Marriage Market Sorting in the U.S.

Anton Cheremukhin*
Paulina Restrepo-Echavarria†
Antonella Tutino‡

Preliminary and Incomplete

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 analyse 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 towards 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.

*Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas TX 75201, chertosha@gmail.com
†Federal Reserve Bank of St. Louis, One Federal Reserve Bank Plaza, Broadway and Locust St., St. Louis MO 63166, paulinares@me.com
‡Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas TX 75201, tutino.antonella@gmail.com

1 Introduction

The increased popularity of online dating websites has transformed how people search for love and marriage partners.[1] Empirical analyses of online dating behavior have unveiled a plethora of important new findings on how potential mates are selected.[2] A pervasive feature of these markets is that the majority of market participants tend to aspire towards potential partners who occupy higher positions on the perceived desirability hierarchy.[3] This observation presents profound complications for the scholarly investigation of marriage markets for two primary reasons. First, the overwhelming majority of existing theoretical models fail to replicate such mate selection behavior in equilibrium, instead positing that individuals seek partners of similar rank or possessing similar characteristics. Second, mate selection choices affect aggregate sorting patterns in ways that remain beyond the reach of most current theoretical models. This limitation extends to both models of individual marital choice and those addressing aggregate marriage market sorting, which often do not specify individual selections.[4]

On the contrary, the targeted search model, developed by Cheremukhin, Restrepo-Echavarria and Tutino (2020), is a unique tool able to simultaneously analyse the selection of marriage partners, consistent with patterns observed in online data, and the implications of mate selection choices on aggregate marriage outcomes. Leveraging solely aggregate matching data, this model allows us to uncover the patterns of mate selection and understand the causes of observed sorting. In this paper, we deploy the targeted search model to delve into the U.S. marriage market,[5] with a specific focus on selection and sorting by education, age, race, income and skill.[6]

To understand the effects of the escalating prevalence of online dating in recent years, as well as the influences of longer-term demographic elements on the U.S. marriage market, we estimate the model on data from different time periods. We look at the

[1]As chronichled by Rosenfeld, Thomas and Hausen (2019), the proportion of couples who met online more than doubled from 2008 to 2017, becoming the dominant form of initial contact and constituting nearly half of all new pairings.

[2]For an overview of online dating behavior see Hitsch et. al. (2010), Finkel et. al. (2012), Brusch and Newman (2019), and Dinh et. al. (2022).

[3]See Brusch and Newman (2018) for a particularly neat measurement of this pattern.

[4]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).

[5]We restrict our attention to traditional marriages.

[6]Unfortunately, physical attributes reflecting looks and healthiness are unavailable in our dataset.

American Community Survey for every year in the 2008-2021 period for the effects of expansion of online dating. For longer-term comparisons, we employ data from the U.S. Census for years 1980 and 1960.

Our findings indicate that both the selection of partners and the resultant aggregate sorting are dictated by the configuration of preferences. If preferences are horizontal, implying an inclination towards similarity,[7] then partners face limited competition and do not have to spread a wide net. Horizontal preferences make people very selective and produce 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 and yield mixed sorting outcomes wherein the most desirable types experience an elevated matching rate from all potential partners, whereas the least desirable types scarcely secure any matches.

We find pronounced horizontal preferences with respect to race, and considerable vertical preferences with respect to income and skill, with education and age exhibiting moderate horizontal tendencies. This estimated preference structure produces considerable assortativeness by race, education and age, while evoking more attenuated sorting by income and skill, consistent with the literature. The model allows us to fully jointly identify preferences towards attributes while taking into account the frequencies and correlations of attributes in the population. This feature allows us to deal with a long-standing issue in the literature: distinguishing interactions of preferences for income, education and skill from their correlational structure in the population. Addressing this issue gains special significance in the context of understanding the effect of sorting on household income inequality, and the distinct effects of sorting by income, skill and education. We provide a comprehensive resolution to this question.

We find minimal change in preferences, mate selection or aggregate sorting over the 2008-2021 period, a perplexing finding. Given the well documented proliferation of online dating we would expect to observe a substantial improvement in the ability to find and meet potential partners, effectively reducing search costs. Assuming a reduction in search costs in the model while maintaining preferences would instigate more focused

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

mate selection, increased selectivity of partner choice, and increased assortativeness of overall matching. We show that the data conradicts these predictions and indicates an absense of change in search costs.

Juxtaposing estimates from the current epoch against those from 1980 and 1960, our findings indicate a trend towards an increased degree of horizontality in preferences, predominantly with respect to income, education, and skill. This shift increased selectivity with respect to income and skill at the expense of selectivity with respect to race. However, overall females became only marginally more selective and males became only slightly less selective. The surge in horizontality raised aggregate assortativeness, with the largest increases observed for assortativeness by income and skill.

The observed increase in assortativeness accounts for approximately half of the increase in household income inequality[8] between 1980 and 2020. The remainder of this disparity is attributable to changes in the socioeconomic composition of prospective brides and grooms and would have occured 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%) playing 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.

In addressing the drivers of these long-term changes, we evaluate and reject two potential explanations: diminished search costs and augmented spacial stratification. First, we find that the reduction in search costs cannot be the explanation of these long-term changes. The recent literature[9] 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

[8]Among households that contain a married couple.

[9]Menzio and Martellini (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 (2021) is that the decline in search frictions has been undone by the endogenous rise in selectivity of workers and firms, consumers and producers respectively.

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 spacial stratification,[10] encompasses a broad range of theories capturing the fact that individuals today face a biased set of possibilities skewed towards encountering similar partners, stemming from shared educational experiences, professional environments, or residential proximity.[11] To encapsulate this biased capacity to locate analogous partners, 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 which have not only influenced the composition of available partners – an aspect accounted for in our model – but also potentially altered the preference structure. In order to understand the influence of this broad factor on our estimates and point out specific mechanisms in 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.[12] In contrast, we find that a substantial proportion of males select partners

[10]links here

[11]This explanation includes the mechanism described by Kalmijn (1998) and Hitsch et al (2010) that the improved ability to find partners while in school or college affects observed sorting in the marriage market.

[12]Females strongly factor in potential partners' current income and skill levels (measured as the

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 towards 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.

---

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 (2006), Chiappori et. al. (2020), and Altonji et. al. (2022) among others.

—

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. To our knowledge, no paper has studied the joint incidence of several attributes on the sorting patterns and document how much time and effort is allocated to the search of individual as well as joint characteristics of a potential match.

As documented in Rosenfeld, Thomas and Hausen (2019), over this period the number of couples meeting online has more than doubled and as of 2017 was the most prevalent means of communication, accounting for almost half of all new couples.

The fourth goal goal of the paper is to evaluate the mechanism proposed by Martellini and Menzio (2020) which emphasizes the possibility that an endogenous increase in selectivity can compensate for the decline in search costs.

Preferences on race seem to be horizontal, so both males and females seem to prefer partners of the same race, consistent with the findings in Rosenfeld (2008).

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 simultaneuously 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.

We find that education together with race 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 attributes. Recent literature[13] found evidence of more interratial marriages in online dating that partially overtunrs the long-lasting horizontal preferences in race. Our results show that much cognitve 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.[14]We find that both men and women value education highly, devoting the

[13]See, e.g., Thomas (2020). Smith et al. (2014) and Lin and Lundquist (2013))

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

most effort among the attributes to this particular category. This focus on education has important implications for inequality in the U.S. and abroad.[15]

In addition, we find that the strategic interations of targeted search significantly reduce inequality across married couples compared with the benchmark of assortative matching typically assumed in the literature. Moreover, we find that strategic interactions of targeted search bring inequality a long way towards the outcome where marriages are assigned randomly. This is because targeted search produces behavior such as reaching up the desirability ladder, which generates a large number of matches between high income and low income individuals. In some cases there are more such matches than would be produced if matching was random. We find that, when compared with random matching, inequality is attributed to sorting by skill and education, with smaller contributions from current income and race.

—

The paper is related to the literature of multidimensional matching in a marriage market, recently surveyed in Chiappori (2020). The first investigation of frictionless matching with unobservable characteristics is due to Choo and Siow (2006). However, by assuming separability of the surplus and restrictions on the distributions of preferences, their model is exactly identified and cannot be tested. To lessen this shortcoming, Dupuy and Galichon (2014) build on Choo and Siow's framework with the additional assumptions of quadratic surplus and normal distribution. They use a survey of Dutch households containing information about education, height, BMI, health, attitude toward risk, and personality traits of the spouses. The estimates of the affinity matrix that defines the quadratic surplus lead to two important empirical conclusions. First, sorting occurs on several dimensions, with individuals trading-off attributes of their spouses according to their characteristics. Second different attributes matter differently for men and women. While we confirm their results on the importance of multidimensionality and differences in preferences between men and women, we do not impose restrictions on the shape and distribution of the surplus. This feature of our model allows us to fully estimate preferences and surplus and measure the contribution of each individual characteristics on the targeting decisions and outcomes.

Using Dupuy and Galichon (2014)'s framework, Ciscato and Weber (2019) use Cur-

[15]See, e.g., Skopek et al. (2011), Greenwod et al. (2014), Lee (2016), Eika et al. (2019), Ciscato and Weber (2020), Chiappori et al (2017), Fagereng et al. (2022).

rent Population Survey data to study the evolution of gains from marriage in the United States from 1964 to 2017. They find that importance of education has increased while that of age has decreased since the 1960s as confirmed in Chiappori et al. (2017). They also report that racial segregation on marriage markets has decreased from the 1960s to the 1970s but recently is slightly increasing. Chiappori et al. (2020) show that assortative matching has increased in education in the U.S. over the last decades. Our 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 relatively high degree of assortativeness, considering them jointly result in low assortativeness in the 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. Focussing also on matching in product markets, Menzio (2021) find 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 the labor markets, Martellini and Menzio (2021) report that the decline in search costs has not been matched by improvements in unemployment, labor productivity growth, vacancies and transition rates. They attribute this finding to an increase in selectivity canceling out the abeting of search frictions. Flashing out the trade-off between selectivity and declining search costs in marriage markets is the recent paper of Antler, Bird and Freshtman (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.

The paper proceeds as follows. Section 2 summarizes the theory used in the empirical part. Section 3 describes data and the empirical results. Section 4 discusses the effects of sorting on income inequality and welfare. Section 5 concludes.

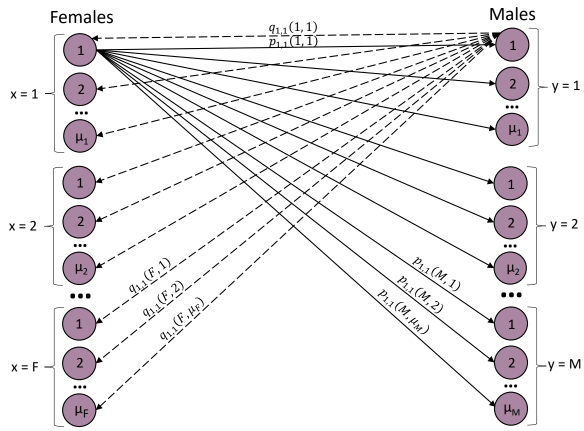

Figure 2.1: Strategies of Males and Females

2 Model of targeted search

In this section we briefly discuss the model we use for the analysis which we borrow from Cheremukhin, Restrepo-Echavarria and Tutino (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 income, age, education, and race. We define a set of types of females and males, which contains 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 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 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 nonnegative 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. We 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 female of type x is defined as follows:

$$\kappa_x = \sum_{y=1}^{M} \mu_y p_x(y) \ln \frac{p_x(y)}{\tilde{p}_x(y)}. \tag{2.1}$$

Likewise, a male's search effort defined as

$$\kappa_y = \sum_{x=1}^{F} \mu_x q_y(x) \ln \frac{q_y(x)}{\tilde{q}_y(x)}. \tag{2.2}$$

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 search effort.

To capture congestion in meetings among identical agents we introduce a congestion function $\phi_{xy} = \phi(p_x(y), q_y(x))$, 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

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

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

$$Y_x = \max_{p_x(y)} \sum_{y=1}^{M} \mu_y \varepsilon_{xy} q_y(x) \phi_{xy} p_x(y) - \theta_x \sum_{y=1}^{M} \mu_y p_x(y) \ln \frac{p_x(y)}{\tilde{p}_x(y)} \tag{2.3}$$

Likewise, a male of type y solves

$$Y_y = \max_{q_{y,j}(x,i) \in S_{y,j}} \sum_{x=1}^{F} \mu_x \eta_{xy} p_x(y) \phi_{xy} q_y(x) - \theta_y \sum_{x=1}^{F} \mu_x q_y(x) \ln \frac{q_y(x)}{\tilde{q}_y(x)} \tag{2.4}$$

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 (2.3) and (2.4) 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:

$$p_x^*(y) = exp\left(\frac{\varepsilon_{xy} q_y^*(x) \phi_{xy}^*}{\theta_x}\right) / \sum_{y'=1}^{M} \mu_{y'} exp\left(\frac{\varepsilon_{xy'} q_{y'}^*(x) \phi_{xy'}^*}{\theta_x}\right), \tag{2.5}$$

$$q_y^*(x) = exp\left(\frac{\eta_{xy} p_x^*(y) \phi_{xy}^*}{\theta_y}\right) / \sum_{x'=1}^{F} \mu_{x'} exp\left(\frac{\eta_{x'y} p_{x'}^*(y) \phi_{x'y}^*}{\theta_y}\right). \tag{2.6}$$

Second, if the congestion function takes the form $\phi_{xy} = p_x^{-\alpha} q_y^{-(1-\alpha)}$, $0 < \alpha < 1$, 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.

In the empirical section we observe the numbers of searchers, μ_x and μ_y, the matching rates, $M_{x,y}$, between each pair of types x and y. We use the model to recover the underlying preferences Φ_{xy}. For identification purposes, we further assume that 1) $\alpha = 0.5$, which implies symmetric congestion and equal split of the surplus, and 2) all agents have the same costs $\theta_x = \theta_y = \theta$, and 3) the smallest element of the matrix Φ_{xy} is normalized to 1. Using a computational algorithm and the properties of the model we can uniquely identify the ratios of preferences to costs Φ_{xy}/θ which in combination produce the empirically observed matching rates as an equilibrium of the model.

Our computational algorithm starts with an initial guess for the unknown surplus matrix Φ_{xy}, computes equilibrium strategies $p_x(y), q_y(x)$ and matching rates which correspond to the proposed surplus, and then computes the likelihood that the empirically observed matching rates are an outcome of the proposed surplus. Several standard likelihood maximization algorithms commonly used in the literature are combined to converge to a local maximum from the initial guess. The procedure is then repeated from 1000 random initial guesses to obtain the global maximum. All the estimated surpluses reproduce the empirical matching rates very closely.

3 Empirical Results

3.1 Data

To study the U.S. marriage market, we use data from the Integrated Public Use Microdata Series (IPUMS) available for 14 years from 2008 to 2021. We take unmarried males and females and (newly) married couples and assign both males and females to bins corresponding to types in the model.

We consider multiple discretizations in several important dimensions. We split the income distribution into tertiles, quintiles or deciles (two bottom deciles are merged representing zero income). We break by education into 3 unequal bins (school, college and post-college) or 2 bins (school, college). The 20-40 age range is broken into 3 or 9 equal bins. The data allow us to distinguish by race into 4 bins (white, asian, hispanic, black) or 2 bins (combining white with asian, and hispanic with black due to similarity of preferences). We also have data on occupations which allows us to sort occupations by average wages to obtain a mapping from occupation to skill level, which we break into 3 or 6 equal bins. The skill bins roughly correspond to white-collar workers (top bins), blue-collar workers (bottom bins) and services (middle bins). We consider uni- and multi-dimensional combinations of attributes and compute the numbers of single adults and marriage rates using the representative sample of the U.S. population for all couples married in the past year and unmarried males and females ages 21-40 for each of the 14 annual samples from 2008 to 2021.

3.2 Methodology

In order to describe the results of the estimation for each breakdown of the data into a combination of attributes, we develop some new measures, as well as employ some concepts and computational techniques proposed in the literature. First, as in Cheremukhin et. al. (2020), we employ measures of assortativeness of the equilibrium matching and of horizontality of preferences.

Let us denote by $P_{xy} = [p_x(y)]$ the matrix of all female strategies and by $Q_{yx} = [q_y(x)]$ the matrix of male strategies. Then let $\xi_x = \left|\{\arg\max_y (P_{xy})\}_{x\in\{1,...,F\}}\right| \in \{1,...,M\}$ be the number of different types of males that females target, and let $\xi_y = \left|\{\arg\max_x (Q_{yx})\}_{y\in\{1,...,M\}}\right| \in \{1,...,F\}$ be the number of different types of females that males target. The Assortativeness Index is then defined as $A(P_{xy}, Q_{yx}) = (\xi_x + \xi_y - 2)/(M + F - 2)$ representing the number of different targets of search relative to the maximum possible number of targets. For an assortative equilibrium where each type has a different target type, the assortativeness index equals 1, while for a mixing equilibrium, where all males have a single target type and all females have a single target type, the assortativeness index equals 0.

We distinguish horizontal and vertical preferences in a similar way. Let $\omega_x = \left|\{\arg\max_y (\varepsilon_{xy})\}_{x\in\{1,...,F\}}\right| \in \{1,...,M\}$ be the number of different types of males who are best matches for at least one type of female. Let $\omega_y = \left|\{\arg\max_x (\eta_{xy})\}_{y\in\{1,...,M\}}\right| \in \{1,...,F\}$ be the number of different types of females who are best matches for at least one type of male. Then the Horizontality Index is defined as $H(\varepsilon_{xy}, \eta_{xy}) = (\omega_x + \omega_y - 2)/(M + F - 2)$ representing the number of different best matches relative to the total number of types. We define preferences to be vertical if every type's best match is the same type, and we define preferences to be horizontal if every type's best match is a different type. Therefore, when preferences are vertical, the horizontality index equals 0, and when preferences are horizontal, the horizontality index equals 1.

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 compute each argmax in the formulas above on the subset of bins corresponding to only e.g. the income dimension to obtain estimates of assortativeness and horizontality of preferences with respect to income alone.

Another concept we introduce into the search and matching literature is a measure of selectivity by agents with respect to attributes. The amount of search effort that each agent exerts in equilibrium, defined in equations (2.1-2.2), represents how targeted towards certain types agents' strategies are, therefore measuring overall selectivity of agents. 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. In each case, selectivity represents 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 probabilitic strategies chosen by agents in equilibrium.

Finally, following Dupuy and Galichon (2014), for each estimated surplus matrix we compute an affinity matrix with respect to attributes. The affinity matrix is a quadratic form approximation of preferences with respect to attributes and captures 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 men and women.

3.3 Uni-dimensional estimates

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 dinstinguishable bins, as shown in summary Table 2. 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, mixed for education and age, and close to vertical for income and skill. Consistent with the idea that only horizontal preferences lead to assortativeness, while vertical preferences lead to looking up the desirability ladder and a mixed 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, affinity and horizontality of preferences are reflected in selectivity of individual strategies which show that people are most selective based on race, and least selective based on income and skill.

It is instructive to compare our results with the existing literature shown in Table 1, also summarized in the last column of Table 2. However, at this point it is important to note a crucial difference of this paper from the existing literature. 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, Dias and Meghir 2020, 2022), 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. Hirsch et al 2010, 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 typically does not contain data on who ended up matching whom. This paper uses a model to break the dichotomy - we are able to use aggregate matching rates to estimate both preferences and strategies - to simultaneously distinguish horizontal from vertical preferences, and infer who targets whom in equilibrium, thus, providing an internally consistent measures of horizontality and assortativeness.

The literature has largely found mixed or horizontal preferences for race, education and age, and vertical preferences for skill and income. [16]

Our findings are mostly consistent with the literature on preferences. However, the literature is largely split arguing about the degree of assortativeness in race, education and skill, and finds some assortativeness in income and age.[17] 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 to evaluate multi-dimensional estimates.

3.4 Multi-dimensional estimates

Ideally we would like to estimate an intersection of the maximum number of bins for all attributes simultaneously. However, estimating a 5832 by 5832 matrix of preferences is not only infeasible, but it would make little sense since the matrix distributing a

[16]See Rosenfeld (2008), Hitsch et. al. (2010, 2010a), Skopek et. al. (2010), Lin and Lundquist (2013), Lee (2016), Lewis (2016), Bruch and Newman (2018), Thomas (2020).

[17]See Kalmijin (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 (2017), Chiappori et al. (2017), Florio and Verzillo (2018), Mansour and McKinnish (2018), Ciscato and Weber (2019), Eika et. al. (2019), Ciscato et al (2020), Gihleb et. al. (2020), Chiappori et al (2022), Guiso et. al. (2022).

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

Table 1: Horizontality and Assortativeness in the Literature

Attribute	Bins	Assortativeness	Horizontality	Affinity	Selectivity	Literature
Income	9	0.16	0.26	0.07	0.05	PAM, vertical
Skill	6	0.24	0.31	0.13	0.06	mixed, vertical
Age	9	0.49	0.53	0.24	0.18	PAM, horizontal
Education	3	0.75	0.63	0.85	0.22	PAM/mixed, mixed
Race	4	0.97	0.99	0.82	0.74	PAM/mixed, mixed

Table 2: Uni-dimensional sorting

few million people into 34 million boxes would be extremely sparse. Therefore, we have to cut on the number of bins along most dimensions. Besides, although our estimation algorithm is very efficient, it has its limitations. In particular, even the BigTex supercomputer that we employ for estimation runs into memory limitations for surplus matrices exceeding 54 by 54. To go around this problem, we intersect various combinations of attributes with various breakdowns into bins and estimate surplus for each such combination.[18] We average the results both across years and across different estimation setups, and combine them all into representative summary Tables 3 and 4.[19]

We find that multidimensional results are in general agreement with uni-dimensional results on horizontality and assortativeness along all five attributes under consideration. Preferences are horizontal in race and education, mixed in age, vertical in skill and income, which produces stronger assortativeness for the more horizontal attributes. However, the selectivity measures are quite different compared with the uni-dimensional case. The multi-dimensional estimation uncovers some striking differences between men and women in their selectivity along income, skill and race. An even more striking result is the large fraction of the selectivity effort that is spent on the interaction between attributes, such as income, skill and education. This interaction, e.g. for men looking at skill and income, is larger than the total effort spent on the two attributes separately, and overall interactions account for roughly one third of the total selectivity effort. On the other hand, selectivity over race is much lower when interactions among attributes are considered, than when race preferences are estimated separately. These results demonstrate that strategies targeting combinations of attributes, such as high income,

[18]For instance, we consider combinations: 3income x 3education x 3skill x 2race, 3income x 3age x 3skill x 2race, 5income x 3 skill x 2race, 5income x 3age x 2race and many others.

[19]The estimation results are very similar and broadly consistent across different estimation setups as can be verified in the appendix.

Attribute	Bins	Assorta-tiveness	Horizon-tality	Selectivity Men	Selectivity Women
Total		0.33	0.35	1.18	1.39
Income	3,5	0.64	0.33	0.10	0.22
Skill	3	0.68	0.39	0.07	0.12
Age	3	0.86	0.51	0.14	0.11
Education	2,3	0.62	0.97	0.24	0.27
Race	2,4	0.98	0.97	0.18	0.25
Interactions				0.45	0.43
Income		Skill		0.18	0.14
Income		Education		0.06	0.04
Skill		Education		0.07	0.05
Income		Age		0.07	0.07

Table 3: Multi-dimensional sorting

high skill and high education, play a profound role in sorting, that cannot be accounted for with uni-dimensional studies of sorting.

The aggregated affinity matrix in Table 4 shows curvature of preferences. The strength of preferences represented by the diagonal elements is generally consistent with uni-dimension results presented in Table 2: preferences are strongest for race, with less importance placed on age, education, income and skill, in that order. 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 follows: 1) returns to skill and age increase with income, 2) returns to education increase with income and skill for women and fall with income for men, 3) race and age exhibit a positive complementarity (old is paired with white).

Since the measures that we constructed are very similar and consistent both across years and across cuts of the dataset along different combinations of dimensions, in what follows we aim at estimating all the measures of preferences and strategies described above jointly for all five attributes. To achieve this, for each dataset we first perform three separate estimations each with dimensions 54x54: 1) 3income x 3education x

	Income	Skill	Age	Education	Race
Income	0.23	0.09	0.16	-0.03	0.06
Skill	0	0.15	0	0	0
Age	0.17	0	0.48	0	0.21
Education	0.04	0.06	0	0.32	0
Race	0	0	0.06	0	0.74

Table 4: Affinity matrix

3skill x 2race; 2) 3age x 3education x 3skill x 2race; 3) 3income x 3age x 3skill x 2race. We then combine the results of these three estimation procedures to obtain results for a 168x168 split with dimenstions 3income x 3education x 3skill x 3age x 2race. This approach allows us to evaluate assorativeness, horizontality, affinity, inequality (see below), and selectivity including interaction terms for all combinations. At this point, this approach does not allow us to fully recover the preferences, however. Therefore, when considering the counterfactuals and welfare implications, we omit age.

3.5 Changes over time

An important question often discussed in the literature is that of changes in assortativeness and preferences over time. Our estimates are uniquely taylored to answer this question. The availability of information on couples that married in the preceding year is limited, however. To further extend the time dimension we could have used all married couples, as some studies do. However, we decided not to do this because then the definition of the numbers of searchers becomes unclear, and the marriages included in each sample start to overlap between samples. As discussed in the previous section, we average indexes of assortativeness and horizontality across samples, but now for each year separately. For selectivity measures, we split the interaction terms equally between attributes, so that now total selectivity equals the sum of contributions of five attributes.

We first describe changes over the period 2008-2021. The averaged series are shown in Figure 3.1. The 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

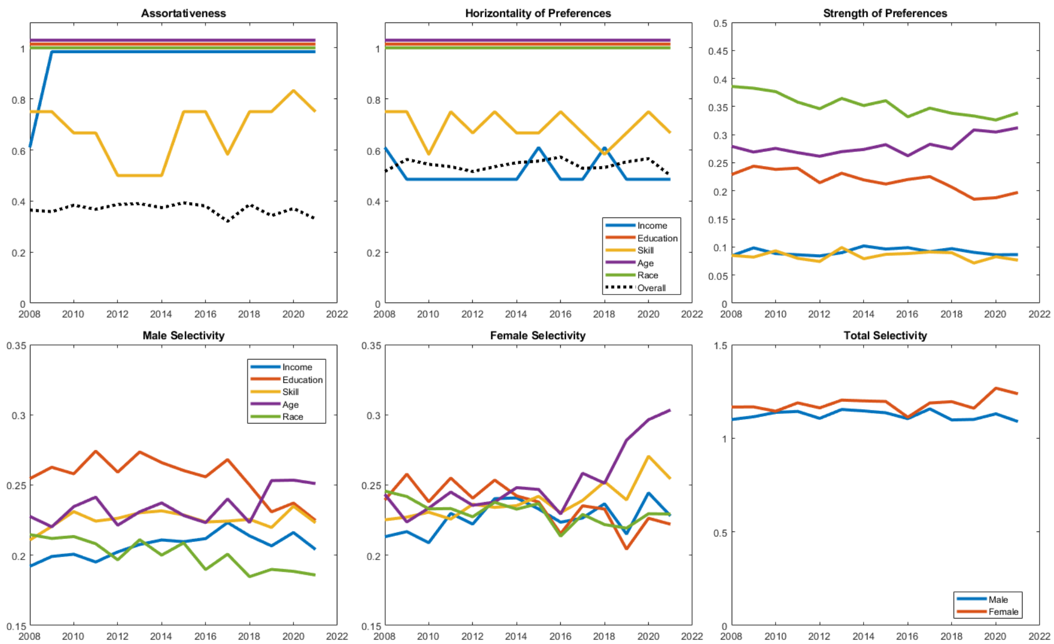

Figure 3.1: Properties of perferences and sorting over time

for age, but reduced somewhat for race and education. The degree to which females target age increased slightly, and the degree to which males target race and education declined slightly, but other components of selectivity saw little change. Although overall selectivity of females is slightly higher than that of males, both remained unchanged throughout the period we consider. This is especially strinking taking into account the fact that the methods of finding a mate changed dramatically between 2008 and 2021. In particular, in 2008 less than one in ten marriages were concieved online, while by 2018 more than half of the marriages originated online. One would expect a profound effect of such a change in the 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.

In the context of our model, the effect of online dating must show up as 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 this period, we might check whether the average values of the elements of the matrix Φ_{xy}/θ has increased over time. We have tried various approaches to computing this value and employed different statistical approaches. The resulting measures all tell the same story. Thus, we simply

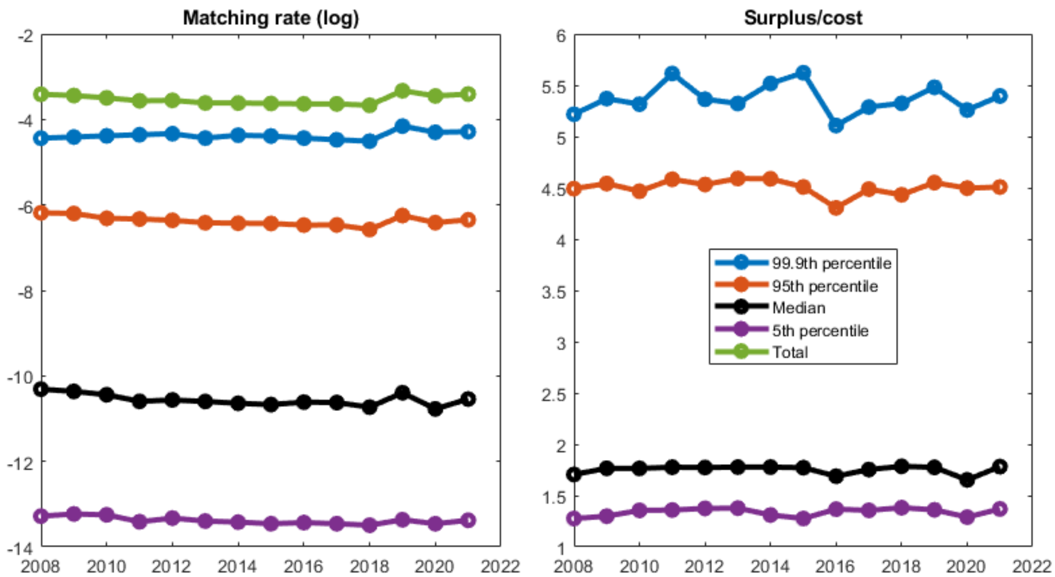

Figure 3.2: Inverse costs and selectivity over time

show quantiles of (log) matching rates and estimated surplus to cost ratios over time in Figure 3.2. We find no evidence of a significant decline in the value of costs. We find no evidence of a significant increase in selectivity, i.e. the precision with which agents are able to identify their best matches. 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.

Another explanation is that online dating platforms are a double-edge sword. On the one hand, the addition of low-cost tools to sort through candidates in an online platform gives agents access to a much wider range of potential matches and makes it easier to sort through them, discarding the ones they do not like. On the other hand, having access to many more potential candidates than previously available increase the complexity faced by the agent seeking the most suitable match.

To see this, consider the case when, prior to online dating, the agent had access to two candidates A and B. The ranking of these candidates, which consitutes the *subjective* state in which she operates, comprises of two states: A first and B second or viceversa. Now suppose that online dating gives access to four potential candidates. The ranking of these candidates expands the states from two potential outcomes to 4!=24.

For a given cost of search θ, this expanded state requires excercising more cognitive effort to establish which of the candidates is the most suitable match. Thus, it may be that the introduction of online dating has proportionally increased the expeced utility and cost of search, leaving Φ_{xy}/θ unchanged.

Martellini and Menzio (2021) suggested that significant improvements in search technology have not resulted in better and more numerous matches due to an increase in selectivity of the agents. Much like complexity, the increase in number of options available to the agents leads to an impasse rather than an increase in the numerosity of matches: in looking for quality candidates, agents are reluctant to settle for the better candidates and look for the best available prospect. The resulting *paradox of choice* manifests in fewer matches, albeit potentially better pairing. We check for evidence of this paradox in our data by measuring selectivity proxied by search effort over time and across genders and other attributes.

The bottom right panel of Figure 3.1 shows the evolution of selectivity in our data from 2008 to 2021 for men and women. As we discussed earlier, our measure of selectivity is based on search effort and its decomposition into components related to attributes and their pairwise interactions. There seems to be no empirical support for increased selectivity in our sample both across race and gender. In fact, selectivity appears to be generally stable throughout the sample. The overall stability portrayed by Figure 3.1 suggests that selectivity is an unlikely explanation for the lack of additional matches that the improved matching technology should have brought about.

Different from Martellini and Menzio (2021)'s selectivity argument, our explanation of increased complexity as defined above is perfectly compatible with a constant search effort throughout our sample. We have defined complexity as the expanded options given by technological improvements. In the example above, given the information-theoretical constraint in our model, going from a ranking of two options (2 rankings) to one of four (24 rankings) increases the initial uncertainty of the space that the agent faces as measured by its entropy from 1 to 3.2 bits of information required to perfectly detect the most suitable match. Thus, the same amount of information processed in a more complex environment leads to a lower reduction in uncertainty about potential matches than that afforded in a world with fewer options.

In our model, uncertainty is captured by the probability on which expected costs and gains from the search are based. So long as the differential in expected gains and

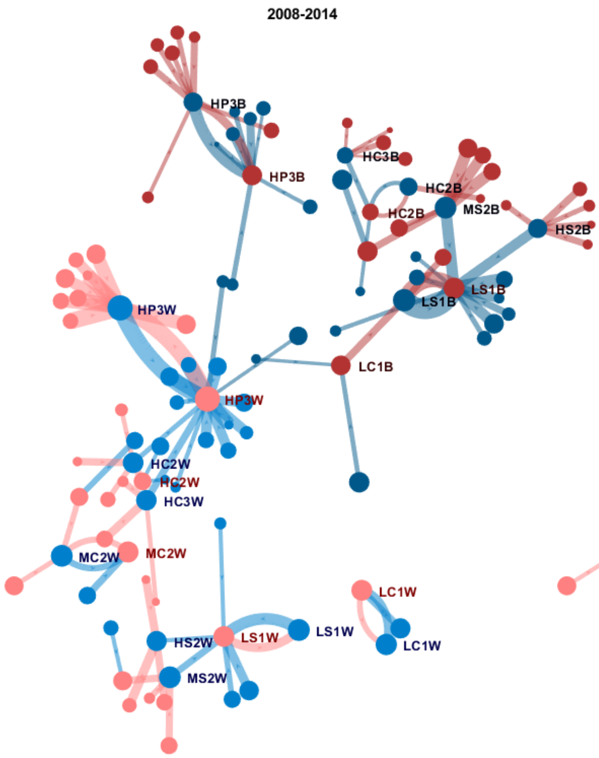

Figure 3.3: Changes in strategies over income, education, skill and race.

costs are equalized, as it appears to be the case in Figure 3.1, our model would not predict a substantial change in matches formed in the marriage market.

Next, 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 3.3. In particular, we estimate preferences and strategies joinly over income, education, skill and race. Income is divided into three levels: low (L), medium (M), 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 agents consider the most worthwhile and the thickness of the arrows portrays the intensity of search. For transparency, we show labels only for the most desirable candidates that become centers of attraction.

Figure 3.3 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 which match certain levels for all attributes simultaneously. No significant changes in horizontality of overall preferences or preferences for attributes can be found, consistent with Figure 3.1.

4 Effects of sorting on inequality and welfare

4.1 Effects 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. Differences in predicted matching rates lead to differences in the number of income draws that are taken from each bin combination. Using this methodology, we compare household income inequality across married couples in the data with alternative 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 more matches formed between top and bottom quintiles along each attribute 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 likes is amplified), blind on a single attribute (one of income, education, skill, age and race) and random matching (blind to all attributes).

Figure 4.1 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 are 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, 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 than themselves (today or in the future). 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.

4.2 Effects on welfare

Using the same counterfactuals used for study of inequality, we can evaluate the distributional impacts on expected welfare, expected matching rates, and expected income across the unmarried individuals. In Figure 4.2 we show distributions of changes in these variables aggregated across all years. Extreme assortativeness increases welfare slightly, with about two thirds of the population gaining, but reduces expected incomes and matching rates with a similar amount of people incurring a loss. Random matching reduces welfare by 7 percentage points, with 95 percent of the population experiencing

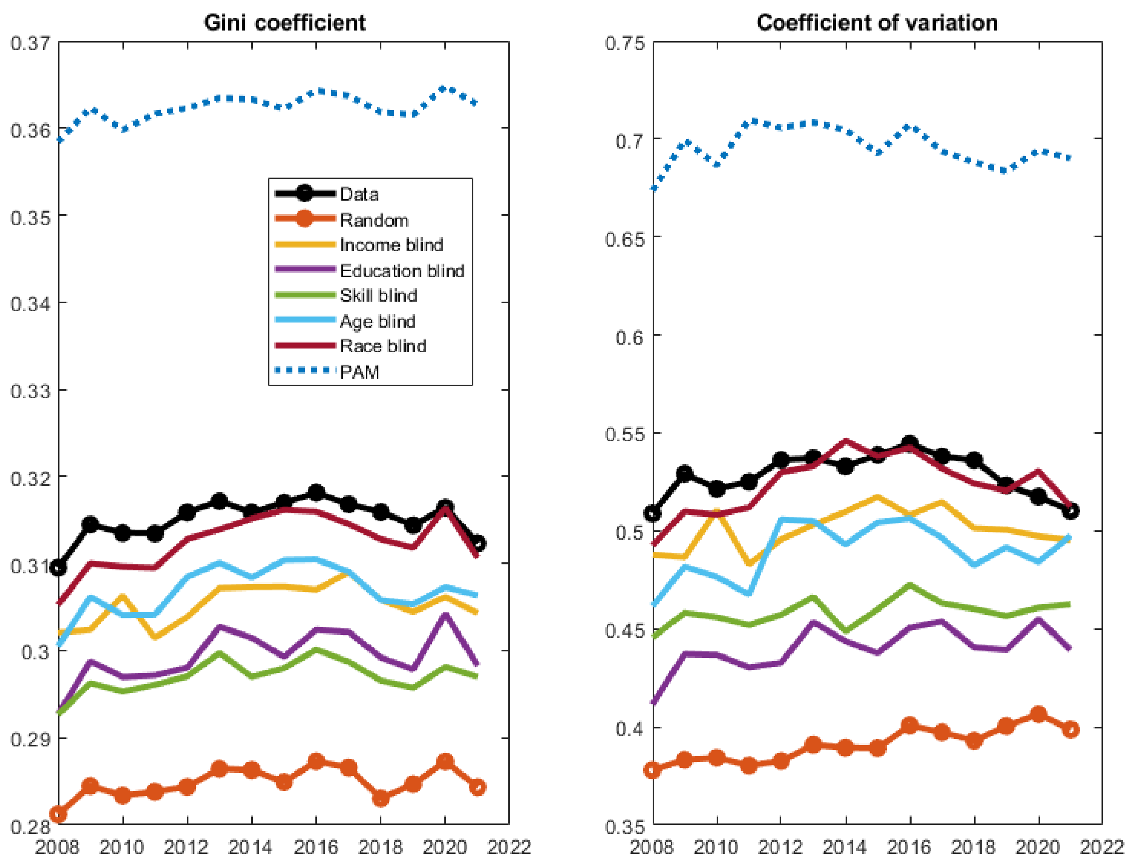

Figure 4.1: The effects of marital sorting on inequality.

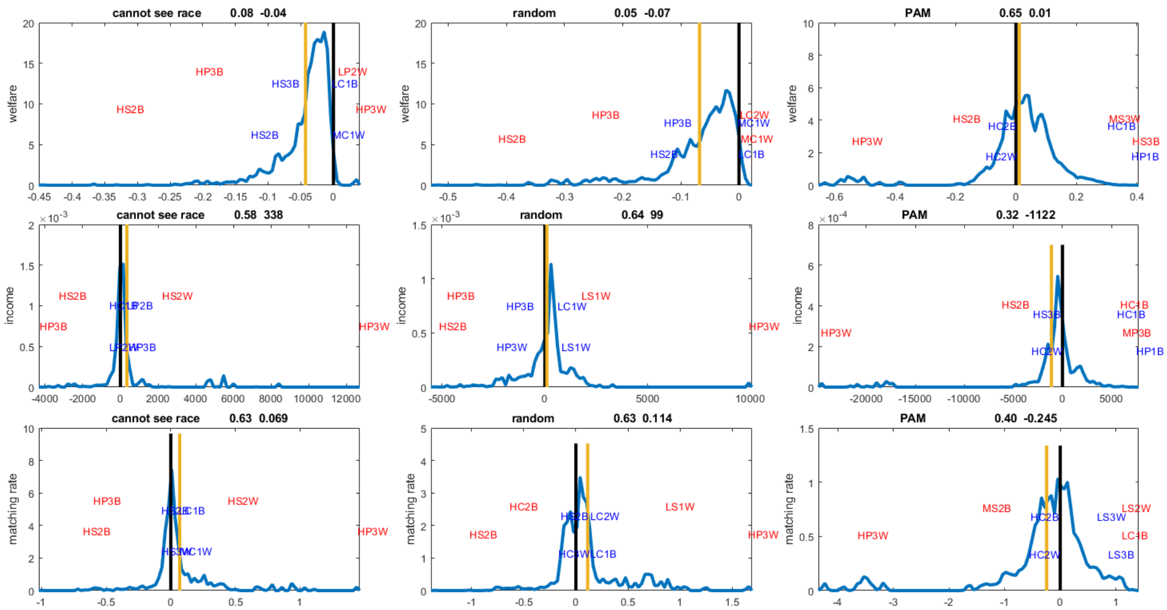

Figure 4.2: The effects of marital sorting on welfare, incomes and matching rates.

a loss, but increases incomes and matching rates for about two thirds of the population. The results of a policy that makes people race-blind (resembling diversity policies at universities and workplaces) shows a 4 percentage point deterioration in welfare and an increase in expected incomes and matching rates for more than half the population. It is notable that highly-desirable white types (e.g. high income, post-college, high-skill white women) lose the most in welfare, income and matching rate from a highly assortative allocation, but gain the most from diversity policies as well as from random matching. At the same time, the main losers from the diversity policies are high-income, high-skill non-white individuals. Because preferences for race are strong and horizontal, diversity policies, by producing more inter-racial marriages, reduce welfare of the otherwise highly-desirable non-white individuals and increase welfare of the otherwise highly-desirable white individuals.

5 Long-term changes

In order 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 couples married in the current or past year from the 1960 and 1980 census data. One important difference of these two samples compared with the more recent period is the occupational coding and therefore the set of occupations that produce higher wages and are thus considered high- or low-skill. As a result of this, it seems that in 1960 the high education low skill (P1), low education medium skill (S2), and medium eduation high skill (C3) types are among the most attractive, while in the later periods the most desirable candidates have a strong correlation between education with skill (S1, C2, P3). Apart from that, the earlier samples are aggregated using the exact same methodology. In order to compare the earlier and later periods and show the results in a transparent way, we average the two halves of the more recent subperiod into two years 2010 and 2020. Thus, all the figures that follow show four observations: 1960, 1980, 2010 and 2020.

Figure 5.1 illustrates the changes in matching rates, surplus to cost ratios, assortativeness, horizontality, affinity, selectivity and inequality, together with their components as described earlier in Sections 3 and 4. When looking at the long-term changes, a number of observations stand out. First, the matching rates have trended down over time, consistent with the decline in overall marriage rates in the US, and so have most of the quantiles of the distribution of matching rates. Consistent with that, the surplus to cost ratios declined across the board between 1960 and 1980, but remained stable ever since. This result suggests that, if anything, search costs have increased rather than declined over the long term, strengthening our conclusions from Section 3.5 regarding the importance of cognitive constraints. Another possibility is highlighted by the substantial trend decline in the strength of preferences for race and education also reflected in the decline in selectivity over both attributes by males and females alike. Moreover, while strength of selectivity on income and skill increased over the longer term, especially for females, total selectivity declined from 1960 to 1980 and stayed largely flat since then. Interestingly, while in the earlier periods males were more selective overall, females being more selective is a recent phenomenon.

Preferences show quite dramatic changes over the longer term. Interestingly, although the strength of preferences has declined markedly, their horizontality increased for all attributes. While in 1960 preferences for education and age could be described as in between vertical and horizontal, they became completely horizontal later on. At the same time preferences for income and skill remained vertical. The increase in horizontality has led to a substantial increase in assortativeness for all attributes, including race and education. This could be related to the near equalization of effort devoted to sorting on different attributes, which was largely focused on race and education at the expense of income in the earlier periods. The increase in assortativeness and the equalization of importance of different attributes may be one reason for the substantial increase in inequality shown in the bottom left panels of Figure 5.1. Indeed, while inequality in the random matching case increased over the long term as well, reflecting changes in factors exogenous to our model, the gap between the data and random matching outcome expanded dramatically over time explaining approximately half of the change in inequality. This implies that our inequality decomposition computed in the previous section can shed light on the causes of half of the increase in inequality. It seems that sorting on skill and education are the main culprits jointly explaining two thirds of the increase in inequality due to changes in sorting, which itself accounts for approximately half of the overall increase in inequality.

Figure 5.2 shows changes in the sorting structure for 1960, 1980 and the most 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. Another 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.

6 Conclusions

We have successfully applied the model of targeted search to analize preferences and sorting of men and women in the U.S. marriage market. For the first time using only aggregate data, we document strong horizontal preferences for race and age, mixed preferences for 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 from 2008-2021, when large changes in the search technology and methods of search have

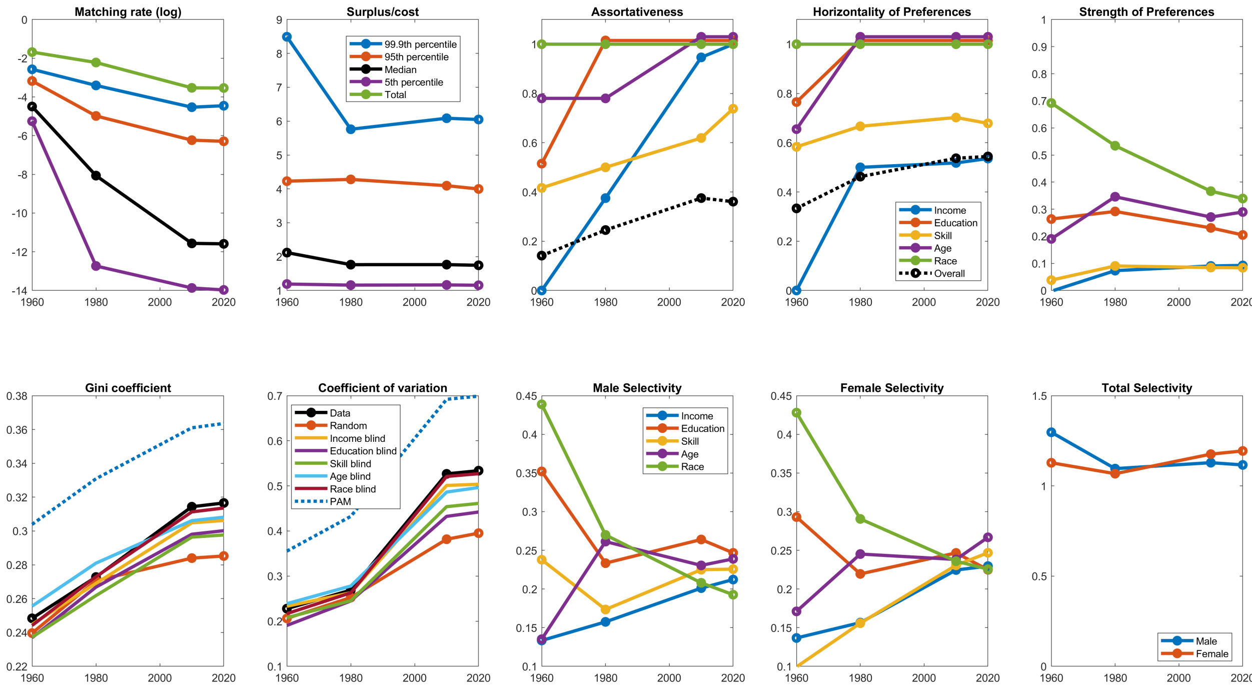

Figure 5.1: Long-term changes in perferences, sorting and inequality

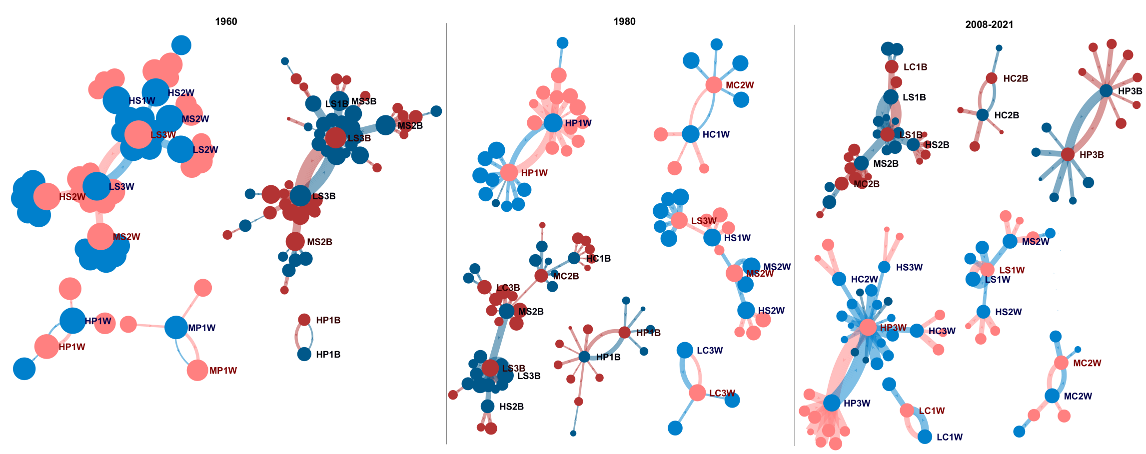

Figure 5.2: Long-term changes in strategies over income, education, skill and race.

been documented. We find no change in strength or structure of preferences, or change in overall selectivity of individuals over this period. This suggests cognitive constraints as the main determinant of selectivity by individuals, and is in stark contrast with theories predicting increased selectivity as a result of technological improvements in labor and product market search, suggested in the literature.

We find that income inequality is mainly accounted for by sorting on vertical characteristics, such as skill, education and income. Nevertheless, verticality of preferences over these characteristics implies reduced assortativeness and much lower inequality than would have prevailed if preferences over the same characteristics were horizontal. We find that diversity policies would reduce welfare for two-thirds of the population, with the main losers from the policy being the otherwise highly-desirable non-white individuals, and the main beneficiaries - the otherwise highly-desirable white/asian individuals.

We find that over the longer term, preferences have become weaker, yet more horizontal across the board, which led to near equalization of effort allocated to different attributes and a substantial increase in assortativeness which has led to an increase in inequality. A large part of the increase in inequality is attributed to increased assortativeness by skill, education and income. However, overall selectivity of individuals has decreased and the shadow cost of search has increased over the longer term, contrary to existing theories, strengthening cognitive constraints as the main theory of the costs of search.

References

[1] **Antler, Y., Bird, D. and Fershtman, D., (2022)**. "Learning in the Marriage Market: The Economics of Dating", unplublished manuscript.

[2] **Becker, G. S. (1973).** "A Theory of Marriage: Part I." *Journal of Political Economy*, 81(4), pp. 813-846.

[3] **Belot, M. and Francesconi, M., (2013**). Dating preferences and meeting opportunities in mate choice decisions. Journal of Human Resources, 48(2), pp.474-508.

[4] **Bertrand, M., E. Kamenica, and J. Pan (2015).** "Gender identity and relative income within households". The Quarterly Journal of Economics 130 (2), 571–614.

[5] **Bruch, E. E. and M. E. J. Newman (2018).** “Aspirational Pursuit of Mates in Online Dating Markets” *Science Advances*, 4(8), pp. 1-6.

[6] **Chade, H., Eeckhout, J. and L. Smith (2017).** “Sorting Through Search and Matching Models in Economics.” *Journal of Economic Literature*, 55(2), 493-544.

[7] **Cheremukhin, A., Popova A. and A. Tutino (2015).** “A Theory of Discrete Choice with Information Costs.” *Journal of Economic Behavior and Organization*, 113, pp. 34-50.

[8] **Cheremukhin, A., Restrepo-Echavarria P. and A. Tutino (2020).** “Targeted Search in Matching Markets.” *Journal of Economic Theory*, vol 185, pp. 1-43.

[9] **Chiaporri, P.-A. and B. Salanie (2015).** “The Econometrics of Matching Models.” *Journal of Economic Literature*, Forthcoming.

[10] **Chiappori, PA. (2020)**. "The Theory and Empirics of the Marriage Market". Annual Review of Economics, Vol. 12, pp. 547-578.

[11] **Chiappori, PA, Costa Dias M., and C. Meghir(2020)**. "Changes in Assortative Matching: Theory and Evidence for the US," Working Papers 2020-033, Human Capital and Economic Opportunity Working Group.

[12] **Chiappori PA, Salanié B, Weiss Y.** (2017). Partner choice, investment in children, and the marital college premium. Am. Econ. Rev. 107(8):2109–67

[13] **Choo E, Siow A. (2006**). Who marries whom and why. J. Political Econ. 114:175–201

[14] **Ciscato E, Weber S. (2019)**. The role of evolving marital preferences in growing income inequality. Work. Pap., Inst. Étud. Politiques Paris, Paris

[15] **Dupuy A, Galichon A. (2014).** Personality traits and the marriage market. J. Political Econ. 122(6):1271–319

[16] **Eika, L., M. Mogstad, and B. Zafar (2019)**. Educational assortative mating and household income inequality. Journal of Political Economy 127.

[17] **Ellison, G., and S. Ellison. (2018)**. "Match Quality, Search, and the Internet Market forUsed Books." NBER Working Paper 24197

[18] **Fagereng, A. Guiso, L. and L. Pistaferri (2022)**. "Assortative Mating in Wealth and Returns to Wealth and Implications for Wealth Inequality" unpublished manuscript

[19] **Fiorio C. and Verzillo S. (2018)** "Looking in Your Partner's Pocket Before Saying 'Yes!' Assortative Mating and Income Inequality". DEMM Working Paper 2018-02

[20] **Galichon, A. and B. Salanié (2012).** "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models." Working Paper hal-01053710

[21] **Gihleb, R. and Lang, K., (2020)**. Educational homogamy and assortative mating have not increased. In Change at Home, in the Labor Market, and On the Job. Emerald Publishing Limited.

[22] **Jepsen, L.K. and Jepsen, C.A., (2002)**. An empirical analysis of the matching patterns of same-sex and opposite-sex couples. Demography, 39(3), pp.435-453.

[23] **Greenwood, J., Guner, N., Kocharkov, G. and Santos, C., (2014)**. Marry your like: Assortative mating and income inequality. American Economic Review, 104(5), pp.348-53.

[24] **Kaplan, G. and G. Menzio (2015)** "The Morphology of Price Dispersion." International Economic Review, 56: 1165-1206.

[25] **Lee, S. (2016).** Effect of online dating on assortative mating: Evidence from South Korea. Journal of Applied Econometrics, 31(6), 1120-1139.

[26] **Lewis, K. (2016).** Preferences in the early stages of mate choice. Social Forces, 95(1), 283-320.

[27] **Lin, K.H. and Lundquist, J., (2013).** Mate selection in cyberspace: The intersection of race, gender, and education. American Journal of Sociology, 119(1), pp.183-215.

[28] **Mansour, H., McKinnish, T. (2018)** "Same-occupation spouses: preferences or search costs?". J Popul Econ 31, 1005–1033.

[29] **Mare, R.D., (2016**). Educational homogamy in two gilded ages: Evidence from inter-generational social mobility data. The ANNALS of the American Academy of Political and Social Science, 663(1), pp.117-139.

[30] **Martellini, P. and G. Menzio (2020)** "Declining Search Frictions, Unemployment, and Growth." Journal of Political Economy, 128: 4387-4437.

[31] **Menzio, G. (2021)** "Optimal Product Design: Implications for Competition and Growth under Declining Search Frictions." unpublished manuscript.

[32] **Qian, Y., (2017).** Gender asymmetry in educational and income assortative marriage. Journal of Marriage and Family, 79(2), pp.318-336.

[33] **Rosenfeld, M., Thomas. R. and S. Hausen (2019).** "Disintermediating your friends: How online dating in the United States displaces other ways of meeting" *Proceedings of the National Academy of Sciences of the United States of America.*

[34] **Rosenfeld, M. J. (2008).** "Racial, Educational, and Religious Endogamy in Comparative Historical Perspective", Social Forces volume 87, issue 1, pages 1-32

[35] **Skopek, J., Schulz, F. and Blossfeld, H.P., (2011**). Who contacts whom? Educational homophily in online mate selection. European Sociological Review, 27(2), pp.180-195.

[36] **Smith, L. (2006).** "The Marriage Model with Search Frictions," *Journal of Political Economy*, 114(6), pp. 1124-1146.

[37] **Smith, J.A., McPherson, M. and Smith-Lovin, L., (2014**). Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. American Sociological Review, 79(3), pp.432-456.

[38] **Thomas, R. J. (2020)**. "Online Exogamy Reconsidered: Estimating the Internet's Effects on Racial, Educational, Religious, Political and Age Assortative Mating," Social Forces

[39] **Williams, P. L., and Beer, R. D. (2010)**. Nonnegative decomposition of multivariate information. arXiv preprint arXiv:1004.2515.