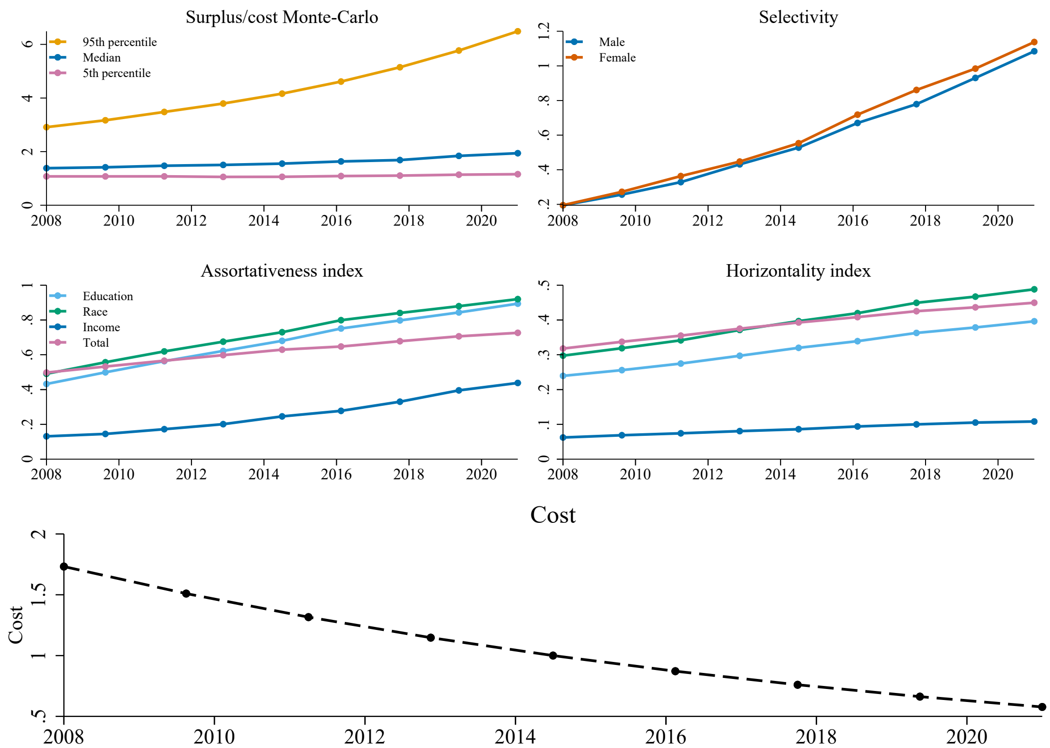

Figure 6.5: Counterfactual experiment: reduction in search costs

6.4.2 Spatial Segregation

Another major development over the past half a century has been the increased spatial segregation by income, skill, education and race. For instance, segregation by income and race could be through sorting on housing costs and hence living in the same neighborhood as people with similar characteristics, while segregation by skill and education could work through a shared working and studying environment. Altogether, these theories encompass a broad range of theories capturing the fact that individuals today face a biased set of possibilities skewed toward encountering similar partners, stemming from shared educational experiences, professional environments, or residential proximity.

In the context of our model, such a biased set of possibilities would manifest itself as prior knowledge making it easier to locate partners that have common characteristics with the person searching. More specifically, we define a prior skewed toward similarity as a step function with higher values corresponding to types that share a larger number of demographic characteristics. This skewed prior makes it easier to locate potential partners that are similar to yourself and harder to locate those that are dissimilar. As in the previous subsection, we run a counterfactual experiment where instead of costs we vary the steepness of the skewed prior. We generate the artificial matching data assuming agents have a skewed prior, but then analyze the matching rates through the lens of a model with a flat prior. We report our findings in Figure 6.6.

We find that an increase in spatial segregation leads to much smaller changes in the estimates than a reduction in costs. The counterfactual experiment predicts an increase in the median and 95th percentile of estimated elements of surplus-to-cost ratios, a mild increase in estimated selectivity, a barely noticeable increase in estimated assortativeness and a notable *decrease* in estimated *horizontality*. The intuition for the decrease in horizontality is quite transparent: as changes in the environment make it easier (cheaper in terms of information costs) to find partners with similar characteristics, agents effectively get a discount on targeting by horizontal attributes and can devote more of their search effort to targeting by vertical attributes. As a result, the estimates of overall horizontality of preferences must decline. Among individual attributes, the only exceptions could be the estimates of horizontality with respect to the most vertical attributes, which would increase.

Recall that we find no evidence of an increase in the top quantiles of the estimated surplus-to-cost ratios, but find *evidence of increased horizontality*, in direct contradiction with the results of the counterfactual experiment. Our findings could not be treated as evidence of the opposite process of de-segregation either, since that would predict a decline in assortativeness and a decrease in horizontality of vertical attributes. Thus, in the context of our model, spatial segregation simply cannot be the main driving force behind the observed trend changes in preferences and sorting that we document in Section 6.3.

6.4.3 Changes in the composition of searchers

There are four fundamental elements of the model that affect the outcomes: search costs, prior information, preferences and the number of searchers. We have excluded the first two as main drivers of long-term changes, leaving preferences and the number of searchers. In this subsection, we evaluate the mechanical effect of changes in the number and composition of searchers on the results. We show that it is not the change

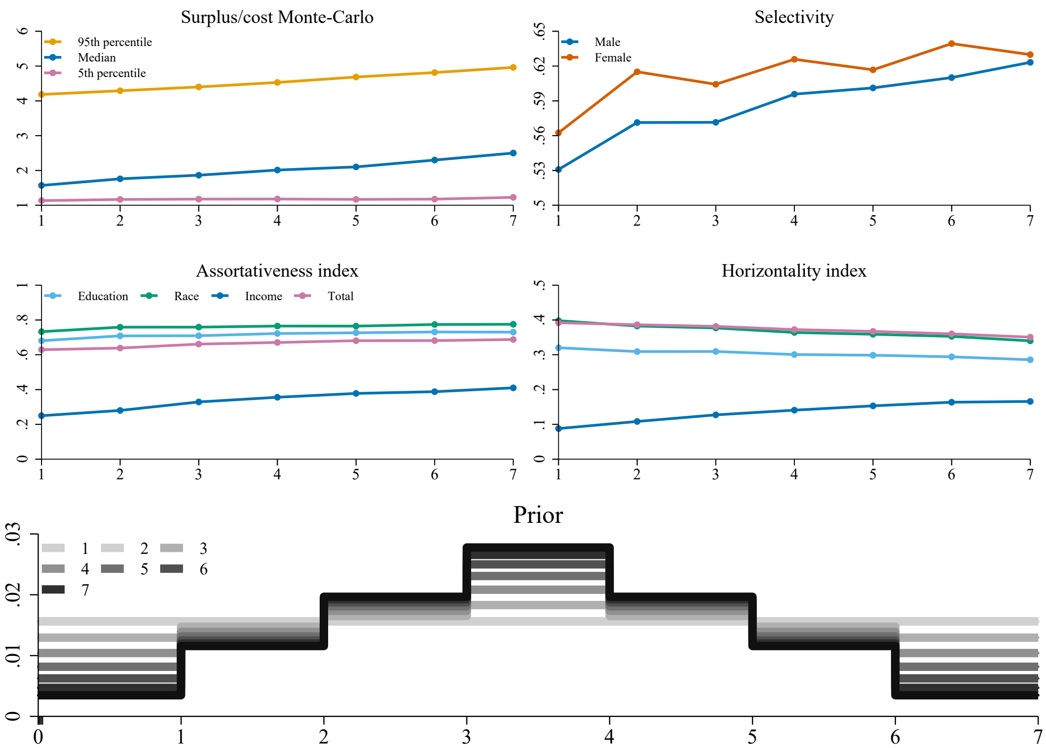

Figure 6.6: Counterfactual experiment: increased spatial segregation

in the availability of potential partners with particular characteristics but the change in their relative desirability that has been the driver of changes in sorting over time.

The long-term changes in horizontality and assortativeness that we documented earlier are suggestive of higher labor force participation and higher education attainment by women as a driving force. More specifically, we investigate how much of that long-term change could be attributed to the mechanical increase in the sheer numbers of, for example, women with higher education and higher skill levels.

To evaluate the contribution of the number and composition of searchers, we conduct a simple counterfactual experiment where we compute all the properties of sorting (in particular, measures of assortativeness and selectivity, as well as inequality as described in the next section) in an imaginary situation where agents have preferences like in the more recent period (2008-2021), but the composition of searchers with respect to one attribute is rescaled to reflect their composition in 1980.[25] The difference in

[25]We do not rescale multiple attributes simultaneously because in that case due to the sparsity of

sorting compared to the benchmark then describes the contribution of a change in the composition of searchers.

We find that although the educational composition of searchers changed substantially between 1980 and 2010, the direct mechanical effect of that change on assortativeness, selectivity and inequality is negligible. We find that the skill and income composition changed little, as we constructed the bins to be of similar size, so the observed changes cannot be attributed to those changes. The racial composition changed the most over the period of interest, but we find that it had a minor effect on the measures of assortativeness and a negligible effect on inequality. The age composition changed somewhat as the searchers became older on average. This composition change is the only one that has a somewhat noticeable effect on the assortativeness decomposition (but not on the total assortativeness) and may explain between 5 and 10 percent of the increase in inequality that we describe in the next section.

Having excluded three of the potential driving forces in the model (costs θ, priors $\tilde{p}, \tilde{q}$, number of searchers μ), leaves only changes in preferences Φ_{xy} as the main driver of the observed changes. We conclude that the vast majority of the observed long-term changes can be attributed to preferences becoming more horizontal. This change in the mutual relative desirability of males and females with different characteristics reflects both the decreased appeal of females assuming domestic roles and the increased preference for an equal or similar partner.

7 Effects of Sorting on Income Inequality

In this section we investigate the effect of marital sorting on household income inequality. To accurately measure household income inequality in the estimated model and for counterfactual matching patterns, for each combination of bins representing male and female attributes, we sample household incomes from the empirical distribution for that bin combination.[26] Differences in predicted matching rates lead to differences in the number of income draws taken from each bin combination. Using this methodology, we compare household income inequality across married couples in the data with alterna-

the matching matrix and the change in definitions of many of the attributes, the set of best match combinations changes too much, and the resulting sorting patterns have little in common.

[26] To have enough observations for each bin, we combine all the observations for the 2008-2019 period together and sample from the combined set.

tive sorting schemes, such as Positive Assortative Matching (matches formed between partners with similar characteristics) and uniform random matching. It is natural to expect PAM to increase inequality compared with random matching. This is because an increased number of matches formed between top and bottom quintiles along each attribute lead to lower inequality.

The model with multiple attributes and their interactions allows us to consider various additional counterfactuals, where people are blind to, i.e. cannot distinguish, particular characteristics. For instance, we can predict the matching rates that would have been observed if people had no information on income and could only base their search on education, skill, age and race. This is a unique feature of our framework that allows us to first estimate the interactions of attributes in search and then evaluate their effects on inequality and welfare.

We evaluate income inequality using two measures used in the literature – the Gini coefficient and the coefficient of variation. We compare household income inequality in the data with seven counterfactual scenarios: positive assortative matching (preference for alikes is amplified), blind on a single attribute (one of income, education, skill, age and race) and random matching (blind to all attributes).

Figure 7.1 illustrates our results for the 2008-2021 period. The ability of people to target their search increases inequality by 3 Gini points compared with blind random matching. About a third of that increase is due to the ability to target based on skill and education each, 18 percent based on income itself, 15 percent based on age, and only 5 percent is explained by targeting based on race. We reach a similar breakdown if we consider the coefficient of variation, which is increased by 14 points due to sorting. These counterfactuals are remarkably stable over time with only a small upward trend, another indication of little change in preferences or selectivity over time, which in turn makes the contributions to inequality stable over time.

Another interesting finding is that in the data, inequality is significantly lower than what would be produced by PAM and only marginally larger than what would be achieved by random matching. The reason for this is the mixing equilibrium of targeted search. When preferences are vertical, a lot of males and females target partners a lot wealthier (today or in the future) than themselves. This increases the number of matches between high- and low-income individuals, in some cases more than would be produced even by matching people randomly.

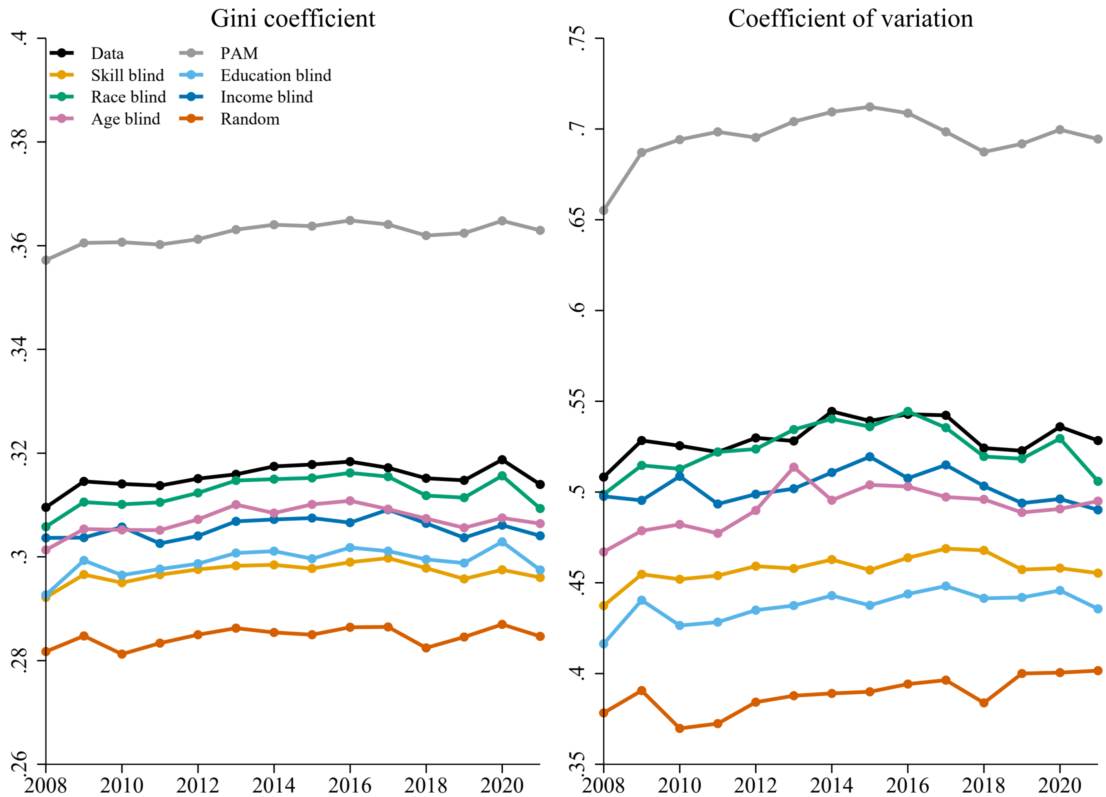

Figure 7.1: The effects of marital sorting on inequality.

We do a similar exercise for long-term estimates, keeping fixed the income sampling distribution for each bin, which represents the 2008-2021 period. We show the evolution of inequality measures for the data and for our counterfactual exercises described earlier in Figure 7.2. The fact that we keep the income sampling distribution fixed over time periods and the only source of change is the preference structure ensures that all the changes in inequality are due to changes in preferences and sorting and not to underlying changes in income or returns to attributes (for the effect of changes in the composition of searchers, see previous section).

We can draw two main lessons from Figure 7.2. First, household income inequality in the 1960s and 1980s was not driven by assortative matching, since randomly reshuffling people in various ways, including uniformly randomly drawing partners, would not lead to any change in income inequality. Second, since uniform random matching serves as a natural counterfactual for no change in preferences and explains about half

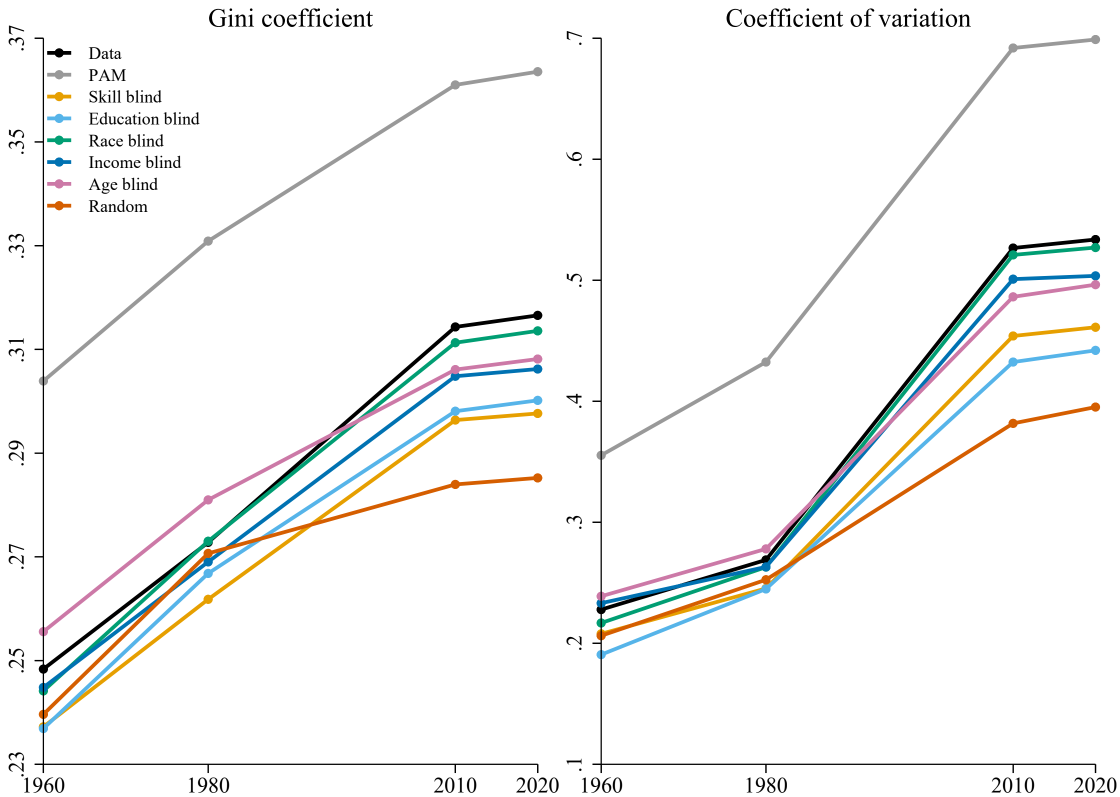

Figure 7.2: Decomposition of long-term changes in income inequality

of the increase in inequality, the other half of the substantial increase in household income inequality from 1960 to 2020 has been due to increased assortativeness. Using the counterfactual exercises described earlier, we can decompose by attribute this overall change in household income inequality, as shown in Table 3. Increased assortativeness on education and skill explains about one third each, increased assortativeness on income and age explains about one sixth each, and increased assortativeness on race plays essentially no role in the household income inequality increase in the past half a century.

8 Conclusions

We have successfully applied the model of targeted search to analyze preferences and sorting of men and women in the U.S. marriage market. For the first time using only

	Gini	CV	%
Education	0.9	5.5	34
Skill	1.0	4.3	32
Income	0.6	1.7	15
Age	0.5	2.3	15
Race	0.2	0.4	4
Total	3.1	14.2	

Table 3: Contributions to Change in Household Income Inequality

aggregate data, we document strong horizontal preferences for race, mixed preferences for age and education, and vertical preferences for income and skill. Our analysis of multidimensional sorting reveals targeting of combinations of vertical characteristics, such as income, skill and education. Effort that is put into identifying such combinations exceeds the effort applied to horizontal characteristics.

We study the evolution of preferences and search strategies over the period 2008-2021, when large changes in the search technology and methods of search have been documented. We find no change in strength or structure of preferences or in overall selectivity of individuals over this period. This suggests cognitive constraints as the main determinant of selectivity by individuals, which is in stark contrast with theories predicting increased selectivity as a result of technological improvements in labor and product market search, as suggested in the literature. Examining data since 1960, we find that overall selectivity of individuals has not increased and the shadow cost of search has not decreased, strengthening cognitive constraints as the main theory of the costs of search.

We find that over the longer term, preferences have become more horizontal, which led to near equalization of effort allocated to different attributes and a substantial increase in assortativeness, which in turn led to an increase in household income inequality. We find that the increase in income inequality over the past half a century is explained to a large extent by sorting on vertical characteristics, such as income and skill, and their interaction with education.

We consider multiple possible explanations for the observed long-term preference changes and highlight women's increased educational attainment and labor force participation as the likely culprit, shifting preferences away from spouses assuming domestic roles and toward equal partnership.

References

Altonji, Joseph G., Disa M. Hynsjo, and Ivan Vidangos, "Individual earnings and family income: Dynamics and distribution," *Review of Economic Dynamics*, 2023, *49*, 225–250.

Antler, Yair, Daniel Bird, and Daniel Fershtman, "Learning in the Marriage Market: The Economics of Dating," CEPR Discussion Papers 17065, C.E.P.R. Discussion Papers February 2022.

Belot, Michele and Marco Francesconi, "Dating Preferences and Meeting Opportunities in Mate Choice Decisions," *The Journal of Human Resources*, 2013, *48* (2), 474–508.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan, "Gender Identity and Relative Income within Households," *The Quarterly Journal of Economics*, 01 2015, *130* (2), 571–614.

Bischoff, Kendra and Sean F. Reardon, "Residential Segregation by Income, 1970-2009," in "Diversity and Disparities: America Enters a New Century," Russel and Sage Foundation, 2014, chapter 7, pp. 208–34.

Bruch, Elizabeth E. and M. E. J. Newman, "Aspirational pursuit of mates in online dating markets," *Science Advances*, 2018, *4* (8), eaap9815.

Bruze, Gustaf, Michael Svarer, and Yoram Weiss, "The Dynamics of Marriage and Divorce," *Journal of Labor Economics*, 2015, *33* (1), 123–170.

Chade, Hector, Jan Eeckhout, and Lones Smith, "Sorting through Search and Matching Models in Economics," *Journal of Economic Literature*, June 2017, *55* (2), 493–544.

Cheremukhin, Anton, Paulina Restrepo-Echavarria, and Antonella Tutino, "Targeted search in matching markets," *Journal of Economic Theory*, 2020, *185*, 104956.

Chiappori, Pierre-Andre, "The Theory and Empirics of the Marriage Market," *Annual Review of Economics*, 2020, *12* (Volume 12, 2020), 547–578.

— , **Bernard Salanie, and Yoram Weiss**, "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, August 2017, *107* (8), 2109–67.

— , **Carlo Fioro, Alfred Galichon, and Stefano Verzillo**, "Assortative Matching on Income," Working Papers 2022-07, Joint Research Centre, European Commission June 2022.

— , **Monica Costa Dias, and Costas Meghir**, "Changes in Assortative Matching: Theory and Evidence for the US," Working Paper 26932, National Bureau of Economic Research April 2020.

— , **Monica Costa Dias, and Costas Meghir**, "The Measuring of Assortativeness in Marriage," Cowles Foundation Discussion Papers 2316, Cowles Foundation for Research in Economics, Yale University December 2021.

Choo, Eugene and Aloysius Siow, "Who Marries Whom and Why," *Journal of Political Economy*, 2006, *114* (1), 175–201.

Ciscato, Edoardo, Alfred Galichon, and Marion Goussé, "Like Attract Like? A Structural Comparison of Homogamy across Same-Sex and Different-Sex Households," *Journal of Political Economy*, 2020, *128* (2), 740–781.

— **and Simon Weber**, "The role of evolving marital preferences in growing income inequality," *Journal of Population Economics*, January 2020, *33* (1), 307–347.

Cortes, Guido Matias, Nir Jaimovich, and Henry E Siu, "The End of Men and Rise of Women in the High-Skilled Labor Market," Working Paper 24274, National Bureau of Economic Research February 2018.

Couture, Victor and Jessie Handbury, "Spatial Sorting within Cities," 2022. mimeo.

Diamond, Rebecca and Cecile Gaubert, "Spatial Sorting and Inequality," *Annual Review of Economics*, 2022, *14* (Volume 14, 2022), 795–819.

Dinh, Rachel, Patrick Gildersleve, Chris Blex, and Taha Yasseri, "Computational courtship understanding the evolution of online dating through large-scale data analysis," *Journal of Computational Social Science*, 2022, *5* (1), 401–426.

Dupuy, Arnaud and Alfred Galichon, "Personality Traits and the Marriage Market," *Journal of Political Economy*, 2014, *122* (6), 1271–1319.

Eika, Lasse, Magne Mogstad, and Basit Zafar, "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, 2019, *127* (6), 2795–2835.

Ellison, Glenn and Sara Fisher Ellison, "Match Quality, Search, and the Internet Market for Used Books," Working Paper 24197, National Bureau of Economic Research January 2018.

Ermisch, John F., *An Economic Analysis of the Family*, Princeton University Press, 2016.

Fan, Jingting and Ben Zou, "The Dual Local Markets: Family, Jobs, and the Spatial Distribution of Skills," Working Paper, SSRN 2021.

Finkel, Eli J., Paul W. Eastwick, Benjamin R. Karney, Harry T. Reis, and Susan Sprecher, "Online Dating: A Critical Analysis From the Perspective of Psychological Science," *Psychological Science in the Public Interest*, 2012, *13* (1), 3–66. PMID: 26173279.

Fiorio, Carlo V. and Stefano Verzillo, "Looking in Your Partner's Pocket Before Saying Yes!; Income Assortative Mating and Inequality," Departmental Working Papers 2018-02, Department of Economics, Management and Quantitative Methods at UniversitÃ degli Studi di Milano February 2018.

Gihleb, Rania and Kevin Lang, "Educational Homogamy and Assortative Mating Have Not Increased," in "Change at Home, in the Labor Market, and On the Job," Vol. 48 of *Research in Labor Economics*, Emerald Group Publishing Limited, October 2020, pp. 1–26.

Goldin, Claudia, "Why Women Won," Working Paper 31762, National Bureau of Economic Research October 2023.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos, "Marry Your Like: Assortative Mating and Income Inequality," *American Economic Review*, May 2014, *104* (5), 348–53.

Grow, Andre and Jan Van Bavel, "The Gender Cliff in the Relative Contribution to the Household Income: Insights from Modelling Marriage Markets in 27 European Countries," *European Journal of Population*, 09 2020, *36*, 711–733.

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

Hitsch, Gunter J., Ali Hortacsu, and Dan Ariely, "Matching and Sorting in Online Dating," *American Economic Review*, March 2010, *100* (1), 130–63.

_ , _ , and _ , "What makes you click? Mate preferences in online dating," *Quantitative Marketing and Economics*, December 2010, *8*, 393–427.

Jepsen, Lisa and Christopher Jepsen, "An empirical analysis of the matching patterns of same-sex and opposite-sex couples," *Demography*, August 2002, *39* (3), 435–453.

Kalmijn, Matthijs, "Assortative Mating by Cultural and Economic Occupational Status," *American Journal of Sociology*, 1994, *100* (2), 422–452.

_ , "Intermarriage and Homogamy: Causes, Patterns, Trends," *Annual Review of Sociology*, 1998, *24*, 395–421.

Kaplan, Greg and Guido Menzio, "The Morphology of Price Dispersion," *International Economic Review*, 2015, *56* (4), 1165–1206.

Lee, Soohyung, "Effect of Online Dating on Assortative Mating: Evidence from South Korea," *Journal of Applied Econometrics*, 2016, *31* (6), 1120–1139.

Lewis, Kevin, "Preferences in the Early Stages of Mate Choice," *Social Forces*, 08 2016, *95* (1), 283–320.

Lin, Ken-Hou and Jennifer Lundquist, "Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education," *American Journal of Sociology*, 2013, *119* (1), 183–215.

Mansour, Hani and Terra McKinnish, "Same-occupation spouses: preferences or search costs?," *Journal of Population Economics*, October 2018, *31* (4), 1005–1033.

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

Martellini, Paolo and Guido Menzio, "Declining Search Frictions, Unemployment, and Growth," *Journal of Political Economy*, 2020, *128* (12), 4387–4437.

Menzio, Guido, "Optimal Product Design: Implications for Competition and Growth Under Declining Search Frictions," *Econometrica*, 2023, *91* (2), 605–639.

Moretti, Enrico, "Chapter 14 - Local Labor Markets," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, 2011, pp. 1237–1313.

_ , *The New Geography of Jobs*, Houghton Mifflin Harcourt, 2012.

Qian, Yue, "Gender Asymmetry in Educational and Income Assortative Marriage," *Journal of Marriage and Family*, 2017, *79* (2), 318–336.

_ , "Gender Asymmetry in Educational and Income Assortative Marriage," *Journal of Marriage and Family*, 2017, *79* (2), 318–336.

Rosenfeld, Michael J., "Racial, Educational and Religious Endogamy in the United States: A Comparative Historical Perspective," *Social Forces*, 09 2008, *87* (1), 1–31.

_ , Reuben J. Thomas, and Sonia Hausen, "Disintermediating your friends: How online dating in the United States displaces other ways of meeting," *Proceedings of the National Academy of Sciences*, 2019, *116* (36), 17753–17758.

Schwartz, Christine and Nikki Graf, "Assortative matching among same-sex and different-sex couples in the United States, 1990-2000," *Demographic Research*, 2009, *21* (28), 843–878.

Skopek, Jan, Florian Schulz, and Hans-Peter Blossfeld, "Who Contacts Whom? Educational Homophily in Online Mate Selection," *European Sociological Review*, 02 2010, *27* (2), 180–195.

Smith, Jeffrey A., Miller McPherson, and Lynn Smith-Lovin, "Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004," *American Sociological Review*, 2014, *79* (3), 432–456.

Thomas, Reuben J, "Online Exogamy Reconsidered: Estimating the Internet's Effects on Racial, Educational, Religious, Political, and Age Assortative Mating," *Social Forces*, 2019, *98* (3), 1257–1286.

Williams, P. L. and R. D. Beer, "Nonnegative decomposition of multivariate information," 2010. arXiv preprint 1004.2515.

Zhang, Hanzhe, "Axiomatic Measures of Assortative Matching," 2024. mimeo.

A Appendix A. Benchmarking the measure of horizontality of preferences

Here we describe how we benchmark our new measure of horizontality of preferences described in Section 4.2. Recall that the measure we use is based on spectral properties of the matrix of preferences. To capture this effective dimensionality of a matrix, we apply the Herfindahl-Hirschman concentration index measure to the distribution of its singular values. Denote the singular values of Φ_{xy} by $\{\lambda_j\}_{j\in\{1,M\}}$. We compute the concentration index as $HHI = \sum_j \frac{\left(\sum_j \lambda_j\right)^2}{\lambda_j^2}$. Then the inverse of HHI captures the effective number of different patterns of preferences in Φ_{xy}, which can take values between 1 and $\min(M, F)$. Properly rescaling this measure gives a better version of the horizontality index $H\left(\Phi_{xy}\right) = \left(1/HHI - 1\right) / \left(\min(M, F) - 1\right)$. Here we describe how we reduce this measure's dimensionality dependence by applying a Monte-Carlo technique to weighted sums of a perfectly vertical and a perfectly horizontal matrix of preferences to benchmark the values of the H-measure for different dimensionalities of the matrix.

Assume attributes take numerical values $x = \{1, ..., F\}$ and $y = \{1, ..., M\}$ and $M = F$. A perfectly horizontal surplus can be described by a diagonal unit matrix I_{xy} with ones on the diagonal and zeros otherwise. Coincidentally, this matrix has exactly M different patterns. This is reflected by M unit eigenvalues and therefore $HHI = 1/M$, which produces a re-scaled horizontality index of 1. As a representation of a perfectly vertical surplus we can use $V_{xy} = \sqrt{x * y}/M$. Since its elements are multiplicative in x and y, the rows and columns of this matrix are all linearly dependent, and it has a single non-zero eigenvalue. Therefore, its $HHI = 1$ and a horizontality index of 0.

We define a weighted matrix $S_{xy} = \Omega I_{xy} + (1 - \Omega)\, V_{xy}$. We tabulate the values of the rescaled horizontality index $H\left(S_{xy}\right)$ for a grid of values of $\Omega \in [0, 1]$ for all useful dimensionalities $M \in 2, 3, ..., 54$. This tabulation is shown in the left panel of Figure A.1. The correspondence is nonlinear and dependent on the dimensionality of the matrix. To make the measure comparable across dimensionalities and easier to interpret, in our estimation section we use this Monte-Carlo tabulation as a lookup table and report the true weight corresponding to the measured horizontality index

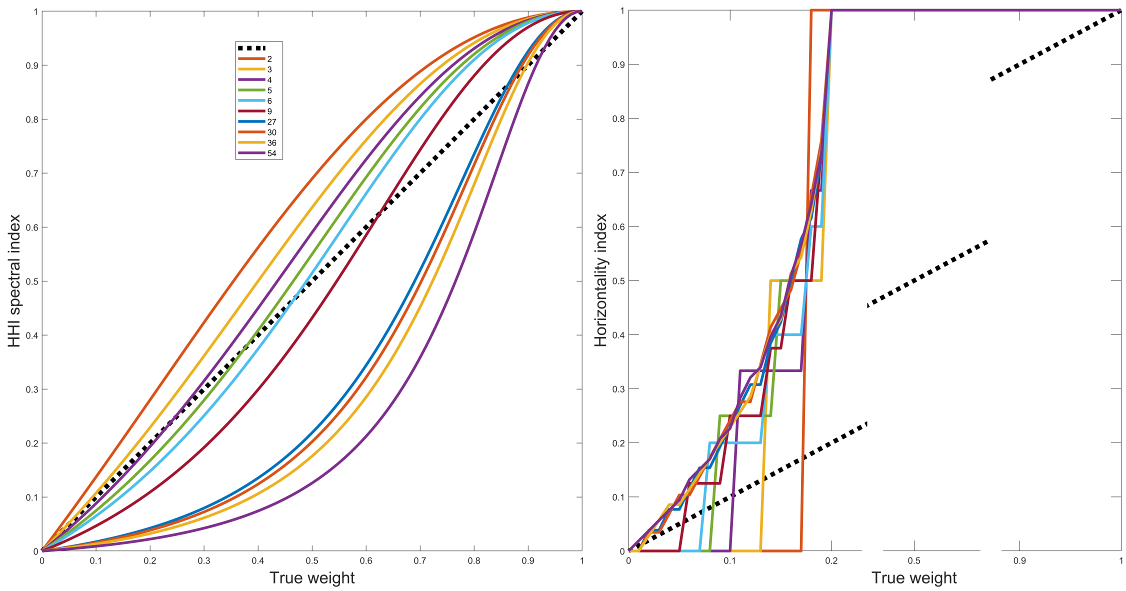

Figure A.1: Monte-Carlo tabulation for the new and old horizontality indexes

from the tabulation with the dimensionality as in the data.

Recall that another measure we proposed in our theoretical paper counted the number of different best matches present in the population. In particular, let $\omega_x = \left|\{\arg\max_y (S_{xy})\}_{x\in\{1,...,F\}}\right| \in \{1,...,M\}$ be the number of different types of males who are best matches for at least one type of female, and let $\omega_y = \left|\{\arg\max_x (S_{xy})\}_{y\in\{1,...,M\}}\right| \in \{1,...,F\}$ be the number of different types of females who are best matches for at least one type of male. Then we can define a measure of horizontality as $(\omega_x + \omega_y - 2) / (M + F - 2)$, representing the number of different best matches scaled by the total number of types. To illustrate the pitfalls of this measure, we show the correspondence between the underlying weight ω and this measure of horizontality in the right panel of Figure A.1. First, the old measure can take only a small number of discrete values and therefore is very coarse when the number of types is small. Second, it does not properly span the whole range of preferences, taking the value of 1 for true weights larger than 0.2. In contrast, the new horizontality index spans the whole set of possible preference matrices in a monotone way, and the inferred weights are comparable for matrices of different dimensions.

B Appendix B. Overview of empirical literature on assortativeness and horizontality

Most studies of the marriage market can be roughly divided into two groups. The first group explores overall matching rates and derives various measures of assortatitveness (see an extended discussion of these in Chiappori et al. (2020) and Chiappori et al. (2021)) but cannot distinguish horizontal from vertical preferences, because both lead to identical predictions of positive assortatitve matching based on existing models. The second group (e.g. Hitsch et al. (2010a), Lee (2016), Bruch and Newman (2018)) explores data from online or in person dating, which shows who is interested in whom and thus sheds light on preferences, but it typically does not contain data on who ended up matching whom. This paper uses a model to break the dichotomy – using aggregate matching rates to estimate both preferences and strategies – to simultaneously distinguish horizontal from vertical preferences and infer who targets whom in equilibrium, along the way providing new internally consistent measures of horizontality and assortativeness.

As shown in Table B.1, the literature has largely found mixed or horizontal preferences for race, education and age, and vertical preferences for skill and income. The evidence on the degree of assortativeness varies by study and by method of analysis. [27]

[27] See Rosenfeld (2008), Hitsch et al. (2010a), Hitsch et al. (2010b), Skopek et al. (2010), Lin and Lundquist (2013), Lee (2016), Lewis (2016), Bruch and Newman (2018), Thomas (2019).

	Not Assortative	Positive Assortative	Horizontal Preferences	Vertical Preferences
Education	Schwartz and Graf (2009); Smith et al. (2014); Gihleb and Lang (2020).	Ciscato et al. (2020) ;Eika et al. (2019); Ciscato and Weber (2020); Chiappori et al. (2017); Greenwood et al. (2014); Jepsen and Jepsen (2002); Mare (2016) ; Qian (2017b); Lee (2016) ; Hitsch et al. (2010a)	Belot and Francesconi (2013); Rosenfeld (2008); Hitsch et al. (2010a); Skopek et al. (2010); Bruch and Newman (2018); Hitsch et al. (2010b); Lee (2016) ; Lewis (2016)	Thomas (2019); Skopek et al. (2010); Lewis (2016)
Income		Bertrand et al. (2015); Chiappori et al. (2022); Fiorio and Verzillo (2018); Guiso et al. (2022) ; Jepsen and Jepsen (2002); Qian (2017b);		Bruch and Newman (2018); Hitsch et al. (2010a); Hitsch et al. (2010b); Lewis (2016)
Age	Ciscato and Weber (2020), CPS data 1964-2017; Chiappori et al. (2017); Schwartz and Graf (2009); Smith et al. (2014)	Ciscato et al. (2020); Ciscato and Weber (2020) ; Choo and Siow (2006); Jepsen and Jepsen (2002); Lee (2016)	Hitsch et al. (2010a); Hitsch et al. (2010b); Thomas (2019); Lee (2016)	
Race		Jepsen and Jepsen (2002) ; Schwartz and Graf (2009); Ciscato et al. (2020)	Lin and Lundquist (2013); Rosenfeld (2008); Bruch and Newman (2018); Hitsch et al. (2010a); Hitsch et al. (2010b); Lewis (2016)	Thomas (2019)
Skills	Kalmijn (1994); Mansour and McKinnish (2018);	Jepsen and Jepsen (2002);		Hitsch et al. (2010b)

Table B.1: Estimates of horizontality and sorting by attribute in the literature.

	Income	Skill	Age	Education	Race
Income	0.09	0.04	0	0	0
Skill	0	0.08	0	0.04	0
Age	0	0	0.28	0	0
Education	0	0.05	0	0.22	0
Race	0	0	0	0	0.35

Table C.1: Affinity matrix

C Appendix C. Affinity estimates

The aggregated affinity matrix in Table C.1 shows curvature of preferences. The strength of preferences represented by the diagonal elements is generally consistent with uni-dimension results presented in Table 1: preferences are strongest for race, with less importance placed on age, education, income and skill. Education is apparently more important when considered separately than when evaluated in combination with other characteristics, which suggests that education often serves as a proxy for other attributes or combinations of attributes, such as income and skill (future income). The off-diagonal elements reflecting complementarities between attributes of men and women are also instructive. We can interpret elements of the table as saying that returns to skill increase with income and returns to education increase with skill. Most of the other possible complementarities are not significantly different from zero.

D Appendix D. Sorting patterns over time

We document whether preferences for suitable candidates have changed between the first half (2008-2014) and the second half (2015-2021) of the sample in Figure D.1. In particular, we estimate preferences and strategies jointly over income, education, skill and race. Income is divided into three levels: low (L), medium (M), and high (H). Education also has three levels: school (S), college (C) or post-college (P). We use skills as proxy for future income and identify three skill levels from lowest (1) to medium (2) to highest (3). Race is split into White/Asian (W) and Black/Hispanic (B) and depicted by lighter and darker color in the Figure. Red circles indicate female searchers, and blue circles indicate male searchers, with the size of each circle reflecting the number of searchers. The arrows starting from each circle indicate the targets that agents consider

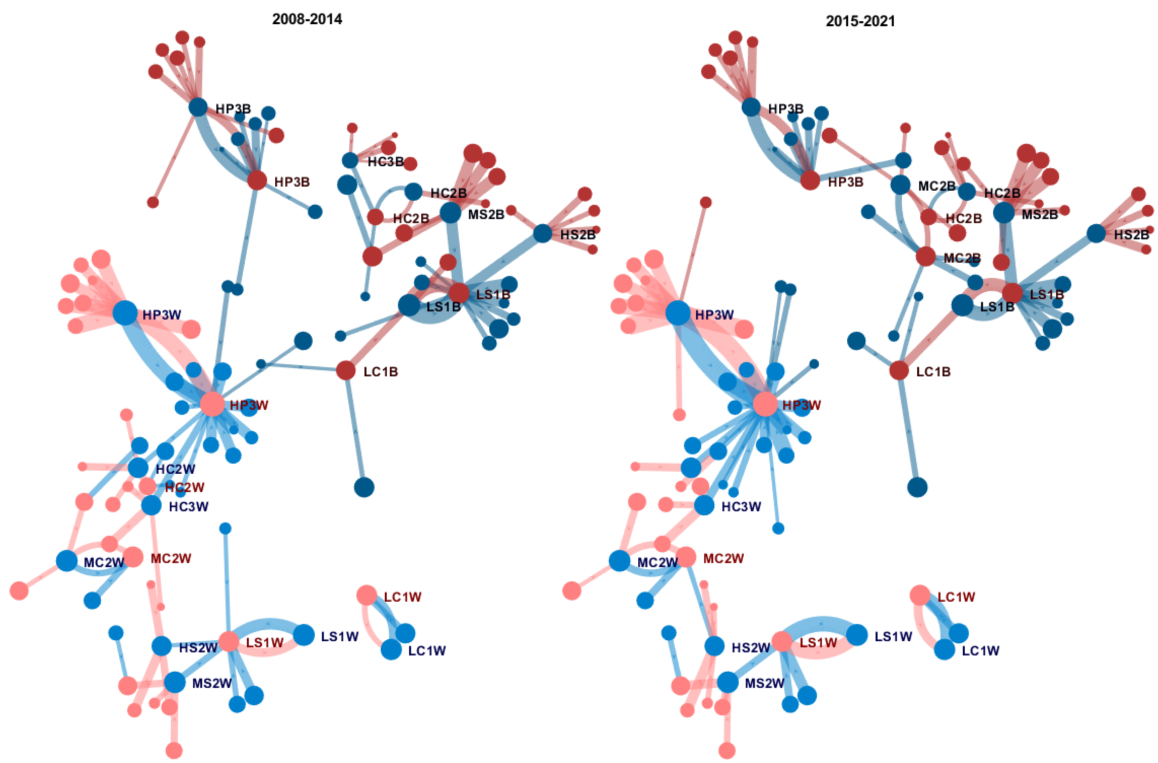

Figure D.1: Changes in strategies over income, education, skill and race

the most worthwhile, and the thickness of the arrows portrays the intensity of search. For transparency, we show labels for only the most desirable candidates that become centers of attraction.

Figure D.1 shows that both the targets and the search intensities have remained remarkably stable from the beginning to the end of the sample. Women and men have unchanged vertical preferences for income and skills: they both prefer to target potential candidates with higher income and skills than their own. Moreover, women constantly appear to put significantly more effort into identifying richer and more skilled potential partners and target their search strategies more than men do. Horizontality in race preferences is also visible throughout the sample. The combinations of attributes that are most attractive (high income, high skill, high education) illustrate how the interaction of attributes works: people search for candidates who match certain levels for all attributes simultaneously. Consistent with Figure 6.1, no significant changes in horizontality of overall preferences or preferences for attributes can be found.

Figure D.2 shows changes in the sorting structure for 1960, 1980 and the most

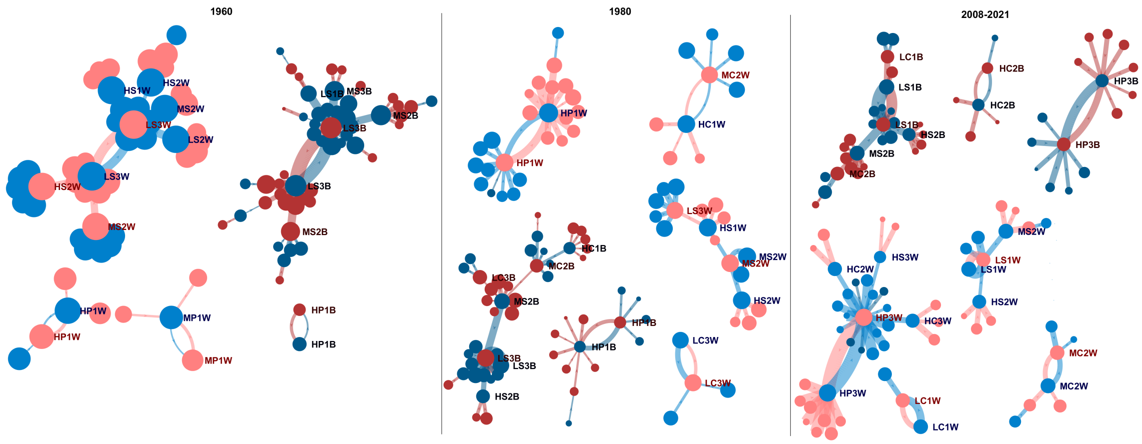

Figure D.2: Long-term changes in strategies over income, education, skill and race

recent period. We can see an increase in the number of clusters of attraction from 5 to 6 and then 7 in the most recent period, reflecting the increase in horizontality and assortativeness over time. We observe an even larger increase in the number of clusters for other combinations of attributes.