1213
2024

SOEPpapers
on Multidisciplinary Panel Data Research

Matching on Gender and Sexual Orientation

Edoardo Ciscato and Marion Goussé

DIW SOEP

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany
Contact: soeppapers@diw.de

Matching on Gender and Sexual Orientation*

Edoardo Ciscato† Marion Gousséǂ

November 19, 2024

Abstract

We study the mating patterns of non-heterosexual individuals, who represent a significant and increasing portion of the population, particularly among the youth. We estimate a multidimensional matching model of the marriage market where partner's gender is endogenously chosen conditional on the agent's sexual orientation, and is subject to trade-offs that depend on both the agents' preferences and the pool of potential partners. We show that same-sex couples experience lower gains from live-in relationships, a "same-sex penalty". Absent this penalty, the share of same-sex couples in the U.S. would increase by about 50% (from 1.36% to 2.05% of all couples).

Keywords: matching, marriage market, homogamy, same-sex households, sexual orientation, gender

JEL Classification: D1, C51, J12, J15.

*We are particularly thankful to Alfred Galichon, who inspired us and helped us through the early stages of the project. We also thank Jad Beyhum, Laurens Cherchye, Joanne Haddad, Etienne Masson-Makdissi, Sonia Oreffice, and all participants at the 2023 WOLFE workshop, the CSQIEP Virtual Seminar on Economics of LGBTQ+ Individuals, and the Erasmus Rotterdam University seminar.

†KU Leuven. Naamsestraat 69, 3000 Leuven, Belgium. Email: edoardo.ciscato@kuleuven.be.

ǂCREST-ENSAI. 51 Rue Blaise Pascal, 35172 Bruz, France. Email: marion.gousse@gmail.com.

Girls who want boys
Who like boys to be girls
Who do boys like they're girls
Who do girls like they're boys
Always should be someone you really love

Blur. *"Girls and Boys"*. 1994.

1 Introduction

A significant and increasing proportion of the American population, particularly among younger individuals, identifies as lesbian, gay, or bisexual (LGB). According to a 2022 survey by the Pew Research Center, 7% of all U.S. adults identify as LGB, with a notable 17% prevalence within the 18-29 age group.[1] Yet, only 1.1% of individuals aged 18 to 60 live with a same-sex partner, and same-sex couples only represent 1.9% of all cohabiting couples, both married and unmarried.[2]

In this paper, we study the mating decisions of LGB individuals and the trade-offs they encounter when looking for a partner. Our first contribution is to describe previously undocumented patterns in partnership status *conditional* on LGB identity. It is already well established that LGB individuals are less likely to be in a cohabiting relationship than straight individuals (Carpenter and Gates, 2008; Badgett, Carpenter, and Sansone, 2021). In our paper, we also show that the vast majority of bisexual men and women who do have a cohabiting partner are actually in a relationship with a person of the opposite gender. Using data from the National Health Interview Survey (NHIS), we show that, in the 2013-2018 period, about 90% of all partnered bisexual men and women in the U.S. were in a different-sex relationship. Moreover, always according to NHIS data, about 10% of all partnered gay men and lesbian women in the U.S. were in a different-sex relationship. We observe similar patterns in Germany using data from the German Socio-Economic Panel (SOEP).

These facts suggest that significant barriers to same-sex relationships still exist, despite rapid changes in societal attitudes and the growing legal recognition of same-sex marriage (Fernández, Parsa, and Viarengo, 2019; Bau and Fernández, 2023). In other

[1]See the survey conducted by Pew Research Center (2023). Similarly, a 2021 Gallup survey (Gallup, 2022) indicates that among Generation Z individuals (born between 1997 and 2003), 15% identify as bisexual, 2% as lesbian, and 2.5% as gay.

[2]These figures were obtained with data from the 2019-2022 National Health Interview Survey (NHIS).

words, LGB individuals may abandon the idea of cohabiting with someone of the same gender despite their romantic and sexual attraction, eventually settling for different-sex partners or forgoing cohabiting relationships altogether.

Our second contribution is to build and estimate an equilibrium model of multidimensional matching with Transferable Utility (TU), based on the work of Choo and Siow (2006) and Galichon and Salanié (2022), in order to study the trade-offs LGB individuals encounter when choosing a partner. We assume all individuals are pooled into one market, and we treat gender and sexual orientation as observable characteristics among others.[3] The novel feature of our model is that individuals self-select into different-sex or same-sex relationships, and their partner's gender is subject to trade-offs with other traits (e.g., age, education, race) as matching is multidimensional.[4] These trade-offs are particularly relevant for people who are attracted to both genders (i.e., bisexual individuals) but also for gay men and lesbian women who want to avoid the extra hurdles same-sex couples have to overcome. On the other hand, these trade-offs also depend on the availability of desirable partners, i.e., they depend on the marginals of the matching problem. Importantly, the share of LGB individuals is higher in certain segments of the population (e.g., among college graduates), whereas other segments are characterized by gender-ratio imbalances (e.g., among Blacks).

We estimate the model with both NHIS and SOEP data, and we leverage the observed variation in partner's gender conditional on LGB identity in order to show that, ceteris paribus, the gains from cohabiting relationships are lower for same-sex couples compared to different-sex couples. We refer to this difference in gains as the "same-sex penalty". This penalty is relatively large for male same-sex couples, both in Germany and the U.S. In contrast, it is generally weaker for female same-sex couples both in Germany and the U.S. In the U.S., the same-sex penalty has decreased in magnitude for male same-sex couples, and the timing of this decrease is consistent with the Obergefell v. Hodges ruling. However, the penalty has remained relatively stable for female couples. We also show that

[3]In the data we use for our analysis, gender is only available as a self-reported variable. Therefore, we cannot distinguish between "sex", referring to "biological sex" or "sex assigned at birth", and "gender", which pertains to a facet of one's identity. As a result of these data limitations, we restrict gender to be binary. However, it is worth noting that the framework we develop could accommodate non-binary gender identities.

[4]The existence of trade-offs among several characteristics such as race and education has also been developed to explain the composition of interracial marriages in the United States (Fryer Jr, 2007; Chiappori, Oreffice, and Quintana-Domeque, 2016).

the same-sex penalty is lower in the North-East of the U.S. and higher in the South. It is also lower in large metropolitan areas than in smaller ones. We argue that this spatial and temporal variation is broadly consistent with differences in the degree of acceptance of homosexuality across U.S. regions and over time, as documented in the World Values Survey, and with the staggered legalization of same-sex marriage.

In order to better understand differences in match surplus across same-sex and different-sex couples, we show that individuals predicted to experience higher match gains according to our model have markedly different family and individual outcomes in the data. In particular, they are more likely to have children, to be legally married, and to be generally better off according to a series of measures of individual well-being. These correlation patterns hold even after controlling for the partners' socioeconomic characteristics (i.e., their education, age, and race), and are thus partly driven by differences in match surplus coming from heterogeneity in couples' gender composition and sexual orientation. This provides suggestive evidence that the same-sex penalty might originate from limited access to legal marriage and fertility for same-sex couples.

Our findings extend beyond the same-sex penalty and also provide insights on the role of sexual orientation in mating. Not surprisingly, we find that agents are much better off in relationships with partners of the gender they are attracted to. However, while bisexual individuals benefit from a larger pool of potential partners, they experience weaker gains from live-in relationships, even after controlling for characteristics such as age and education. With SOEP data, we are also able to document the existence of complementarities between partners' sexual attraction. In other words, the gains from live-in relationship are higher when partners are mutually attracted to each other. Finally, we discuss differences in educational, racial, and age complementarities between same-sex and different-sex couples; our findings confirm those of Ciscato, Galichon, and Goussé (2020) and show that same-sex couples have a weaker taste for racial homogamy and are less concerned by age differences between partners, whereas educational complementarities are comparable across same-sex and different-sex couples.

After estimating our model, we can simulate counterfactual equilibria in order to provide both a more intuitive quantitative assessment of the same-sex penalty and an estimation of the gender-ratio elasticities of the share of same-sex couples. In other words, we first show that, absent the same-sex penalty, the share of same-sex couples would increase by more than 50%, from 1.36% to 2.05% of all couples. This increase is explained by a larger proportion of gay men and lesbian women choosing to move in with a same-sex partner, but also by a substantial increase in the proportion of bisexual individuals opting for a partner of the same gender rather than one of the opposite. In a different counterfactual scenario, we increase the men-to-women gender ratio by 10%. This leads to a modest, but positive increase in the odds that a man is in a same-sex relationship (+1.80%), whereas the odds that a woman is in a same-sex relationship decrease (-1.93%). In general, an increase in the gender ratio is unfavorable for men, who are less likely to find a partner in the counterfactual equilibrium. In particular, as competition stiffens, bisexual men are more likely than straight men to stop looking for female partners, since their gains from different-sex live-in relationships are lower. On the other hand, they become more likely to match with male partners, whose supply has become larger.

Finally, we predict how mating patterns would change if the LGB population kept growing in numbers. As already mentioned at the very beginning of the paper, recent surveys suggest that the share of LGB individuals in younger cohorts is substantially larger than in the past (Pew Research Center, 2023; Gallup, 2022). Hence, we predict how mating patterns would change in the next thirty years if all new cohorts had the same LGB composition as younger individuals in the 2019-2022 NHIS data. In this counterfactual scenario, the share of gay men and lesbian women in the adult population (aged between 21 and 50) would increase by about 50%, whereas the share of bisexual individuals would triple. We find that, in this scenario, the share of same-sex couples would increase by about 73%, from 1.36% to 2.35% of all couples. The odds of finding a partner would increase by 2.85% for lesbian women and by 4.10% for gay men, while cases where gay individuals match with opposite-sex partners would become less frequent.

Our paper is methodologically related to a large literature studying assortative mating among couples. Since Choo and Siow (2006), several papers estimate equilibrium models of the marriage market in order to measure assortative mating, but they focus on different-sex couples (e.g., Chiappori, Oreffice, and Quintana-Domeque, 2012; Dupuy and Galichon, 2014; Chiappori, Salanié, and Weiss, 2017). The paper of Ciscato, Galichon, and Goussé (2020) is one exception and shows that mating patterns differ across same-sex and different-sex couples. However, they assume agents can exclusively look for partners of a single gender, which gives rise to three perfectly segmented markets (different-sex, male same-sex, and female same-sex couples). Moreover, in taking the model to the data, they assume a person's sexual orientation can be uniquely inferred from their partner's gender. Our paper addresses the concern that agents self-select into different-sex and same-sex relationships based on both their preferences and the composition of their pool of potential partners. Masson-Makdissi (2023) also builds a unified model where agents self-select into different types of relationships, although its identification strategy requires additional assumptions due to the lack of matched data on sexual orientation.

Our paper is also related to a strand of literature on the determinants and composition of same-sex households. Many of these works look at the impact of institutional changes. Badgett, Carpenter, Lee, and Sansone (2024) provide an extensive review of papers that study the impact of same-sex marriage legalization on mating decisions; e.g, Chen and van Ours (2020) for the Netherlands, Carpenter (2020) for Massachusetts, and Carpenter, Eppink, Gonzales, and McKay (2021) and Delhommer and Hamermesh (2021) for the entire U.S. Other papers look at the cultural and demographic determinants of same-sex marriage (Aksoy, Carpenter, De Haas, and Tran, 2020; Bau and Fernández, 2023). Baranov, De Haas, and Grosjean (2018), Baranov, De Haas, and Grosjean (2020), and Brodeur and Haddad (2021) point at a link between skewed sex-ratios, the historical commonness of same-sex relationships, and present-day attitudes towards same-sex relationships. Fernández, Parsa, and Viarengo (2019) discuss the importance of the AIDS epidemic in shaping the rapid change in attitudes towards homosexuality. Others discuss differences in household behavior between same-sex and different-sex couples (Oreffice, 2016; Hansen, Martell, and Roncolato, 2020).

Our work is closely connected to this strand of literature, in that we quantitatively assess the overall differences in gains from cohabiting relationships between same-sex and different-sex couples. We show that both the presence of a same-sex penalty in marriage markets and its heterogeneity between genders, regions, and time periods are consistent with previously documented differences between different-sex, male and female same-sex couples, in both family and individual outcomes. While we cannot leverage the staggered legalization of same-sex marriage in the U.S. in our analysis,[5] we discuss how the same-sex penalty has changed before and after the 2015 Obergefell v. Hodges ruling. Finally, we also estimate the elasticity of the share of same-sex households relatively to changes in the

[5]NHIS data do not cover a sufficiently long time span and only contain geographic information on the location of households at the macro-region level.

gender ratio. This provides a quantitative benchmark for studies that leverage temporal and geographic variation in the gender ratio to explain differences in the prevalence of same-sex households.

In the next Section, we present the data we use to first document the discrepancy between sexual orientation and the partner gender, then to estimate our model. In Section 3, we present our model, then in Section 4 and Section 5, we successively present our results and counterfactual simulations. Section 6 concludes.

2 Data

2.1 Databases

Our first objective is to document mating patterns conditional on one's own sexual orientation. This is not an easy task because we need data that provide information about sexual orientation of both partnered and single individuals. Moreover, when a partner is present, we also need information about her/his characteristics, and possibly also about her/his sexual orientation. Finally, we need the sample to contain a sufficient number of LGB individuals. This means we need either a large representative sample or an oversampling of the LGB population. In this Section, we describe the data sources used in the paper, discuss different measures of sexual orientation, and provide descriptive statistics.

In this paper, we use three databases that partly or fully meet our criteria. First, we use data from the National Health Interview Survey (NHIS)[6] The NHIS is a large cross-sectional survey that is representative of the U.S. population, and it has already been used by Badgett, Carpenter, and Sansone (2021) to provide a detailed description of the LGB population. While it samples individuals rather than households, key demographic and socioeconomic variables are available for all members of the respondent's household. However, the survey only collects information on the respondent's sexual orientation, and not the partner's. We will use two different NHIS samples separately: the 2013-2018 and the 2019-2022 samples.[7]

[6] Blewett, Rivera Drew, King, Williams, Del Ponte, and Pat (2022)

[7] The 2013-2018 waves have a homogeneous structure and can be merged. On the other hand, the content and structure of the NHIS survey were updated in 2019, and merging waves before and after 2019 is advised against. Among other changes, since 2019, the race of the partner is not available anymore, although we do know if both partners are of the same race. In addition, we observe the type of metropolitan area the respondent lives in, a variable that was not available before 2019.

Second, we use 2011-2019 data from the National Survey of Family Growth (NSFG).[8] The NSFG is a nationally representative survey of the U.S. population aged between 15 and 49. It does not only ask respondents about their sexual orientation, but it also contains a great amount of additional information on the respondent's sexuality and past sexual experiences. Hence, we can use it to better understand how different measures of sexual orientation compare to each other. However, since the NSFG contains no information on live-in same-sex partners, we cannot use it for the analysis of mating patterns.

Finally, we use the German Socio-Economic Panel (SOEP)[9], a longitudinal survey of approximately 15,000 German households from 1984 to 2021. Since the survey targets households, it contains the same information on all household members, including data on labor supply and earnings. In 2016 and 2019, information on sexual orientation was collected for all sampled adults. Moreover, the 2019 wave contains a boost sample of hard-to-survey population groups, including lesbian women, gay men, bisexual, transgender, queer individuals, and all those who identify as non-binary (LGBTQ+).[10] In our analysis, we only use the 2016 and 2019 waves.

2.2 Measuring sexual orientation

In the NHIS, sampled respondents were asked "Which of the following best represents how you think of yourself?", to which they could answer "Lesbian or gay", "Straight, that is, not gay", "Bisexual", "Something else", or "I don't know". Some would refuse to answer altogether. Similar questions were also asked in the NSFG and in the SOEP, which makes it possible to compare aggregate statistics across different databases. The design of this question allows respondents to express their own sexual orientation using commonly used vocabulary reflecting a person's LGB identity. However, this vocabulary is likely to evolve over time and across different generations (Morgan, Dragon, Daus, Holzberg, Kaplan, Menne, and Spiegelman, 2020). For instance, as bisexual identity has gained legitimization over the last thirty years, people in younger generations are more likely to self-identify as bisexual (MacDowall, 2009).

Throughout the paper, we use answers to this question as our main measure of sexual

[8]CDC National Center for Health Statistics (NCHS)

[9]DIW (2021). See Goebel, Grabka, Liebig, Kroh, Richter, Schröder, and Schupp (2019) for more documentation

[10]835 households were recruited through an approximately 9-month-long telephone screening process. Of these households, 477 participated in the survey between April and November.

orientation. Namely, gay men and lesbian women are assumed to be exclusively attracted to their own gender, straight individuals are assumed to be exclusively attracted to the opposite gender, and bisexual individuals are assumed to be equally attracted to both genders. It is worth noting that, in our baseline sample, we initially exclude respondents who answered “Something else”, “I don’t know”, or who refused to answer.[11] Based on information contained in the NHIS, it is nearly impossible to learn more about the sexual orientation of these respondents.[12] However, NSFG data suggest that, by excluding respondents who answered “Something else”, we under-represent the non-straight population only slightly.[13]

Using the NSFG data, we shed more light on measurement issues by using additional variables related to sexual orientation. Importantly, respondents could not only report their LGB identity as in the other surveys, but they were also asked “Which gender are you the most attracted to?”, to which they could answer “Only attracted to men”, “Mostly to men”, “Equally to both”, “Mostly to women”, and “Only to women”. In the vein of Kinsey, Pomeroy, and Martin (1948) and Kinsey, Pomeroy, Martin, and Gebhard (1953), this question directly inquires about sexual attraction without using labels. It is also consistent with the idea that sexual orientation should be measured on a continuous scale, and not be constrained into pre-established categories. The presence of this additional variable has two main benefits. First, it helps us understand the amount of residual variation in sexual orientation within each LGB category. Second, it helps us better understand the sexual orientation of respondents who do not identify in any of the LGB categories listed as possible answers.

Differences in the way the surveys were administered remain important. While the

[11] Depending on the survey and the year of data collections, these categories represent between 1.3% and 2.9% of individuals in the NHIS, 3.1% of men and 4.9% of women in the NSFG, and 7.3% of men and 7.7% of women in the SOEP. We show these numbers in Table 18 in Appendix.

[12] In the 2013 and 2014 waves of the NHIS, respondents in these two categories were given room to further explain their initial answers. Unfortunately, access to answers to these follow-up questions is restricted, but we do have aggregate statistics for the resulting variables. Dahlhamer, Galinsky, Joestl, and Ward (2014) show that very few respondents can be reclassified into the main LGB categories based on their answers to the follow-up questions. Few utilized other labels (e.g., “Queer”, “Pansexual”) or reported not experiencing any sexual attraction. Conversely, the majority of individuals who initially answered “Something else” prefer not to use labels to define themselves. Instead, the majority of those who answered “I don’t know” either do not understand the vocabulary used or are still in the process of figuring out their own sexuality.

[13] Between 2015 and 2019 the NSFG experimented with the answer list by randomizing the inclusion of the “Something else” category. From this experiment, we learn that 50% of women and the vast majority of men would likely have identified as straight if the category “Something else” had not been present.

NHIS was administered through a face-to-face or phone interview,[14] the most sensitive questions present in the NSFG, including those about sexuality, were administered using Audio Computer-Assisted Self-Interviewing (ACASI). Moreover, while the NHIS only contains one single question on sexual orientation, in the NSFG respondents answer a battery of questions on their sexuality and past sexual experiences before reporting their sexual identity and orientation.[15] As highlighted by Badgett, Carpenter, and Sansone (2021), the percentage of individuals reporting same-sex attraction consistently surpasses those acknowledging same-sex sexual behavior or identifying as LGBTQ. Hence, it is not surprising that the share of LGB individuals, and particularly bisexual individuals, is found to be higher in the NSFG.

As for the SOEP, it is administered through face-to-face interviews. Each respondent of the household is interviewed separately. For the boost Sample Q that supplemented the SOEP core sample by queer households, including gender and sexual minorities such as lesbian, gay, bisexual, transgender, and queer respondents (LGBTQ+), recruitment of these households was performed by a random telephone screening of adults living in Germany.

2.3 Sample construction

In the NHIS database, we restrict our estimation sample to individuals aged between 21 and 50 for whom information on gender, age, sexual orientation, race, and education is available. Individuals are treated as matched if they are observed living with a cohabiting partner or a legal spouse. We only keep in our sample partnered individuals whose partner is also aged between 21 and 50, and for whom we also observe the gender, age, race, and education (but not the sexual orientation, which is never available for the respondent's partner). Non-response for key demographic variables is minimal, whereas we initially exclude individuals who do not provide a clear answer to the sexual orientation question, i.e., who do not identify as either straight, gay, lesbian, or bisexual. Eventually,

[14] In the NHIS face-to-face interviews, flashcards listing the response options were used to administer sensitive questions, including those about sexual orientation.

[15] This is not uncommon in surveys on sexuality. Klein (1978) introduced questions about LGB identity in an extension of the original Kinsey questionnaire, arguing that learning how individuals think of themselves complements blunter questions about sexual attraction. Later studies have shown that the two questions described in the main text mainly measure the same fundamental variation in the respondents' sexual orientation (Weinrich, Snyder, Pillard, Grant, Jacobson, Robinson, and McCutchan, 1993). Since the NSFG contains both types of questions, we compare them in Table 2.

our main estimation sample contains 81,798 households (with 2,925 LGB respondents, and 563 same-sex couples), while the sample size increases to 82,989 households when we impute sexual orientation following the procedure explained in Section 3.4.

In the NSFG database, we restrict our estimation sample to individuals aged between 21 and 50 for whom information on gender, age, LGB identity, sexual attraction towards men/women, race, and education is available. We have no information on same-sex households, so we do not use information on marital status or partner's traits from this dataset. The resulting sample contains 13,144 men and 16,667 women.

In the SOEP database, we pool individuals from waves 2016 and 2019.We restrict our estimation sample to individuals aged between 21 and 50 for whom information on gender, age, sexual orientation, and education is available (race is not available in the SOEP survey). In particular, we only keep couples where both partners answered the sexual orientation question.[16] Our final estimation sample includes 12,069 households.

2.4 Sexual orientation by gender

In Table 1, we present the marginal distribution of sexual orientation across our three databases. The fraction of individuals who identify as gay is extremely close in the NHIS and NSFG (between 2.2% and 2.5% for men and between 1.8% and 2.0% for women) and is stable over time. In the NHIS, in the 2013-2018 period, only 0.6% of men identify as bisexual, whereas for women, this fraction is 1.9%. In the NSFG, in the same period, the share of bisexual individuals is more than three times higher (respectively, 1.9% and 6.5%).[17] In the SOEP, the fraction of gay men and lesbian women (respectively 2.0% and 1.4%) is comparable to the NHIS, whereas the fraction of bisexual individuals (0.9% for men and 2.5% for women) is higher than in the NHIS, but lower than in the NSFG. The NHIS figures for the 2019-2022 period show a strong increase among people identifying as

[16]It means, we select couples where both partners have been interviewed which is demanding. As a result, in our estimation sample, we tend to underestimate the share of individuals in a relationship. In Table 20, we document non-response patterns for couples in the SOEP.

[17]Differences between the two surveys are possibly due to the way the survey is administered, since the perception of one's identity could be influenced by the questions on sexuality, and particularly on same-sex behavior, present in the NSFG (see Section 2.1). Moreover, NFSG respondents are on average three years younger than NHIS respondents, so the difference may partly be explained by a cohort effect. Badgett, Carpenter, and Sansone (2021) also find that the share of bisexual individuals is lower in the NHIS than in the 2014–2018 Behavioral Risk Factor Surveillance System (BRFSS) survey.

bisexual, with a rate of 1.4% for bisexual men and of 4.9% for bisexual women.[18] Table 18 in Appendix shows how these figures would change when including respondents who answered "Something else", "I don't know", or who refused to answer. In particular, it shows that in the 2019-2022 NHIS, 1.6% of men and 2.3% of women answer "Something else" or "I don't know".

The NSFG provides information on sexuality and past sexual experiences, including an alternative measure of sexual orientation, based on a question about sexual attraction. The lower panel of Table 1 shows that there is substantially more heterogeneity in sexual orientation among women. Among men, 92.5% of respondents declare being sexually attracted only to individuals of the opposite sex, whereas 1.6% declare being attracted only to individuals of the same sex. These figures are 79.4% and 1.3% among women, who are much more likely to be attracted to both genders, at least to some extent. These gender-asymmetric patterns are consistent with those observed for the LGB categorization in the upper panel. However, when measuring sexual orientation using LGB categories, a non-negligible amount of variation is lost.

Table 1: Sexual orientation by gender

	NHIS 2013-2018		NHIS 2019-2022		NSFG 2011-2019		SOEP 2016-2019	
	Men	Women	Men	Women	Men	Women	Men	Women
Straight	0.972	0.963	0.962	0.933	0.958	0.915	0.971	0.962
Bisexual	0.006	0.019	0.014	0.049	0.019	0.065	0.009	0.025
Gay or Lesbian	0.022	0.018	0.025	0.019	0.022	0.020	0.020	0.014
N	39,751	46,471	23,123	25,515	14,239	18,032	9,187	10,535

	NSFG	
	Men	Women
Only attracted to opposite sex	0.925	0.794
Mostly attracted to opposite sex	0.042	0.141
Equally attracted to both sexes	0.009	0.041
Mostly attracted to same sex	0.008	0.011
Only attracted to same sex	0.016	0.013
N	14,242	18,023

Notes. In every database, we present aggregate statistics for the population aged between 21 and 50. Weighted results.

In Table 2, we use NSFG data to discuss how our main measure of sexual orientation,

[18]These numbers are close to numbers revealed by the recent surveys from Pew Research Center (2023) and Gallup (2022), cited in our introduction.

based on the LGB categorization, compares with the reported sexual attraction towards men/women. Among individuals who identify as straight, 86.7% of women and 96.4% of men report being exclusively attracted to the opposite gender. Conversely, only 62.1% of all lesbian women and 70.1% of all gay men report being exclusively attracted to the same gender. Among bisexual individuals, there is even more variation in reported sexual attraction. Interestingly, the mode for bisexual women is being equally attracted to both genders, whereas the mode for bisexual men is being mostly attracted to women. Moreover, 2.7% of bisexual women and 9.8% of bisexual men declare being exclusively attracted to the opposite gender. In Appendix, in Table 23, we present additional descriptive results obtained with a series of linear probability models where we regress dummies for being a bisexual man, a gay man, a bisexual woman, or a lesbian woman on several observables. Among the regressors, we include the reported sexual attraction towards men/women. The regressions show that answers to the question "Which gender are you the most attracted to?" are strong, and yet imperfect predictors of whether one identifies as gay or bisexual. In more recent years, men have become more likely to identify as gay and women to identify as bisexual, after controlling for both sexual attraction and demographics. Conditional on sexual attraction and other demographics, older individuals are less likely to identify as bisexual, but more likely to identify as gay, while educated women are less likely to identify as bisexual. Finally, women born in the U.S. are less likely to identify as lesbian, but more likely to identify as bisexual.

Table 2: Gender, LGB identity, and sexual attraction towards men/women (NSFG)

	Women						
	Straight	Bisexual	Lesbian	Something else	Does not know	Refuse	All
Only attracted to opposite sex	86.7	2.7	2.3	35.9	89.5	83.3	79.4
Mostly attracted to opposite sex	12.5	39.9	1.1	34.2	4.0	15.0	14.1
Equally attracted to both sex	0.7	50.4	5.3	18.9	2.6	0.5	4.1
Mostly attracted to same sex	0.0	6.7	29.3	9.1	0.0	0.6	1.1
Only attracted to same sex	0.1	0.3	62.1	1.8	3.9	0.7	1.3
Observations	16,184	1,221	393	147	31	47	18,023

	Men						
	Straight	Bisexual	Gay	Something else	Does not know	Refuse	All
Only attracted to opposite sex	96.4	9.8	3.6	48.1	63.6	98.7	92.5
Mostly attracted to opposite sex	3.3	43.5	0.2	37.1	25.5	0.7	4.2
Equally attracted to both sex	0.2	34.4	1.1	10.4	0.0	0.0	0.9
Mostly attracted to same sex	0.0	12.1	25.1	1.1	0.0	0.6	0.8
Only attracted to same sex	0.1	0.2	70.1	3.4	10.9	0.0	1.6
Observations	13,447	293	372	78	22	30	14,243

Notes. 2011-2019 NSFG data. Sample restricted to 21-50 years-old individuals. Weighted results.

In Table 24 and 25 in the Appendix, we present additional descriptive statistics for our samples. These statistics are in line with Badgett, Carpenter, and Sansone (2021), and show that both in the U.S. and in Germany, bisexual individuals are younger on average, while gay men are more educated than their straight peers. It also shows that Black women are over-represented in the lesbian population, whereas Hispanic women are underrepresented. White individuals are over-represented among bisexual individuals. Finally, Table 25 shows that most couples are formed within the same race. In the 2013-2018 sample, 86.7% among opposite-sex couples, 63.8% among gay couples, and 80.3% among lesbian couples are same-race couples. Interestingly, we also observe that the share of couples where both partners are Black is higher among female same-sex couples (10.7%) than among different-sex couples (6.7%), while the opposite is true among male same-sex couples (4.5%). We observe similar patterns in the 2019-2022 sample.

2.5 Mating outcomes by sexual orientation

In Table 3, we look at mating patterns conditional on sexual orientation in the NHIS. In the first three columns, we notice that LGB individuals, bisexual individuals in particular, are less likely to be in a live-in relationship relatively to straight individuals, with the gap being particularly strong among men. Similar patterns are observed in Germany. Interestingly, Table 3 also shows that only 11% of bisexual women and men are in a same-sex household in the 2013-2018 period. This number increased to 23% for partnered bisexual men in the 2019-2022 period, whereas it did not change among partnered bisexual women. Moreover, 13% of gay men and lesbian women who live with a partner are actually in a different-sex household in the 2013-2018 period. These numbers decreased to 6% for partnered gay men and 10% for partnered lesbian women in the 2019-2022 period. We observe similar numbers in Germany.[19]

Table 3: Sexual attraction/orientation and partner's gender

	In a couple			Share of same-sex couples among couples		
	NHIS		SOEP	NHIS		SOEP
	2013-2018	2019-2022	2016-2019	2013-2018	2019-2022	2016-2019
Straight men	0.62	0.60	0.55	0.00	0.00	0.00
Bisexual men	0.32	0.34	0.30	0.11	0.23	0.25
Gay men	0.40	0.36	0.34	0.87	0.94	0.94
Straight women	0.63	0.63	0.60	0.00	0.00	0.00
Bisexual women	0.44	0.49	0.34	0.11	0.11	0.07
Lesbian women	0.55	0.48	0.58	0.87	0.90	0.84

Notes. In every database, we present aggregate statistics for the population aged between 21 and 50. The NHIS covers the 2013-2018 period and the 2019-2022 period, the SOEP covers wave 2016 and 2019. The first three columns show the rate of individuals living in couple for each gender*sexual orientation type. The last three columns show the rate of same-sex couple among individuals in couple. Weighted results.

Using SOEP data, we document the covariation between the partners' sexual orientations. This is not possible with the other datasets, since the SOEP is the only survey to ask both partners about their sexual orientation. In our SOEP sample, 1.2% of couples are same-sex couples, but the fraction of couples where at least one partner is not straight, is actually 2.6%. Table 4 reports the distribution of the partner's sexual orientation and gender, conditional on the respondent's gender and sexual orientation. The sum of the coefficients in each row is equal to one. The results show that 86.3% of bisexual women are in a relationship with a straight man, and 6.2% with a bisexual man. We also note that 14.5% of lesbian women are in a couple with a straight man, and 5.2% of gay men are in a couple with a straight woman. Finally, it is also worth noting that 1.1% of gay men are in a relationship with a lesbian woman. As the number of observations is sometimes small, we need to interpret these estimates with caution.

[19] In Table 19 in Appendix, we show the number of observations used to compute each average.

Table 4: Matching by gender and sexual orientation (SOEP)

	Straight men	Bisexual men	Gay men	Straight women	Bisexual women	Lesbian women	N
Straight men	0.000	0.000	0.000	0.990	0.008	0.002	5150
Bisexual men	0.082	0.000	0.207	0.600	0.116	0.000	25
Gay men	0.000	0.090	0.847	0.052	0.000	0.011	57
Straight women	0.997	0.003	0.001	0.000	0.000	0.000	5105
Bisexual women	0.863	0.062	0.000	0.000	0.044	0.031	71
Lesbian women	0.145	0.000	0.010	0.000	0.021	0.825	52

Notes. SOEP data, pooled 2016 and 2019 waves. Couples restricted to individuals between 21 and 50 years old. Every row sums to one. Weighted results.

3 Model and empirical methods

3.1 Marriage market equilibrium

While the descriptive statistics in Section 2 are an interesting starting point, we now introduce and estimate a model of the marriage market equilibrium in order to disentangle the role of both mating preferences along different dimensions and marginal distributions of individual characteristics in explaining the above-described mating patterns.

We consider a multidimensional model of matching with Transferable Utility (TU) and additively separable random shocks in the vein of Choo and Siow (2006) and Galichon and Salanié (2022), adapted to the so-called "roommate problem" (Chiappori, Salanié, and Galichon, 2019; Ciscato, Galichon, and Goussé, 2020). There exists a mass of individuals participating in the market. Each individual $i \in \mathcal{I}$ is characterized by his or her type $x_i \in \mathcal{X}$, which consists in a vector of observable characteristics with $\mathcal{X}$ being a finite set of types. The set of marital choices is $\mathcal{X}_0 = \mathcal{X} \cup \{0\}$, where 0 denotes the option of staying single. Let f be the probability mass function describing the distribution of types in the population, so that $f(x)$ denotes the mass of individuals in group x. We normalize the total mass of individuals so that $\sum_{x \in \mathcal{X}} f(x) = 1$.

The matching is described by a measure μ on the product space $(\mathcal{X}_0 \times \mathcal{X}_0) \setminus (0, 0)$, and $\mu(x, x')$ denotes the (unconditional) probability of observing an individual of type x matched to a partner of type x'. However, μ is constrained by the fact that its marginals must coincide with $f(x)$. In other words, the matching is feasible if and only if

$$\mu \in \mathcal{M}(f) := \left\{ \mu \geq 0 : \begin{pmatrix} \mu(x,0) + \sum_{x' \in \mathcal{X}} (\mu(x,x') + \mu(x',x)) = f(x), \ \forall x \in \mathcal{X} \\ \mu(x,x') = \mu(x',x), \ \forall (x,x') \in \mathcal{X} \times \mathcal{X}_0 \end{pmatrix} \right\}. \tag{3.1}$$

The mass of single individuals of type x is denoted by $\mu(x,0)$, and the first set of constraints in (3.1) implies that, for any x, the sum of all matched and unmatched individuals must be equal to the marginal $f(x)$. The symmetry constraint in (3.1) characterizes roommate problems; we will return to its interpretation after introducing the match payoffs.

A match between an individual of type x and an individual of type x' generates a match surplus $\Phi(x,x')$, on top of additional partner-specific random gains. Individuals have to choose between staying single or matching with the best possible partner. The mating preferences of an individual i toward each type of partner are described by a multidimensional random vector $\varepsilon_i = (\varepsilon_{ix'})_{x'\in\mathcal{X}_0} \in \mathbb{R}^{|\mathcal{X}_0|}$, whose elements are i.i.d. standard type-I extreme value distributed, as in Choo and Siow (2006).[20] For an individual i with type x, the utility of choosing a partner of type x' is given by $U(x,x') + \varepsilon_{ix'}$, where $U(x,x')$ is a share of $\Phi(x,x')$, and $\varepsilon_{ix'}$ is the random component of the individual's gains from marriage. The component $U(x,x')$ is endogenously determined in equilibrium, as we will see below, but only depends on the partners' types, and not on their idiosyncratic preferences.[21]

We remark that $\Phi(x,x')$ is symmetric by construction, since $U(x,x') + U(x',x) = \Phi(x,x') = \Phi(x',x)$. This symmetry is a distinctive feature of the roommate problem. Interestingly, it is not incompatible with the existence of asymmetric roles within the couple. This property only implies that, conditional on the partners' type, an individual of type x is rewarded with the same share $U(x,x')$ regardless of the role she/he takes up in a household (x,x'). It also means that, when we look at the equilibrium demographics, couples of type (x,x') should all be treated in the same way, regardless of whether x is listed as the first or second partner in the data. This also explains why a feasible matching should be symmetric, as detailed in the set of constraints (3.1).

A matching μ is stable when there is no pair of individuals who would both prefer being matched together rather than with their current partners. When the matching is stable, all agents maximize utility by choosing the best mate for given surplus shares $U(x,x')$. The latter are equilibrium objects that not only reflect the exogenous quality of a match of type (x,x'), but also result from the endogenous relative bargaining power within a couple of type (x,x'). When demand for a type x increases, the share $U(x,x')$

[20]We relax this constraint in a robustness check using nested logit models.

[21]This is proved by Chiappori, Iyigun, and Weiss (2009).

will increase relatively to the partner's share $U(x', x)$.

Demand and supply functions can be derived from the agents' mate choice problems. The program of individual i with type $x \in \mathcal{X}$ is

$$\max\{\max_{x' \in \mathcal{X}}\{U(x, x') + \varepsilon_{ix'}\}, \varepsilon_{i0}\}. \tag{3.2}$$

where the non-random component of the payoff obtained by singles is normalized to zero.

The distributional assumption on ε implies that an individual of type x will choose an individual of type x' with probability

$$\mu^d(x'|x) = \frac{\exp(U(x, x'))}{1 + \sum_{x''} \exp(U(x, x''))} \tag{3.3}$$

Following Choo and Siow (2006), supply and demand conditions can be combined to obtain the matching function

$$\mu(x, x') = \exp\left(\frac{\Phi(x, x')}{2}\right) \sqrt{\mu(x, 0)\mu(x', 0)}. \tag{3.4}$$

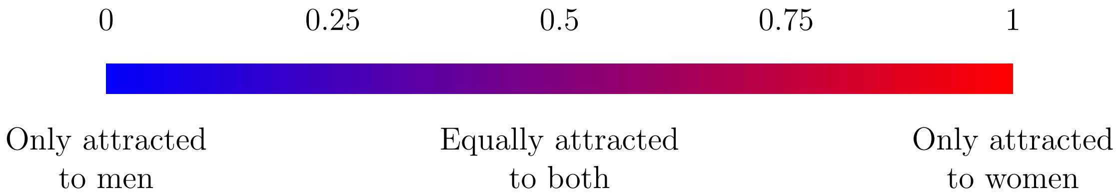
The equilibrium matching must be both stable and feasible. Hence, the set of constraints (3.1) and the matching function constitute a system of $|\mathcal{X}|$ equations in $\{\mu(x, 0)\}_{x=\mathcal{X}}$. This system has a unique solution that characterizes the market equilibrium (Ciscato, Galichon, and Goussé, 2020; Galichon and Salanié, 2022).

3.2 Gains from living together

In our application, the vector of individual characteristics x includes gender, sexual orientation, age, race, and education. These characteristics determine the average match surplus $\Phi(x, x')$ for a pair of individuals with types (x, x'). Notably, we follow the literature and allow for the presence of household complementarities in age, race, and education (Chiappori, Salanié, and Weiss, 2017; Ciscato, Galichon, and Goussé, 2020). While these matching dimensions are present in our model, here we focus on how the match surplus depends on the partners' gender and sexual orientation, a novel aspect of our paper.

As anticipated in footnote 3, gender is restricted to be binary. We let m_i be a dummy variable equal to one if i is a man, and $f_i = 1 - m_i$ a dummy variable equal to one if i is a woman. We assume sexual orientation is measured by a variable o_i, corresponding to the agent's degree of attraction to women. This representation restricts sexual orientation to

Figure 1: One-dimensional measure of sexual orientation



be one-dimensional; the resulting spectrum is shown in Figure 1. In the next Section, we discuss ways of addressing potential limitations of this simplified representation of sexuality.

Whether individuals are attracted to their partner's gender matters for the couple's match surplus. We define a variable of "sexual compatibility" indicating to what extent individual i is attracted to the gender of her/his partner i'. The variable is denoted $s_{ii'} \in [0, 1]$ and is built as follows:

$$s_{ii'} = o_i f_{i'} + (1 - o_i) m_{i'}, \tag{3.5}$$

which means that it is highest (lowest) when a person is matched with a partner of the gender she/he is (not at all) attracted to.

While we expect sexual compatibility to be an important component of the match surplus, same-sex couples may also experience differences in the gains from their relationship relatively to different-sex couples purely due to the couple's gender composition. Social stigma, the lack of legal recognition, and challenges related to bearing children or adopting are all potential reasons that can explain why same-sex couples experience lower gains than different-sex ones. Hence, even individuals who would maximize their sexual compatibility in same-sex relationships will sometimes settle for a partner of the opposite gender to avoid the extra hurdles.

Our baseline specification of the match surplus function is the following:

$$\Phi_A(x_i, x_{i'}) = \lambda_0 + \lambda_1(m_i m_{i'} + f_i f_{i'}) + \lambda_s(s_{ii'} + s_{i'i})$$

where λ_1 measures the difference in gains between same-sex and different-sex couples. As we expect λ_1 to be negative, we refer to it as a "same-sex penalty". The parameter λ_s corresponds to the utility gains from matching with a partner whose gender is deemed attractive. Note that sexual attraction goes both ways. A couple will be better off if both partners are mutually attracted to each other (and thus $s_{ii'} + s_{i'i}$ is equal to two).

When we estimate the model with our baseline sample, we construct the sexual orientation variable using the respondent's LGB identity. We assume that gay men and straight women are only attracted to men ($o_i = 0$), bisexual individuals are equally attracted to both ($o_i = 0.5$), and straight men and lesbian women are only attracted to women ($o_i = 1$). In Table 5, we thus derive the match surplus associated with sorting on gender and LGB identity based on our baseline specification Φ_A.

Table 5: Match surplus for different types of couple with functional form Φ_A

	Partner i'					
Partner i	Straight man	Bisexual man	Gay man	Straight woman	Bisexual woman	Lesbian woman
Straight man	λ_1					
Bisexual man	$\lambda_1 + 0.5\lambda_s$	$\lambda_1 + \lambda_s$				
Gay man	$\lambda_1 + \lambda_s$	$\lambda_1 + 1.5\lambda_s$	$\lambda_1 + 2\lambda_s$			
Straight woman	$2\lambda_s$	$1.5\lambda_s$	λ_s	λ_1		
Bisexual woman	$1.5\lambda_s$	λ_s	$0.5\lambda_s$	$\lambda_1 + 0.5\lambda_s$	$\lambda_1 + \lambda_s$	
Lesbian woman	λ_s	$0.5\lambda_s$	0	$\lambda_1 + \lambda_s$	$\lambda_1 + 1.5\lambda_s$	$\lambda_1 + 2\lambda_s$

Notes. The Table reports the gains from matching when the surplus is given by the function Φ_A and sexual orientation o_i is constructed starting from LGB identity. When $\lambda_1 < 0$, the match surplus is highest ($2\lambda_s$) for different-sex couples with two straight partners. In contrast, gay men and lesbian women maximize the gains from sexual compatibility in same-sex couples but experience a penalty ($\lambda_1 + 2\lambda_s$). By construction of Φ_A, the match surplus for different-sex couples with at least one bisexual partner is lower than $2\lambda_s$; this is a restriction we later relax and test in Section 3.4. Finally, note that a gay man and a lesbian woman can form a different-sex household thereby avoiding the same-sex penalty, but also giving up on the gains from sexual attraction (and thus their gains are zero).

3.3 Identification and estimation

The identification of this class of models is extensively discussed by Choo and Siow (2006) and Galichon and Salanié (2022). In a nutshell, when the TU assumption holds,[22] the matching function (3.4) can be inverted, and the match surplus is pinned down by observed match frequencies $\mu(x, x')$, which can be estimated with cross-sectional data on households. Identification relies on the distribution of random taste shocks ε_i to be known.[23]

An important implication of the model is that $\mu(x, x')$ must have full support in $\mathcal{X} \times \mathcal{X}$. In our specific application, it is key to observe people with similar traits x, and in partic-

[22]If utility is only imperfectly transferable, then data on transfers between partners and further assumptions on the curvature of the Pareto frontier are needed (Galichon, Kominers, and Weber, 2019).

[23]When this assumption is relaxed, data on multiple markets help restore identification (Chiappori, Salanié, and Weiss, 2017). Consistent with this idea, we leverage variation across markets to estimate a nested-logit version of the model presented in Section 3.4.

ular with the same gender and sexual orientation, living with partners of both genders. If gay individuals only matched with partners of the same sex and bisexual individuals only matched with partners of the opposite sex, the assumption that the same-sex and different-sex marriage market are unified would be rejected right away, and it would be impossible to study trade-offs between the partner's gender and other traits. However, we have seen in Section 2.5 that this is not the case, and that there is variation in partner's gender conditional on one's own gender and sexual orientation. In particular, the mating choices of bisexual individuals, the group that displays the largest heterogeneity in terms of partner's gender, are key for the identification of the differences in gains between same-sex and different-sex couples.

Following Galichon and Salanié (2022), the surplus function is specified as a linear combination of J different scalar functions of the partner types x and x', i.e., the match surplus has the structure $\Phi(x, x') = \sum_{j=1}^{J} \lambda_j \phi_j(x, x')$. The vector of linear coefficients $\lambda = (\lambda_j)_{j=1,..,J}$ can be estimated using a Maximum Likelihood Estimator. With a sample of N_c couples and N_s singles, when x_i is observed for every respondent i and x_i' is also observed for every partnered respondent i, the log-likelihood function corresponds to

$$\mathcal{L}(\lambda) = \frac{1}{N_c} \sum_{i=1}^{N_c} \log \frac{\mu^\lambda(x_i, x_i')}{H^\lambda} + \frac{1}{N_s} \sum_{i=1}^{N_s} \log \frac{\mu^\lambda(x_i, 0)}{H^\lambda}, \tag{3.6}$$

where μ^λ is the equilibrium matching corresponding to the surplus function associated to the vector of parameters λ, while $H^\lambda = \sum_{x,x'} \mu^\lambda(x, x') + \sum_x \mu^\lambda(x, 0)$ is the simulated number of households.

In some of our estimation samples, we do not observe all partners' characteristics, and thus we cannot calculate the log-likelihood function (3.6). For instance, in NHIS data, we never observe the respondent's partner's sexual orientation. Let x_i^o and x_i^u, respectively, be the subvector of observed and unobserved characteristics of respondent i's partner, so that $x_i' = (x_i^o, x_i^u)$. In this case, the log-likelihood function we maximize corresponds to:

$$\mathcal{L}(\lambda) = \frac{1}{N_c} \sum_{i=1}^{N_c} \log \sum_{x^u \in \mathcal{X}^u} \frac{\mu^\lambda(x_i, (x_i^o, x_i^u))}{H^\lambda} + \frac{1}{N_s} \sum_{i=1}^{N_s} \log \frac{\mu^\lambda(x_i, 0)}{H^\lambda}, \tag{3.7}$$

where $\mathcal{X}^u$ is the support for the subvector of unobserved variables x^u.

In practice, the most important missing variable we have to deal with is the partner's sexual orientation.[24] However, since we observe the respondents' sexual orientation, we do know how this variable co-varies with other partners' characteristics. Hence, the only missing element of the between-partner covariance matrix is the covariance between the respondents' and partners' sexual orientation. As we lack this data moment, we cannot identify changes in match surplus due to differences in *both* partners' sexual orientation, and therefore we cannot estimate the functional forms Φ_D and Φ_E with NHIS data (but we can do so with SOEP data).

3.4 Extensions

We now propose several extensions of our benchmark model to address some of its limitations. First, we will allow for a gender-specific same-sex penalty. Then, to allow for a more flexible characterization of sexual orientation, we will propose three extensions: (i) we relax that sexual orientation enters linearly in the match surplus; (ii) we allow for taste shocks for potential partners of the same gender to be correlated; (iii) we use NFSG data to build a latent sexual orientation variable over a 5-point support to accommodate greater variation in sexual orientation. Finally, we estimate complementarities in sexual attraction using information of sexual orientation of both partners.

Gender-specific same-sex penalty. In our first deviation from the benchmark model, we allow the penalty to be different for male and female same-sex couples:

$$\Phi_B(x_i, x_{i'}) = \lambda_0 + \lambda_1 m_i m_{i'} + \lambda_2 f_i f_{i'} + \lambda_s (s_{ii'} + s_{i'i}),$$

so that we can test if λ_1 is equal to λ_2. In fact, the hurdles female and male same-sex couples may experience are likely different. Importantly, female same-sex couples are more likely to have children (see Table 27 in Appendix).

Non-linearity in sexual orientation. Next, we can test if sexual orientation is one-dimensional by relaxing the assumption that the gains are linear in sexual compatibility $s_{ii'}$. Our baseline specification Φ_A implies that individuals standing in the middle of the

[24] The lack of partner's race in the 2019-2022 NHIS data does not pose any concrete problem to our estimation, as we do know if the partner is of the same race as the respondent. Since we do not try to estimate race-specific complementarities (e.g., for White and Hispanics, or for Blacks and Others) but only look into differences between same-race and different-race couples, this does not constitute a problem, as we do observe the share of same-race couples also in the 2019-2022 NHIS data.

spectrum (in practice, bisexual individuals) are mildly compatible with both genders, but never experience perfect compatibility ($s_{ii'} = 1$). We can test if this is the case by specifying a more flexible functional form for the surplus function:

$$\Phi_C(x_i, x_{i'}) = \lambda_0 + \lambda_1 m_i m_{i'} + \lambda_2 f_i f_{i'} + \sum_{k \in S} \gamma_k \left(\mathbb{1}\left\{s_{ii'} = k\right\} + \mathbb{1}\left\{s_{i'i} = k\right\}\right),$$

where $S = \{0, 0.5, 1\}$ and with the restriction $\gamma_0 = 0$. Individuals who are in the middle of the spectrum ($o_i = 0.5$) are perfectly compatible with both genders if $\gamma_{0.5} = \gamma_1$.

Nested-logit model. Another potential concern is that LGB identity does not fully capture the amount of variation in sexual orientation among individuals. This is confirmed by Table 2 where, particularly among bisexual respondents, there seem to be differences in the degree of same-sex attraction. Our first solution to address this problem is to consider a nested-logit version of the model where individual taste shocks for potential partners of the same gender are allowed to be correlated. Under this assumption, even after conditioning on (observed) sexual orientation o_i, individuals are heterogeneous in their degree of same-sex attraction. For instance, this means that some bisexual individuals are systematically more attracted to men, and others to women. The technical details of the nested-logit version of the model are provided in Appendix B.1.

Latent sexual orientation. In a separate model extension, we assume a 5-point support for the sexual orientation variable o_i. That is $o_i \in S = \{0, 0.25, 0.5, 0.75, 1\}$, where $o_i = 0$ means individual i is only attracted to men and $o_i = 1$ means individuals i is only attracted to women. Since in the NHIS dataset we only have information about LGB identity, we use NSFG data to construct probability weights $\Pr\{o_i = k|z_i\}$, $k \in S$, conditional on a respondent's characteristics z_i.[25] The latter are observed in both the NHIS and the NSFG datasets and include the respondent's LGB identity.[26] In practice, in order to construct the probability weights, we estimate a battery of ordered probit models with NSFG data, using the degree of sexual attraction towards women as the dependent variable (see Section 2.2 for an extensive description of this variable). After-

[25]We remind the reader that, unfortunately, we cannot directly use NSFG data for our analysis of mating patterns because they lack information on live-in same-sex partners.

[26]The other regressors in z_i include age, education, race, employment, place of birth (in the U.S. or outside), time period of interview, and having kids (only included for women).

wards, we estimate the model by Weighted Maximum Likelihood. The technical details are provided in Appendix B.2.

This approach has two advantages. First, it allows us to accommodate greater variation in sexual orientation that might have gone lost when its measurement only relied on the LGB categorization. Second, it allows us to reintroduce into our estimation sample NHIS respondents who do not identify as either straight, gay/lesbian, or bisexual. In fact, we can estimate probability weights $\Pr\{o_i = k|z_i\}$ also for individuals who answered "Something else", "I don't know", or who refused to answer.

Complementarities in sexual attraction. Finally, when we work with SOEP data, we have information on the sexual orientation of both partners. With such data, we can introduce more flexible functional forms and test if there exist complementarities in sexual attraction. First, we test if $\lambda_{ss} > 0$ in the specification

$$\Phi_D(x_i, x_{i'}) = \lambda_0 + \lambda_1 m_i m_{i'} + \lambda_2 f_i f_{i'} + \lambda_s(s_{ii'} + s_{i'i}) + \lambda_{ss} s_{ii'} s_{i'i}.$$

If $\lambda_{ss} \gg \lambda_s$, partners particularly benefit from the relationship if sexual attraction is mutual. Additionally, we can make the specification even more flexible and estimate the match surplus for all combinations of $s_{ii'}$ and $s_{i'i}$, to better capture the covariation in partners' gender and sexual orientation described in Table 4. The resulting match surplus is given by

$$\Phi_E(x_i, x_{i'}) = \lambda_0 + \lambda_1 m_i m_{i'} + \lambda_2 f_i f_{i'} + \sum_{k \in S} \sum_{k' \in S} \gamma_{kk'} \mathbb{1}\{s_{ii'} = k, s_{i'i} = k'\},$$

with the restrictions $\gamma_{00} = 0$ and $\gamma_{kk'} = \gamma_{k'k}$ for any (k, k'), the latter due to the symmetry of the match surplus function.

3.5 Multiple markets

Similarly to Chiappori, Salanié, and Weiss (2017), when we take the model to the data, we assign households to different marriage markets based on the observed region of residence. In particular, in the NHIS, we assume that each of the four macro-regions (North-East, North-Central & Midwest, South, West) is an independent and perfectly segregated marriage market.[27] While the marginals are region-specific (e.g., gender ratios and fraction of LGB individuals differ across regions), preferences are assumed to be constant, with the exception of a region-specific dummy in the match surplus. To test differences in the same-sex penalty across regions, we also estimate a version of the model where the penalty is region-specific.

With 2019-2022 NHIS data, we can further subdivide each macro-region into separate sub-markets based on the type of metropolitan area the respondent lives in. This additional market subdivision is relevant as sexual orientation is not equally distributed across types of metropolitan areas. Figure 5 in appendix shows for instance that 51.7% of gay men live in a central large metropolitan area in comparison to 34.1% of straight men. In our empirical applications, we will divide each macro-region in two markets: one for central large metropolitan areas and one for other types of areas (non-metropolitan, small metropolitan, fringe large metropolitan). Also in this case, we can test if the same-sex penalty differs across different areas.

4 Results

4.1 Same-sex penalty and sexual orientation

We first estimate the model with NHIS data, using the LGB categorization to build our measure of sexual orientation, as explained in Section 3.2. Results for our benchmark NHIS sample are reported in Table 6. In column (1), our bare-bones three-parameter model, where the surplus function simply corresponds to Φ_A, indicates that sexual orientation largely shapes mating patterns. Not surprisingly, individuals value sexual compatibility highly. Yet, regardless of their sexual orientation, individuals who choose a partner of the same gender face a penalty that might discourage them from engaging in same-sex relationships. In column (2), we separately estimate the penalty for male and female same-sex couples, and we find it to be substantially stronger for male same-sex couples.

Differences in child-rearing patterns between different-sex and same-sex couples can

[27] North-East includes: Connecticut, Maine, Massachusetts, New Hampshire, Rhode island, Vermont, New Jersey, New York, Pennsylvania. North-Central & Midwest includes: Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota. South includes: Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, West Virginia, Alabama, Kentucky, Tennessee, Mississippi, Arkansas, Louisiana, Oklahoma, Texas. West includes: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, California, Hawaï, Oregon, Washington.

potentially explain the presence of such a penalty. Table 27 in Appendix shows that, regardless of the partners' sexual orientation, different-sex households are substantially more likely to raise children. For instance, in the 2013-2018 period, 35% of lesbian women in a same-sex household have children, a non-negligible share. However, the share increases to 71% among lesbian women in a different-sex household. Moreover, child-rearing patterns can also explain the difference in the penalty estimated for male and female same-sex couples. The latter are indeed much more likely to have children, as is already well established in the literature. In Section 4.3, we extend the discussion and relate the estimated match surplus to a larger number of family and individual outcomes (e.g., legal marital status, physical and mental health).

Table 6: Match surplus function parameters - LGB categories

	Specifications					
	(1)	(2)	(3)	(4)	(5)	(6)
Same-sex couple	-1.37					
	(0.10)					
Male same-sex couple		-2.01	-2.32	-2.35	-3.05	-3.00
		(0.15)	(0.15)	(0.15)	(0.18)	(0.14)
Female same-sex couple		-0.70	-0.83	-0.86	-2.20	-2.17
		(0.14)	(0.15)	(0.15)	(0.14)	(0.14)
Sexual compatibility	8.98	9.02	8.82		4.70	
	(0.14)	(0.14)	(0.14)		(0.10)	
Sexual compatibility (intermediate)				3.22		3.31
				(0.18)		(0.11)
Sexual compatibility (high)				8.14		4.91
				(0.15)		(0.10)
Functional form	Φ_A	Φ_B	Φ_B	Φ_C	Φ_B	Φ_C
Other controls	No	No	Yes	Yes	Yes	Yes
Nested logit	No	No	No	No	Yes	Yes
N	81,798	81,798	81,798	81,798	81,798	81,798
BIC	1,664,201	1,664,170	1,518,029	1,518,001	1,515,931	1,515,863

Notes. 2013-2018 NHIS data. "Other controls" refers to the presence of dummies for the partners' age categories (6 groups), levels of education (5 groups), and race (4 groups), as well as polynomial terms measuring complementarities between partners traits. In columns (5) and (6), taste shocks over potential partners are allowed to be correlated within nests. Specifications details in Sections 3.2 and B.1. Standard errors in parentheses.

However, the estimated differences in match surplus between different-sex, male same-sex, and female same-sex couples could partly be explained by a spurious correlation between sexual orientation and socioeconomic characteristics. Hence, in column (3), we introduce flexible controls for the partners' age, education, and race, as well as interaction terms between partners' characteristics to account for complementarities in the match surplus. With these additional controls, the estimated penalty for both male and female same-sex couples is actually slightly stronger than what initially found in column

(2).

In column (4) of Table 6, we test the linearity assumption on the sexual compatibility score $s_{ii'}$ implied by the functional forms Φ_A and Φ_B. The match surplus is found to be slightly convex in sexual compatibility,[28] although this does not alter our conclusions about the same-sex penalty. Given our definition of sexual compatibility, defined in equation (3.5), this result implies that individuals standing in the middle of the spectrum illustrated in Figure 1 are always associated with an intermediate sexual compatibility score. Hence, this convexity might be explained by bisexual individuals having weaker preferences for live-in relationships, also after accounting for the fact that they are on average younger than gay and straight individuals.[29]

In columns (5) and (6), we allow taste shocks over partners to be correlated within nests, with each nest including all partners of the same gender and race. Results from the nested logit model are qualitatively similar to the benchmark i.i.d. case, although the same-sex penalty is found to be stronger for both men and women. We do find that taste shocks are correlated for partners of the same gender, which implies the presence of additional differences in sexual orientation conditional on LGB categories.[30] For instance, some bisexual individuals might systematically be more attracted to men over women, or vice versa. This explains why, in column (5), the weight given to sexual compatibility in the match surplus is lower. Moreover, while column (6) confirms that bisexual individuals enjoy lower match gains than straight or gay people matched with their preferred gender, it also shows a much narrower gap relatively to column (4). Under this specification, the match surplus is found to be concave, and not convex, in sexual compatibility.

The BIC values reported in the last line of Table 6 suggest that the non-linearity introduced in column (4) leads to almost no improvement to the model fit. The model presented in column (3), where sexual compatibility enters the match surplus linearly, al-

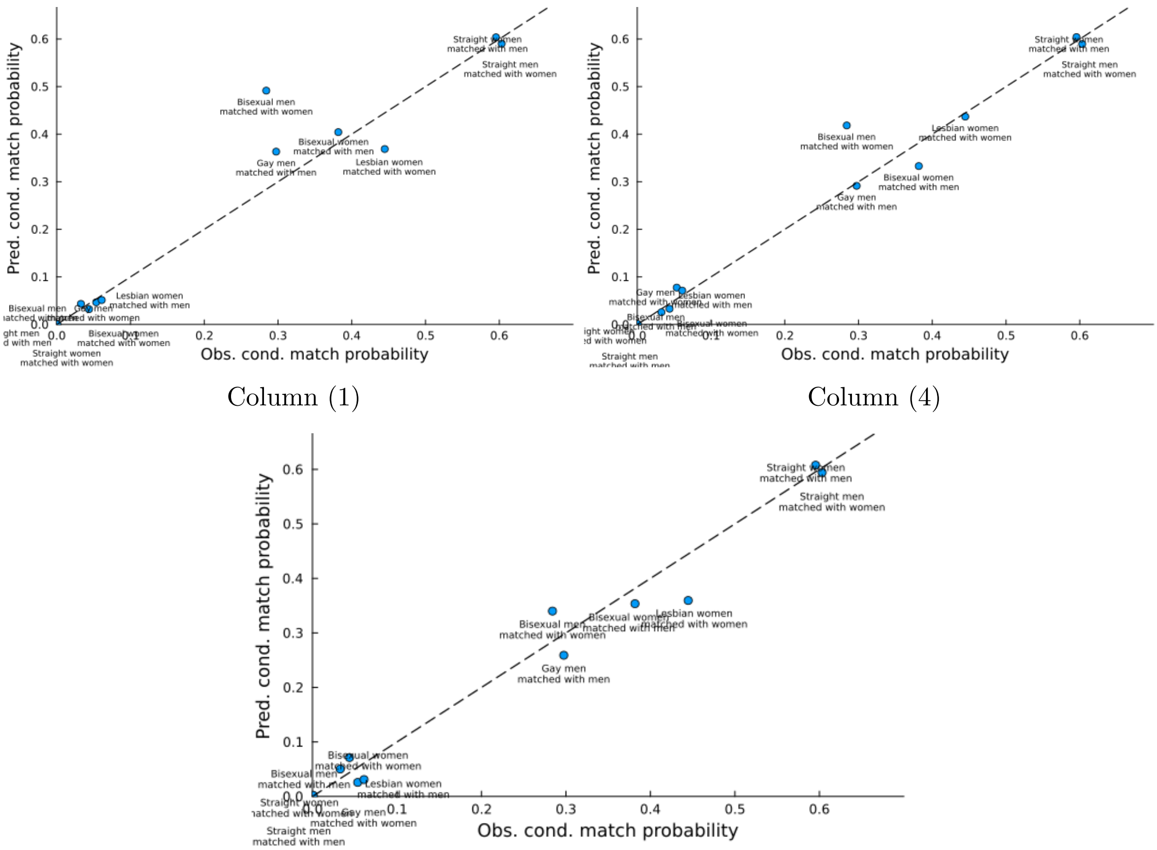
[28] We formally test and reject that the coefficient for intermediate sexual compatibility (estimated to be 3.31) is equal to half the coefficient for high sexual compatibility ($0.5 * 8.14 = 4.07$).

[29] Another explanation is that bisexual individuals no longer identify as such once in a relationship. Hence, we only observe them when single and tend to underestimate their preference for live-in relationships. There is some empirical evidence showing that this might partly explain our result. Using SOEP longitudinal data, Tables 22 and 21 in Appendix show that bisexual individuals are indeed more inclined to changes in LGB identity than other categories. Hu and Denier (2023) also show that, when bisexual individuals enter a relationship, they are slightly more likely to report a different LGB identity.

[30] In other words, we can test and reject that the degree of independence within nest $\mathcal{G}$, named $\rho_{\mathcal{G}}$ in equation (B.1), is equal to one for the vast majority of nests. See the estimated $\rho_{\mathcal{G}}$ in Table 26 in Appendix.

ready fits the data well. The nested logit models in columns (5) and (6) lead to a modest improvement of the model fit, with their BIC values being only slightly lower than the standard logit model in column (4).

Figure 2: Model's fit for mating patterns with respect to sex and sexual orientation



Column (6)

Notes. Results obtained with 2013-2018 NHIS data. Subplots correspond to columns (1), (4), and (6) in Table 6. In each subplot, we plot the predicted probabilities of matching with a partner of a given gender conditional on one's own gender and sexual orientation against their empirical counterparts.

Since we are particularly interested in sorting on gender and sexual orientation, we also document how the fitted models replicate these patterns in Figure 2.[31] The first panel shows that the bare-bones three-parameter model already succeeds in replicating the key features of the mating patterns with respect to gender and sexual orientation. The second panel presents the fit for the model presented in column (4) of Table 6. This model successfully replicates the matching probabilities of individuals with different combinations of gender and sexual orientation, with one exception: the model slightly

[31] In Appendix, in Figures 8 and 9, we provide similar plots documenting the model's fit respectively for racial and educational mating patterns.

over-predicts the share of bisexual men matching with women, while it under-predicts the share of single bisexual men. In the last panel, we show that the nested logit version of the same model (corresponding to column (6) in Table 6) does a better job at explaining the mating patterns of bisexual individuals, but it also underpredicts the share of matched gay men and lesbian women in equilibrium.

4.2 Complementarities in age, education, and race

We also perform a separate model estimation to test if age, education, and race complementarities differ between same-sex and different-sex couples. The findings are displayed in Table 7. In line with a large body of research, the first column of the Table shows that age, education, and race are complements. Among different race groups, Blacks display the strongest taste for same-race relationships. Comparing the different columns, we can see that the degree of educational complementarity for same-sex couples is similar to that of different-sex couples, whereas complementarities with respect to age and race are weaker for same-sex couples. These findings confirm those of Ciscato, Galichon, and Goussé (2020) for California.

Table 7: Estimated complementarities in age, education, and race

	Couple types		
	All	Male same-sex	Female same-sex
	(1)	(2)	(3)
Same-sex penalty		-2.58	-0.78
		(0.22)	(0.21)
Educ. gap	-1.61	0.19	-0.04
	(0.01)	(0.15)	(0.13)
Age gap	-2.07	0.67	0.32
	(0.01)	(0.12)	(0.12)
Both Black	10.69	-1.45	0.35
	(0.12)	(0.84)	(0.42)
Both Hispanic	7.91	-1.48	-1.84
	(0.07)	(0.43)	(0.47)
Both Other	8.38	-1.63	-1.10
	(0.09)	(0.85)	(0.60)
Functional form		Φ_B	
Other controls		Yes	
Nested logit		No	
N		81,798	
BIC		1,518,065	

Notes. Results obtained with 2013-2018 NHIS data. In this Table, we present both the same-sex penalty and the degree of complementarity in education, age, and race for different types of couples, namely different-sex, male same-sex, and female same-sex couples. For education and age, a negative coefficient for the respective *gap* suggests agents are attracted to their likes. Conversely, for different race groups, a positive coefficient indicates that agents are more attracted to same-race partners relatively to different-race ones. In columns (2) and (3), we present the estimated difference in the complementarity parameter between same-sex and different-sex couples. Standard errors in parentheses.

4.3 Understanding differences in the gains from marriage

In order to explore what factors could explain differences in the estimated match surplus across households, we look at how directly observed household outcomes correlate with the structurally estimated payoffs. In other words, for every respondent i in our estimation sample, we calculate both terms of agent i's expected payoff from matching with her/his observed partner:[32]

$$EU(x_i, x_i') = U(x_i, x_i') + \varepsilon^*(x_i'|x_i), \tag{4.1}$$

where $\varepsilon^*(x_i'|x_i)$ denotes the expected value of $\varepsilon_{ix_i'}$ *conditional* on i being matched with a partner of type x_i' in equilibrium. When the shocks are type-I extreme value distributed,

[32]Note that we observe all partner's characteristics except for her/his sexual orientation. Recall the distinction between observed and unobserved partner's characteristics in Section 3.3. We create duplicates for every possible level x^u that the subvector of unobserved partner's traits can take, and we weight duplicate observations by the odds $\Pr\{x_i^u = x^u|x_i, x_i^o\}$, where x_i^o is the subvector of observed partner's traits.

this expected value has a closed-form solution corresponding to $\varepsilon^*(x_i'|x_i) = -\log \mu^d(x_i'|x_i)$. Then, we regress directly observed measures of household outcomes, such as the number of children, the couple's legal status, and the respondent's well-being, on the two terms of $EU(x_i, x_i')$.[33]

The upper panel of Table 8 shows that the match gains predicted by our model are positively correlated with the presence of children and are associated with a higher probability of being legally married. They are also positively correlated with both physical and mental health.[34] This shows that the dimensions of observed heterogeneity we study in the model are important determinants of both households outcomes and individual well-being. In the lower panel, we add flexible controls for the partners' education, age, and race, so that the variation left in both $U(x_i, x_i')$ and $\varepsilon^*(x_i'|x_i)$ is due to differences in the partners' gender and sexual orientation across couples (e.g., same-sex vs different-sex couples, high vs low sexual compatibility). The linear coefficients maintain their signs, most of them stay significant, and some (those related to well-being) actually become larger in size. This suggests that differences in the match surplus due to variation in gender composition and sexual compatibility, including the same-sex penalty, are related to these observed household outcomes.

Several studies have shown how same-sex marriage legalization improves the quality of relationships for same-sex couples in many ways. Same-sex marriage legalization is associated with lower separation rates (Chen and van Ours, 2020), higher income growth and homeownership rates (Delhommer and Hamermesh, 2021), better mental health outcomes (Chen and Van Ours, 2022), and healthcare access (Carpenter, Eppink, Gonzales, and McKay, 2021). Our findings are consistent with this literature; we interpret the positive association between match gains and the odds of being legally married as suggestive that the same-sex penalty might partly be explained by limited access to legal marriage. Similarly, the positive association between match gains and the number of children is suggestive that same-sex couples might experience lower match gains due to difficulties related to childbearing, adoption, and uncertainty over parental rights. This statement is corroborated by the large difference in the estimated penalty across genders,

[33]The presence of $\varepsilon^*(x_i'|x_i)$ can also be understood as a way of taking into account the fact we only observe couples that self-select into cohabiting relationships, who on average benefit from higher draws for $\varepsilon_{x'}$.

[34]Details on how physical and mental health measures are constructed are provided in Appendix A. In Table 28, we provide additional findings using alternative measures of well-being and health.

Table 8: Regressing observed family outcomes on predicted match gains

	Number of children (1)	Legally married (2)	Physical health (3)	Mental health (4)
A. Without controls				
$U(x_i, x_{i'})$	0.167	0.080	0.076	0.032
	(0.006)	(0.002)	(0.006)	(0.003)
$\varepsilon^*(x_{i'} \| x_i)$	0.128	0.053	0.044	0.019
	(0.006)	(0.002)	(0.006)	(0.003)
R^2	0.021	0.066	0.009	0.005
B. Controlling for partners' race, age, and education				
$U(x_i, x_{i'})$	0.111	0.049	0.089	0.067
	(0.010)	(0.003)	(0.011)	(0.005)
$\varepsilon^*(x_{i'} \| x_i)$	0.034	0.013	0.053	0.035
	(0.008)	(0.003)	(0.009)	(0.004)
R^2	0.152	0.145	0.121	0.031

Notes. Results obtained with 2013-2018 NHIS data. In panel A, every column represents a separate weighted OLS regression with only two explanatory variables, $U(x_i, x_{i'})$ and $\varepsilon^*(x_{i'}|x_i)$. These are calculated using the parameters reported in column (4) of Table 6 and are both scaled by the standard deviation of $EU(x_i, x_{i'})$ in our estimation sample. In panel B, we run similar regressions while also adding fixed effects for both partners' race, age, and education. Standard errors in parentheses. More details about the outcome variables are available in Appendix A.

since male same-sex couples are much less likely to have children, as also shown in Table 27 in Appendix. Finally, the positive association between match gains and mental health is not only consistent with a direct impact of social stigma and discrimination on same-sex couples, but also with an indirect impact due to the aforementioned hardships same-sex couples have to face (Carpenter, Eppink, Gonzales, and McKay, 2021; Chen and Van Ours, 2022).

4.4 Same-sex penalty before and after Obergefell v. Hodges

We now estimate the model separately for the 2013-2015 and 2016-2018 periods, after dividing our main NHIS sample in two, and for the 2019-2022 period, using the new version of the NHIS survey data. In particular, we can check if the same-sex penalty changed after 2015, when same-sex marriage became legally available at the federal level with the U.S. Supreme Court ruling *Obergefell v. Hodges*. Table 9 shows that the male same-sex penalty has decreased between 2013-2015 and 2016-2018, while it has remained stable thereafter. Hence, the timing of this decrease coincides with the same-sex marriage legalization in the entire U.S. On the other hand, the same Table shows that the female same-sex penalty has remained stable across the three periods.

Our findings regarding the same-sex penalty should be interpreted in light of two key considerations. First, we quantify the incentives to form a same-sex household, regardless of the legal arrangements between partners. Previous research has shown that a large proportion of same-sex couples do get married when same-sex marriage is made available (Carpenter, 2020; Carpenter, Eppink, Gonzales, and McKay, 2021). However, the same studies have found little effect of same-sex marriage legalization on the overall number of same-sex households. Hence, given that same-sex marriage legalization does not necessarily lead to more same-sex households, we should *not* expect the penalty to have decreased drastically immediately after Obergefell v. Hodges.

Second, in our analysis, we look at the *stocks* of same-sex households across repeated cross-sections rather than the *flows* of newly formed households.[35] Even if new same-sex households are formed at the same rate as before the ruling, it is possible that access to legal same-sex marriage has made both new and existing unions more stable. This is consistent with the strong take-up effects of access to legal same-sex marriage and its positive implications for health and economic outcomes of LGB individuals documented in the literature (Badgett, Carpenter, Lee, and Sansone, 2024). Hence, the decrease in the male same-sex penalty picks up this reinforced commitment among male same-sex couples. Interestingly, we do not see a similar decrease among female same-sex couples, which suggests that access to legal same-sex marriage has not changed their stability patterns. This might be explained by differences in fertility between male and female same-sex couples, as child-rearing is a more prominent commitment mechanism among the latter.[36]

[35]Unfortunately, NHIS data do not contain information on relationship duration, and thus we cannot identify newly formed couples in the data.

[36]Sansone (2019) and Hansen, Martell, and Roncolato (2020) do not find any evidence that access to legal marriage has changed fertility patterns among same-sex couples, while Martin and Rodriguez (2022) do show a positive impact on the number of adoptions.

Table 9: Male and female same-sex penalty by time period

	2013-2015		2016-2018		2019-2022	
	(1)	(2)	(3)	(4)	(5)	(6)
Male same-sex couple	-2.35	-2.40	-1.87	-1.96	-1.88	-1.85
	(0.20)	(0.20)	(0.23)	(0.23)	(0.17)	(0.18)
Female same-sex couple	-0.51	-0.57	-0.69	-0.73	-0.80	-0.79
	(0.21)	(0.20)	(0.22)	(0.22)	(0.17)	(0.18)
Sexual compatibility	8.34		7.72		7.46	
	(0.18)		(0.16)		(0.12)	
Sexual compatibility (intermediate)		2.82		2.78		4.06
		(0.24)		(0.24)		(0.20)
Sexual compatibility (high)		7.71		7.18		7.67
		(0.18)		(0.19)		(0.18)
Functional form	Φ_B	Φ_C	Φ_B	Φ_C	Φ_B	Φ_C
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Nested logit	No	No	No	No	No	No
N	47,316	47,316	34,482	34,482	45,884	45,884
BIC	872,581	872,559	642,626	642,615	843,595	843,603

Notes. Results obtained with 2013-2018 and 2019-2022 NHIS data. We estimate the model separately for each different time period. Standard errors in parentheses.

4.5 Same-sex penalty across the U.S.

In Figure 3, we report the estimated same-sex penalties for different macro-regions of the U.S over the periods 2013-2018 and 2019-2022. The estimates suggest that, for the 2013-2018 period, the penalty for male same-sex couples is comparable in size across regions. On the other hand, the penalty for female same-sex couples is not only lower (in absolute terms) than the men's in every region, but is also not significantly different from zero in the North-East. For the 2019-2022 period, we observe that the decrease in the male same-sex penalty discussed in the previous Section comes mainly from a decrease in the North-East and North-Central regions, while the penalty remains similar in the South and West regions. With regard to the female same-sex penalty, estimates have become barely significant in all macro-regions, with the exception of the South.

These findings are broadly consistent with the 2010-2014 and 2017-2022 waves of the World Values Survey, which reveal that acceptance of homosexuality is increasing over time, although regional differences remain, with the North-East and South respectively being the most and least tolerant region in the U.S.[37] Moreover, this is also consistent with

[37]Figure 6 in appendix shows that less people in the North-East region mention that they would dislike having homosexuals as neighbors, while more people in the South mention it. This Figure also shows a decrease in attitudes against homosexuals over time, consistent with our results on the decrease in the same-sex male penalty. Figure 7 shows that acceptance of same-sex couples as good parents is more homogeneously distributed across regions, although in the North-East, the share of people who think same-sex couples are not as good as parents as other couples is lower while it is higher in the South.

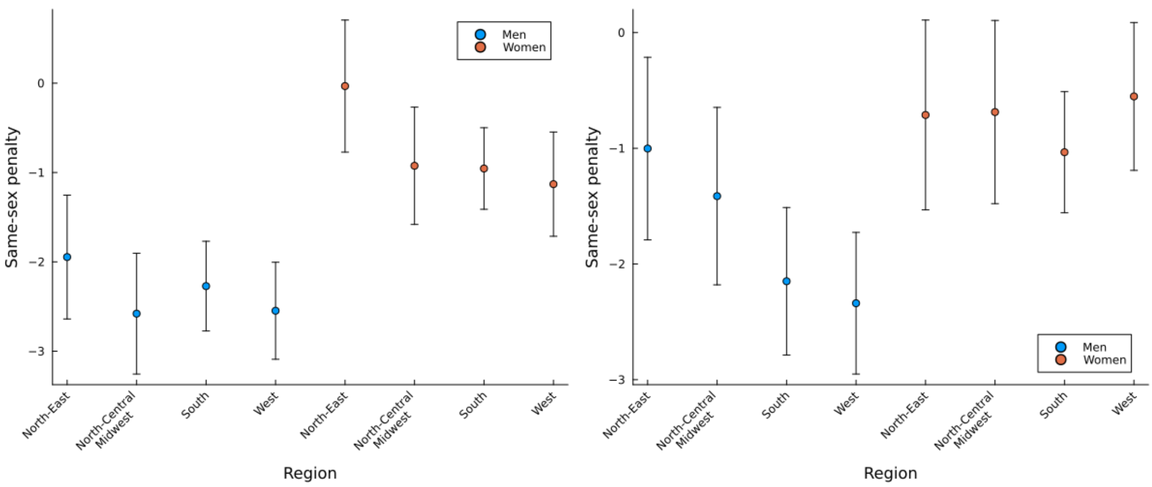
states that legalized same-sex marriage early being almost all located in the North-East region.[38]

Using the 2019-2022 NHIS sample, we estimate an alternative version of the model with 8 separate marriage markets. Each of the four macro-regions is divided into two separate sub-markets, corresponding to a market for central large metropolitan areas, and a market corresponding to all other areas (see Section 3.5). Our estimates, displayed in Figure 4a, suggest that the female same-sex penalty is not significant in large central metropolitan areas. For men, the penalty is smaller in central large metropolitan areas than elsewhere. Yet, it is important to note that, if LGB individuals relocate to larger metropolitan areas to avoid discrimination or to reduce search frictions, we might underestimate differences in the same-sex penalty across different metropolitan areas.

Finally, in Figure 4b and 4c, we show estimates of the male and female same-sex penalty for the 8 distinct markets. Due to the smaller sample size for each market, the estimates are less precise. However, we observe that the male same-sex penalty is much smaller and even not significant in large metropolitan areas in the North-East and North-Central Midwest, while it stays large and negative in the other regions and in smaller areas. We also observe that the female same-sex penalty only persists in non-metropolitan and smaller metropolitan areas of the Southern regions.

[38] The first States that, starting from 2004, legalized same-sex marriage by law through a state court decision are: Massachusetts (NE), Connecticut (NE), Iowa (NC), Vermont (NE), New Hampshire (NE), District of Columbia (S), New York (NE), Maine (NE), Washington (W). Hence, with the exception of Iowa (NC), DC (S) and Washington (W), they are all located in the North-East region.

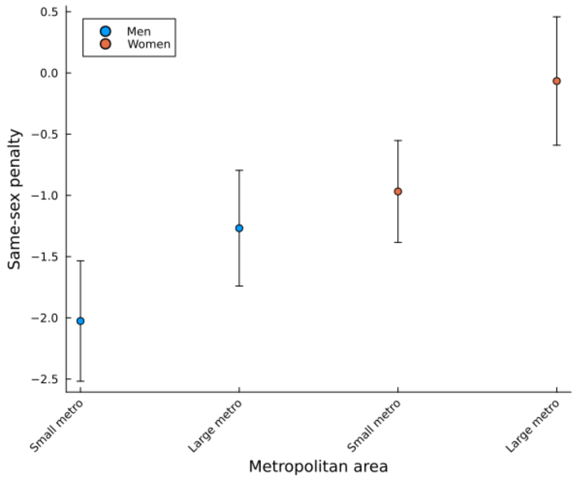
Figure 3: Male and female same-sex penalty by region (NHIS)



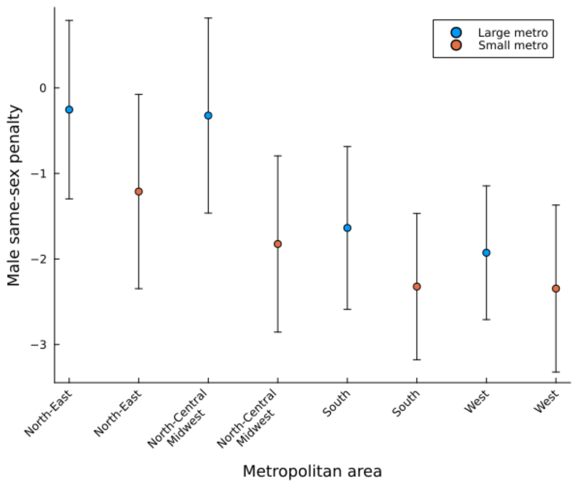
(a) 2013-2018 (b) 2019-2022

Notes. Estimates obtained with 2013-2018 and 2019-2022 NHIS data. In the plot, we report the point-estimates and 95% confidence intervals for region-specific same-sex penalties.

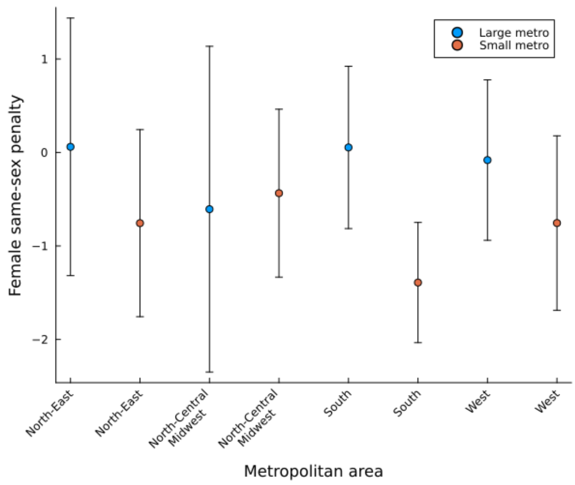
Figure 4: Male and female same-sex penalty by region and metropolitan area (NHIS)



(a) By metropolitan area



(b) By region and metropolitan area (men)



(c) By region and metropolitan area (women)

Notes. Estimates obtained with 2019-2022 NHIS data. In the plot, we report the point-estimates and 95% confidence intervals for metropolitan area and region-specific same-sex penalties.

4.6 Latent sexual orientation

We now want to account for the fact that LGB identity does not fully capture the entire variation in sexual orientation. Hence, using the main NHIS sample for the 2013-2018 period, we estimate the model using a latent measure for sexual orientation. The mapping from LGB categories, present in the NHIS, to the degree of sexual attraction towards women, absent from the NHIS, is possible thanks to the joint presence of both variables in the NSFG, as explained in Section 3.4. This mapping can also be constructed for individuals who answered "Something else", "I don't know", or who refused to answer, thus allowing us to include these individuals in the estimation sample. The results obtained through Maximum Weighted Likelihood Estimation are presented in Table 10. They are both qualitatively and quantitatively similar to our baseline findings, providing validation for the LGB categorization as an appropriate measure of sexual orientation. Yet, they also offer some new insights. First, the same-sex penalties are larger than in Table 6, albeit only slightly. This is explained by LGB categories leading us to underestimate the share of individuals at least partly attracted to partners of the same gender. On the other hand, column (4) also shows that individuals with only a weak sexual compatibility with their partner's gender would rather avoid this type of relationship. Hence, the model suggests that, even if there exists a non-negligible share of women who report being mainly, but not exclusively attracted to men, their chances of seriously considering a same-sex partner for a live-in relationship are actually very slim, and not significantly higher than those of women who report no attraction at all for same-sex partners.

Table 10: Match surplus function parameters - latent sexual orientation

	Latent sexual orientation from NSFG data				LGB identity from NHIS data	
	(1)	(2)	(3)	(4)	(5)	(6)
Same-sex couple	-1.45					
	(0.10)					
Male same-sex couple		-2.17	-2.46	-2.57	-2.32	-2.35
		(0.14)	(0.15)	(0.15)	(0.15)	(0.15)
Female same-sex couple		-0.71	-0.95	-0.97	-0.83	-0.86
		(0.14)	(0.15)	(0.14)	(0.15)	(0.15)
Sexual compatibility	9.14	9.20	8.82		8.82	
	(0.10)	(0.10)	(0.10)		(0.14)	
Sexual compatibility (medium-low)				-0.36		
				(0.25)		
Sexual compatibility (intermediate)				2.66		3.22
				(0.18)		(0.18)
Sexual compatibility (medium-high)				5.14		
				(0.14)		
Sexual compatibility (high)				7.44		8.14
				(0.14)		(0.15)
Functional form	Φ_A	Φ_B	Φ_B	Φ_C	Φ_B	Φ_C
Other controls	No	No	Yes	Yes	Yes	Yes
Nested logit	No	No	No	No	No	No
N	82,989	82,989	82,989	82,989	81,798	81,798
BIC	1,728,708	1,728,664	1,580,952	1,580,850	1,518,029	1,518,001

Notes. Results obtained with 2013-2018 NHIS data. In the first four columns, sexual orientation is not directly observed, and the findings are obtained through Maximum Weighted Likelihood Estimation using a set of weights yielding the conditional probability distribution of the respondent's sexual orientation conditional on her/his own other characteristics (including if she/he identifies as LGB). These weights are calculated using the NSFG survey. In the last two columnes, we report the findings of Table 6 for comparison, where sexual orientation is directly inferred from the respondent's LGB identity. Standard errors in parentheses.

4.7 Complementarities in sexual attraction

We replicate our analysis with SOEP data. The advantage of this survey is that we can observe both partners' sexual orientation, and thus we can estimate complementarities in sexual compatibility. In Table 11, findings from the bare-bones three-parameter model are reported in column (1) and show that the same-sex penalty is comparable to the one estimated for the U.S. On the other hand, sexual orientation matters more for the choice of the partner, with its linear coefficient being higher than in the U.S. In column (2), we show that the same-sex penalty is stronger for male same-sex couples. In column (3), after controlling for the partners' age and education, the difference in penalties for male and female same-sex couples widens. Overall, the estimated penalty for both male and female same-sex couples is close to the one estimated for the U.S. in Table 6. In columns (4) and (5), we estimate the model with more flexible specifications. Column (4) suggests the match surplus is slightly convex in the sexual compatibility variable, similarly to what was found for the U.S. in Table 6. In column (5), we allow couples to enjoy additional gains when sexual compatibility is *mutual*, adopting the functional form Φ_D. Reassuringly, we do find that couples where both partners are sexually attracted to each other enjoy higher gains.

Table 11: Match surplus function parameters

	Specifications				
	(1)	(2)	(3)	(4)	(5)
Same-sex couple	-1.69				
	(0.29)				
Male same-sex couple		-2.11	-2.46	-2.51	-2.50
		(0.40)	(0.41)	(0.41)	(0.41)
Female same-sex couple		-1.15	-0.99	-1.02	-1.12
		(0.43)	(0.45)	(0.45)	(0.45)
Sexual compatibility	11.36	11.38	11.28		2.60
	(0.36)	(0.38)	(0.38)		(1.40)
Sexual compatibility (intermediate)				3.55	
				(0.51)	
Sexual compatibility (high)				9.71	
				(0.45)	
Sexual compatibility (interaction)					18.16
					(0.78)
Functional form	Φ_A	Φ_B	Φ_B	Φ_C	Φ_D
Other controls	No	No	Yes	Yes	Yes
Nested logit	No	No	No	No	No
N	12,069	12,069	12,069	12,069	12,069
BIC	123,268	123,275	116,601	116,600	116,598

Notes. SOEP, waves 2016 and 2019. "Other controls" refers to the presence of dummies for the partners' age categories (6 groups) and levels of education (3 groups), as well as polynomial terms measuring complementarities between partners traits. More details about the functional forms are available in Section 3.2. Standard errors in parentheses.

In Table 12, we present the findings obtained with our most flexible functional form, Φ_E, which provide further insights on the structure of complementarities. The match surplus barely increases if an agent's sexual attraction for her/his partner increases from intermediate to high, but the partner's sexual attraction for her/him is held constant. Yet, matches where sexual compatibility is one-directional (intermediate or high for one partner, low for the other) do produce positive gains (estimated to be 4.13 and 6.15 respectively) relatively to the benchmark, i.e., matches where the partners are not attracted to each other at all.

The last column of Table 11 shows that the same-sex penalty does not change when we allow for interactions between the partners' reciprocal sexual attractions. Similarly, the estimates of the male and female same-sex penalty remain almost unchanged with our most flexible functional form Φ_E; they respectively correspond to -2.53 (0.41) and 1.20 (0.45). Since we are unable to control for such interactions when we work with NHIS data, this finding reassuringly shows that our estimates of the same-sex penalty for the

U.S. are likely unaffected by the lack of such terms in the match surplus.

Table 12: Complementarities in sexual attraction (SOEP)

Partner i \ Partner i'	Low ($s_{i'i}=0$)	Intermediate ($s_{i'i}=1$)	High ($s_{i'i}=2$)
Low ($s_{ii'}=0$)	0.00	4.13	6.15
	(0.00)	(2.69)	(2.40)
Intermediate ($s_{ii'}=0.5$)		9.03	9.77
		(2.53)	(2.37)
High ($s_{ii'}=1$)			16.20
			(2.35)

Notes. Results obtained with SOEP data (2016 and 2019). The parameters correspond to those of function Φ_E, described in Section 3.2. In each cell (k,l), we report the match gains if $(s_{ii'}=k, s_{i'i}=l)$, with the gains for $(s_{ii'}=0, s_{i'i}=0)$ normalized to zero. Standard errors in parentheses. $N=116,612$ and $BIC=116,612$. The estimates of the male and female same-sex penalty are respectively -2.53 (0.41) and 1.20 (0.45).

5 Counterfactuals

5.1 Gender ratio increase

We now run a series of counterfactual simulations in order to understand how individuals adjust their mating strategies in response to changes in the surrounding environment. We start by studying how changes in the gender ratio affect the formation of same-sex households and the mating patterns of LGB individuals. Obviously, there will always be a vast majority of individuals who will not consider partners whose gender they are not attracted to. Yet, as we claim in our introduction, certain individuals - such as bisexual men and women - are expected to be particularly responsive to changes in the surrounding environment.

In Table 13, we increase the men-to-women gender ratio by 10%. With a larger supply of men, we should expect an increase in male same-sex relationships and a decrease in female same-sex relationships, if at least some individuals are ready to be matched with either of the two genders. Through the comparative statics of our model, we can quantify these total elasticities and show that a 10% increase in the men-to-women gender ratio leads to a 1.93% decrease in the probability that a woman is in a same-sex relationship, while it also leads to a 1.80% increase in the probability that a man is in a same-sex relationship. Since men's and women's responses are almost symmetric, the overall share of same-sex couples is basically unchanged in the counterfactual marriage market equilibrium.

We can also delve more into the comparative statics and look at how agents with different gender and sexual orientation respond to the gender-ratio increase on different margins. First, men attracted to women now face a much narrower market, and some will end up staying single. Hence, a man's odds of finding a partner decrease by almost 5% on average. However, the share of bisexual men in a relationship decreases more than for the rest of male population. As competition stiffens, bisexual men are more likely than straight men to stop looking for female partners, since, as we have seen in Section 4, their gains from different-sex live-in relationships are lower. On the other hand, they become more likely to match with male partners; among matched bisexual men, the share in a same-sex relationship increases by 10.66% in the counterfactual scenario. Interestingly, the mating patterns of gay men change little. Yet, we have seen that a non-negligible share of gay men do match with women; in the counterfactual equilibrium, this share shrinks, and gay men become on average slightly less likely to be matched (-0.91%). Among women, we observe almost symmetric changes, with their odds of finding a partner increasing by 4.39% on average. Bisexual women become both more likely to find a partner and, if matched, more likely to be in a different-sex relationship. More lesbian women consider matching with a male partner, which results in a small increase in their overall probability of being matched (0.33%).

Since the Black population has a strongly skewed gender ratio in the U.S. (Caucutt, Guner, and Rauh, 2018), we perform another counterfactual experiment where we increase the gender ratio in the Black population by 10%. The results are available in Appendix C and are summarized in Table 17. The gender-ratio elasticities of same-sex match probabilities are lower for Blacks relatively to the population; a 10% gender-ratio increase leads to a 0.51% decrease in the odds that a Black woman is in a same-sex relationship. This is due to both a lower share of bisexual individuals in the Black population and to adjustments along an additional margin, i.e., individuals can switch from intraracial to interracial relationships (and vice versa). In the counterfactual equilibrium, the gender-ratio increase primarily benefits Black straight women that were either single or in an interracial relationship by increasing their odds of finding a Black male partner.

5.2 Removing the same-sex penalty

In this Section, we perform a different type of counterfactual experiment where we leave the marginal distributions unchanged, but we remove the same-sex penalty in the match

Table 13: Change in overall gender ratio (+10%)

	Data (1)	Fitted model (2)	Counterfactual (3)	% change (4)
Gender ratio (all)	101.99	101.99	112.19	10.00
		(0.00)	(0.00)	(0.00)
Prob of match (women, all)	59.12	59.85	62.47	4.39
		(0.16)	(0.17)	(0.01)
Prob of same-sex match (women, all)	0.93	0.93	0.91	-1.93
		(0.04)	(0.04)	(0.11)
Prob of same-sex match (if matched, women, all)	1.58	1.55	1.46	-6.05
		(0.06)	(0.06)	(0.11)
Prob of match (bisexual women, all)	42.55	36.59	38.62	5.57
		(1.27)	(1.30)	(0.13)
Prob of same-sex match (if matched, bisexual women, all)	10.30	9.01	8.19	-9.10
		(0.47)	(0.43)	(0.08)
Prob of match (lesbian women, all)	50.57	50.81	50.98	0.33
		(1.41)	(1.39)	(0.06)
Prob of same-sex match (if matched, lesbian women, all)	87.94	85.99	84.66	-1.55
		(1.18)	(1.27)	(0.14)
Prob of match (men, all)	59.67	58.45	55.57	-4.93
		(0.16)	(0.15)	(0.01)
Prob of same-sex match (men, all)	0.70	0.68	0.69	1.80
		(0.03)	(0.03)	(0.12)
Prob of same-sex match (if matched, men, all)	1.17	1.17	1.25	7.09
		(0.06)	(0.06)	(0.12)
Prob of match (bisexual men, all)	31.69	44.41	42.11	-5.17
		(1.38)	(1.37)	(0.13)
Prob of same-sex match (if matched, bisexual men, all)	10.36	5.80	6.42	10.66
		(0.35)	(0.38)	(0.10)
Prob of match (gay men, all)	35.10	36.89	36.56	-0.91
		(1.28)	(1.29)	(0.11)
Prob of same-sex match (if matched, gay men, all)	84.74	78.99	80.75	2.23
		(1.67)	(1.56)	(0.19)
Share of same-sex couples (all)	1.37	1.36	1.35	-0.42
		(0.04)	(0.04)	(0.19)

Notes. Results obtained with 2013-2018 NHIS data. The fitted model is the one described in column (4) of Table 6. In the counterfactual, the match surplus parameters stays unchanged, while the gender ratio (ratio of the number of men to the number of women) increases by 10% without altering the composition in terms of observables (age, race, education) within each gender group. All quantities in the Table are expressed in percentage points. Standard errors in parentheses.

surplus. The model we take into consideration is always the one described in column (4) of Table 6, thus the male and female same-sex penalty are respectively -2.35 and -0.86. In the counterfactual scenario, we set both parameters equal to zero. In this way, we provide a more intuitive quantitative assessment of the relevance of the same-sex penalty. Table 14 shows that, absent this penalty, the share of same-sex couples would increase by about 50%, from 1.36% to 2.05%.[39] The increase is more substantial among men, who experience a stronger penalty and whose odds of being in a same-sex relationship almost

[39]The same counterfactual exercise performed with the nested logit version of the same model, whose estimated parameters are presented in column (6) of Table 6, yield qualitatively similar findings. However, in the nested logit model, due to random taste shocks being correlated for partners of the same gender, individuals differ in the degree of attraction to same-sex partners even after conditioning on their LGB identity. Hence, when removing the same-sex penalty, the share of same-sex couples increases by about 125%, reaching 3.14% of all couples.

double, from 0.68% to 1.29%. Conversely, the increase is only moderate among women, whose odds of being in a same-sex relationship increase by 23.27%, from 1.55% to 1.91%.

Looking at Table 14, we can discuss how people respond to the same-sex penalty removal conditional on their sexual orientation. Interestingly, the odds of finding a partner increase only slightly for both bisexual men and women. Yet, all bisexual individuals become much more likely to be in same-sex relationships if they do find a partner. In contrast, the odds of being in a relationship increase by 15.95% for lesbian women and by 58.67% for gay men. Moreover, the share of gay men and lesbian women who match with a partner of the opposite gender is also substantially lower.

Table 14: Removing the same-sex penalty

	Data (1)	Fitted model (2)	Counterfactual (3)	% change (4)
Prob of match (women, all)	59.12	59.85	59.98	0.22
		(0.16)	(0.16)	(0.06)
Prob of same-sex match (women, all)	0.93	0.93	1.15	23.54
		(0.04)	(0.01)	(5.18)
Prob of same-sex match (if matched, women, all)	1.58	1.55	1.91	23.27
		(0.06)	(0.02)	(5.10)
Prob of match (bisexual women, all)	42.55	36.59	37.39	2.18
		(1.31)	(1.34)	(0.37)
Prob of same-sex match (if matched, bisexual women, all)	10.30	9.01	12.04	33.62
		(0.48)	(0.06)	(7.36)
Prob of match (lesbian women, all)	50.57	50.81	58.91	15.95
		(1.42)	(0.35)	(3.33)
Prob of same-sex match (if matched, lesbian women, all)	87.94	85.99	89.45	4.03
		(1.23)	(0.66)	(0.98)
Prob of match (men, all)	59.67	58.45	58.97	0.88
		(0.16)	(0.16)	(0.05)
Prob of same-sex match (men, all)	0.70	0.68	1.29	88.37
		(0.03)	(0.01)	(9.28)
Prob of same-sex match (if matched, men, all)	1.17	1.17	2.18	86.72
		(0.06)	(0.02)	(9.11)
Prob of match (bisexual men, all)	31.69	44.41	46.62	4.99
		(1.43)	(1.45)	(0.26)
Prob of same-sex match (if matched, bisexual men, all)	10.36	5.80	13.30	129.37
		(0.34)	(0.05)	(13.52)
Prob of match (gay men, all)	35.10	36.89	58.54	58.67
		(1.27)	(0.33)	(5.34)
Prob of same-sex match (if matched, gay men, all)	84.74	78.99	90.27	14.28
		(1.67)	(0.61)	(1.88)
Share of same-sex couples (all)	1.37	1.36	2.05	50.51
		(0.04)	(0.02)	(4.83)

Notes. Results obtained with 2013-2018 NHIS data. The fitted model is the one described in column (4) of Table 6. In the counterfactual, the marginal distributions stay unchanged, while the same-sex penalty in the match surplus disappear for both male and female couples. All quantities in the Table are expressed in percentage points. Standard errors in parentheses.

5.3 Increasing LGB population

Finally, we predict how mating patterns would change if the LGB population kept growing in number. Recent surveys that drew attention in the press suggested that the share of LGB individuals in younger cohorts is substantially larger than in the past (Pew Research

Center, 2023; Gallup, 2022). This is also confirmed by the 2019-2022 NHIS data, which show that, among male respondents aged between 21 and 30, 3.22% identify as gay and 2.61% as bisexual, while among female respondents in the same age group, 2.53% identify as lesbian and 8.85% as bisexual. In this Section, we predict how mating patterns would change over the next thirty years if all new cohorts had the same LGB composition as the youth in the 2019-2022 NHIS data. In doing so, we abstract from changes in other demographics (e.g., in the racial and educational composition of the population) and in preferences (e.g., the same-sex penalty stays unchanged).

Table 15 shows that, in this counterfactual scenario, the share of gay men and lesbian women in the adult population (aged between 21 and 50) would increase by about 50%, whereas the share of bisexual individuals would triple. As in the previous counterfactual simulations, our benchmark is the model presented in column (4) of Table 6, estimated with 2013-2018 NHIS data. In the counterfactual equilibrium, preferences stay unchanged and remain anchored to the benchmark. When the LGB population expands, the share of same-sex couples increases by about 73%, from 1.36% to 2.35% of all couples thanks to the increased opportunities of finding partners who reciprocate same-gender attraction. This increase is comparable in size to the one we obtained when removing the same-sex penalty in Section 5.2. However, removing the same-sex penalty leads to an overall increase in the share of partnered individuals, while the LGB population expansion leads to an overall decrease. Due to the presence of the same-sex penalty, the overall fraction of partnered individuals in equilibrium declines by almost 3% when the LGB population increases.

Interestingly, Table 15 also shows that, when the LGB population increases, the odds of finding a partner decrease for bisexual individuals. If one one hand it becomes easier for them to match with same-sex partners, the market for opposite-sex partners shrinks. Since bisexual individuals experience lower gains from different-sex relationships relatively to straight individuals, they are rapidly crowded out of the different-sex market. The effect of an increase in the LGB population on the mating patterns of gay men and lesbian women is instead more straightforward. In the counterfactual scenario, the odds of finding a partner increase by 2.85% for lesbian women and by 4.10% for gay men, while cases where gay individuals match with opposite-sex partners become less frequent.

Table 15: Increasing LGB population

	Data (1)	Fitted model (2)	Counterfactual (3)	% change (4)
% gay men (all)	2.09	2.09	3.22	54.41
		(0.00)	(0.00)	(0.00)
% bisexual men (all)	0.65	0.65	2.61	299.38
		(0.00)	(0.00)	(0.00)
% lesbian women (all)	1.75	1.75	2.53	44.55
		(0.00)	(0.00)	(0.00)
% bisexual women (all)	2.03	2.03	8.85	336.63
		(0.00)	(0.00)	(0.00)
Prob of match (women, all)	59.12	59.85	58.19	-2.78
		(0.14)	(0.15)	(0.11)
Prob of same-sex match (women, all)	0.93	0.93	1.53	65.28
		(0.03)	(0.06)	(1.14)
Prob of same-sex match (if matched, women, all)	1.58	1.55	2.63	69.99
		(0.05)	(0.09)	(0.99)
Prob of match (bisexual women, all)	42.55	36.59	24.62	-32.72
		(1.24)	(1.03)	(0.56)
Prob of same-sex match (if matched, bisexual women, all)	10.30	9.01	11.24	24.75
		(0.43)	(0.53)	(0.46)
Prob of match (lesbian women, all)	50.57	50.81	52.26	2.85
		(1.25)	(1.29)	(0.23)
Prob of same-sex match (if matched, lesbian women, all)	87.94	85.99	88.69	3.14
		(1.16)	(0.98)	(0.26)
Prob of match (men, all)	59.67	58.45	56.72	-2.96
		(0.14)	(0.15)	(0.10)
Prob of same-sex match (men, all)	0.70	0.68	1.17	71.73
		(0.03)	(0.06)	(0.46)
Prob of same-sex match (if matched, men, all)	1.17	1.17	2.07	76.98
		(0.06)	(0.10)	(0.43)
Prob of match (bisexual men, all)	31.69	44.41	29.39	-33.81
		(1.36)	(1.19)	(0.66)
Prob of same-sex match (if matched, bisexual men, all)	10.36	5.80	7.88	35.81
		(0.36)	(0.47)	(0.36)
Prob of match (gay men, all)	35.10	36.89	38.40	4.10
		(1.32)	(1.42)	(0.36)
Prob of same-sex match (if matched, gay men, all)	84.74	78.99	83.77	6.05
		(1.75)	(1.42)	(0.54)
Share of same-sex couples (all)	1.37	1.36	2.35	73.01
		(0.04)	(0.07)	(0.67)

Notes. Results obtained with 2013-2018 NHIS data. The fitted model is the one described in column (4) of Table 6. In the counterfactual, the LGB population grows in number, as shown in the first four lines of the Table. On the other hand, mating preferences and other demographics are held constant. All quantities in the Table are expressed in percentage points. Standard errors in parentheses.

6 Conclusion

Recent large-scale surveys allow us to describe the mating patterns of LGB individuals. In spite of the increase in the share of people who identify as bisexual or report sexual attraction to both genders, the share of LGB individuals in a live-in relationship, be it a legal marriage or an unmarried cohabitation, is still considerably lower than that of straight individuals. Moreover, about 90% of partnered bisexual men and women in the U.S. opt for a partner of the opposite gender, while also about 10% of partnered gay men and lesbian women live in different-sex couples. This suggests that, on marriage markets, LGB individuals face important trade-offs involving their partner's gender, and may eventually choose to remain single or match with a partner of the opposite gender due to the extra hurdles that same-sex couples have to face (e.g., social stigma, lack of legal recognition for same-sex unions, difficulties related to childbearing and adoption).

In this paper, we develop a comprehensive marriage market equilibrium framework to study assortative mating patterns when the partner's gender choice is endogenous. In the model, agents look for a partner in a competitive environment and can transfer utility to each other upon a match, as in Choo and Siow (2006) and Galichon and Salanié (2022). Yet, contrarily to these models, ours is unipartite, i.e., any pair of agents on the market can match. Hence, individuals with different sexual orientations choose their partners among both men and women who also differ along additional dimensions (age, race, and education), and they self-select into same-sex and different-sex relationships.

When we take the model to the data, we show that same-sex couples experience lower gains from live-in relationships. The same-sex penalty is large for men in both Germany and the U.S., although it has been decreasing in recent years. In comparison, the female same-sex penalty is smaller overall, and it is even null in some region of the U.S. These spatial and temporal differences in the same-sex penalty are broadly consistent with patterns of acceptance of homosexuality as documented in the World Values Survey. Moreover, using directly observed family outcomes and individual well-being measures from the NHIS, we show that differences in match gains due to couples' gender and sexual orientation composition correlate with fertility, legal marriage, physical and mental health. Finally, through a series of counterfactual experiments, we show that, absent the same-sex penalty, the share of same-sex couples in the U.S. would increase by about 50%, from 1.36% to 2.05% of all couples.

References

Aksoy, C. G., C. S. Carpenter, R. De Haas, and K. D. Tran (2020): "Do laws shape attitudes? Evidence from same-sex relationship recognition policies in Europe," *European Economic Review*, 124, 103399.

Badgett, M., C. S. Carpenter, and D. Sansone (2021): "LGBTQ economics," *Journal of Economic Perspectives*, 35(2), 141–70.

Badgett, M. L., C. S. Carpenter, M. J. Lee, and D. Sansone (2024): "A review of the effects of legal access to same-sex marriage," *Journal of Policy Analysis and Management.*

Baranov, V., R. De Haas, and P. Grosjean (2020): "Queens of the desert: Convictism and marital attitudes across Australia," in *AEA Papers and Proceedings*, vol. 110, pp. 457–62.

Baranov, V., R. De Haas, and P. A. Grosjean (2018): "Men. Roots and consequences of masculinity norms," .

Bau, N., and R. Fernández (2023): "Culture and the Family," in *Handbook of the Economics of the Family*, vol. 1, pp. 1–48. Elsevier.

Blewett, L. A., J. A. Rivera Drew, M. L. King, K. C. Williams, N. Del Ponte, and C. Pat (2022): "IPUMS Health Surveys: National Health Interview Survey, Version 7.2 [dataset]. Minneapolis, MN: IPUMS," https://doi.org/10.18128/D070.V7.2.

Brodeur, A., and J. Haddad (2021): "Institutions, attitudes and LGBT: Evidence from the Gold Rush," *Journal of Economic Behavior & Organization*, 187, 92–110.

Carpenter, C., and G. J. Gates (2008): "Gay and lesbian partnership: Evidence from California," *Demography*, 45(3), 573–590.

Carpenter, C. S. (2020): "The direct effects of legal same-sex marriage in the United States: evidence from Massachusetts," *Demography*, 57(5), 1787–1808.

Carpenter, C. S., S. T. Eppink, G. Gonzales, and T. McKay (2021): "Effects of access to legal same-sex marriage on marriage and health," *Journal of Policy Analysis and Management*, 40(2), 376–411.

Caucutt, E. M., N. Guner, and C. Rauh (2018): "Is marriage for White people? Incarceration, unemployment, and the racial marriage divide," .

CDC National Center for Health Statistics (NCHS) (2020): "2006-10, 2011-15, 2015-17, and 2017-2019 National Survey of Family Growth Public-Use Data and Documentation," https://www.cdc.gov/nchs/nsfg/index.htm.

Chen, S., and J. C. van Ours (2020): "Symbolism matters: The effect of same-sex marriage legalization on partnership stability," *Journal of Economic Behavior & Organization*, 178, 44–58.

Chen, S., and J. C. Van Ours (2022): "Mental health effects of same-sex marriage legalization," *Health Economics*, 31(1), 42–56.

Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009): "Investment in schooling and the marriage market," *American Economic Review*, 99(5), 1689–1713.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012): "Fatter attraction: anthropometric and socioeconomic matching on the marriage market," *Journal of Political Economy*, 120(4), 659–695.

——— (2016): "Black–White marital matching: race, anthropometrics, and socioeconomics," *Journal of Demographic Economics*, 82(4), 399–421.

Chiappori, P.-A., B. Salanié, and A. Galichon (2019): "On human capital and team stability," *Journal of Human Capital*, 13(2), 236–259.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017): "Partner choice, investment in children, and the marital college premium," *American Economic Review*, 107(8), 2109–67.

Choo, E., and A. Siow (2006): "Who marries whom and why," *Journal of political Economy*, 114(1), 175–201.

Ciscato, E., A. Galichon, and M. Goussé (2020): "Like attract like? A structural comparison of homogamy across same-sex and different-sex households," *Journal of Political Economy*, 128(2), 740–781.

Dahlhamer, J. M., A. M. Galinsky, S. S. Joestl, and B. W. Ward (2014): "Sexual orientation in the 2013 National Health Interview Survey: a quality assessment," .

Delhommer, S. M., and D. S. Hamermesh (2021): "Same-Sex Couples and the Gains to Marriage: The Importance of the Legal Environment," *Journal of Policy Analysis and Management*, 40(4), 1120–1139.

DIW (2021): "Socio-Economic Panel (SOEP), data for years 1984-2019, SOEP-Core v36, 2021," `https://doi:10.5684/soep.core.v36eu`.

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

Fernández, R., S. Parsa, and M. Viarengo (2019): "Coming out in America: AIDS, politics, and cultural change," Discussion paper, National Bureau of Economic Research.

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

Galichon, A., S. D. Kominers, and S. Weber (2019): "Costly concessions: An empirical framework for matching with imperfectly transferable utility," *Journal of Political Economy*, 127(6), 2875–2925.

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

Gallup (2022): "LGBT Identification in U.S. Ticks Up to 7.1%," `https://news.gallup.com/poll/389792/lgbt-identification-ticks-up.aspx`, published on February 17, 2022, visited on December 19, 2023.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019): "The German socio-economic panel (SOEP)," *Jahrbücher für Nationalökonomie und Statistik*, 239(2), 345–360.

Hansen, M. E., M. E. Martell, and L. Roncolato (2020): "A labor of love: The impact of same-sex marriage on labor supply," *Review of Economics of the Household*, 18(2), 265–283.

Hu, Y., and N. Denier (2023): "Sexual Orientation Identity Mobility in the United Kingdom: A Research Note," *Demography*, 60(3), 659–673.

Kinsey, A. C., W. B. Pomeroy, and C. E. Martin (1948): *Sexual behavior in the human male.* Indiana University Press.

Kinsey, A. C., W. B. Pomeroy, C. E. Martin, and P. H. Gebhard (1953): "Sexual behavior in the human female.," .

Klein, F. (1978): *The bisexual option.* Haworth Press.

MacDowall, L. (2009): "Historicising contemporary bisexuality," *Journal of Bisexuality*, 9(1), 3–15.

Martin, J., and Z. Rodriguez (2022): "The effect of same-sex marriage legalization on adoptions and family formation in the US," .

Masson-Makdissi, E. (2023): "Same-Sex and Different-Sex Relationships: A Comprehensive Matching Model," .

Morgan, R. E., C. Dragon, G. Daus, J. Holzberg, R. Kaplan, H. Menne, and M. Spiegelman (2020): "Updates on terminology of sexual orientation and gender identity survey measures," *Federal Committee on Statistical Methodology.*

Oreffice, S. (2016): "Sexual orientation and marriage," *Studies of Applied Economics*, 34(1), 7–34.

Pew Research Center (2023): "5 key findings about LGBTQ+ Americans," https://www.pewresearch.org/short-reads/2023/06/23/5-key-findings-about-lgbtq-americans, published on June 23, 2023, visited on December 19, 2023.

Sansone, D. (2019): "Pink work: Same-sex marriage, employment and discrimination," *Journal of Public Economics*, 180, 104086.

Train, K. E. (2009): *Discrete choice methods with simulation.* Cambridge university press.

Weinrich, J. D., P. J. Snyder, R. C. Pillard, I. Grant, D. L. Jacobson, S. R. Robinson, and J. A. McCutchan (1993): "A factor analysis of the Klein Sexual Orientation Grid in two disparate samples," *Archives of Sexual Behavior*, 22(2), 157–168.

For Online publication for
"Matching on Gender and Sexual Orientation",
Edoardo Ciscato and Marion Goussé

A Data appendix

In addition to sexual orientation, whose measurement is extensively described in Section 2.2, our main analysis with NHIS data uses information on the sex, age, race, and education of respondents (and, when present, their partners'). These variables are constructed as follows:

- **Sex**: It indicates whether the person is male or female. Until 2018, due to the survey design, all respondents are classified as either male or female. In the 2019-2022 period, a negligible fraction of respondents (less than 0.01%) refused to answer or did not know, while all others are classified as either male or female.
- **Age**: Starting from self-reported age, we build the following age categories: 21-25, 26-30, 31-35, 36-40, 41-45, and 46-50.
- **Race**: Based on their self-reported race, individuals are initially divided into three groups: White only; Black or African American only; a residual group including American Indian only, Alaska Native only, Asian only, or Multiple Race. In a second step, we reassign all those who identify as Hispanic to a separate group, so that our final variable has four categories.
- **Education**: Based on their self-reported educational attainment, individuals are divided into five groups: no high school diploma; high school diploma or equivalent; associate degree or attended college without graduating; college degree; graduate degree.

Our analysis with SOEP data uses similar definitions, although race is not present in the analysis. Moreover, we only consider three educational categories: high school diploma or less; post-secondary non-tertiary degree; bachelor degree or higher.

The NHIS also contains information about the family composition and the respondent's well-being. We use these variables in Section 4.3, and study their correlation patterns with the estimated match surplus. Please note that these variables are only available for respondents, but not for their partners. They include:

- **Number of children**: Number of own children present in the household.
- **Legally married**: Dummy variable indicating if the respondent is married to her/his partner.
- **Mental health**: It is measured as the average of six indicators of stress and despair described below: 'Effort', 'Hopeless', 'Nervous', 'Restless', 'Sad', and 'Worthless'.

- **Physical health**: It is measured as the first principal component of four variables described below: 'Physical activity', 'Smoker', 'Self-assessed health', and 'BMI'.

Respondents were asked how often they express the different feelings stated below in the 30 days before the interview. They could choose between five ordered answers, from "None of the time" to "All of the time".

- **Effort**: "Felt that everything is an effort".
- **Hopeless**: "How often they felt hopeless".
- **Nervous**: "How often they felt nervous" .
- **Restless**: "How often they felt restless".
- **Sad**: "How often they felt sad".
- **Worthless**: "How often they felt worthless".

Other variables include

- **Physical activity**: Frequency of moderate, vigorous or strengthening physical activities per week in hours.
- **Smoking**: Dummy variable indicating if the respondent is an everyday or someday smoker.
- **Self-assessed health**: Respondents were asked to rate their general health on a five-point Likert scale, ranging from "Poor" to "Excellent".
- **Body Mass Index (BMI)**: It is measured using self-reported height and weight.
- **Worried**: "Feelings interfered with their life" in the 30 days before the interview. They could choose between four ordered answers, from "A lot" to "Not at all".
- **Hours of sleep**: Average number of hours of sleep per day.
- **Quality of sleep**: It is measured as an average of four indicators of quality of sleep: number of times having trouble falling asleep in previous week, number of times having trouble staying asleep in previous week, number of times taking medication for sleep in previous week, days respondent woke up feeling rested. The sign of these variables is switched when needed, so that a higher value corresponds to better quality.

Table 16 presents descriptive statistics for these variables on individuals in couple, by sexual orientation and gender.

Table 16: Distribution of well-being variables among couples by sexual orientation

	Men				Women			
	Straight	Bisexual	Gay	Total	Straight	Bisexual	Lesbian	Total
Age	37.2	34.7	37.5	37.2	36.6	31.6	36.1	36.6
Legally married	0.82	0.64	0.40	0.82	0.84	0.54	0.51	0.83
Number of own children	1.40	0.56	0.26	1.38	1.48	0.84	0.75	1.46
Mental health [0 ... 4]	3.65	3.25	3.31	3.64	3.56	3.02	3.50	3.55
Physical health [-3 ... 2]	-0.011	-0.63	0.15	-0.011	0.14	-0.73	-0.21	0.12
Usual hours sleep per day	6.86	7.18	7.00	6.87	7.04	6.87	6.94	7.03
Sleep quality [0 ... 7]	5.32	4.87	5.14	5.31	5.17	4.61	4.99	5.15
Physical activity (hours per week)	1.49	1.35	1.59	1.49	1.42	1.45	1.35	1.42
Smoker	0.15	0.24	0.17	0.15	0.12	0.22	0.14	0.12
Self reported health status [1 ... 5]	1.93	2.37	1.87	1.93	1.94	2.43	2.06	1.95
Body Mass Index	28.4	30.3	27.2	28.4	27.0	31.5	29.4	27.1
Felt everything an effort, past 30 days [0 4]	0.47	1.04	0.88	0.48	0.53	1.25	0.67	0.55
How often felt hopeless, past 30 days [0 4]	0.15	0.36	0.36	0.15	0.20	0.54	0.30	0.20
How often felt nervous, past 30 days [0 4]	0.54	1.20	1.09	0.55	0.76	1.49	0.70	0.77
How often felt restless, past 30 days [0 4]	0.62	0.99	1.07	0.63	0.68	1.32	0.73	0.69
How often felt sad, past 30 days [0 4]	0.22	0.62	0.49	0.23	0.33	0.76	0.39	0.34
How often felt worthless, past 30 days [0 4]	0.11	0.29	0.23	0.11	0.15	0.50	0.20	0.15
How often feel worried, nervous, or anxious [0 4]	3.93	3.24	3.33	3.92	3.59	2.58	3.34	3.57

Notes. 2013-2018 NHIS sample restricted to 21-50 years-old respondents living in couple. The Table shows sample means for each variable, conditional on gender and sexual orientation.

B Technical appendix

B.1 Nested logit

Following Galichon and Salanié (2022), we also estimate a version of the model with a nested logit structure where potential partners of different types x' belong to different nests. Within each nest $\mathcal{G}$, the random taste shocks $(\varepsilon_{ix'})_{x'\in\mathcal{G}}$ are allowed to be correlated. In practice, partners are grouped into different nests based on their gender and racial background.In the presence of within-nest correlation, Galichon and Salanié (2022) show that the matching function becomes

$$\mu(x,x') = \exp\left(\frac{\Phi(x,x')}{\rho_{\mathcal{G}_x}+\rho_{\mathcal{G}_{x'}}}\right)\left(\mu(x,0)\mu(x',0)\right)^{\frac{1}{\rho_{\mathcal{G}_x}+\rho_{\mathcal{G}_{x'}}}}\mu(x,\mathcal{G}_{x'})^{\frac{\rho_{\mathcal{G}_{x'}}-1}{\rho_{\mathcal{G}_x}+\rho_{\mathcal{G}_{x'}}}}\mu(x',\mathcal{G}_x)^{\frac{\rho_{\mathcal{G}_x}-1}{\rho_{\mathcal{G}_x}+\rho_{\mathcal{G}_{x'}}}} \tag{B.1}$$

where $\mu(x,\mathcal{G}_{x'}) = \sum_{x''\in\mathcal{G}_{x'}}\mu(x,x'')$ indicates the probability that an individual of type x matches with a partner in nest $\mathcal{G}_{x'}$. Moreover, $\rho_{\mathcal{G}} \in [0,1]$ measures the *degree of independence* between random taste shocks within a nest $\mathcal{G}$ (Train, 2009). When both $\rho_{\mathcal{G}_x} = 1$ and $\rho_{\mathcal{G}_{x'}} = 1$, then the matching function (B.1) boils down to its simpler i.i.d. multinomial logit version (3.4).

To derive (B.1), let $x \in \mathcal{X}$ be the agent's type and $\mathcal{G}_x = \mathcal{G}$ her nest. Her optimal mate problem (3.2) now gives the following probability of choosing type x' within nest $\mathcal{G}_{x'} = \mathcal{G}'$

$$\frac{\mu(x,x')}{f(x)} = \frac{\exp(U_{xx'}/\rho_{\mathcal{G}'})}{\sum_{x'\in\mathcal{G}'}\exp(U_{xx'}/\rho_{\mathcal{G}'})}\frac{\mu(x,\mathcal{G}')}{f(x)}, \tag{B.2}$$

where $\mu(x, \mathcal{G}')/f(x)$. Moreover, the probability that x chooses from within nest $\mathcal{G}'$ is given by

$$\frac{\mu(x, \mathcal{G}')}{f(x)} = \frac{\left(\sum_{x' \in \mathcal{G}'} \exp(U_{xx'}/\rho_{\mathcal{G}'})\right)^{\rho_{\mathcal{G}'}}}{1 + \sum_{\mathcal{G}'} \left(\sum_{x' \in \mathcal{G}'} \exp(U_{xx'}/\rho_{\mathcal{G}'})\right)^{\rho_{\mathcal{G}'}}}. \tag{B.3}$$

Using $U_{xx'} + U_{x'x} = \Phi_{xx'} = \Phi_{x'x}$, we can derive the matching function (B.1). Let $\kappa_{xx'} = \exp(\Phi_{xx'}/(\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}))$ and rewrite

$$\mu(x, \mathcal{G}') = \mu(x, 0)^{\frac{1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \mu(x, \mathcal{G}')^{\frac{\rho_{\mathcal{G}'} - 1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \sum_{x' \in \mathcal{G}'} \kappa_{xx'} \mu(x', 0)^{\frac{1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \mu(x', \mathcal{G})^{\frac{\rho_{\mathcal{G}} - 1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \tag{B.4}$$

$$= \mu(x, 0)^{\frac{1}{\rho_{\mathcal{G}} + 1}} \left[\sum_{x' \in \mathcal{G}'} \kappa_{xx'} \mu(x', 0)^{\frac{1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \mu(x', \mathcal{G})^{\frac{\rho_{\mathcal{G}} - 1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \right]^{\frac{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}{\rho_{\mathcal{G}} + 1}}. \tag{B.5}$$

Due to the symmetry constraint in (3.1), the feasibility constraints are given by

$$f(x) = \mu(x, 0) + 2 \sum_{x'} \mu(x, x') = \mu(x, 0) + 2 \sum_{\mathcal{G}'} \mu(x, \mathcal{G}') \tag{B.6}$$

$$= \mu(x, 0) + 2\mu(x, 0)^{\frac{1}{\rho_{\mathcal{G}} + 1}} \sum_{\mathcal{G}'} \left[\sum_{x' \in \mathcal{G}'} \kappa_{xx'} \mu(x', 0)^{\frac{1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \mu(x', \mathcal{G})^{\frac{\rho_{\mathcal{G}} - 1}{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}} \right]^{\frac{\rho_{\mathcal{G}} + \rho_{\mathcal{G}'}}{\rho_{\mathcal{G}} + 1}}. \tag{B.7}$$

This set of equations, one for every type x, can be solved for $\{\mu(x, 0)\}_{x \in \mathcal{X}}$ using a variant of the Iterative Projection Fitting Procedure (IPFP) for given parameters κ and ρ. The IPFP algorithm and its convergence properties are discussed in Galichon and Salanié (2022).

B.2 Latent sexual orientation

Let z_i be a vector of respondent i's characteristics observed both in the NHIS and the NSFG. The vector z_i includes the respondent's LGB identity, age, education, race, employment, place of birth (in the U.S. or outside), time period of interview, and having kids (only included for women). Using NSFG data, we fit a series of ordered probit models in order to estimate the conditional probabilities $\Pr\{o_i = k | z_i\}$, where $k \in S = \{0, 0.25, 0.5, 0.75, 1\}$ represents the degree of attraction to women measured on a five-level scale through the question "Which gender are you the most attracted to?". We estimate a separate order probit model for each combination of gender (men and women) and LGB category (gay/lesbian, bisexual, straight, something else, other response).

Now let us deviate slightly from the notation in the main text and name $\tilde{x}_i = (x_i, o_i)$ the agent's type in the model. The type $\tilde{x}_i$ is *latent* in that it includes her/his sexual orientation o_i, now treated as unobserved to the econometrician, and her/his other traits x_i, still observed. When we estimate the model, we must average the likelihood contribution of an individual with observed traits z_i across all possible levels of o_i, using weights $\Pr\{o_i = k | z_i\}$. For the sake of clarity, we ignore the fact that some partners' traits may be unobserved, an issue we discussed in Section 3.3. The log-likelihood function writes:

$$
\begin{aligned}
\mathcal{L}(\lambda) =& \frac{1}{N_c} \sum_{i=1}^{N_c} \sum_{k \in S} \sum_{l \in S} \Pr\{o_i = k | z_i\} \Pr\{o_i' = l | z_i'\} \log \frac{\mu^\lambda((x_i, k), (x_i', l))}{H^\lambda} + \\
&+ \frac{1}{N_s} \sum_{i=1}^{N_s} \sum_{k \in S} \Pr\{o_i = k | z_i\} \log \frac{\mu^\lambda((x_i, k), 0)}{H^\lambda}.
\end{aligned}
\tag{B.8}
$$

C Gender ratio increase in the Black population

We perform a similar counterfactual experiment but we focus on the Black population, whose gender ratio in the data is strongly imbalanced due to the widely documented scarcity of adult Black men (Caucutt, Guner, and Rauh, 2018). In the counterfactual equilibrium, the men-to-women gender ratio in the Black population is 10% higher than what observed in the data. The comparative statics is qualitatively similar to the previous Section, although now we can also look at an additional margin, namely how the odds of being in an interracial relationship change. Table 17 shows that the gender-ratio elasticity of a Black woman's odds of being in a same-sex relationship is lower than in the overall population; a 10% increase in the men-to-women gender ratio in the Black population leads to a mere 0.51% decrease in the probability that a Black woman is in a same-sex relationship. This lower elasticity is due to a lower share of bisexual individuals in the Black population (see Table 24) and to part of the action happening along the intra- vs interracial margin. On the other hand, we do not observe a statistically significant change in a Black man's odds of being in a same-sex relationship.

Table 17 also shows that the 10% gender-ratio increase in the Black population leads to a 3.95% increase in a Black woman's odds of finding a partner. Black bisexual women have a particular hard time finding a partner, but the gender-ratio increase improves their chances by 4.04%, thanks to an increased probability of finding a male partner. A small share of Black lesbian women now opt for a male partner. Since the marriage market is clustered by race - due to both a pronounced preference for racial homogamy and an uneven spatial distribution of race groups - many Black single women are now able to find a same-race partner following the increase in the supply of Black men. On the other hand, Black men now face a narrower market, and their odds of finding a partner decrease by 4.05%. Black bisexual men, who are much more likely to match with female rather than male partners, particularly suffer from this increased competition; while some of them succeed in finding a male partner, their overall odds of finding a partner decrease by 4.57%.

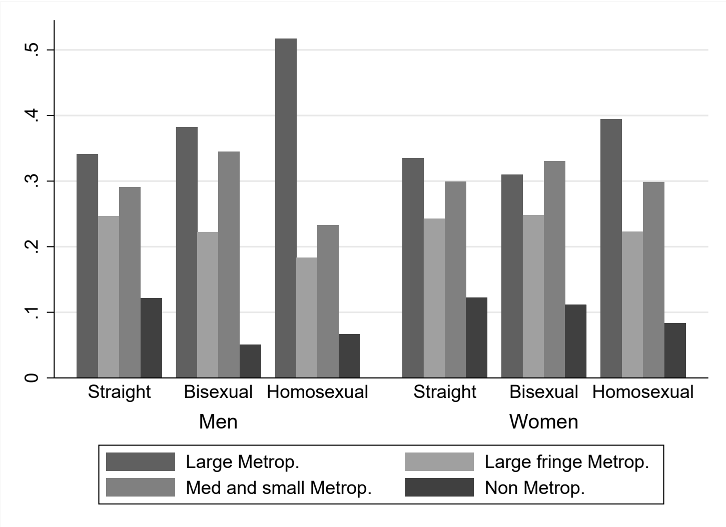
Table 17: Change in Black population gender ratio (+10%)

	Data (1)	Fitted model (2)	Counterfactual (3)	% change (4)
Gender ratio (Black)	86.11	86.11	94.72	10.00
		(0.00)	(0.00)	(0.00)
Prob of match (women, Black)	30.08	35.77	37.18	3.95
		(0.43)	(0.45)	(0.03)
Prob of same-sex match (women, Black)	0.79	0.64	0.64	-0.51
		(0.03)	(0.03)	(0.03)
Prob of same-sex match (if matched, women, Black)	2.64	1.79	1.71	-4.29
		(0.09)	(0.09)	(0.04)
Prob of match (bisexual women, Black)	16.93	18.43	19.17	4.04
		(0.86)	(0.88)	(0.06)
Prob of same-sex match (if matched, bisexual women, Black)	5.78	10.54	10.02	-4.92
		(0.59)	(0.56)	(0.06)
Prob of match (lesbian women, Black)	33.04	28.32	28.42	0.36
		(1.28)	(1.28)	(0.04)
Prob of same-sex match (if matched, lesbian women, Black)	85.85	88.08	87.47	-0.69
		(1.06)	(1.11)	(0.06)
Prob of same-race match (women, Black)	27.04	29.11	30.61	5.17
		(0.41)	(0.43)	(0.01)
Prob of same-race match (if matched, women, Black)	89.92	81.38	82.33	1.17
		(0.46)	(0.44)	(0.03)
Prob of match (men, Black)	46.50	39.75	38.14	-4.05
		(0.48)	(0.46)	(0.01)
Prob of same-sex match (men, Black)	0.40	0.30	0.30	-0.19
		(0.02)	(0.02)	(0.05)
Prob of same-sex match (if matched, men, Black)	0.86	0.75	0.78	4.02
		(0.04)	(0.05)	(0.05)
Prob of match (bisexual men, Black)	2.47	23.92	22.83	-4.57
		(1.08)	(1.04)	(0.05)
Prob of same-sex match (if matched, bisexual men, Black)	0.00	4.52	4.75	5.20
		(0.28)	(0.30)	(0.06)
Prob of match (gay men, Black)	17.33	16.12	15.84	-1.70
		(0.76)	(0.76)	(0.09)
Prob of same-sex match (if matched, gay men, Black)	82.02	74.34	75.35	1.37
		(1.89)	(1.84)	(0.10)
Prob of same-race match (men, Black)	35.50	33.40	31.98	-4.26
		(0.47)	(0.45)	(0.01)
Prob of same-race match (if matched, men, Black)	76.35	84.03	83.85	-0.21
		(0.41)	(0.41)	(0.01)
Share of same-sex couples (Black)	1.62	1.28	1.25	-2.19
		(0.05)	(0.05)	(0.16)

Notes. Results obtained with 2013-2018 NHIS data. The fitted model is the one described in column (4) of Table 6. In the counterfactual, the match surplus parameters stays unchanged, while the gender ratio in the Black population (ratio of the number of Black men to the number of Black women) increases by 10% without altering the composition in terms of observables. All quantities in the Table are expressed in percentage points. Standard errors in parentheses.

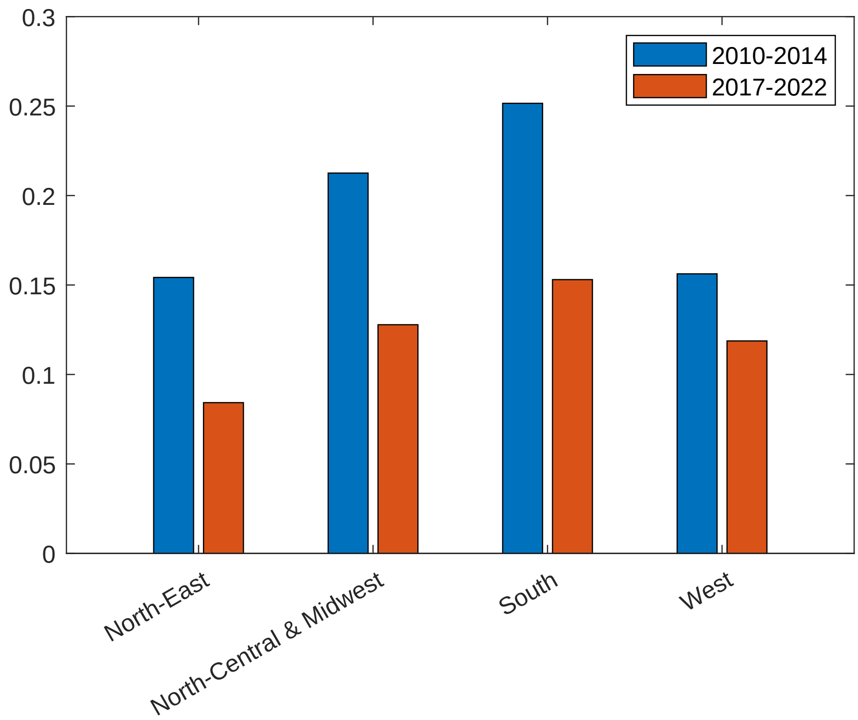
D Additional Figures

Figure 5: Distribution of individual by sexual orientation and type of areas



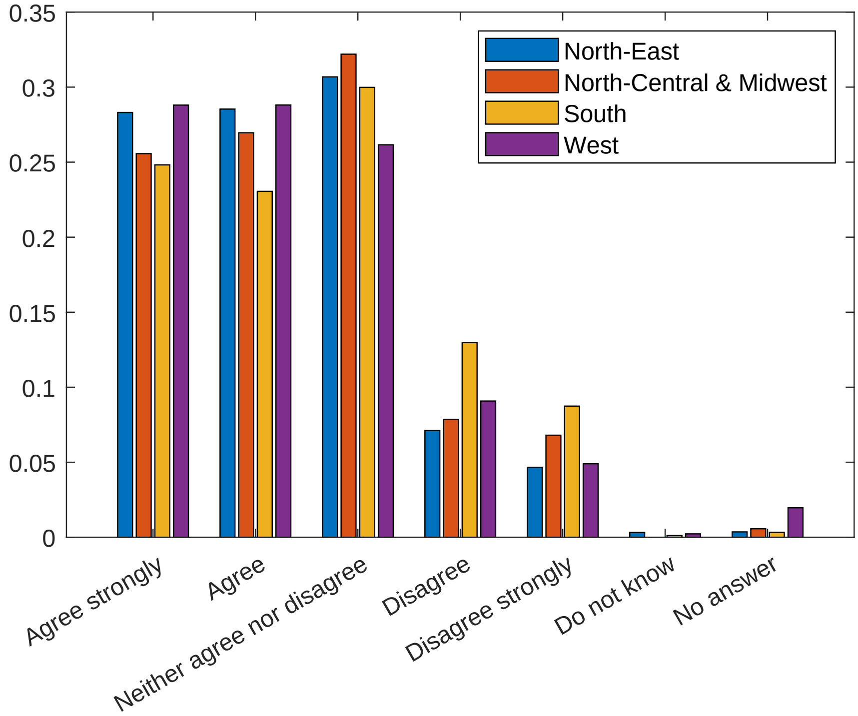
Notes. NHIS 2019-2022. For each of the six sexual orientation groups, we plot the probability of living in different types of metropolitan areas. Population restricted to 21-50 years-old individuals. Weighted results. Exact numbers are displayed in Table 24

Figure 6: % of respondents who would not like homosexuals as neighbors by region and year



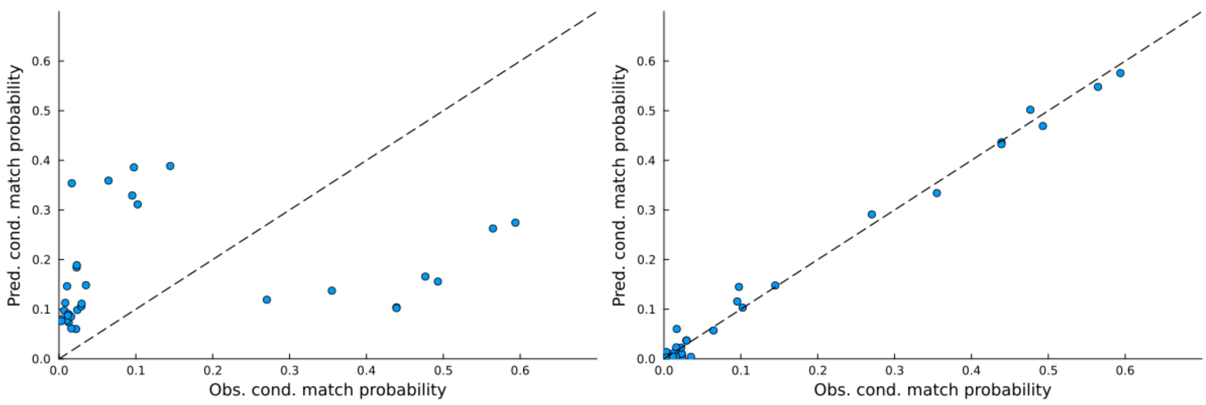
Notes. World Values Survey. Waves 2010-2014 and 2017-2022. Weighted results.

Figure 7: Agreement rates with "Homosexual couples are as good parents as other couples" by region



Notes. World Values Survey. Waves 2017-2022. Weighted results.

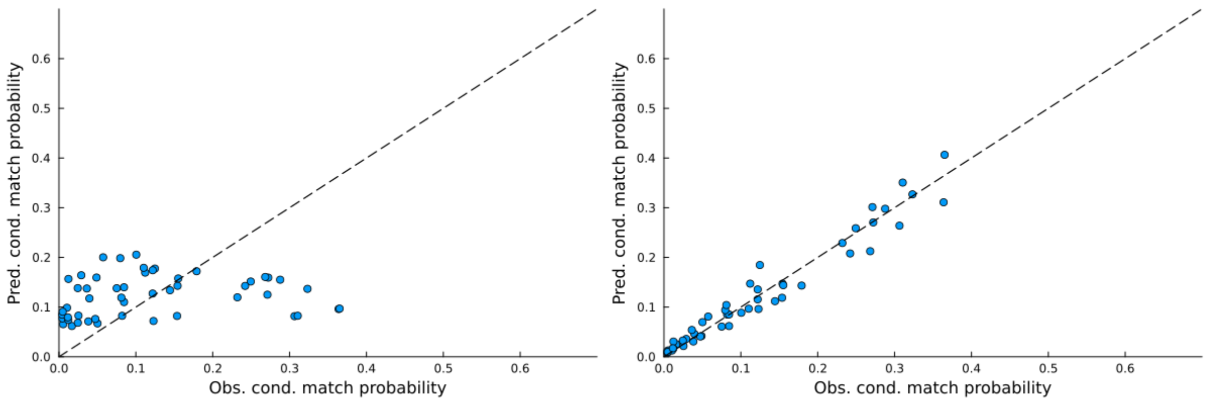
Figure 8: Model's fit for mating patterns with respect to gender and race



Column (1) Column (4)

Notes. Results obtained with 2013-2018 NHIS data. The different plots correspond to the model specifications presented in the columns of Table 6. In each subplot, we plot the predicted probabilities of matching with a partner of a given race conditional on one's own sex and race against their empirical counterparts. Points that lie on the 45° line indicate a perfect fit.

Figure 9: Model's fit for mating patterns with respect to gender and education



Column (1) Column (4)

Notes. Results obtained with 2013-2018 NHIS data. The different plots correspond to the model specifications presented in the columns of Table 6. In each subplot, we plot the predicted probabilities of matching with a partner of a given education conditional on one's own gender and education against their empirical counterparts. Points that lie on the 45° line indicate a perfect fit.

E Additional Tables

Table 18: Sexual orientation by gender

	NHIS 2013-2018		NHIS 2019-2022		NSFG (a) 2011-2019		NSFG (b) 2015-2019		SOEP 2016		SOEP 2019	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Straight	0.959	0.949	0.943	0.906	0.951	0.909	0.920	0.859	0.909	0.892	0.929	0.924
Bisexual	0.006	0.019	0.013	0.047	0.019	0.061	0.019	0.073	0.003	0.015	0.014	0.032
Gay or Lesbian	0.022	0.017	0.024	0.018	0.020	0.019	0.028	0.019	0.015	0.016	0.023	0.010
Something else	0.003	0.004	0.006	0.011			0.025	0.037	0.044	0.049		
Does not know	0.006	0.007	0.010	0.012	0.003	0.003	0.001	0.002				
Refuse	0.004	0.004	0.005	0.006	0.007	0.007	0.007	0.010	0.029	0.028	0.033	0.032
Observations	40,317	47,167	23,580	26,208	10,761	13,774	3,714	4,640	4,942	5,853	4,802	5,317

Notes. In every database, we present aggregate statistics for the population aged between 21 and 50. While the NSFG covers the 2011-2019 period, part (a) includes respondents for whom the category "Something else" was not available in the answer list, while part (b) includes respondents for whom this category was available. Hence, part (a) of the sample includes all respondents from 2011 to 2015, and half of the respondents from 2015 to 2019, whereas part (b) only includes half of the respondents from 2015 to 2019.

Table 19: Sexual attraction/orientation and partner's gender

	In a couple			Share of same-sex couples among couples		
	NHIS		SOEP	NHIS		SOEP
	2013-2018	2019-2022	2016-2019	2013-2018	2019-2022	2016-2019
Straight men	0.62	0.60	0.55	0.00	0.00	0.00
	(N=38,444)	(N=22,133)	(N=8,900)	(N=20,327)	(N=12,223)	(N=5,289)
Bisexual men	0.32	0.34	0.30	0.11	0.23	0.25
	(N=301)	(N=316)	(N=92)	(N=77)	(N=95)	(N=29)
Gay men	0.40	0.36	0.34	0.87	0.94	0.94
	(N=1006)	(N=674)	(N=195)	(N=291)	(N=212)	(N=57)
Straight women	0.63	0.63	0.60	0.00	0.00	0.00
	(N=44,718)	(N=23,860)	(N=10,121)	(N=24,083)	(N=13,708)	(N=5,245)
Bisexual women	0.44	0.49	0.34	0.11	0.11	0.07
	(N=923)	(N=1,173)	(N=282)	(N=333)	(N=496)	(N=74)
Lesbian women	0.55	0.48	0.58	0.87	0.90	0.84
	(N=830)	(N=482)	(N=132)	(N=372)	(N=205)	(N=55)

Notes. In every database, we present aggregate statistics for the population aged between 21 and 50. The first three columns show the rate of individuals living in couple for each gender and sexual orientation type. The last three columns show the rate of same-sex couple among individuals in couple for gender and sexual orientation type.

Table 20: Observations with missing partner's sexual orientation

	Share of couples with missing info on the partner	Single rate before selection	Single rate after selection	% Change
Straight Men	0.18	0.45	0.50	+ 11.1%
Bisexual Men	0.12	0.70	0.73	+ 4.3%
Gay men	0.30	0.66	0.74	+ 12.1%
Straight Women	0.30	0.40	0.49	+ 22.5%
Bisexual Women	0.53	0.66	0.81	+ 22.7%
Lesbian Women	0.23	0.42	0.48	+ 14.3%

Notes. SOEP data, pooled 2016 and 2019 waves. Weighted results.

Table 21: Sexual identity mobility

	Sexual Identity in 2019			
Sexual Identity in 2016	Heterosexual (12,644)	Bisexual (200)	Gay or Lesbian (136)	Refuse (447)
Straight (12,252)	0.970	0.008	0.002	0.020
Bisexual (115)	0.254	0.609	0.052	0.085
Gay or Lesbian (115)	0.116	0.009	0.719	0.157
Other (601)	0.830	0.011	0.000	0.158
Refuse (344)	0.701	0.060	0.009	0.229

Notes. SOEP data, 13,427 respondents, aged between 21 and 50 years old in 2016, each observed twice. Row percentages are reported in the cross-tabulation of mobility patterns. Data are weighted statistics with unweighted sample sizes. Figures in parentheses are the size sample of each subgroup. The category "other" does not exist in the 2019 survey.

Table 22: Patterns of partnership mobility

	Partnership Status in 2019		
Partnership Status in 2016	No Partner (5,241)	Different-Sex Partner (8,507)	Same-Sex Partner (49)
No Partner (5,350)	0.889	0.111	0.001
Different-Sex Partner (8,396)	0.057	0.943	0.000
Same-Sex Partner (51)	0.330	0.000	0.670

Notes. SOEP data, 13,797 respondents, aged between 21 and 50 years old in 2016, each observed twice. Row percentages are reported in the cross-tabulation of mobility patterns. Data are weighted statistics with unweighted sample sizes. Figures in parentheses are the size sample of each subgroup.

Table 23: Linear regression of reporting being bisexual or gay/lesbian

	Men				Women			
	Bisexual		Gay		Bisexual		Lesbian	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mostly attracted to opposite sex		0.190***		-0.000		0.174***		0.001
		(0.005)		(0.003)		(0.004)		(0.002)
Equally attracted to both sexes		0.710***		0.026***		0.772***		0.025***
		(0.010)		(0.006)		(0.007)		(0.003)
Mostly attracted to same sex		0.283***		0.691***		0.387***		0.535***
		(0.011)		(0.006)		(0.014)		(0.006)
Only attracted to opposite sex		0.000		0.927***		0.010		0.918***
		(0.007)		(0.004)		(0.012)		(0.005)
25-29 years old	-0.017***	-0.010***	-0.006	0.000	-0.024***	-0.015***	0.004	0.002
	(0.004)	(0.003)	(0.004)	(0.002)	(0.006)	(0.005)	(0.003)	(0.002)
30-34 years old	-0.015***	-0.004	-0.009**	-0.000	-0.038***	-0.018***	0.010***	0.003
	(0.004)	(0.003)	(0.004)	(0.002)	(0.006)	(0.005)	(0.004)	(0.002)
35-39 years old	-0.022***	-0.006*	-0.014***	0.001	-0.050***	-0.016***	0.006	0.005**
	(0.004)	(0.003)	(0.004)	(0.002)	(0.007)	(0.005)	(0.004)	(0.002)
40-44 years old	-0.025***	-0.008**	-0.008*	0.004**	-0.059***	-0.016***	-0.000	0.002
	(0.004)	(0.003)	(0.004)	(0.002)	(0.007)	(0.005)	(0.004)	(0.002)
45-49 years old	-0.035	-0.008	-0.014	0.001	-0.072	-0.007	-0.013	0.000
	(0.028)	(0.023)	(0.030)	(0.013)	(0.046)	(0.034)	(0.026)	(0.014)
High school	0.002	0.002	0.008	-0.002	-0.008	-0.011**	0.003	0.002
	(0.004)	(0.003)	(0.005)	(0.002)	(0.007)	(0.006)	(0.004)	(0.002)
Some college	0.001	-0.001	0.013***	-0.003	-0.018**	-0.017***	0.002	0.004*
	(0.005)	(0.004)	(0.005)	(0.002)	(0.008)	(0.006)	(0.004)	(0.002)
Bachelor	0.004	-0.004	0.017***	-0.002	-0.034***	-0.020***	-0.012***	0.000
	(0.004)	(0.004)	(0.005)	(0.002)	(0.007)	(0.006)	(0.004)	(0.002)
Master, doctorate	-0.002	-0.007	0.036***	0.003	-0.048***	-0.033***	-0.002	0.001
	(0.005)	(0.004)	(0.006)	(0.002)	(0.009)	(0.006)	(0.005)	(0.003)
Black	-0.014***	-0.008**	-0.008*	-0.001	0.007	0.015***	0.006*	0.001
	(0.004)	(0.004)	(0.005)	(0.002)	(0.007)	(0.005)	(0.004)	(0.002)
Hispanic	-0.002	0.001	0.008**	-0.001	-0.012**	-0.004	0.002	-0.001
	(0.004)	(0.003)	(0.004)	(0.002)	(0.005)	(0.004)	(0.003)	(0.002)
Others	-0.004	-0.005	0.009**	0.000	-0.003	-0.003	0.002	0.001
	(0.004)	(0.003)	(0.004)	(0.002)	(0.006)	(0.005)	(0.004)	(0.002)
2014-2016	0.002	-0.001	0.006**	0.000	0.013***	0.004	0.008***	0.002
	(0.003)	(0.002)	(0.003)	(0.001)	(0.004)	(0.003)	(0.002)	(0.001)
2017-2019	0.005	-0.003	0.008**	0.002*	0.038***	0.012***	0.005*	0.000
	(0.003)	(0.002)	(0.003)	(0.001)	(0.005)	(0.004)	(0.003)	(0.001)
Have kids	-	-	-	-	-0.031***	-0.002	-0.030***	-0.001
					(0.004)	(0.003)	(0.002)	(0.001)
Employed	-0.007	0.008*	-0.010*	0.001	-0.001	-0.010**	-0.000	0.001
	(0.005)	(0.004)	(0.005)	(0.002)	(0.005)	(0.004)	(0.003)	(0.002)
Born in U.S.	-0.001	0.003	0.002	0.001	0.041***	0.012***	0.005	-0.004***
	(0.004)	(0.003)	(0.004)	(0.002)	(0.006)	(0.004)	(0.003)	(0.002)
Constant	0.040***	0.002	0.018**	-0.001	0.095***	0.028***	0.027***	-0.001
	(0.007)	(0.006)	(0.008)	(0.003)	(0.010)	(0.008)	(0.006)	(0.003)
Observations	13,350	13,144	13,350	13,144	17,029	16,667	17,029	16,667
R-squared	0.005	0.353	0.005	0.821	0.027	0.460	0.013	0.723

Notes. 2011-2019 NSFG sample restricted to 21-50 years-old respondents. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 24: Descriptive statistics

	Men			Women		
	Straight	Bisexual	Gay	Straight	Bisexual	Lesbian
NHIS (2013-2018)						
Age	35.3	32.1	34.4	35.5	30.6	35.3
High school or less	0.367	0.329	0.235	0.298	0.316	0.260
Some college	0.303	0.348	0.314	0.322	0.363	0.345
College degree	0.330	0.323	0.451	0.379	0.322	0.395
White	0.590	0.662	0.594	0.576	0.681	0.575
Black	0.114	0.098	0.1242	0.133	0.113	0.163
Hispanic	0.205	0.166	0.200	0.193	0.143	0.171
Other	0.091	0.073	0.082	0.098	0.063	0.091
North East	0.161	0.171	0.183	0.169	0.149	0.162
North Central/ Midwest	0.228	0.235	0.197	0.217	0.249	0.180
South	0.359	0.233	0.341	0.375	0.345	0.416
West	0.251	0.361	0.279	0.239	0.257	0.242
NHIS (2019-2022)						
Age	35.3	29.8	33.5	35.7	30.0	33.0
High school or less	0.374	0.271	0.221	0.315	0.313	0.248
Some college	0.285	0.398	0.290	0.299	0.369	0.350
College degree	0.341	0.331	0.489	0.387	0.318	0.402
White	0.586	0.630	0.594	0.545	0.648	0.543
Black	0.111	0.083	0.104	0.134	0.117	0.220
Hispanic	0.204	0.196	0.185	0.218	0.157	0.162
Other	0.099	0.091	0.117	0.102	0.079	0.075
North East	0.165	0.145	0.190	0.165	0.155	0.167
North Central / Midwest	0.211	0.256	0.174	0.204	0.233	0.162
South	0.372	0.323	0.324	0.390	0.346	0.425
West	0.252	0.277	0.312	0.241	0.265	0.247
Large central metrop. areas	0.341	0.383	0.517	0.335	0.310	0.395
Large fringe metrop. areas	0.247	0.223	0.183	0.243	0.248	0.223
Medium and small metrop. areas	0.291	0.345	0.233	0.300	0.330	0.299
Non metrop. areas	0.122	0.050	0.067	0.123	0.111	0.084
SOEP (2016-2019)						
Age	35.7	31.7	36.0	35.3	32.5	35.2
High school or less	0.561	0.667	0.499	0.548	0.584	0.511
Some college	0.169	0.133	0.165	0.161	0.093	0.227
College degree	0.271	0.200	0.336	0.291	0.324	0.262

Notes. The Table shows sample means for each variable, conditional on gender and sexual orientation. In each data set, the population is restricted to the 21-50 years old.

Table 25: Distribution of partners' race for different- and same-sex couples

NHIS (2013-2018)		White	Black	Hispanic	Other	NHIS (2019-2022)		White	Black	Hispanic	Other
Different-sex	White	0.567	0.016	0.058	0.039	Different-sex	White	0.573	-	-	-
	Black		0.067	0.008	0.004		Black		0.061	-	-
	Hispanic			0.163	0.008		Hispanic			0.180	-
	Other				0.070		Other				0.066
Male same-sex	White	0.470	0.051	0.201	0.060	Male same-sex	White	0.540	-	-	-
	Black		0.045	0.017	0.003		Black		0.010	-	-
	Hispanic			0.104	0.030		Hispanic			0.131	-
	Other				0.019		Other				0.031
Female same-sex	White	0.596	0.015	0.107	0.038	Female same-sex	White	0.545	-	-	-
	Black		0.107	0.006	0.011		Black		0.118	-	-
	Hispanic			0.063	0.019		Hispanic			0.138	-
	Other				0.037		Other				0.028

Notes. In each of the six panels, relative frequencies sum to one. E.g., the first panel on the top left reads, "In 56.7% of all different-sex couples, both partners are White." In the panels of the right, only the numbers on the diagonal are available.

Table 26: Estimated degree of independence in nested logit models

	Specifications	
	(1)	(2)
White men	0.18	0.15
	(0.01)	(0.01)
White women	0.27	0.21
	(0.02)	(0.01)
Black men	0.34	0.27
	(0.02)	(0.02)
Black women	0.20	0.16
	(0.01)	(0.01)
Hispanic men	0.21	0.18
	(0.01)	(0.01)
Hispanic women	0.28	0.22
	(0.02)	(0.01)
Other men	0.17	0.14
	(0.01)	(0.01)
Other women	0.30	0.24
	(0.02)	(0.02)
Functional form	Φ_B	Φ_C
Demographics	Yes	Yes
Interactions	Yes	Yes
Nested logit	Yes	Yes
BIC	1,515,931	1,515,863

Notes. Results obtained with 2013-2018 NHIS data. Columns (1) and (2) refer to the same models whose match surplus parameters are presented respectively in columns (5) and (6) in Table 6. We report the estimated degree of independence $\rho \in [0, 1]$ for the nested logit versions of our model. We remind that a lower value of ρ indicates a stronger positive correlation between random taste shocks for partners in the demographic group indicated on the left of this table. Standard errors in parentheses.

Table 27: Share of households with children by couple type

	Men						Women					
	Straight		Bisexual		Gay		Straight		Bisexual		Lesbian	
In couple with	Man	Woman	Man	Woman	Man	Woman	Man	Woman	Man	Woman	Man	Woman
NHIS (2013-2018)												
Share with Children	0.38	0.72	0.00	0.46	0.04	0.76	0.73	0.35	0.49	0.42	0.71	0.35
N	24	20,151	7	70	250	40	23,837	29	298	34	48	323
NHIS (2019-2022)												
Share with Children	0.53	0.67	0.05	0.31	0.06	0.45	0.68	0.62	0.41	0.19	0.80	0.31
N	40	12,196	20	75	203	11	13,693	24	447	49	18	287
SOEP												
Share with Children	0.00	0.67	0.23	0.63	0.11	0.63	0.67	0.30	0.52	0.22	0.93	1.00
N	1	5,315	5	24	52	5	5,263	10	65	47	8	1

Notes. In the NHIS (2013-2018) data, 4% of gay men in couple with man have children, whereas 76% of gay men in couple with women have children.

Table 28: Additional regressions of observed family outcomes on predicted match gains

	Physical activity (1)	Smoking (2)	Self-assessed health (3)	BMI (4)	Tired (5)	Worried (6)	Hopeless (7)	Nervous (8)	Restless (9)	Sad (10)	Worthless (11)	Hours of sleep (12)	Quality of sleep (13)
A. Without controls													
$U(x_i, x_{i'})$	-0.004	-0.021	0.037	-0.256	-0.043	0.065	-0.032	-0.018	-0.034	-0.042	-0.021	0.000	0.013
	(0.007)	(0.002)	(0.005)	(0.031)	(0.005)	(0.008)	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)	(0.006)	(0.005)
$\varepsilon^*(x_{i'}\|x_i)$	0.002	-0.010	0.021	-0.142	-0.023	0.047	-0.022	-0.010	-0.016	-0.030	-0.015	-0.014	0.002
	(0.007)	(0.002)	(0.005)	(0.032)	(0.005)	(0.009)	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)	(0.006)	(0.005)
R^2	0.000	0.009	0.003	0.004	0.004	0.006	0.004	0.001	0.003	0.005	0.002	0.001	0.001
B. Controlling for partners' race, age, and education													
$U(x_i, x_{i'})$	0.006	-0.004	0.050	-0.375	-0.074	0.101	-0.053	-0.092	-0.081	-0.061	-0.040	-0.004	0.040
	(0.013)	(0.003)	(0.008)	(0.053)	(0.008)	(0.014)	(0.005)	(0.008)	(0.009)	(0.006)	(0.005)	(0.011)	(0.009)
$\varepsilon^*(x_{i'}\|x_i)$	0.009	0.002	0.025	-0.248	-0.035	0.057	-0.029	-0.052	-0.040	-0.032	-0.021	-0.002	0.016
	(0.011)	(0.002)	(0.006)	(0.043)	(0.007)	(0.012)	(0.004)	(0.007)	(0.007)	(0.005)	(0.004)	(0.009)	(0.008)
R^2	0.010	0.106	0.083	0.074	0.020	0.034	0.026	0.028	0.030	0.028	0.022	0.017	0.026

Notes. Results obtained with 2013-2018 NHIS data. This Table completes Table 8 in the main text with findings on additional individual well-being measures. Standard errors in parentheses. More details about the outcome variables are available in Appendix A.

Table 29: Testing differences in the same-sex penalty by gender and region

	Men				Women			
	North-East	North-Central & Midwest	South	West	North-East	North-Central & Midwest	South	West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
North-Central & Midwest (men)	-0.6334							
	(0.1994)							
South (men)	-0.3253	0.3081						
	(0.4560)	(0.4729)						
West (men)	-0.6010	0.0325	-0.2756					
	(0.1806)	(0.9414)	(0.4644)					
North-East (women)	1.9140	2.5475	2.2394	2.5150				
	(0.0002)	(0.0000)	(0.0000)	(0.0000)				
North-Central & Midwest (women)	1.0216	1.6550	1.3470	1.6226	-0.8924			
	(0.0359)	(0.0005)	(0.0014)	(0.0002)	(0.0762)			
South (women)	0.9912	1.6246	1.3165	1.5921	-0.9229	-0.0305		
	(0.0191)	(0.0001)	(0.0001)	(0.0000)	(0.0368)	(0.9401)		
West (women)	0.8161	1.4496	1.1415	1.4171	-1.0979	-0.2055	-0.1750	
	(0.0770)	(0.0014)	(0.0036)	(0.0005)	(0.0219)	(0.6448)	(0.6412)	

Notes. Results obtained with 2013-2018 NHIS data. The estimates correspond to those displayed in Figure 3a. In every cell, we report the difference in point-estimates between the estimated same-sex penalty for the group indicated in the first column and the top row. For instance, the first cell reads, "Male same-sex couples in the North-Central & Midwest region experience a lower surplus (equivalently, a stronger same-sex penalty) relatively to men in the North-East". In parentheses, we report the corresponding p-value.

Table 30: Testing differences in the same-sex penalty by gender and metropolitan area

	Men		Women	
	Non-metro & small metro	Central large metro	Non-metro & small metro	Central large metro
	(1)	(2)	(3)	(4)
Central large metro (men)	0.7578			
	(0.0279)			
Non-metro & small metro (women)	1.0578	0.3000		
	(0.0010)	(0.3444)		
Central large metro (women)	1.9599	1.2020	0.9020	
	(0.0000)	(0.0007)	(0.0067)	

Notes. Results obtained with 2019-2022 NHIS data. The estimates correspond to those displayed in Figure 4a. In every cell, we report the difference in point-estimates between the estimated same-sex penalty for the group indicated in the first column and the top row. For instance, the first cell reads, "Male same-sex couples in central large metropolitan areas experience a higher surplus (equivalently, a weaker same-sex penalty) relatively to men in non-metropolitan and small metropolitan areas". In parentheses, we report the corresponding p-value.