Assessing racial and educational segmentation in large marriage markets

Edoardo Ciscato*

April 25, 2023

Abstract

Complementarities between partners' characteristics are often held responsible for the patterns of assortative mating observed in marriage markets along different dimensions, such as race and education. However, when the marriage market is segmented into racially and educationally homogeneous clusters, people naturally have more match opportunities with their likes. In this paper, we build an empirically tractable dynamic matching model with endogenous separation and remarriage. In every period, agents participate in a competitive assignment game in the vein of Choo and Siow (2006), where mating strategies depend on both the expected match gains and search frictions in the form of meeting costs. We leverage panel data on the duration of both non-cohabiting and cohabiting relationships to jointly estimate both determinants of assortative mating with a nationally representative sample of the U.S. population. We show that, in the absence of search frictions, the share of matches between people of the same race (education) would decrease from 88.2% (49.2%) to 55.5% (40.8%), as opposed to 53.3% (33.5%) if singles were randomly matched. As a result, search frictions explain nearly all the racial homogamy observed in the data, but only approximately half of the observed educational homogamy, with the other half attributed to match complementarities. In a counterfactual exercise, we show that minority groups experiencing an unfavorable gender ratio when marriage markets are segmented, such as Hispanic men and black women, would benefit from access to a broader and more diverse pool of partners.

Keywords: marriage markets, divorce, assortative mating, matching models, cultural segregation.

JEL Classification: D13, J11, J12.

*KU Leuven. Email: edoardo.ciscato@kuleuven.be. Website: https://sites.google.com/site/ciscatoedoardo.
I thank Pierre-André Chiappori, Alfred Galichon and Jean-Marc Robin for their guidance. I am grateful to Fabio Blasutto, Laurens Cherchye, Arnaud Dupuy, Marion Goussé, Costas Meghir, Frederic Vermeulen, and Simon Weber for their useful suggestions. I also thank all participants at various seminars and conferences. I acknowledge financial support from the Research Fund KU Leuven through the grant STG/19/019. All errors are my own.

1. Introduction

Marriages between people with similar traits, such as education and race, are predominant (Schwartz, 2013). These patterns of assortative mating have been linked with cross-sectional and intergenerational economic inequality (Fernández and Rogerson, 2001), as well as cultural segregation (Bisin and Verdier, 2000). Marital homogamy was famously rationalized by Becker (1973) in his economic theory of the family. Becker claims that technological complementarities in the household lead to positive assortative mating in marriage markets. For instance, families where parents cooperate in educating their children can enjoy higher gains if both parents are highly educated, thus leading to positive assortative mating in this dimension (Chiappori et al., 2017). Similarly, when parents wish to pass on their preferences to their children, they can benefit from sharing a similar social, religious, ethnic, or racial background, thus leading to homogamy along additional dimensions (Bisin et al., 2004; Dohmen et al., 2012). Hence, the nature of the gains from marriage, and changes therein, can explain the patterns of assortative mating we observe in the data.

However, marital homogamy can also result from the presence of frictions that make meeting opportunities between likes more frequent, a hypothesis that has been advanced by social scientists in different fields (Kalmijn and Flap, 2001). Since Becker's analysis, many have emphasized that searching for a mate takes time and effort.[1] When marriage markets are segmented into homogeneous clusters, people will be more likely to consider a homogamous relationship even if it does not produce additional gains. In concrete terms, when neighborhoods, workplaces, schools, and friendship networks are characterized by low diversity, people will naturally have more match opportunities with their likes. Understanding to what extent assortative mating is shaped by these barriers has important policy implications. A segmented marriage market can reinforce assortative mating, and therefore strengthen economic inequality and cultural segregation, without generating any additional welfare gains from optimal sorting, which are often associated with technological complementarities within the household.

In this paper, we ask to what extent marriage markets are segmented along multiple dimensions, including education and race, and to what extent patterns of positive assortative mating are shaped by search frictions as opposed to complementarities between partners' inputs. This question is challenging because, in the data, we typically observe realized matches only, and we do not have information about the set of potential partners that individuals consider in their search for a mate. To overcome this limitation, previous studies have collected

[1]Mortensen (1988) and Oppenheimer (1988) were among the first to insist on the importance of search in marriage markets. Burdett and Coles (1997), Shimer and Smith (2000), and Smith (2006) are seminal papers of a rich theoretical literature discussing the implications of introducing search frictions for mating patterns.

data from speed-dating experiments or online dating platforms in order to better understand the respective roles of mate preferences and search mechanisms (Fisman et al., 2006; Belot and Francesconi, 2013). This literature, reviewed in Section 2, documents that individuals do have preferences for their likes, although these preferences alone may not entirely explain the high degree of assortativeness observed in large-scale data (Hitsch et al., 2010). On the other hand, other papers use administrative data on college education to show that educational institutions play an important role in shaping mating patterns, and that the marriage market is actually segmented into a constellation of local markets (Nielsen and Svarer, 2009; Kirkebøen et al., 2021).

While there is evidence that both the nature of the gains from marriage and meeting opportunities shape the patterns of assortative mating, assessing the relative importance of these two forces along multiple dimensions and in nationally representative samples remains a difficult task. In addressing this question, this paper provides three contributions: first, it discusses the identification of search frictions in a broad class of matching models leveraging panel data on relationship duration. Second, it builds an empirically tractable dynamic bipartite one-to-one matching model with Transferable Utility (TU), endogenous separations, and remarriage. Third, it uses nationally representative U.S. data on both non-cohabiting and cohabiting relationships to estimate such a model and measure the degree of marriage market segmentation with respect to age, race, and education.

Matching models that explicitly take frictions into account are not identified with data on the matching outcome only (e.g., data on newly hired workers or newly formed couples). This has been well understood in the search and matching literature at least since Flinn and Heckman (1982). In Section 3, we argue that this is a pervasive problem, regardless of the assumptions made about mate search (e.g., random vs. directed, sequential vs. simultaneous). To address this issue, the structural estimation of frictional matching models of the marriage market often relies on relationship duration data for identification (Bruze et al., 2015; Goussé et al., 2017). On the other hand, a large body of literature measuring assortative mating has remained silent about the role of search frictions in the determination of sorting patterns (Choo and Siow, 2006; Dupuy and Galichon, 2014; Chiappori et al., 2017).

Our first contribution is to combine insights from these two strands of literature and argue that panel data can also be exploited to disentangle the two competing mechanisms behind assortative mating. How this works can be understood through the following example. In our data, we observe that, in a large number of new couples, partners have the same education. Hence, either these couples enjoy high match gains, or it is easier for people with similar education to meet each other. Yet, if we also observe the stability of realized matches, we can disentangle the two explanations. If educationally homogamous couples display low separation rates, then the high match rates are explained by large match gains. Alternatively, if homogamous couples display high separation rates, then these matches do not produce particularly large gains, and their prevalence is due to weak frictions among likes.

Two key aspects of this approach are discussed in the paper. First, the longer partners stay together, the more their stakes in the relationship are likely to diverge from those of couples who have just matched. As couples move in together, get married, and have children, both the value they attach to the relationship and their outside options in case of separation change. Hence, the behavior of recently formed couples who have not yet progressed in their relationship is the most informative about the initial match gains, and thus about search frictions intervening in the matching process. This explains why we use data on the stability of non-cohabiting relationships in our empirical application. Second, we can expect couples with better chemistry to be more stable and self-select into later stages of the relationship. When unobserved match quality is persistent, low-quality matches dissolve at a faster rate. Using existing results from the literature on finite mixture models (Kasahara and Shimotsu, 2009; Arcidiacono and Miller, 2011), we take account of this dynamic selection process and discuss how to infer the latent structure of unobserved heterogeneity from the residual duration dependence patterns observed in the data.

Our second contribution is to build a dynamic matching model of the marriage market where agents are heterogeneous with respect to traits such as age, education, and race, and can match and split with different partners throughout the life-cycle. In every period, singles are free to meet and match in a one-to-one bipartite assignment game with TU à la Choo and Siow (2006). Yet, differently from their canonical setup, sorting is determined by both expected match gains, which account for potential complementarities between partners' inputs, and search frictions. As in Jaffe and Weber (2019), the latter take the form of meeting costs and may discourage individuals from crossing barriers across socioeconomic and demographic groups, causing the market to be imperfectly segmented. If an agent does not match in a given period, he/she will be able to continue his/her search in the next. Two individuals who start dating learn about their match quality, which has both a persistent and a volatile component. Couples may decide to move in together to enjoy larger gains from their relationship and make it more stable. However, they may also decide to break up if their relationship deteriorates, in which case they rejoin the pool of singles and can look for a new partner. Rematch prospects shape outside options and partly determine breakup decisions. Aging plays a key role in these intertemporal trade-offs. While relationships between individuals of different ages are possible, agents will typically experience a narrower market as they get older.

While the majority of empirically tractable dynamic matching models assume mate search to be random and sequential (Wong, 2003b; Greenwood et al., 2016; Goussé et al., 2017), our model studies the role of search frictions in an environment where agents can simultaneously screen multiple potential partners. Hence, mating strategies are not limited to acceptance/rejection decisions, but result from trade-offs between the expected match gains, net of meeting costs, across all types of partners.[2] The model also allows search frictions to be pair-specific and possibly weaker for similar individuals. On the other hand, our paper extends the dynamic matching models of Choo (2015) and Chen and Choo (2023) by introducing frictions, endogenous separations, and remarriage.[3]

Our third contribution is empirical. We use data from the How Couples Meet and Stay Together (HCMST) survey, a nationally representative panel study of American adults run between 2009 and 2014 by Rosenfeld and Falcon (2018), where respondents are asked not only about their current relationship status and their partner's traits, but also where and how they met him/her. Importantly, respondents provide information about both cohabiting and non-cohabiting relationships. In our model, agents differ in terms of age, race, and education. Our estimates show that interracial couples experience large meeting costs, while evidence in favor of same-race preferences is weak. On the other hand, educational complementarities are strong, while meeting costs are only slightly higher for agents with the same education. These conclusions are unchanged when we use a more refined categorization for race and education.

Using additional data from the HCMST survey, we show that the structure of search frictions partly reflects segmentation across space, educational institutions, and friendship networks. Interestingly, not all forms of segmentation favor the formation of high-surplus matches. While the intermediation of friends and social life in college campuses both reduce search costs and help agents find better matches, looking for a partner among acquaintances from the same hometown or high school is inexpensive, but not necessarily rewarding in terms of relationship gains. We also show that certain channels, such as looking for a partner at a party or on an online platform, are costly but lead to high-surplus matches, reflecting the agents' deliberate efforts to look for their preferred types.

[2]While similar in many aspects, our model and canonical directed search models differ in the way uncertainty is resolved, and thus in how market clearing works (Eeckhout and Kircher, 2010; Chade et al., 2017). In the former, market clearing conditions are imposed at the aggregate level. In the latter, every matching round is divided into two stages, information about partners is only fully revealed in the second stage, and market clearing is local.

[3]The works of Fox (2008) and Bruze et al. (2015) go in the same direction, but model frictions as switching costs and rule out persistent match quality. Corblet et al. (2023) deals with repeated TU games but does not consider frictions.

Further, in order to quantify the role of search frictions, we can calculate how individuals would sort if search frictions were absent and agents chose their partners entirely based on the expected match gains. In this counterfactual scenario, the share of racially homogamous matches would drop from the observed 88.2% to 55.5%. In comparison, if matching were random, this share would be 53.3%, which suggests that racial homogamy could almost entirely be explained by search frictions. On the other hand, in the absence of search frictions, the share of same-education matches would drop from 49.2% to 40.8%. If matching were random, the share of same-education matches would be 33.5%, which suggests that search frictions alone could account for half the racial homogamy observed in the data.

Finally, we take the counterfactual analysis one step further and compute the counterfactual steady-state equilibrium (SSE) of the marriage market after removing racial segmentation, i.e. after equalizing meeting costs for same-race and different-race pairs. At the new SSE, agents adjust their mating strategies taking into account changes in the size and composition of the pool of available partners. Due to weaker frictions overall, the new SSE is characterized by higher turnover, shorter relationships, and a lower share of partnered individuals. However, these changes are not equal for all and are shown to be beneficial for groups that experience an unfavorable gender ratio within their race group, such as Hispanic men and black women. In the presence of racial segmentation, these groups suffered from the scarcity of potential same-race partners, which also affected their bargaining power in relationships. When racial segmentation is removed, they benefit from a broader and more diverse pool of potential partners, can afford to be more selective, and are empowered in relationships. In the counterfactual steady-state equilibrium, the gap in the odds of having a partner between white and black women goes from 13.7% to 7.8%, i.e. 43.1% of such a gap is closed after removing racial segmentation.

The paper is organized as follows. Section 2 briefly reviews the literature that has empirically investigated the driving forces behind assortative mating and the literature that has dealt with the structural estimation of search frictions in marriage markets. Section 3 discusses the identification of the match surplus function and search frictions with data on both relationship formation and dissolution in a broad class of matching models. Section 4 introduces an empirically tractable dynamic matching model with endogenous separations. Section 5 describes the dataset used in our analysis, the parametric specification, and the estimation method. Section 6 presents the empirical findings and counterfactual simulations. Section 7 concludes.

2. Literature review

Demographers, economists, and sociologists have extensively documented the patterns of positive assortative mating (PAM) in the U.S. and around the world; Schwartz (2013) provides a multidisciplinary review of this literature. In recent years, the link between assortative mating and economic inequality has sparked interest in the measurement of the strength of sorting - particularly on education - among economists. In a seminal paper, Choo and Siow (2006) develop a metric for assortativeness that can easily be computed with cross-sectional matched data and compared across markets and over time. Dupuy and Galichon (2014) extend their method in order to measure the strength of sorting in multiple dimensions, while Chiappori et al. (2017) use it to measure the time trends of educational assortativeness in the U.S. Eika et al. (2019) and Chiappori et al. (2020) compare different measures of assortativeness and discuss the relationship between educational assortativeness and economic inequality.

This applied literature is grounded in previous theoretical work: both Becker (1973) and Oppenheimer (1988) stress that changes in mating patterns reveal changes in the nature of the gains from marriage. Hence, researchers have looked at households in order to understand the origins of PAM. For instance, Greenwood et al. (2016), Cherchye et al. (2017), and Chiappori et al. (2017) stress the importance of technological complementarities in the home production of public goods, whereas Goussé et al. (2017) jointly look at time use, home production, and non-economic gains. In the same spirit, building on the theoretical work of Bisin and Verdier (2000), Bisin et al. (2004) and Bisin and Tura (2019) look at technological complementarities in the cultural socialization of children in order to rationalize homogamy along ethnic and religious dimensions.

In this paper, we ask whether search frictions lead to assortative mating through marriage market segmentation. While reviewing theoretical explanations for the trends of interracial marriage in the U.S., Fryer Jr (2007) lists search frictions as a potential channel. However, quantifying their role as opposed to other channels (such as technological complementarities) is challenging because we usually lack data on who meets whom. Without these data, we cannot elicit both mate preferences and search frictions using only cross-sectional data about partner choices, as argued by Jaffe and Weber (2019). Hence, drawing insight from the psychology and sociology literature,[4] Fisman et al. (2006) measure mate preferences about traits such as intelligence and physical appearance using data collected during a speed dating experiment, where the researcher observes who is paired with whom in the rotation. Their findings point to gender asymmetries in mate preferences, and provide evidence that men are less attracted to women who are more ambitious or intelligent than them. Using the same data, Fisman et al. (2008) show that their speed-dating participants prefer partners of the same race. Similarly, Hitsch

[4]Kalmijn and Flap (2001) use survey data in order to document when and where couples meet, while Rosenfeld and Thomas (2012) collects new data of the same kind for the internet era. Eastwick et al. (2014) provides a meta-analysis of the empirical findings about mate preferences in the psychology literature, which mainly relies on data collected through dedicated surveys about *ideal partner preferences*.

et al. (2010) rely on user data from a dating website to elicit mate preferences and, among many results, find that both men and women prefer partners with similar race and education. They also show that, in the absence of search frictions, mate preferences alone can generate patterns of assortative mating quantitatively comparable to those observed in nationally representative marriage data along many dimensions, but not education and race. Hence, they suggest that search frictions could explain the gap between their model predictions and the data. Belot and Francesconi (2013) also elicit mate preferences using data from speed-dating sections, but leverage the fact that daters go through a random and sequential meeting process, and only meet a relatively small number of potential partners (if compared to dating websites). They show that individual mating strategies are largely determined by the composition of the pool of potential partners rather than by their own preferences.

Another approach consists of looking at the role played by institutions, and in particular educational institutions, in shaping mating patterns by acting as local marriage markets. Nielsen and Svarer (2009) use Danish administrative data on educational achievements in order to assess how search frictions, measured by the geographic distance between institutions, have an impact on the choice of the partner. They find that geographic proximity between individuals with similar levels of education can account for about half of the observed degree of educational assortativeness in the data. Kirkebøen et al. (2021) compare the marital decisions of individuals that ended up in different educational institutions as a result of the centralized admission process used by Norwegian universities, but are otherwise similar. They find that the choice of a partner is largely driven by the meeting opportunities in the local campus rather than by an individual's background traits, consistently with the coexistence of multiple local marriage markets. Using very different data, Goñi (2022) studies a highly institutionalized marriage market with well-defined and observed boundaries, that of the British aristocracy in the 19th century. He shows that, when these boundaries are suddenly removed, sorting on socioeconomic status becomes weaker.

Search frictions play a key role in equilibrium search-and-matching models of the marriage market. When agents only differ in one dimension, stronger frictions are usually associated with weaker sorting (Shimer and Smith, 2000). In applied work, search frictions have been used to rationalize the overall marriage market turnover observed in the data (Wong, 2003b; Greenwood et al., 2016; Goussé et al., 2017). We build on this literature and discuss in Section 3 how data about both marital inflows and outflows, i.e. both marriage and divorce rates, jointly identify the gains from marriage and search frictions. Differently from the previous literature, we allow for heterogeneous search frictions in our model and employ a similar identification strategy to study the role of frictions in shaping assortative mating along multiple dimensions.[5] For

[5] A few papers have already introduced heterogeneous search frictions. Díaz-Giménez and Giolito (2013)

instance, after estimating a random search-and-matching model of racial sorting, Wong (2003a) finds that removing same-race preferences would increase the share of interracial couples from 5.5% to 64%, but assumes search frictions to be homogeneous for all pairs. In this paper, we relax this assumption and show that search frictions are indeed much stronger across racial groups, and thus play a key role in explaining racial homogamy.

3. Identification

In this section, we first provide a general characterization of the matching function, where the matching depends on both expected gains and search frictions. As argued by Flinn and Heckman (1982), in the presence of search frictions, the matching function is underidentified if only cross-sectional data on the matching outcome are available. Hence, we discuss how panel data on relationship duration can be used to solve this underidentification problem and provide sufficient conditions on the couple's joint dynamic model structure in order to recover the initial match surplus from observed relationship histories. We highlight two critical aspects of this approach. First, we argue that it is essential to observe incumbent couples at the very early stages of their relationship, as their match gains are the most comparable with those of two partners who have just matched. Second, when unobserved match quality is persistent, dynamic selection on unobservables needs to be accounted for. We draw from the literature on finite mixture models to show that identification is restored with a longer panel (Kasahara and Shimotsu, 2009; Arcidiacono and Miller, 2011).

3.1. Matching function and search frictions

We are interested in two-sided matching problems where a mass of agents form couples across two groups, men and women, in one or multiple *matching rounds*. We let $i \in \mathcal{I}$ ($j \in \mathcal{J}$) denote a vector of characteristics for male (female) individuals, with $\mathcal{I}$ ($\mathcal{J}$) being a finite set. We refer to i (j) as a man's (woman's) *type*. At the beginning of a matching round, there is an (uncountably) infinite number of agents actively seeking on the market, and we let $\tilde{n}^m$ and $\tilde{n}^f$ denote the probability mass functions, respectively with support $\mathcal{I}$ and $\mathcal{J}$, describing the distribution of types among male and female seekers.[6] The size of these two populations is

discuss how heterogeneous meeting rates across birth cohorts contribute to the age patterns of marriage and divorce observed in the data. Shephard (2018) and Ciscato (2019) estimate flexible meeting technology and find that people with similar age and education are more likely to meet. However, both papers focus on the relationship between wages and marriage market equilibrium outcomes, and do not explore the question of marriage market segmentation further.

[6]In models with multiple matching rounds, like the one considered later in Section 4, the frequencies $\tilde{n}^m$ and $\tilde{n}^f$ can change from one round to the next, and should be indexed by time. We do not introduce this indexation for two reasons: first, at this stage, it is not necessary to have data on couples formed across multiple matching

given by $N^m = \sum_i \tilde{n}_i^m$ and $N^f = \sum_j \tilde{n}_j^f$ with N^m and N^f not necessarily the same.

A *match* between a man i and a woman j generates a total match surplus $\bar{S}_{i,j}$. This may depend on expectations about how the relationship will unfold in the future. On the other hand, agents may not find a partner by the end of the matching round, in which case their payoff is normalized to zero. Hence, $\bar{S}_{i,j}$ represents the total gains for a couple (i, j) over the partners' reservation values, determined by the value attached to staying single. In models with multiple matching rounds, expectations about future match opportunities partly determine the reservation value. The normalization implies that $\bar{S}_{i,j}$ may be low (high) because agents are optimistic (pessimistic) about their future match opportunities.

A *matching function* maps the marginals $\tilde{n}^m$ and $\tilde{n}^f$ into $MF_{i,j}$, the mass of matches between men of type i and women of type j at the end of the round. The mapping depends on both demographics and the entire structure of the match surplus, $\bar{S} = \{\bar{S}_{i,j}\}_{i\in\mathcal{I},j\in\mathcal{J}}$. Moreover, Λ represents a set of additional parameters measuring search frictions intervening during the matching process. Hence, we can characterize the matching function as

$$MF_{i,j} = M_{i,j}(\bar{S}, \Lambda; \tilde{n}^m, \tilde{n}^f) \quad \forall (i, j) \in \mathcal{I} \times \mathcal{J}. \tag{3.1}$$

Since this is a one-to-one assignment problem, the following feasibility constraints must hold (Chen et al., 2021):

$$\tilde{n}_i^m - \sum_j MF_{i,j} \geq 0 \quad \forall i \in \mathcal{I}, \tag{3.2}$$

$$\tilde{n}_j^f - \sum_i MF_{i,j} \geq 0 \quad \forall j \in \mathcal{J}. \tag{3.3}$$

If Λ were known and the mapping implied by the matching function M were invertible, then $\bar{S}$ would be identified provided the matching outcome $MF = \{MF_{i,j}\}_{i\in\mathcal{I},j\in\mathcal{J}}$ is observed. However, in many models, MF is at least partly determined by parameters other than the match surplus, and it becomes impossible to tell to what extent sorting is explained by the match surplus without imposing further restrictions.

The general characterization of the matching function given by (3.1) is consistent with different assumptions about how mate search takes place. Below, we discuss three classes of models where similar underidentification issues arise.

1. In the classic assignment problem with TU of Shapley and Shubik (1971) and Becker (1973), frictions are ruled out, and sorting between types is entirely determined by the

rounds for identification; it is sufficient to observe the outcome of one matching round and follow the resulting cohort of couples over time. Second, in Section 4, we focus on an environment where $\tilde{n}^m$ and $\tilde{n}^f$ are stationary.

match surplus. In particular, PAM arises when $\bar{S}_{i,j}$ is supermodular. Choo and Siow (2006) introduce random taste shocks that help rationalize why dissimilar types are sometimes matched in the data.[7] In their setup, $\bar{S}_{i,j}$ is recovered by inverting the matching function, provided MF is observed. Jaffe and Weber (2019) generalize their model by introducing search frictions, either as meeting costs or as constraints to the set of potential partners, and show that the model is underidentified with only cross-sectional data on the matching outcome MF.

2. Shimer and Smith (2000) study sorting in a model with sequential and random search, where men meet a woman of type j with probability $\lambda \tilde{n}_j^f$ in a given period. Upon a meeting, a match is realized if the total surplus $\bar{S}_{i,j}$ is positive. Hence, the matching function is given by $MF_{i,j} = \lambda \tilde{n}_i^m \tilde{n}_j^f \mathbb{1}\{\bar{S}_{i,j} > 0\}$, with the arrival rate of meetings λ being the only element of the set Λ. When the match gains partly depend on a random component z, as in Greenwood et al. (2016) or Goussé et al. (2017), then $MF_{i,j} = \lambda \tilde{n}_i^m \tilde{n}_j^f \Pr\{\bar{S}_{i,j}(z) > 0\}$. Clearly, if we only observe the matching outcome MF, it is impossible to determine if a high number of matches is due to low search frictions or to a high surplus (Flinn and Heckman, 1982).

3. In models of directed search, agents self-select into local markets based on expected match opportunities. The matching occurs locally, but search frictions may prevent agents from matching with the chosen type (Eeckhout and Kircher, 2010; Chade et al., 2017). On local markets, the matching outcome depends on both local market tightness and the matching technology. For instance, in Arcidiacono et al. (2016) the number of matches between men and women who have self-selected into a given market is determined by a CES function. The parameters of this matching technology, the CES function in the example, belong to the set Λ, and contribute to the aggregate sorting patterns together with the entire surplus structure $\bar{S}$.

In Section 4, we develop a dynamic matching model where every matching round is an assignment game with partly random payoffs, as in Choo and Siow (2006) and Galichon and Salanié (2022). Search frictions are introduced in the form of meeting costs as in Jaffe and Weber (2019). Because mating patterns are determined by both the match surplus and search frictions, the model is not identified with only data on the matching outcome MF. Hence, in

[7] The distributions of these random taste shocks contribute to sorting next to the match surplus $\bar{S}$, and thus also belong to the set Λ. Galichon and Salanié (2022) and Gualdani and Sinha (2023) emphasize how the Choo-Siow approach crucially relies on a parametric assumption on these distributions, and discuss identification issues when this assumption is relaxed. Chiappori et al. (2017) leverage variation across markets in order to estimate a heteroskedastic version of the Choo-Siow model.

the next section, we discuss how relationship duration data can provide further information about the match surplus.

3.2. Match gains and relationship instability

Couples formed by a man of type i and a woman of type j are further characterized by an additional state variable $x \in \mathcal{X}$, with $\mathcal{X}$ being a finite set. The variable x describes other match-specific traits, such as the type of union the partners agreed upon (e.g., informal cohabitation vs. legal marriage) or additional factors influencing the couple's Pareto weights beyond the individual types (i, j). At the beginning of a new period, an incumbent couple draws a match-quality shock z from a distribution G and decides if to continue the relationship. Match-quality shocks are independent across periods and couples, their distribution G has full support and is continuous.

We let $S^*_{i,j}(x, z)$ define the total match surplus of an incumbent couple and assume that, whenever $S^*_{i,j}(x, z) > 0$, partners are able to share the surplus so that each receives a positive share.[8] This means that, conditional on (i, j, x) and the last draw z, the partners may need to renegotiate their terms of agreement if, under the previous terms, only one of the two partners is unsatisfied with the match.[9] Hence, the probability of continuing a relationship conditional on (i, j, x) is given by

$$\alpha_{i,j}(x) = \Pr\{S^*_{i,j}(x, z) > 0 | i, j, x\}. \tag{3.4}$$

Conditional on individual types (i, j), all relationships are assumed to start from the same state $\bar{x}_{i,j}$, a "clean slate" that only depends on the partners' individual endowments.[10,11] A couple's initial match surplus is given by $\bar{S}_{i,j} = S_{i,j}(\bar{x}_{i,j})$, where $\bar{S}_{i,j}$ determines the attractiveness of a match between types i and j on the marriage market. Following the match, couples

[8] This requires that partners can transfer utility to each other, although transferability does not need to be perfect. While in the model presented in Section 4 utility is assumed to be perfectly transferable (TU), the identification strategy would also apply to models with Imperfectly Transferable Utility (ITU).

[9] Empirical evidence supports the assumption that partners renegotiate the terms of their relationship (Mazzocco, 2007; Lise and Yamada, 2019). In particular, the assumption that partners break up only if $S^*_{i,j}(x, z) < 0$ is consistent with ex-ante efficient Limited Commitment Models, where partners only renegotiate their terms of agreement whenever one partner's individual rationality constraint is violated (Voena et al., 2015; Chiappori and Mazzocco, 2017; Shephard, 2018).

[10] Notice that this is consistent with competitive matching models where initial Pareto weights are pinned down by stability conditions.

[11] This can easily be relaxed to let couples self-select into a desired state x from the start, as in Section 4. However, it is reasonable to think that certain states, particularly those associated with a strong degree of commitment, can only be reached later in the relationship.

may diverge from their initial state due to both exogenous shocks and endogenous decisions.

We let $\tilde{m}_{i,j}(x)$ denote the mass of couples in state (i,j,x) at the beginning of a period, i.e. just before deciding if to stay together or to break up. We also define $DF_{i,j}(x)$ as the mass of couples in state (i,j,x) who decide to break up. When all state variables but z are observed, we can consistently estimate $1-\alpha_{i,j}(x)$ as the fraction $DF_{i,j}(x)/\tilde{m}_{i,j}(x)$.

Lemma 1. *Assume $S^*_{i,j}(x,z) = S_{i,j}(x)+z$ with $z \sim G$ and couples break up only if $S_{i,j}(x)+z < 0$. Further, assume z is i.i.d. across couples and periods, and G is continuous and has full support. Conditional on G, the initial match surplus $\bar{S}_{i,j}$ for a man i and a woman j is identified from $DF_{i,j}(\bar{x}_{i,j})/\tilde{m}_{i,j}(\bar{x}_{i,j})$ if $\tilde{m}_{i,j}(\bar{x}_{i,j}) > 0$.*

Proof. Due to additive separability, $\alpha_{i,j}(x) = \Pr\{z > -S_{i,j}(x)|i,j,x\}$, and $\alpha_{i,j}(x) = 1 - G(-S_{i,j}(x))$. When G is continuous and with full support, $S_{i,j}(x) \in \mathbb{R}$ is uniquely determined as $S_{i,j}(x) = -G^{-1}(1-\alpha_{i,j}(x))$. The mass of breakups is given by $DF_{i,j}(x) = (1-\alpha_{i,j}(x))\tilde{m}_{i,j}(x)$. Assuming $\tilde{m}_{i,j}(\bar{x}_{i,j}) > 0$, we have $1-\alpha_{i,j}(\bar{x}_{i,j}) = DF_{i,j}(\bar{x}_{i,j})/\tilde{m}_{i,j}(\bar{x}_{i,j})$, so that $\bar{S}_{i,j} = S_{i,j}(\bar{x}_{i,j}) = -G^{-1}(DF_{i,j}(\bar{x}_{i,j})/\tilde{m}_{i,j}(\bar{x}_{i,j}))$. □

Lemma 1 uses standard arguments for discrete choice models (Hotz and Miller, 1993; Magnac and Thesmar, 2002).[12] Without further parametric restrictions, the distribution G is not identified when the number of possible states is finite.[13] Most importantly, since the initial match surplus $\bar{S}_{i,j}$ is the object we want to identify, we must observe some incumbent couples who are still in (or have reverted to) state $(i,j,\bar{x}_{i,j})$. These couples have comparable stakes in the relationship to those who have just matched, and their breakup patterns are essential for identifying of the initial surplus. In contrast, incumbent couples who have been together for long enough will likely experience different, and often higher gains from the relationship. These differences may reflect returns to investment in match-specific capital (e.g., raising a child together or buying a home) and changes in the partners' outside options (e.g., divorce laws shape the outside options of legally married couples).

In our empirical application, when individuals match, they start a non-cohabiting romantic relationship, which may later grow into a more committed one if and when the partners decide

[12] Additive separability may be replaced by other parametric restrictions as long as the choice of breaking up can be recast as a binary threshold crossing model, i.e. as long as couples break up if $S^*_{i,j}(x,z) = S_{i,j}(x) + \sigma_{i,j}(x)z > 0$. For instance, in Goussé et al. (2017), shocks z are multiplicatively separable in the couple's per-period match gains. Hence, the scale of the shocks is proportional to the match gains, and both $\sigma_{i,j}(x)$ and $S_{i,j}(x)$ are determined by the same primitives.

[13] Matzkin (1992) provides sufficient conditions for the non-parametric identification of both S_x and G when there is a continuum of possible states.

to move in together. The initial gains for these relationships are inferred from the breakup decisions of couples who stay together in later periods, but who have *not* changed the terms of their relationship yet, i.e. who have not decided to move in together yet.

3.3. Persistent match quality

When unobserved match quality is correlated over time, we lose the straightforward relationship between observed separation rates and conditional choice probabilities established in the previous section (Heckman et al., 1984). Yet, this allows us to introduce duration dependence through a dynamic selection process whereby low-quality matches are dissolved at a faster rate. In what follows, we assume relationships are characterized by a latent time-invariant match quality component $k \in \{1, ..., K\}$ assigned from the start of their relationship. We let $\pi_{i,j,k}$ be the probability that a relationship with initial individual types (i, j) is of type k when the match is formed. Introducing a finite number of time-invariant unobserved types next to i.i.d. random shocks is a popular solution to account for selection on unobservables in dynamic discrete choice models (Arcidiacono and Miller, 2011).[14]

In this setup, the data generating process is therefore a finite mixture model whose identification is extensively discussed in the literature (Anderson, 1954; Hall and Zhou, 2003; Kasahara and Shimotsu, 2009). For the sake of exposition, we focus on the case where couples only have to choose between staying together and breaking up, but the analysis extends to the case where couples can choose between a finite number of actions (Kasahara and Shimotsu, 2009). The likelihood of observing a couple staying together for d periods and experiencing a sequence of states $\{i_t, j_t, x_t\}_{t=1}^{d}$, conditional on the initial types (i_0, j_0) and with $x_0 = \bar{x}_{i,j}$, can be written as follows:

$$\Pr\{\{i_t, j_t, x_t\}_{t=1}^{d} | i_0, j_0\} = \sum_{k=1}^{K} \pi_{i_0,j_0,k} \prod_{t=1}^{d} \alpha_{i_t,j_t,k}(x_t) \Pr\{i_t, j_t, x_t | i_{t-1}, j_{t-1}, x_{t-1}\} \tag{3.5}$$

where $\pi_{i,j,k} \in (0, 1)$ for any (i, j, k), as well as $\alpha_{i,j,k}(x) \neq \alpha_{i,j,k}(x)$ for any (x, k, k') s.t. $k \neq k'$. Conditional choice probabilities $\alpha_{i,j,k}(x)$ are now k-specific, and so is the underlying surplus. The transition probabilities $\Pr\{i_t, j_t, x_t | i_{t-1}, j_{t-1}, x_{t-1}\}$ are assumed not to depend on k nor on past draws of z, and are non-parametrically identified as in Rust (1987).[15]

[14] Other structures on the nature of latent match quality have been explored in the literature, and relationship quality is sometimes modeled as a unidimensional variable following a Markov process (Miller, 1984; Brien et al., 2006; Greenwood et al., 2016). The argument made in this section also applies to cases where match quality is unidimensional but persistent. Likewise, the assumption of a discrete support for time-invariant match quality can be relaxed.

[15] Kasahara and Shimotsu (2009) and Hu and Shum (2012) show that this Conditional Independence assumption can be relaxed to allow the transition probabilities to be k-specific, although in this case identification requires a longer panel.

Lemma 2. *Assume $S^*_{i,j,k}(x,z) = S_{i,j,k}(x) + z$ with $z \sim G$ and couples break up only if $S_{i,j,k}(x) + z < 0$. Further, assume k is drawn from a K-point support and does not change over time, while z is i.i.d. across couples and periods, and G is continuous and has full support. Conditional on G, the initial match surplus $\{\bar{S}_{i,j,k}\}_{k=1}^{K}$ and probabilities $\{\pi_{i,j,k}\}_{k=1}^{K}$ are identified for any (i,j) from the relationship survival profile*

$$\left\{\Pr\{\{i_t = i, j_t = i, x_t = \bar{x}_{i,j}\}_{t=1}^{d} | i_0 = i, j_0 = j\}\right\}_{d=1}^{D} \tag{3.6}$$

if there exists a $D \geq 2K - 1$ s.t. each element of (3.6) is positive.

Proof. See Appendix A. □

Lemma 2 states that the finite mixture model is identified if there exists a positive mass of couples who have stayed together and remained in the initial state $(i, j, \bar{x}_{i,j})$ for at least $2K - 1$ periods following the match. This argument extends the one of Lemma 1, in that the breakup decisions of incumbent couples who still find themselves in the initial state are informative about the match surplus. In the absence of time-persistent unobserved heterogeneity, the hazard of separation conditional on $(i, j, \bar{x}_{i,j})$ was enough to pin down the match surplus. With time-persistent unobserved heterogeneity, how the conditional hazard of separation changes with duration is informative about the latent structure.

Finally, it is worth noting that transitions across states can in principle be leveraged as an additional source of identifying variation. Kasahara and Shimotsu (2009) show that identification with a shorter panel ($D = 3$) can be achieved, but only when transitions between any two states x and x' are possible in both directions. However, as time goes by and couples go through the different stages of a relationship, many of these transitions are irreversible, rendering this approach not always viable.

3.4. Inverting the matching function

Proposition 1. *Assume that either $K = 1$ and Lemma 1 holds or $K > 1$ and Lemma 2 holds. Conditional on M, up to $|\mathcal{I}| \times |\mathcal{J}|$ parameters in the set Λ are identified from MF.*

Proof. When $K = 1$, the matching function provides $|\mathcal{I}| \times |\mathcal{J}|$ additional degrees of freedom provided the matching outcome MF is observed. These additional degrees of freedom identify up to $|\mathcal{I}| \times |\mathcal{J}|$ parameters of the model M. The case $K > 1$ is discussed in Appendix A. □

Proposition 1 closes our identification argument. If panel data on incumbent couples guarantees the identification of the initial match surplus, then the matching function can be used to learn about search frictions. In our setup, the number of additional restrictions coming from the matching function depends on cross-sectional variation in the number of types, so that only up to $|\mathcal{I}| \times |\mathcal{J}|$ parameters of the matching model M can be identified. Data on multiple markets can potentially provide additional degrees of freedom.

4. Model

We now outline a full-fledged and empirically tractable model in order to study the role of search frictions in shaping assortative mating. In this model, time is discrete, and in every period, a new generation of young individuals enters the marriage market. Every agent is endowed with a given set of traits, including race and education in our application, which, together with the agent's age, constitute his (or her) type i (j). Over time, agents get older and eventually die when they reach a terminal age. Different generations overlap, and the population of agents constitutes a mass whose size is assumed to be constant over time and normalized to one. The stationary probability mass functions ℓ^m and ℓ^f respectively describe the composition of the male and female populations.

All individuals are single at the onset of adult life. In every period, singles can look for a partner on the marriage market. Agents who do not find a partner can participate in the next matching round. A matching round is an assignment game with random payoffs, as in Choo and Siow (2006) and Galichon and Salanié (2022). However, as in Jaffe and Weber (2019), singles face meeting costs that depend on the partners' types and that may discourage them from crossing socioeconomic barriers when looking for a mate. This results in an imperfectly segmented marriage market based on characteristics such as age, race, and education. Notice that, since the distribution of singles is endogenous, so are match opportunities. We focus on a stationary environment where this distribution does not change over time.

After the match, a newly formed couple learns about the time-invariant match quality k. In every new period, incumbent couples draw random preference shocks in each alternative living arrangement a. In our application, living arrangements include non-cohabitation and cohabitation. All relationships are initially non-cohabiting, and moving in together is costly and irreversible. However, cohabitation is possibly associated with extra gains that make the relationship more rewarding and stable. If the partners are unsatisfied with any of these living arrangements, they can break up and look for a new partner. Rematch opportunities crucially shape the partners' outside options, and thus the total match surplus.

4.1. Gains from relationship

In every period, incumbent couples choose living arrangements $a' \in \{0, 1, ..., A\}$ upon observing $z = (z_0, z_1, ..., z_A)$, conditional on the partners' types (i, j) and the past living arrangements $a \in \{1, ..., A\}$. By convention, the option $a' = 0$ corresponds to the choice of breaking up. The random taste shocks are extreme value type I distributed. Each alternative yields per-period total match gains $H_{i,j,a',k} - \kappa_{a,a'} + z_{a'}$, where $\kappa_{a,a'}$ is the switching cost associated with changing living arrangements from a to a', with $\kappa_{a,a} = 0$ for any a. Since $H_{i,j,0,k} - \kappa_{a,0} = 0$ is normalized to zero, $H_{i,j,a',k} - \kappa_{a,a'} + z_{a'} - z_0$ can be understood as a measure of the partners' net gains over singlehood.

We assume that, conditional on choosing a', utility is perfectly transferable and partners share the match gains through Nash bargaining.[16] The Nash bargaining weight θ reflects the male partner's bargaining power. As we will see in the next section, θ is endogenously determined on the marriage market at the start of a relationship, and is assumed to remain constant thereafter.[17] Let $V^m_{i,0}$ and $V^f_{0,j}$ denote the reservation values of the male and female partner. These correspond to the value of being single, and depend on the expected (re)match opportunities. Let also $C^m_{i,j,a,k}(\theta)$ and $C^f_{i,j,a,k}(\theta)$ denote the continuation values when the Nash-bargaining parameter is θ. Nash bargaining implies that the male partner obtains a share

$$\begin{aligned} &v^m_{i,j,a,k}(a', \theta, z) = \\ &= \arg\max_v \left\{ \left[v + C^m_{i,j,a',k}(\theta) - V^m_{i,0}\right]^\theta \left[H_{i,j,a',k} - \kappa_{a,a'} + z_{a'} - v + C^f_{i,j,a',k}(\theta) - V^f_{0,j}\right]^{1-\theta} \right\} \end{aligned} \tag{4.1}$$

of the per-period gains, whereas the wife receives the remaining part,

$$v^f_{i,j,a,k}(a', \theta, z) = H_{i,j,a',k} - \kappa_{a,a'} - v^m_{i,j,a,k}(a', \theta, z). \tag{4.2}$$

When utility functions are quasilinear in the transferable good,[18] the total match surplus, $S_{i,j,a,k}(a') + z_{a'}$, does *not* depend on θ. Under the same assumption, Nash bargaining ensures

[16]Nash bargaining has found applications in household economics since McElroy and Horney (1981). Shimer and Smith (2000) and Goussé et al. (2017) embed it in their search-and-matching framework to split rents generated by the presence of search frictions. Since the optimal household allocation lies on the Pareto frontier, this household model is a special case of the general collective model (Chiappori, 1988).

[17]Repeated Nash bargaining is consistent with an intertemporal collective model where the Pareto weight is updated in every period and is jointly determined by the Nash-bargaining parameter θ and the reservation utilities $V^m_{i,0}$ and $V^f_{0,j}$. Hence, even if θ is held constant throughout the relationship, the Pareto weight does change.

[18]When utility is perfectly transferable, preferences always admit such a cardinal representation (Demuynck and Potoms, 2020).

that the match surplus is split between the two partners proportionally to θ:

$$\begin{aligned} v^m_{i,j,a,k}(a',\theta,z) + C^m_{i,j,a',k}(\theta) - V^m_{i,0} = \\ = \theta \underbrace{\left[H_{i,j,a',k} - \kappa_{a,a'} + z_{a'} + C^m_{i,j,a',k}(\theta) + C^f_{i,j,a',k}(\theta) - V^m_{i,0} - V^f_{0,j}\right]}_{= S_{i,j,a,k}(a') + z_{a'}}. \end{aligned} \tag{4.3}$$

Individual rationality requires that, upon a new draw of z, there exists at least one alternative so that both partners enjoy positive gains in order for a relationship to stand. Condition (4.3) ensures that both partners are willing to continue their relationship as long as there exists an alternative a' s.t. $S_{i,j,a,k}(a') + z_{a'} - z_0 > 0$. Similarly, decisions about living arrangements maximize the total surplus. Hence, when the extreme value type I shocks z are independent,[19] the expected match surplus conditional on state (i, j, a, k) is

$$\begin{aligned} \tilde{S}_{i,j,a,k} &= \mathbb{E}\left[\max_{a'\in\{0,1,\ldots,A\}} S_{i,j,a,k}(a') + z_{a'} \middle| i,j,a,k\right] \\ &= \ln\left[1 + \sum_{a'=1}^{A} \exp S_{i,j,a,k}(a')\right] \end{aligned} \tag{4.4}$$

where the expectation is taken over z. Finally, the conditional probability of choosing alternative a' is given by

$$\alpha_{i,j,a,k}(a') = \frac{\exp S_{i,j,a,k}(a')}{1 + \sum_{a''=1}^{A} \exp S_{i,j,a,k}(a'')}. \tag{4.5}$$

4.2. Mating

Singles can meet and match with agents of the opposite sex. Both the time-invariant match quality component k and the initial random taste shocks $(z_1, ..., z_A)$ are assumed to be unobserved at the time of the match and are only revealed when two agents have committed to a match with each other. The probability of drawing a time-invariant component k conditional on individual types (i, j) is $\mu_{i,j,k}$. All relationships start from state $a = 1$, which corresponds to non-cohabitation in our implementation, but couples can self-select into their preferred living arrangement upon matching. The expected match surplus for a pair (i, j) is given by

$$\bar{S}_{i,j} = \sum_{k=1}^{K} \mu_{i,j,k} \mathbb{E}\left[\max_{a'\in\{1,\ldots,A\}} S_{i,j,1,k}(a') + z_{a'} \middle| i,j,1,k\right]. \tag{4.6}$$

A match is also associated with meeting costs $\Lambda_{i,j}$, which measure the strength of search frictions for pairs (i, j). A high value of $\Lambda_{i,j}$ means that it is costly for men of type i and women of type j to reach out to each other. In addition, every male single draws a vector of random

[19] We later estimate the model both under this assumption and allowing for correlation between $\{z_1, ..., z_A\}$.

taste shocks $\varepsilon^m \in \mathbb{R}^{|\mathcal{J}|+1}$ from a distribution F^m. These shocks are assumed to be independent and extreme value type I distributed, and can be interpreted as random components of the meeting costs, introducing an element of chance in the process.[20]

A male single chooses his mating strategies by solving the following discrete choice problem:

$$\bar{V}_i^m = \max \left\{ \max_{j \in \mathcal{J}} \left\{ \theta_{i,j}(\bar{S}_{i,j} - \Lambda_{i,j}) + \varepsilon_j^m \right\}, \varepsilon_0^m \right\}, \tag{4.7}$$

where 0 denotes the option of staying single. By solving (4.7), the agent trades off expected match gains and meeting costs, while taking into account that women whose type is in high demand will be associated with a lower bargaining weight $\theta_{i,j}$. On the other side of the market, every female single draws a vector of random taste shocks $\varepsilon^f \in \mathbb{R}^{|\mathcal{I}|+1}$ from a distribution F^f and chooses her mating strategies by solving a symmetric problem. These shocks are also assumed to be independent and extreme value type I distributed.

Under the extreme value assumption on the distribution of the random taste shocks, a matching round corresponds to the assignment game of Choo and Siow (2006). The aggregate matching function arises from individual-maximizing behavior, namely from the aggregation of individual demand functions given by problem (4.7). Yet, supplies of different types are limited, and are given by the marginal distributions $\tilde{n}^m$ and $\tilde{n}^f$; the bargaining weights adjust so that supply and demand meets. In this setup, we obtain the following result, due to Galichon and Salanié (2022).

Proposition 2. *Let $\tilde{n}^m$ and $\tilde{n}^f$ be probability mass functions over the finite supports $\mathcal{I}$ and $\mathcal{J}$ and rescaled so that $N^m = \sum_i \tilde{n}_i^m$ and $N^f = \sum_j \tilde{n}_j^f$. Mating decisions are described by problem (4.7) for single men and by an analogous problem for single women, with the shocks ε^m and ε^f being independent and extreme value type I distributed.*

1. *The aggregate matching function and the male partner's bargaining weight are given by:*

$$MF_{i,j} = \exp\left[(\bar{S}_{i,j} - \Lambda_{i,j})/2\right] \sqrt{c_i^m c_j^f} \quad \forall (i,j) \in \mathcal{I} \times \mathcal{J} \tag{4.8}$$

$$\theta_{i,j} = \frac{1}{2}\left[1 + \left(\ln c_j^f - \ln c_i^m\right)\left(\bar{S}_{i,j} - \Lambda_{i,j}\right)^{-1}\right] \quad \forall (i,j) \in \mathcal{I} \times \mathcal{J}. \tag{4.9}$$

2. *There exists a unique pair of functions c^m and c^f, respectively defined over $\mathcal{I}$ and $\mathcal{J}$, that*

[20] When the random components ε_j^m and ε_j^f are additively separable, it can be proved that they are nontransferable in that they do not affect how the match surplus is shared, but only shifts the utility shares of each partner separately. This is thoroughly discussed in Chiappori et al. (2009) and Chiappori et al. (2020).

satisfy the feasibility constraints

$$c_i^m = \tilde{n}_i^m - \sum_j MF_{i,j} \geq 0 \quad \forall i \in \mathcal{I} \tag{4.10}$$

$$c_j^f = \tilde{n}_j^f - \sum_i MF_{i,j} \geq 0 \quad \forall j \in \mathcal{J} \tag{4.11}$$

where $MF_{i,j}$ is given by (4.8).

Proposition 2 determines how the matching outcome, given by both the mass of new matches (4.8) and the bargaining weight (4.9), depends on the inputs in the process, namely the match surplus $\bar{S}_{i,j}$, the meeting costs $\Lambda_{i,j}$, and the current supplies of male and female singles. In particular, the supplies enter the matching function through the type fixed effects c_i^m and c_j^f, defined by (4.10) and (4.11). As explained by Chen et al. (2021), solving for the matching function given the primitives is equivalent to finding two functions c^m and c^f that solve a system of $|\mathcal{I}| + |\mathcal{J}|$ equations given by (4.10) and (4.11), after replacing $MF_{i,j}$ for (4.8).

Embedding this assignment game into a dynamic matching model grants the following properties to the matching function (4.8). First, a stable and feasible matching exists and is unique; its numerical computation relies on a fast iterative algorithm to recover c^m and c^f.[21] Second, competition between singles shapes bargaining power; when the excess supply of women of type j is larger than the excess supply of men of type i, then the male partner in a couple (i, j) obtains more favorable terms of trade, as measured by his bargaining weight $\theta_{i,j}$. Third, when observed types can be ordered along one dimension (e.g., age or human capital), meetings are positive (negative) assortative when $\bar{S}_{i,j} - \Lambda_{i,j}$ is supermodular (submodular).[22] Notice that supermodularity in the expected match surplus $\bar{S}_{i,j}$ is not enough to guarantee PAM, since it can be dampened or exacerbated by the structure of the cost function $\Lambda_{i,j}$.

[21] See Section 5.2 and Theorem 5 in Galichon and Salanié (2022). The vectors c^m and c^f are found using an Iterative Proportional Fitting Procedure (IPFP).

[22] Chiappori et al. (2020) proves this result in a one-dimensional framework where types correspond to human capital levels. When there exist multiple dimensions of sorting, PAM with respect to one trait can be discussed while holding other traits constant. In fact, PAM is a local property: e.g., PAM with respect to education can be stronger for couples within a certain race group, and weaker or absent for other groups. Lindenlaub (2017) extends the definition of PAM to a multidimensional setting and provides conditions on the match surplus for PAM to hold.

4.3. Bellman equations

We can write down the Bellman equations for male and female partners with a bargaining weight θ as:

$$V^m_{i,j,a,k}(a', \theta, z) = v^m_{i,j,a,k}(a', \theta, z) + \beta \mathbb{E}_{i'} V^m_{i',0} + \theta \beta \mathbb{E}_{i',j'} \tilde{S}_{i',j',a',k} \tag{4.12}$$

$$V^f_{i,j,a,k}(a', \theta, z) = v^f_{i,j,a,k}(a', \theta, z) + \beta \mathbb{E}_{j'} V^f_{0,j'} + (1-\theta) \beta \mathbb{E}_{i',j'} \tilde{S}_{i',j',a',k}, \tag{4.13}$$

where $\beta \in (0,1)$ is the discount factor. Using (4.12) and (4.13), we can write down the Bellman equations for the match surplus net of $z_{a'}$:

$$S_{i,j,a,k}(a') = H_{i,j,a',k} - \kappa_{a,a'} + \beta \mathbb{E}_{i'} V^m_{i',0} + \beta \mathbb{E}_{j'} V^f_{0,j'} + \beta \mathbb{E}_{i',j'} \tilde{S}_{i',j',a'k} - V^m_{i,0} - V^f_{0,j} \tag{4.14}$$

where θ vanishes as a result of the TU assumption.

When the distribution of singles is stationary, the value functions $V^m_{i,0}$ and $V^f_{0,j}$ for single agents are given by the following Bellman equations:

$$V^m_{i,0} = \beta \mathbb{E}_{i'} V^m_{i',0} + \beta \mathbb{E}_{i'} \bar{V}^m_{i'} \tag{4.15}$$

$$V^f_{0,j} = \beta \mathbb{E}_{j'} V^f_{0,j'} + \beta \mathbb{E}_{j'} \bar{V}^f_{j'} \tag{4.16}$$

where the last term, defined in (4.7), represents the expected surplus from search on the marriage market in the next period. Under the logit assumption, it corresponds to

$$\bar{V}^m_i = \ln \left[1 + \sum_j \exp(\theta_{i,j}(\bar{S}_{i,j} - \Lambda_{i,j})) \right] \tag{4.17}$$

$$\bar{V}^f_j = \ln \left[1 + \sum_i \exp((1-\theta_{i,j})(\bar{S}_{i,j} - \Lambda_{i,j})) \right] \tag{4.18}$$

where $\theta_{i,j}$ is given by (4.9), and thus depends on the marginals $\tilde{n}^m$ and $\tilde{n}^f$.

4.4. Population dynamics

In order to describe how the stock of relationships evolves over time, we characterize aggregate inflows and outflows. While the mass of new matches $MF_{i,j}$ for pairs (i,j) is given by the matching function (4.8), the mass of separations $DF_{i,j,a,k}$ for couples (i,j,a,k) is given by:

$$DF_{i,j,a,k} = \alpha_{i,j,a,k}(0) \tilde{m}_{i,j,a,k}, \tag{4.19}$$

where $\tilde{m}_{i,j,a,k}$ represents the mass of couples in state (i,j,a,k) at risk of separation at the beginning of the period, before matching decisions are made. We let $m_{i,j,a,k}$ denote the mass of couples in state (i,j,a,k) at the end of a period. This mass results from the following law of motion:

$$m_{i,j,a',k} = \pi_{i,j,a',k} MF_{i,j} + \sum_{a=1}^{A} \alpha_{i,j,a,k}(a') \tilde{m}_{i,j,a,k}, \tag{4.20}$$

where $\pi_{i,j,a',k}$ is the fraction of couples who draw match-quality k and choose living arrangement a' upon matching. In our setup, the fraction $\pi_{i,j,a',k}$ is endogenous, and corresponds to

$$\pi_{i,j,a',k} = \frac{\alpha_{i,j,1,k}(a')}{1 - \alpha_{i,j,1,k}(0)} \mu_{i,j,k}. \tag{4.21}$$

We let n^m and n^f denote the mass of male and female singles at the end of each period. These functions result from the following accounting restrictions, which must hold in any period:

$$\ell_i^m = n_i^m + \sum_{j,a,k} m_{i,j,a,k} \tag{4.22}$$

$$\ell_j^f = n_j^f + \sum_{i,a,k} m_{i,j,a,k}. \tag{4.23}$$

where ℓ_i^m and ℓ_j^f are the (exogenous) marginal distributions.[23] We complete the description of the population dynamics by linking the start-of-period with end-of-period frequencies:

$$\tilde{m}_{i',j',a,k} = \sum_{i,j} \Pr\{i', j' | i, j\} m_{i,j,a,k} \tag{4.24}$$

$$\tilde{n}_{i'}^m = \sum_{i} \Pr\{i' | i\} n_i^m \tag{4.25}$$

$$\tilde{n}_{j'}^f = \sum_{j} \Pr\{j' | j\} n_j^f. \tag{4.26}$$

4.5. Steady-state equilibrium

A Steady-State Equilibrium (SSE) is defined by stationary matching outcomes, match values, and strategies that are consistent with the Bellman and population equations outlined in the previous sections. While Proposition 2 establishes the existence and uniqueness of a matching equilibrium conditional on the size and composition of the pool of unmatched agents in a given period, this pool is only stationary at the SSE defined below.

Definition 1 (Steady-State Equilibrium, SSE)**.** *A SSE is given by measures n^m, n^f and m, a matching outcome MF and θ, a match surplus S, and reservation values $V^m_{\cdot,0}$ and $V^f_{0,\cdot}$ s.t.:*

- *The matching function MF and the bargaining weight function θ are given by (4.8) and (4.9), and are s.t. constraints (4.10) and (4.11) hold.*

- *The match surplus S is given by the Bellman equation (4.14).*

[23] When individual types change as a consequence of mating decisions, these marginals are endogenous. In this case, conditions (4.22) and (4.23) are replaced by flow-balance conditions that pin down the stationary marginal distributions.

- *Decisions about living arrangements and separations maximize the match surplus.*
- *The reservation utilities $V^m_{\cdot,0}$ and $V^f_{0,\cdot}$ are given by the Bellman equations (4.15) and (4.16).*
- *The mass of end-of-period couples m is given by the law of motion (4.20).*
- *The mass of end-of-period single men and women, respectively n^m and n^f, are given by the accounting restrictions (4.22) and (4.23).*
- *Beginning-of-period measures are given by the transition rules (4.24), (4.25), and (4.26).*

A SSE can be computed as a fixed point (n^m, n^f) in a $|\mathcal{I}| + |\mathcal{J}|$-dimensional space, with n^m_i bounded into $[0, \ell^m_i]$ for any $i \in \mathcal{I}$ and n^f_j bounded into $[0, \ell^f_j]$ for any $j \in \mathcal{J}$. Fixed-point theorems are employed by Manea (2017) and Shephard (2018) to generalize the original proof of SSE existence by Shimer and Smith (2000) for random search models. In our case, the fixed-point operator defined by the structural equations always converges to the same fixed point (n^m, n^f) for a given set of primitive parameters, regardless of the initial values chosen.

Leveraging the identification strategy discussed in Section 3, we can now establish that the model parameters are identified from data on the matching outcome MF and panel data on relationships, provided the discount factor β and the distributions (F^m, F^f, G). Under the assumption that the marriage market is at a SSE, the rational expectation problem simplifies greatly, and the per-period match gains and switching costs can be recovered from the Bellman equation.[24] Notice that, as in Galichon et al. (2019), identification does not hinge on a specific parametric family for (F^m, F^f, G).

Proposition 3. *Assume that the marriage market is at a SSE. Conditional on the parameters (β, F^m, F^f, G), with $\beta \in (0, 1)$ and F^m, F^f, and G continuous and with full support, i) the per-period match gains H, ii) the switching costs κ, iii) the meeting costs Λ, and iv) the initial distribution of time-invariant match quality μ are identified from the matching outcome MF and relationship survival profiles (3.6) if, for any (i, j), there exists a $D \geq 2K - 1$ s.t. each element of (3.6) is positive.*

Proof. See Appendix A. □

[24] Chen and Choo (2023) discuss identification of dynamic matching models in a non-stationary environment with repeated cross-sections $\{MF_t\}_{t=1}^T$.

5. Data

The model is estimated with data from the How Couples Meet and Stay Together (HCMST) survey, a nationally representative study of American adults run in 2009 by Rosenfeld and Falcon (2018). What makes this survey particularly attractive for our study is the presence of information about both cohabiting and non-cohabiting romantic relationships, along with detailed data about the characteristics of the partner. Moreover, it contains detailed retrospective information about the beginnings of a relationship, including its starting date and, for cohabiting couples, the date when the two partners moved in together. Hence, the data allow us to describe mating patterns in the very early stages of relationships. Another crucial aspect of the survey is its longitudinal dimension. Respondents who had a partner in 2009 were interviewed again on a yearly basis until the end of their relationship, with the last follow-up survey taking place in 2014. Hence, we can measure the hazard rate of a breakup at different relationship stages.

5.1. Descriptive statistics

Panel A in Table 1 presents our estimation sample, which includes 545 respondents who were single and aged between 19 and 62 when interviewed in 2009. Moreover, it includes 1,162 couples where both partners were aged between 19 and 62 and had been dating (without necessarily cohabiting) for at most 15 years in 2009. Couples were interviewed again in the follow-up survey, so that, on average, the sample contains slightly more than three data points per couple.[25]

Table 1 also presents descriptive statistics detailing the sample composition in 2009. Since the HCMST survey is representative of the adult non-institutionalized U.S. population, women outnumber men. Hence, the share of singles among female respondents in our estimation sample is 34.8%, as opposed to only 28.6% among male respondents. In particular, in panel B we see that the black population has a strongly imbalanced gender ratio, which results in black women being overrepresented among singles. In contrast, Hispanic men outnumber Hispanic women in the population, which may be explained by gender-asymmetric immigration patterns. Panel C shows that individuals in a relationship tend to be more educated than singles. Finally, panel

[25]Since couples are followed up to their separation by survey design, we observe 22% of all observed couples breaking up between 2009 and 2014, and thus we know their completed duration. Right-censoring due to both limited panel length and random attrition does not represent a problematic issue for the analysis of duration data. On the other hand, since we include in our sample couples formed before 2009, we have to deal with left-censoring: in the estimation, it is assumed that the distribution of time-persistent unobserved match-quality at different stages of the relationship and conditional on other observables is the same for couples formed in 2009 and in earlier years (Heckman and Singer, 1984; Keane and Wolpin, 2001).

D shows that, since we only look at relationships whose observed duration is less than 15 years, partnered individuals are actually younger than singles. Interestingly, women are more likely than men to have a partner when they are young, but they are also more likely to be single when they are older.

[Table 1 about here.]

Table 2 describes cross-sectional mating patterns for the year 2009 among all couples in our estimation sample. In panel A, we present the joint distribution of partners' educational levels. The relative frequencies are highest on the diagonal, suggesting that partners sort on education. Yet, while certain combinations remain extremely rare (e.g., a high school dropout dating someone with graduate education), a non-negligible share of individuals who never went to college still match with partners who did attend college. Notably, about 44% of partnered women with a bachelor's degree date men without any college degree, a number that is partly explained by a shortage of male college graduates. In panel B, we present the joint distribution of partners' racial backgrounds. On the diagonal, we notice strong homogamy patterns, particularly for whites and blacks. Finally, in panel C, we show that, in about 50% of all couples, the age gap between the partners is no larger than two. On the other hand, in about 65% of all couples, the male partner is older.

[Table 2 about here.]

In Figure 1, we plot the yearly separation rate conditional on relationship duration. In panel (a), the rate is found to be extremely high in the first stages of a relationship, with almost 40% of all couples breaking up in the first year. The separation rate decreases sharply and stabilizes around 5% after 10 years. Interestingly, in panel (b), the profiles for same-race and different-race couples are found to be very similar, with the separation rate for different-race couples being only slightly higher on average. On the other hand, in panel (c), we see that, in the first 10 years of a relationship, different-education couples face a higher separation risk than same-education couples, with their profile being steeper. Table 8 in appendix completes the description of our data with average matching and separation rates for different population groups. As previously documented in the literature (McLanahan, 2004), individuals without a college degree and certain minorities (black men, Hispanic women) experience a relatively higher separation risk.

[Figure 1 about here.]

5.2. Parameterization and estimation

A period in the model corresponds to one calendar year. It is assumed that individuals start looking for a mate at age 19 and stop when they are 62. Once past this age, they can continue their relationship with their current partner, but cannot rematch in the event of a separation. Aging is modeled stochastically, and age groups span four years.[26] Race and education are exogenous; while race does not change over time, a fraction of individuals graduate by age 23.[27] As anticipated, we allow for two living arrangements: non-cohabitation and cohabitation, with the latter being an irreversible state. In our baseline model, the shocks (z_0, z_1, z_2) are assumed to be independent from each other, although we also estimate a nested logit version of the model.[28] The cost of moving in together is given by a constant κ. The annual discount factor β is set to 0.95.

Meeting costs are parameterized as $\Lambda_{i,j} = x_{i,j}^{\intercal}\lambda$. The terms of $x_{i,j}$ include each partner's age, dummies for educational attainment and race, and interaction terms between partners' traits. The vector λ contains the corresponding linear coefficients to be estimated. Similarly, we parameterize the per-period match gains as $H_{i,j,a,k} = x_{i,j}^{\intercal}\delta + \zeta_a + \eta_k$, where δ is a vector of linear coefficients, while ζ and η represent a-specific and k-specific fixed effects. Time-invariant match quality has a two-point support ($K = 2$) and, in our baseline model, its distribution is assumed constant across individual types ($\mu_{i,j,k} = \mu_k$). We relax the latter assumption in Section 6.3.

In order to estimate the parameters $(\lambda, \delta, \zeta, \eta, \mu, \kappa)$, we implement a method of moments estimator. We build moments from the empirical distributions of new matches and separations. The moments include the overall average match and separation rate; the average match and separation rates by gender, education, and race; men's and women's average age at separation; the average length of a relationship. We then simulate the model for every draw of the parameters, calculate the predicted moments, and minimize the distance with their empirical counterparts. Additional technical details and the full list of moments are provided in Appendix B.

[26] A more computationally intensive version of the model with deterministic aging (i.e. age changes in every calendar year) yields very similar findings.

[27] Graduation rates are gender- and race-specific. They are calibrated straight from HCMST data.

[28] Allowing for correlation between z_1 and z_2, so that continuing the relationship constitutes a nest with two alternatives, only changes the estimates for the cost of moving and for the value assigned to cohabitation, leaving all other findings almost unchanged.

6. Results

In this section, we first report our baseline structural estimates of search frictions and per-period match gains. We then use additional data from the HCMST survey in order to ease the interpretation of our structural estimates of search frictions. Next, in order to show that our baseline findings are robust, we present additional estimation results obtained after relaxing certain assumptions. In particular, we present the findings obtained for a model with a more refined categorization of the race and education variables. Finally, we present results from two different types of counterfactual exercises. First, we show how mating patterns would change if meeting costs were disregarded (or, alternatively, if they were the only determinants of mating strategies) while holding the pool of singles constant. This helps quantify the degree of marriage market segmentation and the importance of relationship complementarities. Second, we compute counterfactual SSE under different scenarios and discuss how agents adjust their mating strategies following a change in either the meeting cost or the match gain structure.

6.1. Search frictions and match gains

[Table 3 about here.]

We start from a simple model with two racial groups (white, non-white) and three educational groups (high school, some college, college degree), and we later show that our results are robust to a more refined categorization of race and education. Table 3 reports the estimated parameters, whereas the model fit is discussed in Appendix C. We start by discussing interaction terms in meeting costs and couples' per-period gains, which respectively measure marriage market segmentation and relationship complementarities.

Individuals of similar age face both lower search frictions and higher match gains. Both forces contribute to the strong assortativeness with respect to age observed in the data. Individuals of the same race face significantly lower search frictions but do not benefit from additional complementarities throughout the relationship. This suggests that racial homogamy mainly results from barriers in the marriage market that hinder interracial matches. On the other hand, same-education couples experience significantly higher per-period match gains but do not benefit from lower meeting costs. Hence, educational homogamy mainly results from educational complementarities that play out throughout the relationship.

Table 3 also shows that non-white women experience higher meeting costs and lower match gains relative to their white peers. In contrast, non-white men experience lower meeting costs and match gains comparable to their white peers. This places non-white women at a significant disadvantage: not only is it harder for them to find a partner and enjoy stable relationships, but their bargaining power is also undermined by the lack of other (re)match opportunities. This situation is made worse by an unfavorable gender ratio, since women outnumber men in the non-white population. We explore this further when we present our findings with more refined racial categories.

Women who have attended college face mildly lower search frictions than those who have not. Yet, male college graduates experience stronger frictions than those without a college degree. Most importantly, having a college degree is associated with stronger relationship gains, particularly for women. Hence, the highly educated rely less than the poorly educated on good chemistry (a high type k) and are less exposed to changes in match quality (the z shocks).

Search frictions also become stronger with age, particularly for women. At the same time, the age of both partners has a negative impact on match gains, which is consistent with a negative effect of declining health and fertility (Low, forthcoming). However, as the market is clustered with respect to age, all individuals experience a narrower market as they get older and their peers are more likely to be in stable relationships. Worsening outside options raise the match surplus among older people, reinforcing commitment among incumbent couples and facilitating the matching between older singles who successfully meet in spite of the strong frictions.

More generally, it is worth noting that, while $H_{i,j,a,k}$ is the primitive object we estimate, it is the expected initial match surplus $\bar{S}_{i,j}$ that ultimately shapes sorting patterns next to the meeting costs $\Lambda_{i,j}$. Because of the dynamic and nonlinear nature of the household problem, $\bar{S}_{i,j}$ does not have the same complementarity structure as $H_{i,j,a,k}$, although the two are closely related. In order to describe the initial match surplus function $\bar{S}_{i,j}$ and compare it with $H_{i,j,a,k}$, in Table 9 in appendix, we present linear coefficients obtained by projecting $\bar{S}_{i,j}$ on a polynomial that depends on the partners' inputs (i, j). There we can see that, since $H_{i,j,a,k}$ is increasing in both partners' education, $\bar{S}_{i,j}$ exhibits stronger educational complementarities than $H_{i,j,a,k}$.

Finally, Table 3 shows that cohabiting relationships produce larger gains, but moving in together represents a significant sunk cost. More than half of all new couples turn out to be of high quality. High-quality relationships produce substantially larger gains than low-quality ones, with the differential comparable to almost twice the extra gains associated with cohabitation. As shown in Figure 2, low-quality relationships are not sustainable in the long term, the partners are less likely to move in together, and most of them break up within five years of the match. This rationalizes the duration profile of separation rates presented in Figure 1, which is initially steep but flattens after five years from the match.

[Figure 2 about here.]

6.2. Interpreting the structural estimates

[Table 4 about here.]

The HCMST survey collects information on how couples met, in particular on the meeting place and the channels that led to the match. In order to shed light on the actual nature of search frictions, we use our structural estimates of meeting costs and expected match gains to run a series of regressions of the following type:

$$y_{i,j,l} = \beta_0 + \beta_1 \Lambda_{i,j} + \beta_2 \bar{S}_{i,j} + \xi_l, \tag{6.1}$$

where $y_{i,j,l}$ is a variable describing how match l took place (e.g., in school? in a bar?), while i and j indicate the partners' observed types at the time of the match. The regression results describe how the matching takes place in practice for different types of couples. For instance, they allow us to describe through which channels pairs associated with high costs effectively meet and match. Similarly, they allow us to describe in what context and through which channels high-surplus matches take place.

The results are reported in Table 4, and suggest that partners who come from the same town and went to the same high school benefit from comparatively low meeting costs, but their expected match gains are also low. On the other hand, partners who went to the same college and/or met through the intermediation of friends not only face low meeting costs, but also benefit from higher expected match gains. In Table 10, we provide supplementary descriptive statistics showing that these matches usually occur earlier in the life cycle, particularly for partners who went to the same high school or college. These findings suggest that the marriage market is primarily segmented across space, educational institutions, and friendship networks. Yet, they also suggest that not all types of segmentation favor the formation of high-surplus matches. While the college campus and friendship network represent cost-effective channels to generate rewarding and stable relationship opportunities, the hometown community and high school are only interesting channels because of the low associated costs.

Table 4 shows that finding a partner through online dating, in a bar, on vacation, in the workplace, or at a party is associated with higher costs. However, partners who met online or at a party can expect higher gains, consistent with the idea that this behavior, while costly in terms of money and/or time, reflects the agents' deliberate efforts to search for their preferred type of partner. Table 10 shows that partners who met in a bar or online are on average older. This suggests that these channels may be particularly exploited by agents who happen to be single after completing their studies or moving from their hometown. Interestingly, in our 2009 estimation sample, couples who met in a bar are more likely to be racially homogamous than the average, whereas couples who met online are neither more racially nor more educationally homogamous than the average.

On the other hand, our results suggest that agents avoid looking for a partner in the workplace or through the intermediation of coworkers. Finding a partner at work tends to be expensive and does not lead to more rewarding relationships. Similarly to online dating platforms and bars, the workplace is also a channel more often used by older agents. Yet, couples who met at work are less well sorted in terms of educational achievements, which suggests that the workplace is not an ideal place to look for a partner. Finally, matches formed thanks to family intermediaries are also associated with comparatively low expected match gains, possibly because the agents' preferences do not align with those of their family members.

As an additional check, we can also correlate the estimated match surplus for all incumbent couples with directly observed couple outcomes, e.g., if the partners are legally married and how satisfied they are with the relationship.[29] We use weighted OLS to account for the probability that a couple with observables (i, j, a, d) is of type k after d years of relationship. The findings can be found in Table 11 in the appendix, but they are briefly summarized here. A higher match surplus is associated with a higher probability that the partners are legally married, that their parents approve the relationship, and that the male partner earns more than the female. It is also associated with higher relationship satisfaction as measured on a five-level Likert scale, a higher household income, and a lower probability that the respondent is unemployed. This provides further evidence that the dimensions of observed heterogeneity studied in the model are important determinants of relationship outcomes.

6.3. Robustness checks

Next, we look at potential differences in match quality dynamics across race and education. First, we estimate an alternative specification of the model where returns from cohabitation are allowed to differ for homogamous couples. Table 12 shows that both racial and educational complementarities only materialize after couples move in together. This is consistent with the idea that traits such as race and education are inputs in the production of household public goods (Chiappori et al., 2017). Since agents take the intertemporal value of a match into account, these complementarities partly drive sorting in marriage markets. Yet, in our analysis, allowing for heterogeneous returns to cohabitation does not alter the conclusions on

[29] Notice that almost all questions about the status of the relationship are only asked once, in the first wave of the HCMST survey.

the structure of meeting costs, which are still found to be lower for homogamous couples.

Second, we estimate a heteroskedastic version of our model to test if the volatility of the idiosyncratic match-quality component z differs between homogamous and mixed couples. We let the variance of z differ for same-race and different-race couples, and find that the latter experience stronger volatility. This suggests that, while racial homogamy is not associated with higher per-period gains $H_{i,j,a,k}$, it may benefit partners through lower match-quality volatility. In an alternative specification, we let the variance of z differ for same-education and different-education couples, and find that the latter experience less volatility. In this case, this may reflect a positive correlation between labor market shocks. These findings are summarized in Table 13 in the appendix, where we also show that, in neither case, allowing for heteroskedasticity affects our conclusions about marriage market segmentation and complementarities.

Third, we estimate another version of our model where the probability of drawing a high match quality upon a match is allowed to be different between homogamous and mixed couples. The findings are summarized in Table 14 in the appendix. When $\mu_{i,j,2}$ is allowed to differ for same-education and different-education couples, we find that 68% of the former are high quality upon a match, compared to 44% of the latter. In contrast, educational complementarities in the per-period match gains $H_{i,j,a,k}$ are non-significant. This sheds a different light on why people prefer to match with partners with the same education, and it explains why different-education couples are more likely to break up in the very first years of the relationship, as shown in Figure 1, while in the longer run their odds of separation are comparable to those of same-education couples. However, our estimates of the meeting cost function $\Lambda_{i,j}$ are robust to this more flexible specification.

An analogous exercise yields opposite findings when focusing on racial homogamy. Table 14 shows that 52% of all new same-race couples are high quality, compared to 86% of all new different-race couples. Moreover, in opposition to our baseline findings, we do find evidence of racial complementarities in the per-period gains $H_{i,j,a,k}$, and their size is comparable to that of educational complementarities. These results suggest that, while interracial couples face higher search frictions and lower match gains, they may benefit from better chemistry from the start of the relationship. Yet, once again, our estimates of the meeting cost function $\Lambda_{i,j}$ are very close to the baseline, and confirm that racial segmentation in the marriage market is strong.

6.4. More heterogeneity

[Table 5 about here.]

Table 5 reports the findings obtained with a refined version of the model with four racial groups (white, black, Hispanic, other) and five educational groups (high school dropout, high school diploma, some college, college degree, graduate degree). Both qualitatively and quantitatively the findings are very similar to those obtained with our baseline model, which suggests that the main fault lines in the market are between individuals with and without college degree when it comes to education and between the white majority and non-white minorities when it comes to race.

Interestingly, Table 5 shows that black women and Hispanic men experience the strongest search frictions. In contrast, black men face comparatively low frictions. As shown in Table 1, black women outnumber black men in the population, while Hispanic men outnumber Hispanic women. In racial groups with a gender ratio imbalance, individuals whose gender is in excess supply may have to undergo costly efforts to sample potential partners when the latter are scarce. This adds to the presence of racial segmentation, which limits their ability to look for partners elsewhere. Hence, mate search is particularly costly for population groups like black women and Hispanic men, a cost that, according to the predictions of our model, also reflects in an unfavorable sharing rule at the start of a new relationship.

We do not find strong differences in match gains across racial groups, although non-white individuals seem to enjoy slightly lower gains overall, with the only exception being men in the category "other", which mainly includes Non-Hispanic Native Americans and Asians. Such differences may partly reflect a human capital divide between racial groups beyond that already captured by the educational categories included in the model. Both this human capital divide and gender ratio imbalances have been linked with differences in marriage, separation, employment, and fertility patterns across racial groups (Seitz, 2009; Beauchamp et al., 2018).

6.5. Search frictions and assortative mating

[Table 6 about here.]

In order to quantify how frictions and preferences contribute to the observed sorting patterns in our estimation sample, we can compare them with the mating patterns obtained if sorting were entirely driven by meeting costs, and not by match surplus. Alternatively, we can simulate the mating patterns obtained if individuals disregarded meeting costs and based their mating decisions uniquely on match gains. In Table 6, we calculate the share of homogamous pairs among couples who have just started dating under these two counterfactual scenarios, as well as the same shares if singles were paired randomly. In this exercise, when simulating these counterfactual assignments, the supplies of partners are held fixed and identical to those observed in the data, i.e. we do *not* calculate a new SSE, an exercise left for the next section.

We first consider our baseline model with two racial and three educational groups. If the pairing were random, only 53.3% of all matches in a given year would be mixed, while in our sample 88.2% of all new couples are racially homogamous. This is to a very large extent explained by meeting costs; if the latter were the only determinant of mating strategies, 86.9% of all couples would be racially homogamous. If meeting costs were absent and mating strategies were entirely determined by expected match gains, the share of same-race couples would drop to 55.5%. Hence, search frictions alone could almost entirely explain the strong racial homogamy observed among newly formed couples.

The situation is quite different for education; if the pairing were random, the share of same-education couples would be 33.5%, whereas it is observed to be 49.2% in the data. If meeting costs alone shaped mating patterns, then the share would drop to 42.0%. On the other hand, if meeting costs were absent and mating strategies were entirely determined by expected match gains, the share would drop to 40.8%. Hence, in this case, search frictions could only explain about half the educational homogamy observed in the data.

When we increase the number of racial and educational categories and repeat the exercise, we find that the share of same-race couples would drop from 77.5% to 39.5% in the absence of meeting costs, close to the same share under random matching (39.1%), while the share of same-education couples would drop from 36.8% to 28.3% (24.0% under random matching). If meeting costs were the only determinant of mating strategies, 73.1% of all new couples would be racially homogamous, while 32.0% would be educationally homogamous.

6.6. Counterfactual experiments

[Table 7 about here.]

In Section 6.5, we calculated matching outcomes while holding the size and composition of the pool of agents available on the market constant. However, the comparative statics of our model suggest that this pool is endogenous and results from mating and separation strategies, which in turn are influenced by the (re)match opportunities available to the agents. In this section, we modify certain parameters of either the meeting cost function or the per-period match gains, compute the resulting SSE, and compare it with the SSE obtained for the fitted model.

In Experiment I, we remove the penalty faced by different-race pairs in the meeting cost function, corresponding to 3.90 in our augmented model (see Table 5), so that meeting costs are the same for same-race and different-race pairs. Table 7 shows that the steady-state share of same-race couples would drop from 82.8% to 59.8%. Removing racial barriers reduces the share of matched individuals by about 6pp at the SSE. The overall decrease in frictions leads to a joint increase in match and separation rates, as agents become more likely to quit matches they are unhappy with and look for better ones. Agents expect a higher level of bliss from a relationship, and the average volatile match quality z more than doubles. With a higher turnover, agents are also more often able to find a partner within their age group, and the share of same-age couples increases by 2pp. In contrast, because of the unequal educational distribution across racial groups, after removing racial segmentation, agents have access to a more educationally diverse pool of partners, and the share of same-education couples slightly declines, albeit by only 0.7pp. Finally, since agents experiment with more partners and relationships tend to be shorter, the fraction of couples cohabiting and with a high persistent match quality k decreases.

In Figure 3, we can see that removing racial segmentation decreases the odds of being in a relationship at the SSE for all groups expect Hispanic men, who experience an unfavorable gender ratio in the Hispanic population and thus benefit from access to a broader and more diverse pool of partners. In panel (c), we can see how an improvement in their match prospects also translates into greater bargaining power. In contrast, the share of partnered black men decreases from 60.4% to 50.1%, more than the corresponding share for white men, which decreases from 66.3% to 58.6%. In fact, black men benefit from a particularly favorable gender ratio in the black population, and removing racial segmentation makes the market more competitive for them.

On the other side of the market, panel (b) in Figure 3 shows that, while white women are still the most likely to be in a relationship at the counterfactual SSE, they experience the strongest decline relative to the benchmark SSE, from 66.7% to 58.8%. In contrast, the share of partnered black women only decreases by 2pp, from 53.0% to 51.0%. The gap in the odds of being in a relationship between white and black women goes from 13.7% to 7.8%, i.e. 43.1% of such a gap is closed after removing racial segmentation. The remaining gap still reflects the fact that black women experience higher meeting costs and lower match gains (see Table 5) and are on average less educated. While they experienced a narrow market due to the scarcity of same-race partners at the benchmark SSE, black women now benefit from access to a broader and more diverse pool of partners, and their bargaining power in relationships increases, as shown in panel (d).

[Figure 3 about here.]

In Experiment II, we remove the penalty faced by different-education couples in the meeting cost function, corresponding to 0.43 in our augmented model with five educational levels (see Table 5). Table 7 shows that, qualitatively, the changes are analogous to those we see in

Experiment I. Quantitatively, the changes are smaller. Most importantly, the steady-state share of same-education couples only drops from 56.9% to 52.4%. The change is small relative to Experiment I not only because the estimated penalty is lower, but also because, in the presence of educational complementarities, agents still prefer to court partners with the same education. When frictions weaken, agents become more selective, and thus they will more often turn down (or break up with) different-education partners, while actively seeking same-education ones. Such adjustments in strategies partly curb the decrease in educational homogamy spurred by the removal of educational segmentation.

In Experiments III and IV, we respectively remove racial and educational complementarities in the per-period match gains $H_{i,j,a,k}$. In both experiments, removing a source of complementarities reduces the incentives to start a relationship, and leads to lower match rates, higher separation rates, and a lower share of matched individuals at the SSE. Match quality takes on a more important role, and agents are ready to wait longer for a high-quality match. These effects are stronger when removing educational complementarities, as the latter are found to be larger. In Experiment III, after removing racial complementarities, the share of same-race couples decreases from 89.8% to 86.8%, a much smaller effect than what is obtained when removing racial segmentation in Experiment I. In contrast, in Experiment IV, removing educational complementarities leads to a decrease in the share of same-education couples from 56.9% to 39.4%, a much larger effect than what is obtained in Experiment II.

7. Conclusion

In this paper, we ask to what extent marriage markets are segmented with respect to education and race, and to what extent patterns of positive assortative mating are shaped by search frictions as opposed to complementarities between partners' inputs. We complement the existing literature on this topic with a novel structural approach to the question. Our contribution is threefold. First, we build on the search-and-matching literature and discuss how panel data on relationship duration can be used to identify search frictions in the matching function in a broad class of models. In particular, we draw attention to the fact that search frictions are only identified if we can observe the continuation/separation decisions of incumbent couples whose stakes are still comparable to those of couples who have just formed. Second, we build an empirically tractable dynamic matching model where, in every period, singles participate to an assignment game à la Choo and Siow (2006). Their mating strategies depend on both expected match gains and search frictions in the form of meeting costs, as in Jaffe and Weber (2019). Couples experience changes in match quality, although the latter is partly persistent, and can decide to break up and look for a new partner. Third, we estimate such a model and show that racial segmentation is strong relative to racial complementarities. On the other hand, educational complementarities play an important role in shaping assortative mating with respect to education. If search frictions were absent and agents chose their partners entirely based on the expected match gains, the share of racially homogamous couples would drop from the observed 88.2% to 55.5%. In comparison, if matching were random, this share would be 53.3%. On the other hand, if search frictions were absent, the share of same-education couples would drop from 49.2% to 40.8%, well above the same share obtained if matching were random (33.5%). Our counterfactual experiments show that removing racial segmentation could benefit minority groups who experience an unfavorable gender ratio when marriage markets are segmented into racially homogeneous clusters, such as Hispanic men and black women.

A. Proofs

Lemma 2. The proof is an application of Proposition 2 in Kasahara and Shimotsu (2009). We exploit the Conditional Independence assumption to rewrite (3.5) as

$$\widetilde{\Pr}\{\{i_t, j_t, x_t\}_{t=1}^{d} | i_0, j_0\} = \sum_{k=1}^{K} \pi_{i_0,j_0,k} \prod_{t=1}^{d} \alpha_{i_t,j_t,k}(x_t) \tag{A.1}$$

where the lhs can be consistently estimated with a panel of length D. For any $(i, j) \in \mathcal{I} \times \mathcal{J}$, we define

$$p_d = \widetilde{\Pr}\{\{i_t = i, j_t = j, x_t = \bar{x}_{i,j}\}_{t=1}^{d} | i_0 = i, j_0 = j\} > 0, \tag{A.2}$$

which indicates the probability for a couple with initial individual types (i, j) of staying together and remaining in the initial state $\bar{x}_{i,j}$ for d periods, with $p_d > 0$ by assumption. The subscripts (i, j) and argument x of $\alpha_{i,j,k}(x)$ can be omitted in what follows as they are held constant over time. The probability p_d simplifies to

$$p_d = \sum_{k=1}^{K} \pi_k \alpha_k^d. \tag{A.3}$$

We define the following matrices of unobservables

$$L = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_1^{K-1} \\ 1 & \alpha_2 & \cdots & \alpha_2^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_K & \cdots & \alpha_K^{K-1} \end{bmatrix} \tag{A.4}$$

$$V = diag(\pi_1, ..., \pi_K), \quad Q = diag(\alpha_1, ..., \alpha_K). \tag{A.5}$$

We can relate unobservables and observables through the relationships

$$P = L'QL = \begin{bmatrix} 1 & p_1 & \cdots & p_{K-1} \\ p_1 & p_2 & \cdots & p_K \\ \vdots & \vdots & \ddots & \vdots \\ p_{K-1} & p_K & \cdots & p_{2K-2} \end{bmatrix} \tag{A.6}$$

$$P^* = L'VQL = \begin{bmatrix} p_1 & p_2 & \cdots & p_K \\ p_2 & p_3 & \cdots & p_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ p_K & p_{K+1} & \cdots & p_{2K-1} \end{bmatrix}. \tag{A.7}$$

The conditional probabilities $(\alpha_1, ..., \alpha_K)$ correspond to the eigenvalues of

$$P^{-1}P^* = L^{-1}V^{-1}(L')^{-1}L'VQL = L^{-1}QL \tag{A.8}$$

where L is invertible since $\alpha_k \neq \alpha_{k'}$ for any $k \neq k'$, and V is invertible since $\pi_k \in (0,1)$ for any k. Once recovered $(\alpha_1, ..., \alpha_K)$, we can reconstruct L and obtain $diag(\pi_1, ..., \pi_K)$ from $V = (L')^{-1} P L^{-1}$. Once obtained $\alpha_{i,j,k}(\bar{x}_{i,j})$ for any (i,j,k), the proof continues as in Lemma 1. $\square$

Proposition 1. When $K = 1$ and Lemma 1 holds, S is recovered from separation rates and the identification of the set of parameters Λ follows immediately from the matching function (3.1), as discussed in the main text. When $K > 1$ and Lemma 2 holds, both S and π are recovered from relationship duration data. Recall that $\pi_{i,j,k}$ denotes the probability that a new match between types i and j is of type k. Hence, $MF_{i,j}\pi_{i,j,k}$ yields the mass of new matches between types i and j of quality k. Ex ante, the probability that a pair (i,j) draws match quality k is denoted $\mu_{i,j,k}$, so that $\sum_{k=1}^{K} \mu_{i,j,k} = 1$. The matching function (3.1) can be rewritten as

$$MF_{i,j}\pi_{i,j,k} = M_{i,j,k}(S, \Lambda, \mu; \tilde{n}^m, \tilde{n}^f) \quad \forall (i,j,k) \in \mathcal{I} \times \mathcal{J} \times \{1, ..., K\}. \tag{A.9}$$

Since the lhs of (A.9) is known, the matching function provides $|\mathcal{I}| \times |\mathcal{J}| \times K$ restrictions. However, there are $|\mathcal{I}| \times |\mathcal{J}| \times (K-1)$ additional parameters to estimate since μ is unknown. This limits the number of parameters of model M in the set Λ that can be identified through (A.9) to $|\mathcal{I}| \times |\mathcal{J}|$. $\square$

Proposition 2. Assumptions 1 and 2 in Galichon and Salanié (2022) are respected since the total match surplus is given by

$$\bar{S}_{i,j} - \Lambda_{i,j} + \varepsilon_j^m + \varepsilon_i^f \tag{A.10}$$

and the payoffs for singles are given by ε_0^m and ε_0^f, with the elements of ε^m and ε^f being independent and extreme value type I distributed. From problem (4.7), since we have a continuum of agents of type i with mass $\tilde{n}_i^m$, we can derive their aggregate demand for women of type j, $MF_{i,j}^d$, defined as

$$\begin{aligned} MF_{i,j} =& \tilde{n}_i^m \exp\left(\theta_{i,j}(\bar{S}_{i,j} - \Lambda_{i,j})\right) \left[1 + \sum_{j'} \exp\left(\theta_{i,j'}(\bar{S}_{i,j'} - \Lambda_{i,j'})\right)\right]^{-1} && \text{(A.11)} \\ =& c_i^m \exp\left(\theta_{i,j}(\bar{S}_{i,j} - \Lambda_{i,j})\right) && \text{(A.12)} \end{aligned}$$

where c_i^m corresponds to the mass of agents who stay single at the end of the matching round. We can derive a symmetric expression for women's aggregate demand for men, $MF_{i,j}$, defined as

$$MF_{i,j} = c_j^f \exp\left((1 - \theta_{i,j})(\bar{S}_{i,j} - \Lambda_{i,j})\right). \tag{A.13}$$

The market clears when (A.12) equals (A.13), so that the matching function is given by

$$MF_{i,j} = \exp\Big(\frac{\bar{S}_{i,j} - \Lambda_{i,j}}{2}\Big) \sqrt{c_i^m c_j^f}. \tag{A.14}$$

Moreover, we can derive the bargaining weight $\theta_{i,j}$ from

$$\frac{c_j^f}{c_i^m} = \exp\left((2\theta_{i,j} - 1)(\bar{S}_{i,j} - \Lambda_{i,j})\right). \tag{A.15}$$

The assignment described by the matching function $MF_{i,j}$ is only feasible if the following constraints are met:

$$\tilde{n}_i^m = c_i^m + \sum_{j'} \exp\Big(\frac{\bar{S}_{i,j} - \Lambda_{i,j}}{2}\Big)\sqrt{c_i^m c_{j'}^f} \;\; \forall i \in \mathcal{I} \tag{A.16}$$

$$\tilde{n}_j^f = c_j^f + \sum_{i'} \exp\Big(\frac{\bar{S}_{i,j} - \Lambda_{i,j}}{2}\Big)\sqrt{c_{i'}^m c_j^f} \;\; \forall j \in \mathcal{J}. \tag{A.17}$$

There exist a unique solution to this system of $|\mathcal{I}| + |\mathcal{J}|$ equations. This is established by Theorem 5 in Galichon and Salanié (2022). □

Proposition 3. The parameters (β, F^m, F^f, G) are taken as given. The constructive proof consists of the following steps:

1. Since there exists a $D \geq 2K - 1$ s.t. each element of (3.6) is positive, we can recover $\alpha_{i,j,a,k}(a)$ and $\pi_{i,j,a,k}$ for any (i,j,a,k) from the survival profile (3.6), exactly as explained in the proof of Lemma 2. We can now recover $\alpha_{i,j,a,k}(a')$ for any (i,j,a,k) and any $a' \neq a$ from the survival profile

$$\left\{\widetilde{\Pr}\{\{i_t = i, j_t = i, a_t = a\}_{t=1}^{d-1}, i_d = i, j_d = i, a_d = a' | i_0 = i, j_0 = j\}\right\}_{d=1}^{D}, \tag{A.18}$$

   which gives the share of couples who switched from living arrangement a to a' after $d = 1, ..., D$ periods. The d-th element of (A.18) corresponds to

$$\sum_{k=1}^{K} \pi_{i,j,a,k}\alpha_{i,j,a,k}^{d-1}(a)\alpha_{i,j,a,k}(a'). \tag{A.19}$$

2. For any (i,j,a,k), the probabilities $\{\alpha_{i,j,a,k}(a')\}_{a'=0}^{A}$ uniquely determine the conditional match surplus $\{S_{i,j,a,k}(a')\}_{a'=1}^{A}$, since by construction the conditional match surplus is zero in case of separation. This follows from Proposition 1 in Hotz and Miller (1993).

3. Conditional on G, the probability of drawing match quality k conditional on types i and j, $\mu_{i,j,k}$, is identified from the restriction

$$\mu_{i,j,k} \Pr\{a = \arg\max_{a' \in \{1,\ldots,A\}} S_{i,j,1,k}(a') + z_a\} = \pi_{i,j,a,k}. \tag{A.20}$$

   where the rhs was obtained at step 1, while the match surplus $\{S_{i,j,1,k}(a')\}_{a'=1}^{A}$ was obtained at step 2. Notice that, when the shocks are extreme value type I, (A.20) simplifies to (4.21).

4. For given (F^m, F^f), Theorem 4 in Galichon and Salanié (2022) establishes that the difference $\bar{S}_{i,j} - \Lambda_{i,j}$ and the male partner's share $\theta_{i,j}$ are identified for any (i,j) from the matching outcome MF. Since $\bar{S}_{i,j}$ corresponds to (4.6) for given G, the identification of $\Lambda_{i,j}$ follows immediately. When shocks are extreme value type I, for given $\bar{S}$, $\tilde{n}^m$ and $\tilde{n}^f$,

$$MF_{i,j} = \exp\left[(\bar{S}_{i,j} - \Lambda_{i,j})/2\right]\sqrt{c_i^m c_j^f} \tag{A.21}$$

$$\Lambda_{i,j} = \bar{S}_{i,j} + \ln c_i^m + \ln c_j^f - 2\ln MF_{i,j} \tag{A.22}$$

$$\theta_{i,j} = \frac{1}{2}\left[1 + \left(\ln c_j^f - \ln c_i^m\right)\left(2\ln MF_{i,j} - \ln c_i^m - \ln c_j^f\right)^{-1}\right] \quad \forall (i,j) \in \mathcal{I} \times \mathcal{J}, \tag{A.23}$$

where $c_i^m = \tilde{n}_i^m - \sum_j MF_{i,j}$ and $c_j^f = \tilde{n}_j^f - \sum_i MF_{i,j}$.

5. For given (F^m, F^f), as a by-product of the previous step, we also obtain male and female agents' expected value from participating to the marriage market, respectively $\bar{V}_i^m$ and $\bar{V}_j^f$, defined in (4.7). Hence, since $\beta \in (0,1)$ is given and the transition probabilities $\Pr\{i'|i\}$ and $\Pr\{j'|j\}$ are non-parametrically identified, we can recover the reservation values by finding the unique solution of the linear system given by the singles' Bellman equations (4.15) and (4.16).

6. From the Bellman equation for the match surplus (4.14), we obtain the per-period match gains net of the switching costs for any (i,j,a,k,a'):

$$H_{i,j,a',k} - \kappa_{a,a'} = S_{i,j,a,k}(a') - \beta \mathbb{E}_{i'} V^m_{i',0} - \beta \mathbb{E}_{j'} V^f_{0,j'} - \beta \mathbb{E}_{i',j'} \tilde{S}_{i',j',a'k} + V^m_{i,0} + V^f_{0,j}, \tag{A.24}$$

where the transition probability $\Pr\{i',j'|i,j\}$ is non-parametrically identified and $\tilde{S}_{i',j',a'k}$ is given by (4.4) conditional on G. The normalization $\kappa_{a,a} = 0$ helps recover $H_{i,j,a,k}$, and the identification of $\kappa_{a,a'}$ for any $a' \neq a$, with $a, a' \neq 0$, follows.

□

B. Estimation

All parameters $\vartheta \equiv (\lambda, \delta, \zeta, \eta, \mu, \kappa)$ are estimated jointly using the method of moments. The estimator is the global minimum of the following program:

$$\min_{\vartheta} \sum_i (h_i(\vartheta) - \hat{h}_i)^2 \tag{B.1}$$

where $h(\vartheta)$ is a vector of simulated moments and $\hat{h}$ is its empirical counterpart. The length of $h(\vartheta)$ corresponds to the number of parameters in ϑ. For a given draw ϑ, we can calculate the predicted distribution of new and incumbent couples over their full support by solving the assignment game described in Section 4.2 and using the demographic equations described in

Section 4.4. After obtaining estimates ϑ^* and computing the Jacobian matrix $J^* \equiv J(\vartheta^*)$, the covariance matrix of ϑ^* corresponds to:

$$(J^*)^T \Sigma J^* \tag{B.2}$$

where Σ is the covariance matrix of $\hat{h}$, where Σ is computed with 1,000 bootstrap replications from the relevant sample.

In both our baseline and augmented model, whose findings are respectively reported in Table 3 and 5, the moments include: the overall match and separation rate, the match and separation rates by gender and race, the match and separation rates by gender and education, the match and separation rates for same age, same-race, and same-education couples, men's and women's average age at match and at separation, the average rate of transition from non-cohabitation to cohabitation, the separation rate for cohabiting couples, the average duration, and the variance of duration at separation.

C. Model fit

We now discuss how our baseline model, whose estimated parameters are presented in Table 3, fits the data patterns. Panels (a) and (b) in Figure 4 show that the estimated model replicates well the entire age profiles of match formation even though only the average age of men and women in new couples is targeted in the estimation. Similarly, panel (c) shows that the model fits the entire duration profile of separation rates. Panel (d) shows that the model replicates the duration profile of the fraction of cohabiting couples well. In Figure 5, we can see that the model replicates very well the sorting patterns with respect to race and education. In particular, the simulated and observed match rates are very close. The fit is also excellent for separation rates, except for few infrequent types of couples; e.g., the model underpredicts the separation rate of college graduates matched with high school diplomas, but these couples are rare in the data (see Table 2).

In Figure 6, we report the model predictions about the cumulative number of partners by age and gender. Unfortunately, HCMST data do not contain information about the number of past relationships. Hence, we looked at the 2011-2013 sample from the National Survey of Family Growth (NSFG), which is nationally representative for the population aged between 15 and 44. The model generates life-cycle patterns remarkably close to what observed in the data. According to NSFG data, about 95% of both men and women aged between 39 to 42 have already had a cohabiting partner, in line with the model predictions in panels (c) and (d). The average number of past cohabiting relationships for women in this age group is 1.5, whereas it is 1.9 for men. These number are extremely close to the model predictions for men (1.9), whereas they are less than what the model predicts for women (1.8). In other words, the model tends to slightly overpredict the rematch rates of women and, even though it does predict that men tend to have more relationships than women, it cannot fully account for the gender asymmetries we observe in the data.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

D. Additional figures and tables

[Figure 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

[Table 14 about here.]

References

Theodore W Anderson. On estimation of parameters in latent structure analysis. *Psychometrika*, 19(1):1–10, 1954.

Peter Arcidiacono and Robert A Miller. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867, 2011.

Peter Arcidiacono, Andrew Beauchamp, and Marjorie McElroy. Terms of endearment: An equilibrium model of sex and matching. *Quantitative Economics*, 7(1):117–156, 2016.

Andrew Beauchamp, Geoffrey Sanzenbacher, Shannon Seitz, and Meghan M Skira. Single moms and deadbeat dads: The role of earnings, marriage market conditions, and preference heterogeneity. *International Economic Review*, 59(1):191–232, 2018.

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

Michèle Belot and Marco Francesconi. Dating preferences and meeting opportunities in mate choice decisions. *Journal of Human Resources*, 48(2):474–508, 2013.

Alberto Bisin and Giulia Tura. Marriage, fertility, and cultural integration in italy. Working paper 26303, National Bureau of Economic Research, 2019.

Alberto Bisin and Thierry Verdier. "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, 115 (3):955–988, 2000.

Alberto Bisin, Giorgio Topa, and Thierry Verdier. Religious intermarriage and socialization in the united states. *Journal of Political Economy*, 112(3):615–664, 2004.

Michael J Brien, Lee A Lillard, and Steven Stern. Cohabitation, marriage, and divorce in a model of match quality. *International Economic Review*, 47(2):451–494, 2006.

Gustaf Bruze, Michael Svarer, and Yoram Weiss. The dynamics of marriage and divorce. *Journal of Labor Economics*, 33(1):123–170, 2015.

Ken Burdett and Melvyn G Coles. Marriage and class. *The Quarterly Journal of Economics*, 112(1):141–168, 1997.

Hector Chade, Jan Eeckhout, and Lones Smith. Sorting through search and matching models in economics. *Journal of Economic Literature*, 55(2):493–544, 2017.

Liang Chen and Eugene Choo. Identification and parametric estimation of empirical dynamic marriage matching models. *The Economic Journal*, 133(651):1005–1035, 2023.

Liang Chen, Eugene Choo, Alfred Galichon, and Simon Weber. Matching function equilibria with partial assignment: existence, uniqueness and estimation. 2021.

Laurens Cherchye, Thomas Demuynck, Bram De Rock, and Frederic Vermeulen. Household consumption when the marriage is stable. *American Economic Review*, 107(6):1507–34, 2017.

Pierre-André Chiappori, Monica Costa Dias, Sam Crossman, and Costas Meghir. Changes in assortative matching and inequality in income: Evidence for the uk. Working paper 26933, National Bureau of Economic Research, 2020.

Pierre-André Chiappori. Rational household labor supply. *Econometrica*, 56(1):63–90, 1988.

Pierre-Andre Chiappori and Maurizio Mazzocco. Static and intertemporal household decisions. *Journal of Economic Literature*, 55(3):985–1045, 2017.

Pierre-André Chiappori, Murat Iyigun, and Yoram Weiss. Investment in schooling and the marriage market. *American Economic Review*, 99(5):1689–1713, 2009.

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

Eugene Choo. Dynamic marriage matching: An empirical framework. *Econometrica*, 83(4): 1373–1423, 2015.

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

Edoardo Ciscato. The changing wage distribution and the decline of marriage. Working paper, 2019.

Pauline Corblet, Jeremy Fox, and Alfred Galichon. Repeated matching games: an empirical framework. Working paper, 2023.

Thomas Demuynck and Tom Potoms. Weakening transferable utility: The case of non-intersecting pareto curves. *Journal of Economic Theory*, 188:105035, 2020.

Javier Díaz-Giménez and Eugenio Giolito. Accounting for the timing of first marriage. *International Economic Review*, 54(1):135–158, 2013.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, 79(2):645–677, 2012.

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

Paul W Eastwick, Laura B Luchies, Eli J Finkel, and Lucy L Hunt. The predictive validity of ideal partner preferences: a review and meta-analysis. *Psychological bulletin*, 140(3):623, 2014.

Jan Eeckhout and Philipp Kircher. Sorting and decentralized price competition. *Econometrica*, 78(2):539–574, 2010.

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

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

Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121(2):673–697, 2006.

Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Racial preferences in dating. *The Review of Economic Studies*, 75(1):117–132, 2008.

Christopher Flinn and James Heckman. New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, 18(1):115–168, 1982.

Jeremy T Fox. An empirical, repeated matching game applied to market thickness and switching. 2008.

Roland G Fryer Jr. Guess who's been coming to dinner? trends in interracial marriage over the 20th century. *Journal of Economic Perspectives*, 21(2):71–90, 2007.

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

Alfred Galichon, Scott Duke Kominers, and Simon Weber. Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy*, 127(6):2875–2925, 2019.

Marc Goñi. Assortative matching at the top of the distribution: Evidence from the world's most exclusive marriage market. *American Economic Journal: Applied Economics*, 14(3): 445–487, 2022.

Marion Goussé, Nicolas Jacquemet, and Jean-Marc Robin. Marriage, labor supply, and home production. *Econometrica*, 85(6):1873–1919, 2017.

Jeremy Greenwood, Nezih Guner, Georgi Kocharkov, and Cezar Santos. Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics*, 8(1):1–41, 2016.

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

Peter Hall and Xiao-Hua Zhou. Nonparametric estimation of component distributions in a multivariate mixture. *The annals of statistics*, 31(1):201–224, 2003.

James Heckman, Burton Singer, et al. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320, 1984.

James J Heckman and Burton Singer. Econometric duration analysis. *Journal of Econometrics*, 24(1-2):63–132, 1984.

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

V Joseph Hotz and Robert A Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

Yingyao Hu and Matthew Shum. Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, 171(1):32–44, 2012.

Sonia Jaffe and Simon Weber. The effect of meeting rates on matching outcomes. *Economic Theory*, 67(2):363–378, 2019.

Matthijs Kalmijn and Henk Flap. Assortative meeting and mating: Unintended consequences of organized settings for partner choices. *Social forces*, 79(4):1289–1312, 2001.

Hiroyuki Kasahara and Katsumi Shimotsu. Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77(1):135–175, 2009.

Michael P Keane and Kenneth I Wolpin. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103, 2001.

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

Ilse Lindenlaub. Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, 84(2):718–789, 2017.

Jeremy Lise and Ken Yamada. Household sharing and commitment: Evidence from panel data on individual expenditures and time use. *The Review of Economic Studies*, 86(5):2184–2219, 2019.

Corinne Low. The human capital - reproductive capital tradeoff in marriage market matching. *Journal of Political Economy*, forthcoming.

Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002.

Mihai Manea. Steady states in matching and bargaining. *Journal of Economic Theory*, 167: 206–228, 2017.

Rosa Matzkin. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica*, 60(2):239–70, 1992.

Maurizio Mazzocco. Household intertemporal behaviour: A collective characterization and a test of commitment. *The Review of Economic Studies*, 74(3):857–895, 2007.

Marjorie B McElroy and Mary Jean Horney. Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review*, 22(2):333–49, 1981.

Sara McLanahan. Diverging destinies: How children are faring under the second demographic transition. *Demography*, 41(4):607–627, 2004.

Robert A Miller. Job matching and occupational choice. *Journal of Political economy*, 92(6): 1086–1120, 1984.

Dale T Mortensen. Matching: Finding a partner for life or otherwise. *American Journal of Sociology*, 94:S215–S240, 1988.

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

Valerie Kincade Oppenheimer. A theory of marriage timing. *American Journal of Sociology*, 94(3):563–591, 1988.

Michael J Rosenfeld and Reuben J Thomas. Searching for a mate: The rise of the internet as a social intermediary. *American Sociological Review*, 77(4):523–547, 2012.

Reuben J. Thomas Rosenfeld, Michael J. and Maja Falcon. How couples meet and stay together, 2018. Waves 1, 2, and 3: Public version 3.04, plus wave 4 supplement version 1.02 and wave 5 supplement version 1.0 and wave 6 supplement ver 1.0 [Computer files].

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033, 1987.

Christine R Schwartz. Trends and variation in assortative mating: Causes and consequences. *Annual Review of Sociology*, 39:451–470, 2013.

Shannon Seitz. Accounting for racial differences in marriage and employment. *Journal of Labor Economics*, 27(3):385–437, 2009.

Lloyd S Shapley and Martin Shubik. The assignment game i: The core. *International Journal of Game Theory*, 1(1):111–130, 1971.

Andrew Shephard. Marriage market dynamics, gender, and the age gap. Working paper, 2018.

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

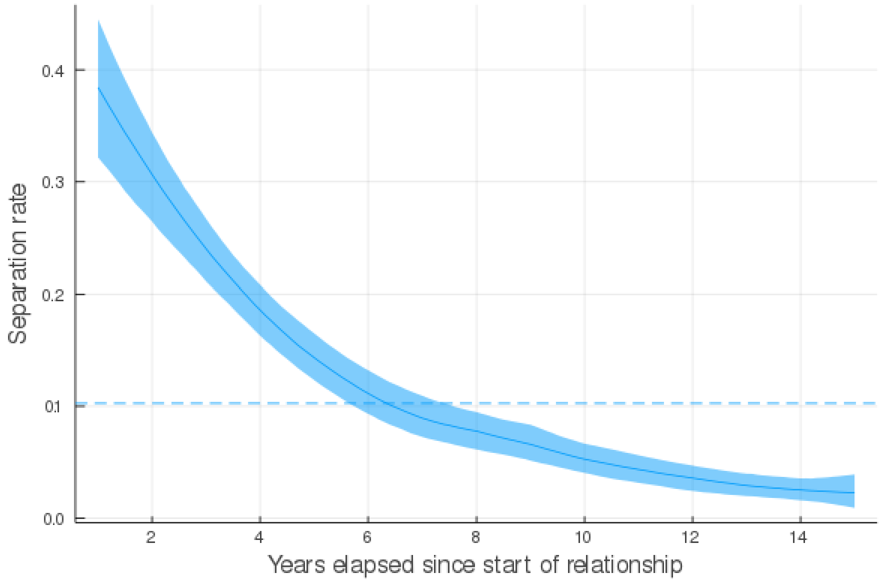
Lones Smith. The marriage model with search frictions. *Journal of Political Economy*, 114(6): 1124–1144, 2006.

Alessandra Voena et al. Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples? *American Economic Review*, 105(8):2295–2332, 2015.

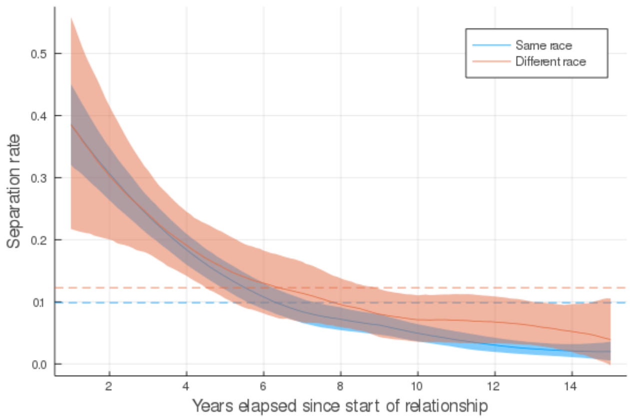
Linda Y Wong. Why so only 5.5% of black men marry white women? *International Economic Review*, 44(3):803–826, 2003a.

Linda Y Wong. Structural estimation of marriage models. *Journal of Labor Economics*, 21(3): 699–727, 2003b.

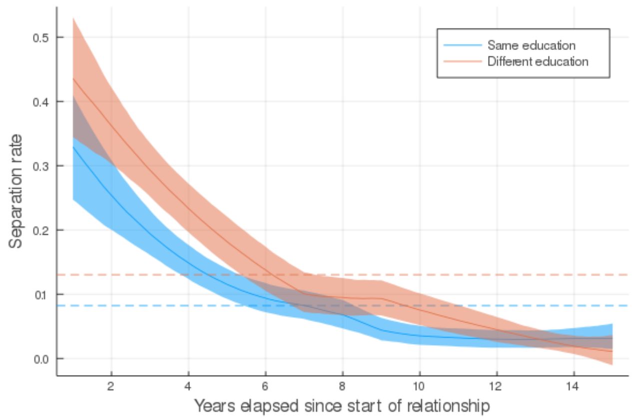
Figure 1: Separation rate by duration



(a)



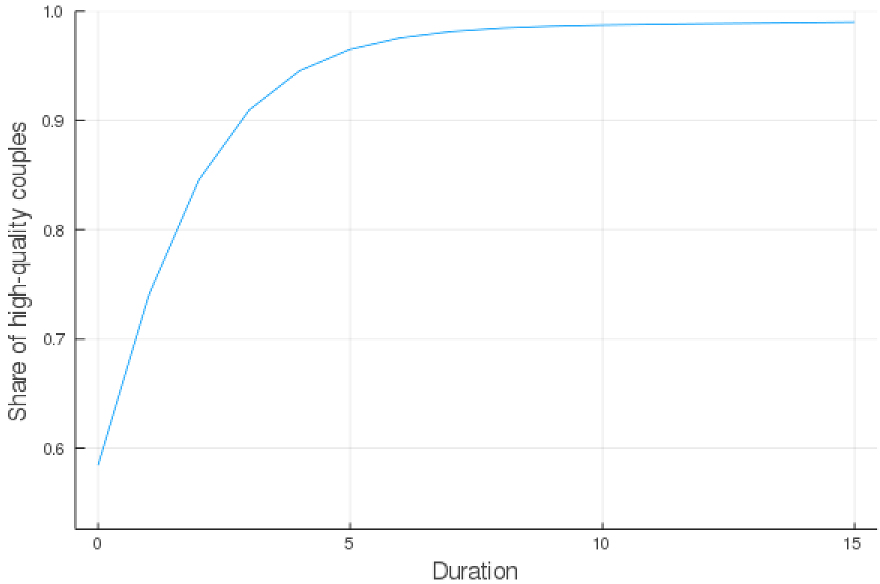
(b)



(c)

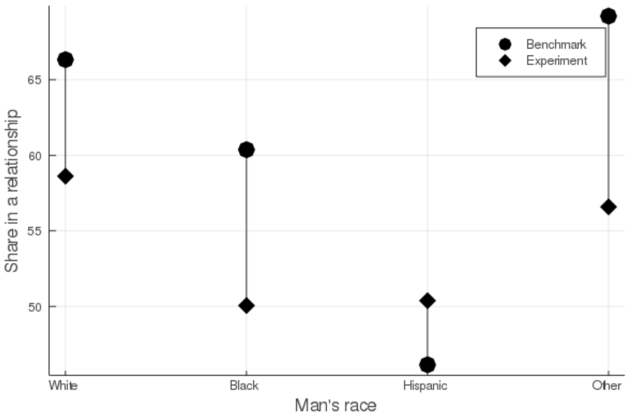
Notes. The separation rate corresponds to the fraction of couples that break up in a given year. The solid lines represent the rate conditional on relationship duration with its 90% two-sided confidence interval for the entire sample (panel (a)), the subsample of same-race vs different-race couples (panel (b)), and the subsample of same-education vs different-education couples (panel (c)). The dashed lines represent the corresponding average separation rates throughout the first 15 years of the relationship. Notice that in panel (b) we use two categories for race (white, non-white), while in panel (c) we use three categories for education (high school diploma or less, some college, college degree). In Figure 7, we show that these facts are robust to different categorizations of the race and education variables.

Figure 2: Predicted share of high-quality matches by duration

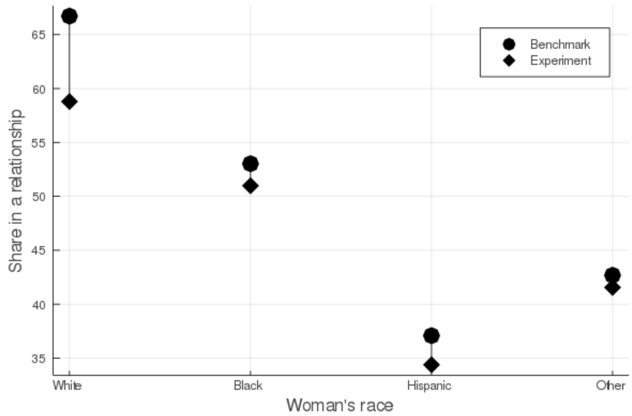


Notes. We report the predicted share of high-quality matches by relationship duration implied by the estimates presented in Table 3.

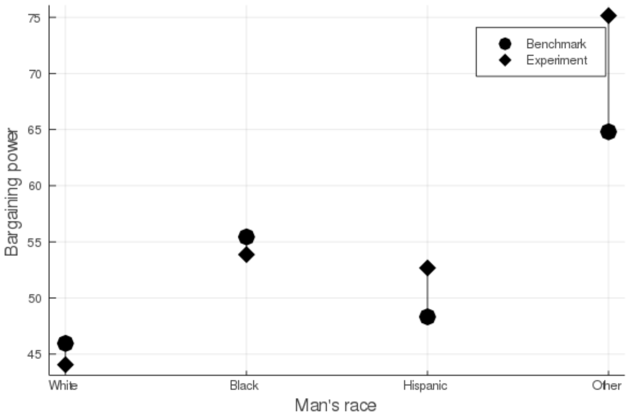
Figure 3: Experiment I



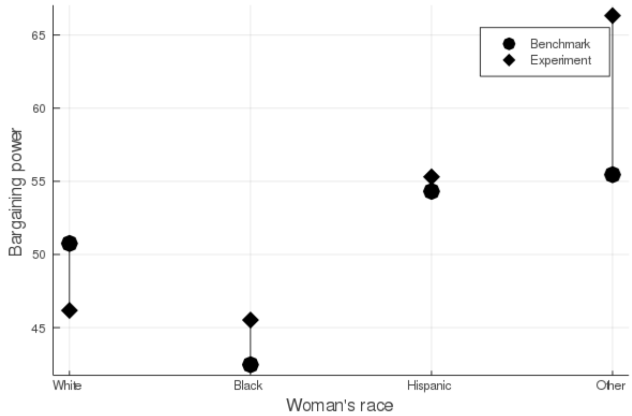
(a) Changes in % of partnered men



(b) Changes in % of partnered women



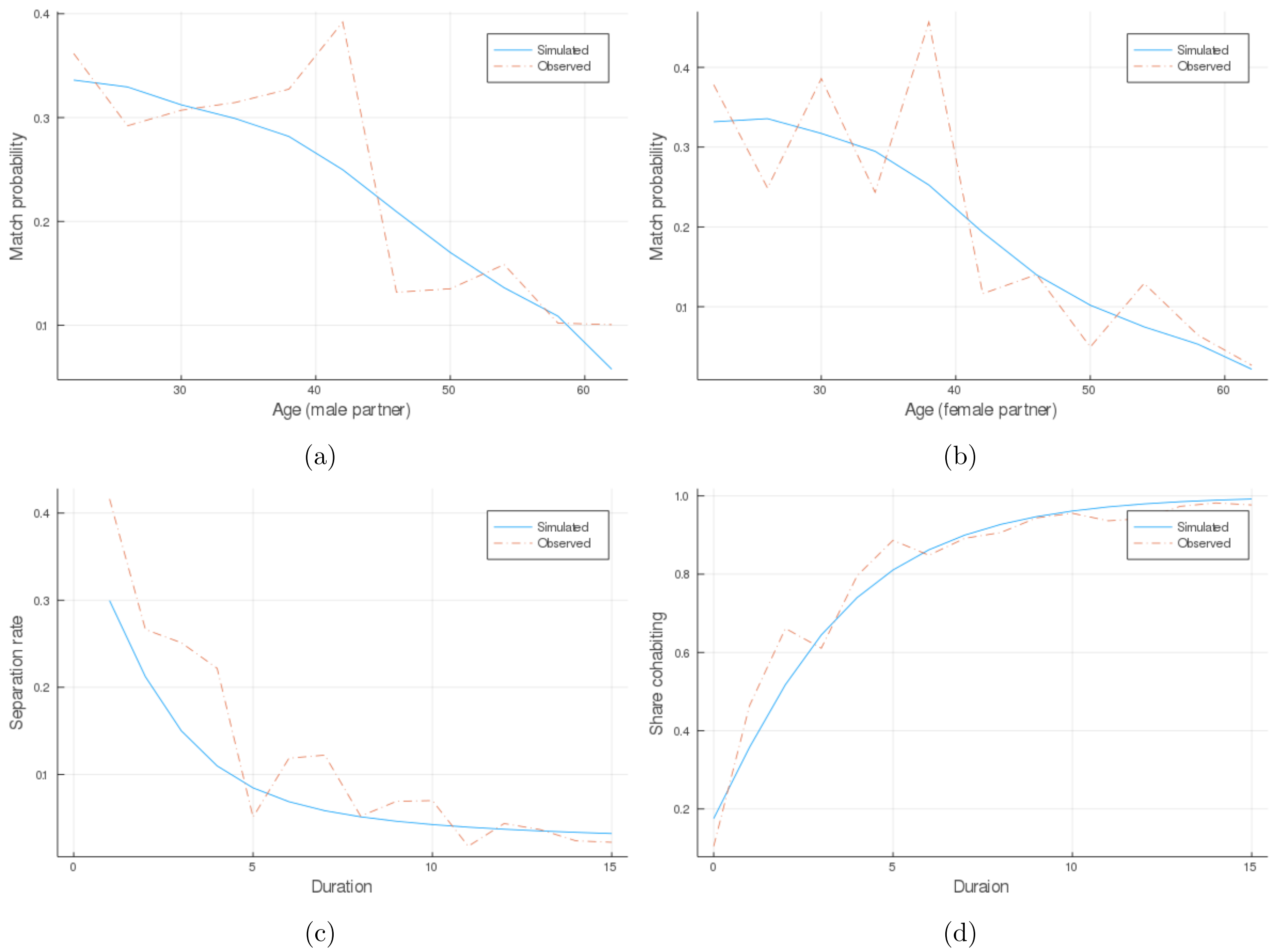
(c) Changes in men's bargaining power (θ)



(d) Changes in women's bargaining power ($1 - \theta$)

Notes. The benchmark corresponds to the SSE of our augmented model, whose parameter estimates are in Table 5. Experiment I is obtained by calculating the SSE after equalizing meeting costs for same-race and different-race couples. In panels (a) and (b), we report changes in the share of partnered individuals in the working-age population (19 to 62). In panels (c) and (d), we report changes in the bargaining power of partnered individuals in the working-age population (19 to 62), by averaging $\theta_{i,j}$ (for men, $1 - \theta_{i,j}$ for women) among couples conditional on the race of one partner. See Proposition 2 for more details on the determination of the Nash-bargaining weights in marriage markets.

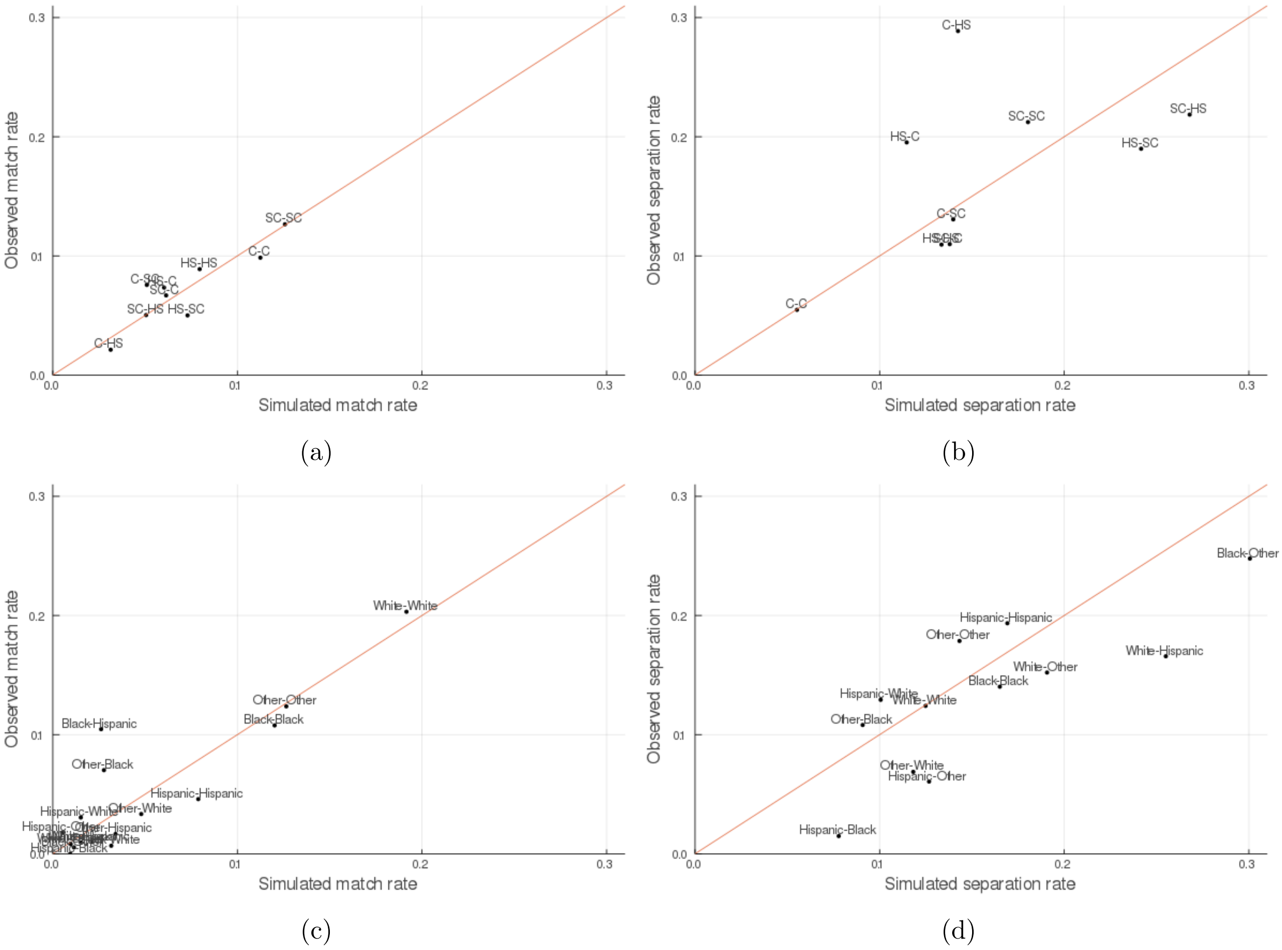
Figure 4: Age and duration profiles - model vs data



(a) (b) (c) (d)

Notes. In panels (a) and (b), the match probability corresponds to the odds of finding a partner over the next year. In panel (c), the separation rate corresponds to the odds that a couple breaks up over the next year. In panel (d), we plot the fraction of cohabiting couples conditional on relationship duration. In all panels, the solid lines represent the simulated moments, while the dashed lines represents the empirical moments estimated straight from the data.

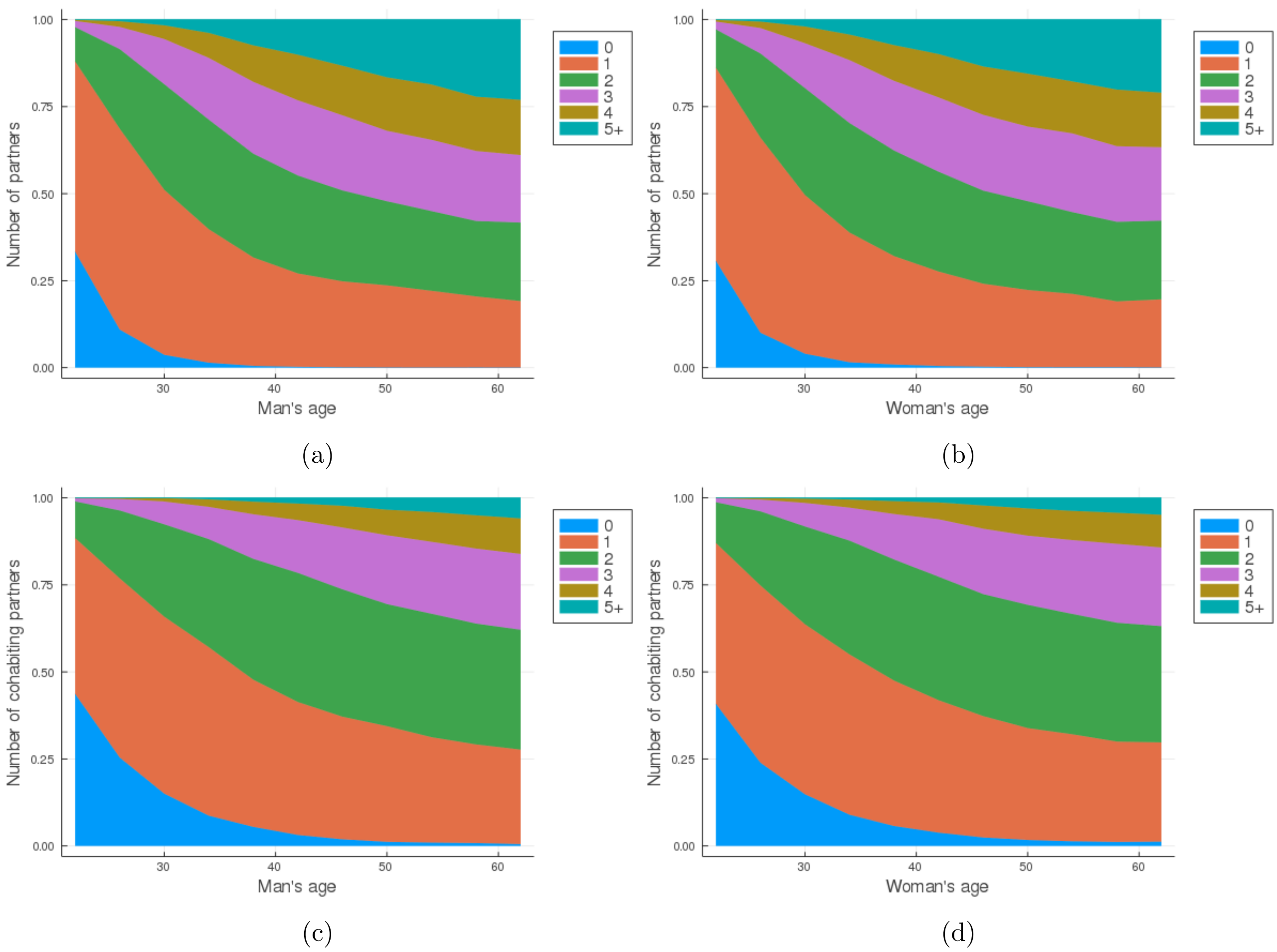
Figure 5: Racial and educational sorting - model vs data



(a) (b) (c) (d)

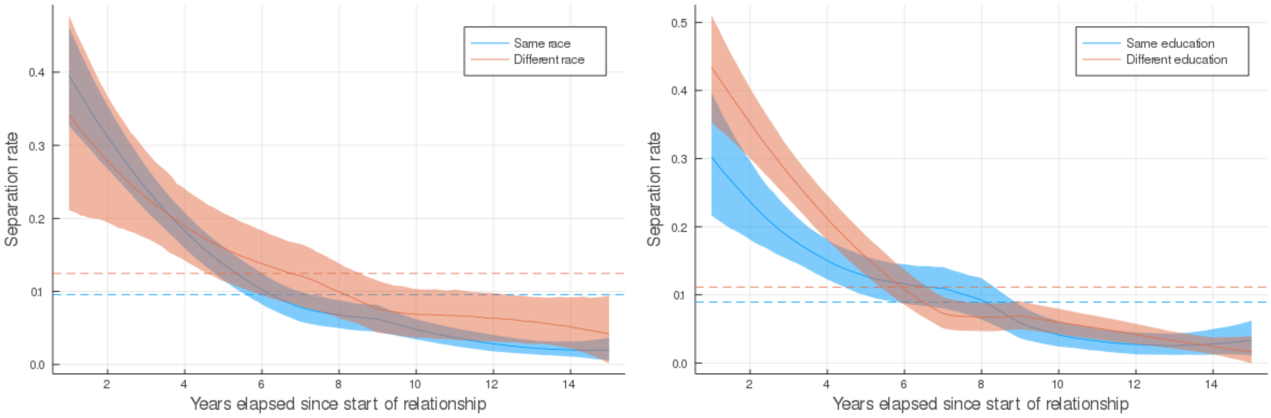
Notes. In panel (a), we plot the simulated match rate conditional on partners' education on the x-axis, and the corresponding empirical moment on the y-axis. The match rate is given by the number of yearly matches between a man of a given type and a woman of given type, divided by the square root of the product of the number of male and female singles of the corresponding types. Points that lie on the 45-degree line indicate a perfect fit. In panel (b), we plot the empirical vs the simulated separation rates conditional on both partners' education. Analogously, in panel (c) and (d), we plot the match and separation rates conditional on the partners' race.

Figure 6: Number of partners by age and gender



Notes. The different panels show the simulated distribution of the total number of partners by age, gender, and type of relationship. Panels (a) and (b) respectively show men's and women's total number of partners by age. Panels (c) and (d) respectively show men's and women's number of cohabiting partners by age.

Figure 7: Separation rate by duration



(a) (b)

Notes. We replicate panels (b) and (c) in Figure 1 with a more refined categorization of the race and education variables. In panel (b), we use four categories for race (white, black, Hispanic, other), while in panel (c) we use five categories for education (high school dropout, high school diploma, some college, college degree, graduate degree).

Table 1: Descriptive statistics - estimation sample

	All		Singles		Partnered	
	Women	Men	Women	Men	Women	Men
A. Sample composition						
# of respondents	1,465	1,408	301	244	1,164	
# of observations	3,833	3,776	301	244	3,532	
In a relationship (%)	65.2	71.4				
B. Race (%)						
White	68.7	70.7	62.5	73.2	72.1	69.7
Black	13.2	10.4	19.9	9.9	9.7	10.6
Hispanic	11.3	13.4	12.0	12.4	11.0	13.9
Other	6.7	5.5	5.7	4.5	7.3	5.8
C. Education (%)						
High-school dropout	8.3	6.6	9.4	3.4	7.8	7.9
High-school diploma	28.3	28.0	34.7	28.7	24.8	27.7
Some college	31.0	35.1	29.1	40.9	32.0	32.8
Bachelor's degree	22.8	20.9	18.4	19.2	25.1	21.7
Graduate degree	9.6	9.3	8.4	7.9	10.3	9.9
D. Age and relationship characteristics						
Age (years)	36.9	37.1	42.5	39.9	33.9	36.0
Cohabiting (%)					76.8	
Duration (years)					6.8	

Notes. In panel B, each column details the racial composition of the sample in 2009 conditional on gender and relationship status, with every column summing to 100. Similarly, panel C details the educational composition of the sample in 2009, again with every column summing to 100. By construction of the sample, age is capped at 62 and relationship duration at 15. The categories "white" and "nlack" only include respondents who do not identify as Hispanic, while the category "other" mainly includes respondents who identify as non-Hispanic Native American or Asian.

Table 2: Mating patterns

A. Education (%) Man \ Woman	High-school dropout	High-school diploma	Some college	College degree	Graduate degree
High-school dropout	2.4	2.7	1.8	0.9	0.1
High-school diploma	2.8	11.8	8.0	3.8	1.3
Some college	2.2	8.2	14.2	6.3	1.9
College degree	0.2	1.5	6.4	9.7	3.9
Graduate degree	0.1	0.6	1.6	4.4	3.2

B. Race (%) Man \ Woman	White	Black	Hispanic	Other
White	62.1	0.6	3.7	3.3
Black	1.0	7.5	1.9	0.2
Hispanic	6.8	1.1	5.1	0.8
Other	2.1	0.5	0.3	3.0

C. Age (%)

				Age gap				
< -10	-10 to -6	-5 to -3	-2 to -1	0	1 to 2	3 to 5	6 to 10	> 11
2.0	4.2	7.8	10.3	13.0	22.7	19.6	14.0	6.5

Notes. The different panels describe the composition of all couples in our estimation sample at the time of the first interview in 2009. In each panel, all relative frequencies sum to 100.

Table 3: Estimates of frictions and gains - baseline model

	Meeting costs $\Lambda_{i,j}$		Match gains $H_{i,j,a,k}$	
	Women	Men	Women	Men
Constant	12.66		0.36	
	(0.98)		(0.13)	
Age	1.05	0.65	-0.09	-0.07
	(0.19)	(0.17)	(0.02)	(0.01)
Non-white	1.00	-0.89	-0.18	0.06
	(0.59)	(0.70)	(0.07)	(0.06)
Some college	-0.67	-0.20	0.06	-0.03
	(0.59)	(0.57)	(0.05)	(0.05)
College degree	-0.34	0.90	0.22	0.10
	(0.73)	(0.75)	(0.07)	(0.07)
Same age group	-2.82		0.25	
	(0.45)		(0.11)	
Same race	-3.75		0.05	
	(0.69)		(0.06)	
Same education	-0.46		0.11	
	(0.53)		(0.04)	
Cohabiting			0.37	
			(0.06)	
Low quality			-0.70	
			(0.18)	
	Additional parameters			
Cost of moving in	-2.56			
	(0.18)			
Prob. of low-quality match	0.42			
	(0.07)			

Notes. Standard errors in parentheses. Meeting costs are given by $\Lambda_{i,j} = x_{i,j}^{\intercal}\lambda$. In the first two columns in the upper panel, we report the estimated λ. Per-period match gains are given by $H_{i,j,a,k} = x_{i,j}^{\intercal}\delta + \zeta_a + \eta_k$. In the upper panel, we report the estimated δ, ζ, and η. In the lower panel, we report the estimated switching cost κ and the initial distribution of time-invariant match quality μ. White is the reference category for racial groups, high-school diploma or less is the reference category for education. The base unit for age is 10 years. Age groups span four years.

Table 4: Correlation between meeting costs and observed meeting circumstances

	OLS estimates	
	Meeting costs $\Lambda_{i,j}$	Expected match surplus $\bar{S}_{i,j}$
Same town	-3.20	-2.40
	(0.40)	(1.50)
Same high school	-2.80	-1.59
	(0.35)	(1.28)
Same college	-1.14	3.94
	(0.26)	(0.98)
Met online	0.99	2.63
	(0.34)	(1.24)
Met through friends	-0.69	2.36
	(0.50)	(1.84)
Met through family	-0.33	-2.12
	(0.38)	(1.40)
Met through neighbors	-0.09	0.20
	(0.28)	(1.05)
Met through coworkers	0.90	-0.04
	(0.40)	(1.47)
Met at work	0.99	-1.38
	(0.40)	(1.50)
Met at church	-0.56	0.07
	(0.21)	(0.80)
Met on vacation	0.27	0.10
	(0.15)	(0.55)
Met in a bar	0.36	-0.32
	(0.29)	(1.07)
Met in an association	0.02	0.21
	(0.17)	(0.64)
Met at a party	0.62	4.18
	(0.32)	(1.18)
Standard deviation	3.22	0.87

Notes. Each line displays the OLS estimates obtained by regressing the variable in the first column on both the meeting costs $\Lambda_{i,j}$ and the expected match surplus $\bar{S}_{i,j}$, which are implied by our structural estimates in Table 3. Standard errors in parentheses. The variables in the first column are all dummies, and all coefficients are multiplied by 100. Both regressors, $\Lambda_{i,j}$ and $\bar{S}_{i,j}$, are scaled by the variance of the bliss shock $Var[z]$, and in the last line we report their sample standard deviation. E.g., the very first line in our table reads: "a one-unit increase in meeting costs is associated with a 3.20% lower probability that the partners come from the same town" and "a one-unit increase in expected match surplus is associated with a 2.40% lower probability that the partners come from the same town". The variables in the first column all describe directly observed meeting circumstances for our sample of couples taken from the HCMST survey. In this survey, respondents answer an open-ended question about *how* and *where* they met their partner; the dummy variables were generated ex post according to the guidelines detailed by Rosenfeld and Falcon (2018). Questions about intermediation (online, friends, family, neighbors, coworkers) are based on text answers to question 24 in the original survey, whereas questions about where couples have met (at work, at church, on vacation, in a bar, in an association, at a party) are based on answers to question 31.

Table 5: Estimates of frictions and gains - augmented model

	Meeting costs $\Lambda_{i,j}$		Match gains $H_{i,j,k}$	
	Women	Men	Women	Men
Constant	14.38		0.35	
	(1.00)		(0.14)	
Age	1.07	0.66	-0.10	-0.08
	(0.16)	(0.18)	(0.03)	(0.02)
Black	1.58	-1.60	-0.06	0.00
	(0.85)	(0.87)	(0.08)	(0.08)
Hispanic	0.16	1.45	-0.24	-0.05
	(0.93)	(1.20)	(0.08)	(0.07)
Other	1.51	-2.27	-0.18	0.15
	(1.24)	(1.28)	(0.11)	(0.16)
High-school dropout	-2.63	1.51	0.22	0.32
	(1.11)	(1.08)	(0.11)	(0.18)
Some college	-2.01	-0.26	0.09	-0.00
	(0.75)	(0.64)	(0.06)	(0.05)
Bachelor's degree	-1.28	0.97	0.22	0.19
	(0.93)	(0.77)	(0.09)	(0.08)
Graduate degree	-0.73	1.95	0.39	0.10
	(1.08)	(1.16)	(0.15)	(0.08)
Same age group	-2.79		0.26	
	(0.42)		(0.10)	
Same race	-3.90		0.02	
	(0.63)		(0.06)	
Same education	-0.43		0.10	
	(0.55)		(0.05)	
Cohabiting			0.37	
			(0.06)	
Low quality			-0.69	
			(0.18)	
	Additional parameters			
Cost of moving in	-2.57			
	(0.20)			
Prob. of low-quality match	0.39			
	(0.07)			

Notes. Standard errors in parentheses. White is the reference category for racial groups, high-school diploma is the reference category for education. The base unit for age is 10 years. Age groups span four years. See Table 3 for more details.

Table 6: Homogamy patterns among new couples - data vs counterfactuals

	Data	Counterfactuals		
		Random assignment	No matching gains	No meeting costs
A. Baseline model				
Share of same-race couples (%)	88.2	53.3	86.9	55.5
Share of same-education couples (%)	49.2	33.5	42.0	40.8
B. Augmented model				
Share of same-race couples (%)	77.5	39.1	73.1	39.5
Share of same-education couples (%)	36.8	24.0	32.0	28.3

Notes. We present the share of homogamous matches (in terms of race and education) in the data and in three counterfactuals. In each counterfactual, we keep the pool of singles identical to what is observed in the data and we calculate the optimal assignment under three scenarios: first, if couples were randomly formed; second, if individuals only took meeting costs $\Lambda_{i,j}$ into account when looking for a mate; third, if individuals disregarded meeting costs and only took the expected match gains $\bar{S}_{i,j}$ into account when looking for a mate.

Table 7: Counterfactual experiments

	Benchmark	No segmentation		No match complementarities	
		Experiment I Race	Experiment II Education	Experiment III Race	Experiment IV Education
Share in a relationship (%, men)	63.1	56.5	61.3	61.9	60.4
Share in a relationship (%, women)	60.2	54.0	58.5	59.0	57.4
Match rate (%, men)	20.7	27.7	21.8	20.3	19.8
Match rate (%, women)	17.8	24.2	18.8	17.5	17.0
Separation rate (%)	11.9	22.3	13.7	12.3	12.9
Share same-age couples (%)	23.4	25.5	23.8	23.5	23.6
Share same-race couples (%)	82.8	59.8	82.6	81.4	82.7
Share same-education couples (%)	39.9	39.2	36.0	39.9	27.0
Share cohabiting (%)	73.0	60.4	70.2	72.4	71.6
Share high k (%)	90.5	85.7	89.5	90.2	90.0
Average z ($\times$100)	14.9	32.4	17.6	15.5	16.2

Notes. The benchmark corresponds to the SSE of our augmented model with four racial groups and five educational categories, whose parameter estimates are in Table 5. In Experiment I (II), we calculate the SSE obtained after equalizing meeting costs respectively for racially (educationally) homogamous and mixed couples. In Experiment III (IV), we calculate the SSE obtained after equalizing the match gains respectively for racially (educationally) homogamous and mixed couples. The moments reported in the table describe the working-age population (19 to 62).

Table 8: Match and separation rates

	Match rate		Separation rate	
	Women	Men	Women	Men
A. Education (%)				
High-school dropout	25.5	36.5	21.8	9.1
High-school diploma	9.1	22.7	12.1	15.0
Some college	23.9	24.1	17.4	17.3
Bachelor's degree	21.4	24.2	8.8	7.6
Graduate degree	24.5	19.7	6.4	8.6
B. Race (%)				
White	21.2	22.4	11.7	12.0
Black	13.2	30.3	13.0	18.1
Hispanic	16.4	15.0	20.0	13.0
Other	19.4	45.6	13.9	12.9
C. Age (%)				
19-25	32.4	31.7	24.1	24.0
26-34	37.2	35.1	8.5	11.4
35-42	23.6	33.3	12.1	7.1
43-50	8.5	14.4	9.1	15.1
51-62	6.7	9.8	13.4	12.6

Notes. Match rates are calculated as the ratio between the number of couples formed in a year and the number of singles at the beginning of the year. Separation rates correspond to the fraction of couples that break up in a given year. Both rates are multiplied by 100.

Table 9: Projection of expected match surplus on agents' types

	Expected match surplus $\bar{S}_{i,j}$		Net expected match surplus $\bar{S}_{i,j} - \Lambda_{i,j}$	
	Women	Men	Women	Men
Intercept	0.30		-12.36	
Age	-0.06	0.08	-2.57	-1.48
Non-white	-0.46	-0.01	-1.45	0.88
Some college	-0.01	-0.12	0.66	0.08
College degree	0.80	0.68	1.14	-0.22
Same age group	0.55		3.38	
Same race	0.21		3.96	
Same education	0.65		1.11	
R^2	0.81		0.99	

Notes. In the first two columns, we report OLS estimates obtained by regressing the expected match surplus $\bar{S}_{i,j}$ implied by the estimates of Table 3, on a vector of polynomial terms conditional on the agents' types (i, j). We repeat the exercise for the expected match surplus net of meeting costs $\Lambda_{i,j}$. In the last line, we report the R^2 indicating how well the chosen polynomial fits the expected match surplus structure. White is the reference category for racial groups, high-school diploma or less is the reference category for education. The base unit for age is 10 years. Age groups span four years.

Table 10: Match characteristics conditional on meeting circumstances

	Same race (%)	Same education (%)	Man's age (years)	Woman's age (years)
Overall	82.5	55.2	29.1	27.1
	-	-	-	-
Same town	87.2	61.2	25.6	23.8
	(1.6)	(3.2)	(0.0)	(0.0)
Same high school	86.5	62.0	23.6	22.2
	(9.5)	(4.1)	(0.0)	(0.0)
Same college	84.5	83.3	22.4	20.9
	(44.8)	(0.0)	(0.0)	(0.0)
Met online	78.8	52.9	33.2	30.8
	(34.7)	(49.8)	(0.0)	(0.0)
Met through friends	80.5	59.2	27.8	26.0
	(28.5)	(6.6)	(0.2)	(0.3)
Met through family	83.6	47.5	28.9	27.5
	(54.8)	(1.6)	(82.6)	(65.0)
Met through neighbors	84.1	64.1	27.1	24.9
	(54.2)	(6.3)	(3.6)	(2.1)
Met through coworkers	78.2	45.3	30.5	28.4
	(14.6)	(0.1)	(2.0)	(3.5)
Met at work	80.6	48.5	31.6	28.3
	(44.6)	(2.9)	(0.0)	(4.9)
Met at church	87.8	62.1	24.0	22.0
	(19.6)	(25.7)	(0.0)	(0.0)
Met on vacation	90.4	46.7	27.5	27.4
	(23.2)	(43.0)	(45.5)	(89.3)
Met in a bar	88.2	53.2	32.6	31.4
	(9.0)	(66.7)	(0.0)	(0.0)
Met in an association	81.6	61.9	26.3	24.6
	(88.8)	(42.6)	(5.7)	(5.9)
Met at a party	76.6	63.1	29.2	27.3
	(8.1)	(5.1)	(92.3)	(78.0)

Notes. Every line provides the fraction of racially and educationally homogamous couples, along with the male and female partner's age, conditional on specific meeting circumstances. In parentheses, we report p-values (multiplied by 100) corresponding to the probabilities that a given estimate is equal to the corresponding estimate for the entire estimation sample, reported at the first line. More details about the variables listed in the first column are available below Table 4.

Table 11: Correlation between match surplus and observed match characteristics

	OLS estimates
	Match surplus $S_{i,j,a,k}(a)$
Legally married	10.83
	(0.68)
Relationship approved by parents	5.02
	(0.61)
Male partner earns more	1.66
	(0.74)
Female partner earns more	-0.31
	(0.65)
Relationship satisfaction	0.06
	(0.01)
Log household income	0.10
	(0.01)
Unemployed (respondent)	-1.04
	(0.33)
Standard deviation	1.99

Notes. Each line displays the weighted OLS estimates obtained by regressing the variable in the first column on the match surplus $S_{i,j,a,k}(a)$ implied by the structural estimates in Table 3, net of switching costs and of the random match quality component z. Standard errors in parentheses. Weights are constructed from the same structural estimates to determine the probability that a couple with observed traits (i, j, a, d) is of type k after d years together. The regressor $S_{i,j,a,k}(a)$ is scaled by the variance of the bliss shock $Var[z]$, and in the last line we report its sample standard deviation. All variables in the first column are dummies, and all coefficients are multiplied by 100, with the exception of relationship satisfaction and log-household income. Relationship satisfaction is measured on a five-level Likert scale ranging from "poor" to "excellent".

Table 12: Estimates of frictions and gains - heterogeneous returns to cohabitation

	Meeting costs $\Lambda_{i,j}$		Match gains $H_{i,j,k}$	
	Women	Men	Women	Men
Constant	12.84		0.62	
	(0.88)		(0.12)	
Age	1.04	0.65	-0.09	-0.06
	(0.16)	(0.17)	(0.02)	(0.02)
Non-White	0.93	-0.89	-0.19	0.06
	(0.74)	(0.71)	(0.09)	(0.08)
Some college	-0.64	-0.20	0.06	-0.04
	(0.50)	(0.56)	(0.05)	(0.06)
College degree	-0.04	1.21	0.26	0.14
	(0.68)	(0.89)	(0.10)	(0.08)
Same age group	-2.77		0.28	
	(0.40)		(0.10)	
Same race	-3.84		-0.04	
	(0.63)		(0.08)	
Same education	-0.64		-0.05	
	(0.52)		(0.07)	
Cohabiting			0.05	
			(0.08)	
Cohabiting, same race			0.09	
			(0.08)	
Cohabiting, same educ.			0.21	
			(0.08)	
Low-quality			-0.74	
			(0.15)	
	Additional parameters			
Cost of moving in	-2.86			
	(0.81)			
Cost of moving in, same race	0.31			
	(0.84)			
Cost of moving in, same educ.	-0.53			
	(0.54)			
Prob. of low-quality match	0.46			
	(0.05)			

Notes. The model is similar to our baseline, but both returns to cohabitation and the cost of moving in together are allowed to be heterogeneous across couples. Standard errors in parentheses. White is the reference category for racial groups, high-school diploma is the reference category for education. The base unit for age is 10 years. Age groups span four years. See Table 3 for more details.

Table 13: Robustness check - heteroskedasticity

		Heteroskedastic specifications	
	Baseline	Education	Race
A. Meeting costs $\Lambda_{i,j}$			
Same race	-3.75	-3.79	-4.80
	(0.57)	(0.58)	(1.28)
Same education	-0.46	-0.04	-0.16
	(0.46)	(0.49)	(0.46)
B. Match gains $H_{i,j,a,k}$			
Same education	0.11	0.11	0.15
	(0.04)	(0.04)	(0.06)
Same race	0.05	0.04	-0.01
	(0.06)	(0.06)	(0.09)
C. Log-variance of z			
Same education	0.00	0.40	0.00
	-	(0.29)	-
Same race	0.00	0.00	-0.41
	-	-	(0.29)

Notes. In the second column, we report estimates from Table 3 for comparison with two alternative specifications. In the third column, we report the findings obtained by estimating a model identical to our baseline, but with $Var[z|i,j]$ allowed to be different for same-education and different-education couples; for the latter, $Var[z|i,j]$ is normalized to one. Similarly, in the fourth column, we report the findings obtained by estimating a model identical to our baseline, but with four instead of two race categories, and with $Var[z|i,j]$ allowed to be different for same-race and different-race couples; for the latter, $Var[z|i,j]$ is normalized to one. Standard errors in parentheses.

Table 14: Robustness check - heterogeneous odds of high match quality

	Baseline	Heterogeneous odds of high match quality	
		Education	Race
A. Meeting costs $\Lambda_{i,j}$			
Same education	-0.46	-0.47	-0.32
	(0.46)	(0.43)	(0.46)
Same race	-3.75	-3.75	-3.98
	(0.57)	(0.60)	(0.65)
B. Match gains $H_{i,j,a,k}$			
Same education	0.11	0.01	0.13
	(0.04)	(0.04)	(0.05)
Same race	0.05	0.04	0.15
	(0.06)	(0.06)	(0.10)
C. Probability of low-quality match $\mu_{i,j,1}$			
Same education	0.42	0.32	-
	(0.07)	(0.12)	-
Different education	0.42	0.56	-
	(0.07)	(0.09)	-
Same race	0.42	-	0.48
	(0.07)	-	(0.08)
Different race	0.58	-	0.14
	(0.07)	-	(0.24)

Notes. In the second column, we report estimates from Table 3 for comparison with two alternative specifications. In the third column, we report the findings obtained by estimating a model identical to our baseline, but with the probability of drawing a low-quality match, $\mu_{i,j,1}$, allowed to be different for same-education and different-education couples. Similarly, in the fourth column, we report the findings obtained by estimating a model identical to our baseline, but with four instead of two race categories, and with $\mu_{i,j,1}$ allowed to be different for same-race and different-race couples. Standard errors in parentheses.