



Like Attract Like? A Structural Comparison of Homogamy Across Same-Sex and Different-Sex Households

Edoardo Ciscato, Alfred Galichon, Marion Goussé

► To cite this version:

Edoardo Ciscato, Alfred Galichon, Marion Goussé. Like Attract Like? A Structural Comparison of Homogamy Across Same-Sex and Different-Sex Households. 2015. hal-03393232

HAL Id: hal-03393232

<https://sciencespo.hal.science/hal-03393232>

Preprint submitted on 21 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LIKE ATTRACT LIKE? A STRUCTURAL COMPARISON OF HOMOGAMY ACROSS SAME-SEX AND DIFFERENT-SEX HOUSEHOLDS

EDOARDO CISCATO^b, ALFRED GALICHON[†], AND MARION GOUSSE[§]

ABSTRACT. In this paper, we extend Gary Becker’s empirical analysis of the marriage market to same-sex couples. Becker’s theory rationalizes the well-known phenomenon of homogamy among heterosexual couples: individuals mate with their likes because many characteristics, such as education, consumption behaviour, desire to nurture children, religion, etc., exhibit strong complementarities in the household production function. However, because of asymmetries in the distributions of male and female characteristics, men and women may need to marry “up” or “down” according to the relative shortage of their characteristics among the populations of men and women. Yet, among homosexual couples, this limit does not exist as partners are drawn from the same population, and thus the theory of assortative mating would boldly predict that individuals will choose a partner with nearly identical characteristics. Empirical evidence suggests a very different picture: a robust stylized fact is that the correlation of characteristics is in fact weaker among the homosexual couples. In this paper, we build an equilibrium model of the same-sex marriage market which allows for straightforward identification of the gains to marriage. We estimate the model with 2008-2012 ACS data on California and show that *positive assortative mating* is weaker for homosexuals than for heterosexuals with respect to age and race. Yet, contrarily to previous empirical findings, our results suggest that positive assortative mating with respect to education is stronger on the same-sex marriage market. As regards labor market outcomes, such as hourly wages and working hours, we find that the process of specialization within the household mainly applies to heterosexual couples.

Keywords: sorting, matching, marriage market, homogamy, same-sex households, roommate problem.

JEL Classification: D1, C51, J12, J15.

Date: First circulated version: November 21, 2014. This version: September 2015. Galichon’s research has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreements no 313699. The authors would like to thank Arnaud Dupuy, Sonia Oreffice, Bernard Salanié and Simon Weber for useful discussions.

1. INTRODUCTION

How individuals sort themselves into marriage has important implications for income distribution, labor supply, and inequality (Becker, 1973). Strong evidence shows that the rise in income inequality across households over the last fifty years is partly due to assortative mating, as individuals have been sorting into increasingly assortative marriages (Greenwood, Guner, Kocharkov, and Santos, 2014; Eika, Mogstad, and Zafar, 2014).

Individuals tend to mate with their likes, although, because of asymmetries between the distributions of characteristics in male and female populations, homogamy cannot be perfect among heterosexual couples. In other words, heterosexuals cannot always find a “clone” of the opposite gender to match with. A large body of the literature has noticed that, up until recently, “men married down, women married up” due to the gender asymmetry in educational achievement that has only recently started to fade (Goldin, Katz, and Kuziemko, 2006). Gender asymmetries exist in other dimensions such as biological characteristics (windows of fertility¹, life expectancy, bio-metric characteristics), psychological traits, economic attributes (due to the gender wage gap), ethnic and racial characteristics (immigration is not symmetric across gender, see Weiss, Yi, and Zhang, 2013) or demographic characteristics (some countries, such as China, have a comparatively more imbalanced gender ratios).

Homogamy has been famously rationalized by Becker’s theory of positive assortative mating (PAM), arguably the simplest structural model of homogamy: men and women are characterized by some socio-economic “ability” index and the marriage market clears so that men are matched with women that are as close as possible to them in terms of an index accounting for all the characteristics that matter on the marriage market². The (strong)

¹Women’s fertility rapidly declines with age, whereas men’s fertility does not. Biologists and anthropologists argue that this dissymmetry could explain the well-documented preference of men for younger women (Hayes, 1995; Kenrick and Keefe, 1992). Low (2013) evaluates this young age premium for women and names it “reproductive capital,” as it gives them an advantage on the marriage market over older women.

²Becker (1973) expects most of non-labor market traits, such as “intelligence, height, skin color, age, education, family background or religion”, to be complementary. However, he also suggests that some attributes (e.g. some personality traits) could be substitutes. As concerns labor market traits, Becker suggests that we should observe a negative correlation between wage rates, because of partners maximizing marriage surplus. More recently, Chiappori, Oreffice, and Quintana-Domeque (2012) model a Becker-like

prediction of this model is that the rank of the husband's index in the men's population is the same as the wife's in the women's population. However, this does not imply that the partners' indices are identical: they would be so only if the distributions of the indices were the same for both men and women's populations.

This analysis of the marriage market has attracted wide attention in the economic literature, in spite of its shortcomings. One shortcoming is that it originally refers to heterosexual unions only. However, in a growing number of countries, same-sex couples have gained legal recognition and the institutions of civil partnership and marriage no longer require that the partners must be of opposite sex. This official recognition is the result of several legal disputes and social activism by the gay and lesbian communities³. The issue of whether to recognize same-sex unions has long been a topical subject in many countries, since it challenges the traditional model of family. From both an economic and a legal point of view, the definition of what "family" means has relevant political implications as long as this term is present - and is generally central - in many modern constitutions and legal systems. Consequently, family households benefit from a special attention of policy-makers. Therefore, a discussion of the issues related to the same-sex marriage - remarkably at policy level - requires a good understanding of similarities and differences in the household dynamics among same-sex and different-sex couples. Besides, it is important to remember that the legal recognition of same-sex couples is only one of many transformations that the institution of the family has gone through in the last decades (Stevenson and Wolfers, 2007; Stevenson, 2008). Finally, since more and more data on same-sex unions have been made available, the extension of the economic analysis of family to the homosexual population can now be taken to data.

While it is natural to consider an extension of Becker's theory of PAM to same-sex households, it is worthwhile noting that the previous considerations on asymmetries between marriage market with sorting on a unidimensional index. The estimation of such index reveals that high values in some attributes can compensate for poor values in others, thus showing that sorting is based on complex interactions between traits.

³Public actions for homosexual rights acknowledgment are often considered to have started in 1969, in New York City. See Eskridge Jr (1993) and Sullivan (2009) for a detailed history and a full overview of the arguments in favor and against same-sex marriage.

men and women distributions only hold as long as each partner comes from a separate set according to his/her sex. On the same-sex marriage market, the two partners are drawn from the same population and the distributions of characteristics is the same for each of them. Hence, the assortative mating theory pushed to its limits implies that, in this setting, partners should be exactly identical, i.e. each individual will choose to marry someone with identical characteristics.

In spite of such theoretical predictions, facts suggest a very different picture. Recent empirical results on the 1990 and 2000 American Census show that same-sex couples have less correlated attributes than different-sex ones, at least in terms of a variety of non-labor traits, including racial and ethnic background, age and education (Jepsen and Jepsen, 2002; Schwartz and Graf, 2009). Studies on Norway, Sweden (Andersson, Noack, Seierstad, and Weedon-Fekjær, 2006) and Netherlands (Verbakel and Kalmijn, 2014) brought to similar findings. In order to explain this heterogeneity, the literature has suggested several possible reasons. A first consideration is that homosexuals might be forced to pick from a restricted pool because of their small numbers in the population (Harry, 1984; Kurdek and Schmitt, 1987; Andersson, Noack, Seierstad, and Weedon-Fekjær, 2006; Schwartz and Graf, 2009; Verbakel and Kalmijn, 2014), thus having a narrower choice when choosing their partner, resulting in a more diverse range of potential matches. Furthermore, homosexuals have been found to be more likely to live in urban neighborhoods than heterosexuals, and since diversities in socio-economic traits are stronger in cities, this facilitates the crossing of racial and social boundaries (Black, Gates, Sanders, and Taylor, 2002; Rosenfeld and Kim, 2005; Black, Sanders, and Taylor, 2007). In light of these observations, one could argue that the homosexual marriage market is faced with stronger search frictions. Nevertheless, this might not necessarily be the case if the potential partners gather in specific locations, as it is case for the choice of “gay-friendly” cities and neighborhoods. Other analysts argue that homosexuals may have different preferences than heterosexuals, as they tend to be less conservative than straight individuals. Some explanations in this regard point out that, since homosexuality is still considered in some cultures as at odds with prevailing social norms, homosexuals might grow less inclined to passively accept social conventions, and consequently they would end up choosing their partner with fewer concerns about

his/her background traits⁴(Blumstein, Schwartz, et al., 1983; Meier, Hull, and Ortyl, 2009; Schwartz and Graf, 2009). The detachment from the community of origin and the research for more tolerant surroundings have an influence both on values and social norms and on the heterogeneity of interpersonal ties.

A part of these explanations has to do with individual preferences, whereas another part has to do with demographics, i.e. the distribution of characteristics in the population. It is clear that the explanations listed in the former paragraph, while different in nature, are not mutually exhaustive, but all contribute to a better understanding of the equilibrium patterns. For instance, a high correlation in education may arise from individual tastes (as individuals could find more desirable to match with a partner with similar educational background), but also from demographics (indeed, if some educational category represents a large share of individuals, this will increase odds of unions within this category, thus mechanically increasing the correlation in education).

In this paper, we focus on individual preferences: we would like to compare the “preference for homogamy” across same-sex and different-sex households. In order to do so, we need a methodology to interpret the observation of matching patterns which disentangles preferences from demographics. This is achieved by a structural approach, which captures the preference parameters leading to equilibrium matching patterns that exhibit the closest fit with the patterns actually observed. This approach hence will require an equilibrium model of matching.

In the wake of Becker (1973, 1981), the economic literature has classically modeled the marriage market as a bipartite matching game with transferable utility. A couple consists of two partners coming each from a separate or identical subpopulation (respectively, in the case of heterosexual and homosexual unions). Both partners are characterized by vectors of

⁴Note that household location choice and social norms are strictly related: it has been reported that homosexuals often leave their town of origin and escape social pressure exerted by relatives and acquaintances and go living in larger cities reputed to be gay/lesbian-friendly (Rosenfeld and Kim, 2005). Analogously, homosexuals are aware that they have more probabilities of avoiding discrimination by achieving higher educational levels and orienting their professional choices toward congenial working environments (Blumstein, Schwartz, et al., 1983; Verbakel and Kalmijn, 2014).

attributes, such as education, wealth, age, physical attractiveness, etc. It is assumed that, when two partners with respective attributes x and y form a pair, they generate a surplus equal to $\Phi(x, y)$, which is shared endogenously between them. In the case of separate sub-populations (heterosexual marriage), the landmark contribution of Choo and Siow (2006) showed that the surplus function Φ can easily be estimated based on matching patterns modulo a distributional assumption on unobservable variations in preferences, and was followed by a rich literature (Fox, 2010; Galichon and Salanié, 2014; Chiappori, Salanié, and Weiss, 2011, to cite a few). Dupuy and Galichon (2014) extended Choo and Siow’s model to the case of continuous attributes and propose the convenient bilinear parameterization $\Phi(x, y) = x' Ay$, where A is a matrix called “affinity matrix” whose terms reflect the strength of assortativeness between two partners’ attributes. However, the bipartite assumption is restrictive and does not allow to estimate the surplus on same-sex marriage markets, and, to the best of our knowledge, no such estimation procedure is proposed in the literature. This problem is addressed in the present paper, using the observation by Chiappori, Galichon, and Salanié (2012) that, when the population to be matched is large, the same-sex marriage problem or “unipartite matching problem” can be theoretically reformulated as a heterosexual matching problem or “bipartite matching problem”. As a consequence, the empirical tools developed to perform estimation of preferences on the heterosexual marriage market, including those cited above, can be adapted to estimate preferences on the homosexual marriage market.

A few papers already deal with the issue of assortativeness among same-sex households, although none of them allows to draw conclusion on preference parameters. The most relevant benchmarks for the empirical results of this work are the aforementioned Jepsen and Jepsen (2002) and Schwartz and Graf (2009). Both papers make use of the American census data (1990 and 1990/2000 respectively) and their most important result is that members of different-sex couples are more alike than those of same-sex ones with respect to non-labor market traits. The heterogeneity in assortativeness is measured in a logit framework containing dedicated parameters for homogamy. In general, in a logit framework individuals choose their best option among all possibilities. However, this fails to take into account the fact that matching takes place under scarcity constraint on the various characteristics.

In this paper, we fully describe the equilibrium matching pattern in respect of market conditions. We estimate the true preference parameters for each type of couple (same-sex and different-sex ones): the following cross-comparison turns out to be very insightful for the understanding of heterogeneity in assortativeness.

The contributions of the present paper are twofold. On a methodological level, this paper is the first to propose a structural estimator of the matching surplus which applies to same-sex households, or, more generally, to instances of the unipartite matching problem. On an empirical level, we provide evidence through the means of a structural analysis that, as concerns age and ethnicity, the heterosexual population has a stronger preference for homogamy than the homosexual one. Nevertheless, in contrast with previous empirical findings (Jepsen and Jepsen, 2002; Schwartz and Graf, 2009), we find that sorting on education is stronger on the same-sex marriage market. Further, we also look at labor market traits such as hourly wages and working hours. Comparing assortativeness on labor market outcomes between homosexual and heterosexual couples allows us to infer different family dynamics and differences in the household specialization process. Finally, we briefly discuss the estimates of the mutually exhaustive affinity indices obtained through saliency analysis.

The rest of the paper is organized as follows. Section 2 will present the model and section 3 the estimation. We describe our data in section 4 and our results in section 5. Section 6 concludes.

2. THE MODEL

In what follows, it is assumed that the full type of each individual, i.e. the complete set of all individual characteristics that matter for the marriage market (physical attributes, psychological traits, socio-economic variables, gender, sexual orientation, etc.), is fully observed by market participants. Each individual is characterized by a vector of observable characteristics $x \in \mathcal{X} = \mathbb{R}^K$, which constitutes his observable type. However, we allow for a certain degree of unobserved heterogeneity by assuming that agents experience unobserved variations in tastes that are not observable to the analyst, following Choo and Siow (2006). In this paper, types are assumed to be continuous, as in Dupuy and Galichon

(2014), hereafter DG, and Menzel (2013). Assume that the distribution of characteristics x has a density function f with respect to the Lebesgue measure. Without loss of generality, the marginal distribution of the attributes is assumed to be centered, i.e. $\mathbb{E}[X] = 0$.

2.1. Populations. A *pair* is an ordered set of individuals, denoted $[x_1, x_2]$ where $x_1, x_2 \in \mathcal{X}$, in which the order of the partner matters, which implies that the pair $[x_1, x_2]$ will be distinguished from its inverse twin $[x_2, x_1]$. In empirical datasets, x_1 will often be denominated “head of the household” and x_2 “spouse of the head of the household” even though this denomination is used mainly for practical reasons and cannot be fully representative of the actual roles in the household. A *couple* is an unordered set of individuals (x_1, x_2) , so that the couple (x_1, x_2) coincides with the couple (x_2, x_1) . A *matching* is the density of probability $\pi(x_1, x_2)$ of drawing a couple (x_1, x_2) . Pairs $[x_1, x_2]$ and $[x_2, x_1]$ stand for the same couple, so that one has $\pi(x_1, x_2) := \pi[x_1, x_2] + \pi[x_2, x_1]$, hence the symmetry condition $\pi(x_1, x_2) = \pi(x_2, x_1)$ holds. This symmetry constraint means that the position of the individual must not matter and thus that there are no predetermined “roles” within the couple that would be relevant for the analysis⁵.

We shall impose assumptions that will ensure that everyone is matched at equilibrium, hence the density of probability of type $x \in \mathcal{X}$ in the population is given by $\int_{\mathcal{X}} \pi(x, x') dx' = \int_{\mathcal{X}} \pi[x, x'] dx' + \int_{\mathcal{X}} \pi[x', x] dx'$, where the right hand side counts the number of individuals of type x matched either as the head of household (first term), or as the spouse of the head (second term). Thus, we are led to assume:

Assumption 1 (Populations). *The density $\pi(x, x')$ over couples satisfies $\pi \in \mathcal{M}^{sym}(f)$, where*

$$\mathcal{M}^{sym}(f) = \left\{ \pi \geq 0 : \left(\int_{\mathcal{X}} \pi(x, x') dx' = f(x) \ \forall x \in \mathcal{X} \right) \right\}.$$

⁵Candelon and Dupuy (2014) extend Chiappori, Galichon, and Salanié (2012)’s analysis to a model where agents form couples with endogenously assigned roles according to their characteristics. The model is applied to professional cycling, where teams usually choose one leader, typically the best rider, and some helpers. However, a similar hierarchical structure would be simplistic in a model of marriage.

In contrast, in the classical bipartite problem, we try to match optimally two distinct populations (men and women) which are characterized by the same set of observable variables \mathcal{X} , and it is assumed that the distribution of the characteristics among the population of men has density f , while the density of the characteristics among the population of women is g . In this setting, the feasibility constraints take on the typical following form:

$$\mathcal{M}(f, g) = \left\{ \pi \geq 0 : \left(\int_{\mathcal{X}} \pi(x, y) dy = f(x) \quad \forall x \in \mathcal{X} \right) \right\}$$

Hence, $\pi \in \mathcal{M}^{sym}(f)$ if and only if $\pi \in \mathcal{M}(f, f)$ and $\pi(x_1, x_2) = \pi(x_2, x_1)$. Thus the feasibility set in the unipartite problem and in the bipartite problem differ only by the additional symmetry constraint in the unipartite problem.

2.2. Preferences. We now model preferences. Following DG, it is assumed that a given individual x_1 does not have access to the whole population, but only to a set of acquaintances $\{z_k^x : k \in \mathbb{Z}_+\}$, randomly drawn, which is described below.

Assumption 2 (Preferences). *An individual of type x matched to an individual of type x' enjoys a surplus which is the sum of three terms:*

(i) *the systematic part of the pre-transfer matching surplus enjoyed by x from his/her match with x' , denoted $\alpha(x, x')$.*

(ii) *the equilibrium utility transfer from x' to x , denoted $\tau(x, x')$. This quantity can be either positive or negative; we assume utility is fully transferable, hence feasibility imposes $\tau(x, x') + \tau(x', x) = 0$.*

(iii) *a “sympathy shock” $(\sigma/2)\varepsilon^x$, which is stochastic conditional on x and x' , and whose value is $-\infty$ if x is not acquainted with an individual x' . The quantity $\sigma/2$ is simply a scaling factor. More precisely, the set of acquaintances is an infinite countable random subset of \mathcal{X} ; it is such that (z_k^x, ε_k^x) are the points of a Poisson process on $\mathcal{X} \times \mathbb{R}$ of intensity $dz \times e^{-\varepsilon} d\varepsilon$.*

The utility of unmatched individuals is $-\infty$ for all types, so that every market participant is matched at equilibrium.

While the stochastic structure of the unobserved variation in preference described in part (iii) of Assumption 2 may appear complex, it is in fact a very natural extension of the logit framework to the continuous case, as we now argue. Indeed, it will imply that the individual maximization program of an agent of type x with this set of acquaintances is

$$\max_{k \in \mathbb{Z}_+} \alpha(x, z_k^x) + \tau(x, z_k^x) + \frac{\sigma}{2} \varepsilon_k^x, \quad (2.1)$$

where the utility of matching with acquaintance k yields a total surplus which is the sum of three terms, the systematic pre-transfer surplus, the transfer, and the sympathy shock. Define the systematic quantity of surplus at equilibrium U by

$$U(x, x') := \alpha(x, x') + \tau(x, x')$$

thus an individual of type x hence maximizes $U(x, z_k^x) + (\sigma/2) \varepsilon_k^x$ over the set of his acquaintances, which are indexed by k . This induces an aggregate demand over the type space. Indeed, it follows from the continuous logit theory initiated in Dagsvik (1994) that the conditional probability density of an individual of type x of matching with a partner of type x' is

$$\pi(x'|x) = \frac{\exp \frac{U(x, x')}{\sigma/2}}{\int_{\mathcal{X}} \exp \frac{U(x, x')}{\sigma/2} dx'}. \quad (2.2)$$

It is clear from expression (2.2) that this is a generalization of the logit framework to the continuous case.

2.3. Equilibrium. Next, we introduce our equilibrium concept. Denote

$$\Phi(x, x') := \alpha(x, x') + \alpha(x', x) = U(x, x') + U(x', x)$$

the systematic part of the joint surplus⁶ between x and x' . It follows from (2.2) and symmetry of π that

$$(\sigma/2) \ln \pi(x, x') = U(x, x') - a(x) = U(x', x) - a(x'), \quad (2.3)$$

$$\text{where } a(x) : = \frac{\sigma}{2} \log \int_{\mathcal{X}} \frac{1}{f(x)} \exp \frac{U(x, x')}{\sigma/2} dx'. \quad (2.4)$$

⁶Note that Φ is symmetric by definition, but α has no reason to be symmetric. Mathematically speaking, Φ is (twice) the symmetric part of α .

Substituting out for U in (2.3) yields the following equation, which expresses optimality in individual decisions:

$$\log \pi(x, x') = \frac{\Phi(x, x') - a(x) - a(x')}{\sigma}, \quad (2.5)$$

At equilibrium, the value of $a(\cdot)$ is determined by market-clearing condition $\int_{\mathcal{X}} \pi(x, x') dx' = f(x)$, that is

$$\int_{\mathcal{X}} \exp\left(\frac{\Phi(x, x') - a(x) - a(x')}{\sigma}\right) dx' = f(x). \quad (2.6)$$

After all these preparations, we can define our equilibrium matching concept.

Definition 1. π is an equilibrium matching if and only if there is a function $a(\cdot)$ such that both optimality equations (2.5) and market clearing equations (2.6) are satisfied.

The main results on equilibrium characterization are summarized in the following statement, whose proof is given below:

Theorem A. Under Assumptions 1 and 2:

(i) The equilibrium matching $\pi(x, x')$ is the unique solution to

$$\max_{\pi \in \mathcal{M}(f, f)} \iint_{\mathcal{X} \times \mathcal{X}} \Phi(x, x') \pi(x, x') dx dx' - \sigma \mathcal{E}(\pi), \quad (2.7)$$

where $\mathcal{E}(\pi)$ is defined by

$$\mathcal{E}(\pi) = \iint_{\mathcal{X} \times \mathcal{X}} \pi(x, x') \ln \pi(x, x') dx dx'. \quad (2.8)$$

(ii) The expression of $\pi(x, x')$ is given by

$$\pi(x, x') = \exp\left(\frac{\Phi(x, x') - a(x) - a(x')}{\sigma}\right), \quad (2.9)$$

where $a(\cdot)$ is a fixed point of F , which is given by

$$F[a](x) = \sigma \log \int_{\mathcal{X}} \exp\left(\frac{\Phi(x, x') - a(x')}{\sigma}\right) dx' - \sigma \log f(x). \quad (2.10)$$

Proof. By DG, Theorem 1, Problem (2.7) has a unique solution which can be expressed as

$$\pi(x, x') = \exp\left(\frac{\Phi(x, x') - a(x) - b(x')}{\sigma}\right)$$

for some $a(x)$ and $b(x')$ determined by $\pi \in \mathcal{M}(f, f)$. By the symmetry of Φ and by the symmetry of the constraints implied by $\pi \in \mathcal{M}(f, f)$, then $\tilde{\pi}(x', x) := \pi(x, x')$ is also solution to (2.7). By uniqueness, $\tilde{\pi} = \pi$, thus $\pi(x, x') = \pi(x', x)$. As a result, $b(x) = a(x)$, where a is determined by

$$\int \exp\left(\frac{\Phi(x, x') - a(x) - a(x')}{\sigma}\right) dx' = f(x)$$

QED. □

This result deserves a number of comments. First, we should note that there is an interesting interpretation of (2.7). While the first term inside the maximum tends to maximize the sum of the observable joint surplus, and hence draws the solution toward assortativeness, the second term $\mathcal{E}(\pi)$ is an entropic term which draws the solution toward randomness. The trade-off between assortativeness and randomness is expressed by the ratio Φ/σ . If this ratio is large, the assortative term predominates, and the solution will be close to the assortative solution. If this ratio is small, the entropic term predominates, and the solution will be close to the random solution. At the same time, note that the model parameterized by (Φ, σ) is scale-invariant: if $k > 0$, then the equilibrium matching distribution π when the parameter is (Φ, σ) is unchanged when the parameter is $(k\Phi, k\sigma)$. This will have important consequences for identifications, which are discussed in the next paragraph.

As a consequence of this result, we can deduce the equilibrium transfers and the utilities at equilibrium. Indeed, note that combining the expression of π as a function of U and a and Equation (2.5) yields the following expression of U as a function of a :

$$U(x, x') = (\Phi(x, x') + a(x) - a(x')) / 2. \tag{2.11}$$

which is the systematic part of utility that an individual of type x obtains at equilibrium from a match with an individual of type x' . It is equal to the half of the joint surplus, plus an adjustment $(a(x) - a(x'))/2$ which reflects the relative bargaining powers of x and

x' . These bargaining powers depend on the relative scarcity of their types; indeed, $a(x)$ is to be interpreted as the Lagrange multiplier of the scarcity constraint which imposes that $\pi(\cdot, x)$ should sum to $f(x)$. Hence, the equilibrium transfer $\tau(x, x')$ from x to x' is given by

$$\tau(x, x') = (\alpha(x', x) - \alpha(x, x') + a(x) - a(x')) / 2. \quad (2.12)$$

Next, note that an interesting feature of Theorem A is that, while it characterizes equilibrium in the homosexual marriage problem, it highlights at the same time the equivalence with the heterosexual marriage problem: indeed, as argued in DG, Theorem 1, the equilibrium matching in the heterosexual marriage problem is given by the same expression as (2.7), with the only difference that $\mathcal{M}(f, f)$ is replaced by $\mathcal{M}(f, g)$, where f and g are respectively the distribution of men and women's characteristics.

We will use this characterization of the equilibrium matching as the solution of an optimization problem in order to estimate the joint surplus Φ based on the observation of the matching density π . As it is classical in the literature on the estimation of matching models with transferable utility, the primitive object of our investigations will be the joint surplus Φ rather than the individual pre-transfer surplus α ; indeed, without observations on the transfers, Φ is identified but α is not. In other words, if we estimate that there is a high level of joint surplus in the (x, x') relationship, we will not be able to determine if this is due to the fact that “ x likes x' ” or “ x' likes x ”. We will only be able to estimate that there is a high affinity between x and x' .

3. ESTIMATION

3.1. Estimation of the affinity matrix. Following Dupuy and Galichon (2014), we assume a quadratic parametrization of the surplus function Φ to focus on a limited number of parameters which could characterize the matching pattern. We parametrize Φ by an *affinity matrix* A so that

$$\Phi_A(x, y) = x' A y = \sum_{ij} A_{ij} x^i y^j$$

where A has to be symmetric ($A_{ij} = A_{ji}$) in order for Φ to satisfy the symmetry requirement. Then the coefficients of the affinity matrix are given by $A_{ij} = \partial^2 \Phi(x, y) / \partial x^i \partial y^j$ at any value (x, y) . Matrix A has a straightforward interpretation: A_{ij} is the marginal increase (or decrease, according to the sign) in the joint surplus resulting from a one-unit increase in the attribute i for the first partner, in conjunction with a one-unit increase in the attribute j for the second. Hence, this approach is arguably the most straightforward one to model pairwise positive or negative complementarities for any pair of characteristics. It does, however, not preclude nonlinear functions of the x_i 's and the y_j 's, which can always be appended to x and y .

Recall equation (2.7), the optimal matching π maximizes the social gain

$$\mathcal{W}(A) = \max_{\pi \in \mathcal{M}(f, f)} \mathbb{E}_{\pi} [x' A y] - \sigma \mathbb{E}_{\pi} [\ln \pi(x, y)] \quad (3.1)$$

which yields likelihood $\pi^A(x, x')$ of observation (x, x') , where π^A is the solution to (3.1). By the Envelope theorem, $\partial \mathcal{W}(A) / \partial A_{ij} = \mathbb{E}_{\pi^A} [x^i y^j]$. Hence, our empirical strategy, following DG, is to look for \hat{A} satisfying

$$\partial \mathcal{W}(A) / \partial A_{ij} = \mathbb{E}_{\hat{\pi}} [x^i y^j], \quad (3.2)$$

where $\hat{\pi}$ is empirical distribution associated with the observed matching.

As noted before, the model with parameters (A, σ) is equivalent to the model with parameters $(kA, k\sigma)$ for $k > 0$. Hence, a choice of scale normalization should be imposed without loss of generality; a simple choice for single market observation is $\sigma = 1$, in which case the estimator A is meant as the estimator of the ratio of the affinity matrix over the scale parameter. The observation and comparison of multiple markets leads to slightly different normalizations choices, which are discussed at the end of this section.

If a sample of size n $\{(x_1, y_1), \dots, (x_n, y_n)\}$ is observed, then $\hat{\pi}(x, y) = n^{-1} \sum_{t=1}^n \delta(x - x_t) \delta(y - y_t)$. In DG, an estimator of A is obtained by solving the following concave optimization problem

$$\min_{A \in M_K} \mathcal{W}(A) - \mathbb{E}_{\hat{\pi}} \left[\sum_{ij} A_{ij} x^i y^j \right], \quad (3.3)$$

where M_K is the set of real $K \times K$ matrices. Indeed, the first order conditions associated to (3.3) is exactly given by (3.2). However, in the present case, symmetry of A is a requirement of the model. The population cross-covariance matrix $\mathbb{E}_\pi[x^i y^j]$ is symmetric, as π satisfies the symmetry restriction $\pi(x, x') = \pi(x', x)$ in the population. Yet, in the sample, $\hat{\pi}$ does not need to verify the symmetry restriction, as the first variable typically designates the surveyed individual, while the second variable designates the partner of the surveyed individual. Hence, the empirical matrix of co-moments $\mathbb{E}_{\hat{\pi}}[x^i y^j]$ will only be approximately symmetric. Thus, we symmetrize the sample by adding the symmetric households, that is, if household ij individual i was surveyed and reports partner j , we add a symmetric household ji as if the surveyed individual was j and had reported a partner i . In other words, we replace the empirical distribution $\hat{\pi}(x, x')$ by its symmetric part $(\hat{\pi}(x, x') + \hat{\pi}(x', x))/2$. In the sequel, $\hat{\pi}$ will denote that symmetric part. This leads us to propose the following definition:

Definition 2. *The estimator \hat{A} of the affinity matrix is obtained by*

$$\hat{A} = \arg \min_{A \in M_K} \{\mathcal{W}(A) - \mathbb{E}_{\hat{\pi}}[\sum_{1 \leq i, j \leq K} A_{ij} X^i Y^j]\}, \quad (3.4)$$

where M_K is the set of real $K \times K$ matrices.

3.2. Categorical variables. The previous analysis can be slightly adapted to deal with the case of categorical variables, such as race. Assume that the set of categories is denoted $\mathcal{R} = \{1, \dots, r\}$. Assume that the individuals are characterized by $x = (\tilde{x}, \bar{x}_1, \dots, \bar{x}_r)$, where $\tilde{x} \in \mathbb{R}^K$ are socio-economic characteristics, and \bar{x}_i ($1 \leq i \leq r$) is a dummy variable which is 1 if individual x is of category i , and zero otherwise. We work with the following specification of the surplus

$$\Phi(x, y) = \tilde{x}' \tilde{A} \tilde{y} + \lambda 1\{\bar{x} = \bar{y}\}$$

where λ is a term that reflects assortativeness on the categorical variable, which provides a utility increment λ if the partners belong to the same category. Of course, this surplus function can be expressed multiplicatively as $\Phi(x, y) = x' A y$, where A can be written

blockwise as

$$A = \begin{pmatrix} \tilde{A} & 0 \\ 0 & \lambda I_r \end{pmatrix} \quad (3.5)$$

and hence, A is obtained by running optimization problem (3.4) subject to constraint (3.5). Note that the envelope theorem implies that λ is indentified by the moment matching condition

$$\Pr_{\pi}(\bar{x} = \bar{y}) = \Pr_{\hat{\pi}}(\bar{x} = \bar{y})$$

which states that the predicted frequency of interracial couples should match the observed one.

3.3. Saliency analysis. The rank of the affinity matrix is informative about the dimensionality of the problem, that is, how many indices are needed to explain the sorting in this market. To answer this question, DG introduced *saliency analysis*, which consists in looking for successive approximations of the K -dimensional matching market by p -dimensional matching markets ($p \leq K$). Assume (without loss of generality as one can always rescale) that $\text{var}(X_i) = \text{var}(Y_j) = 1$. Then saliency analysis consists of a singular value decomposition of the affinity matrix $A = U' \Lambda V$, where U and V are orthogonal loading matrices, and Λ is diagonal with positive and decreasing coefficients on the diagonal. This idea is found in Heckman (2007), who interprets the assignment matrix as a sum of Cobb-Douglas technologies using a singular value decomposition in order to refine bounds on wages. This allows to introduce new indices $\tilde{x} = Ux$ and $\tilde{y} = Vy$ which are orthogonal transforms of the former, and such that the joint surplus reflects diagonal interactions of the new indices, i.e. $\Phi(x, y) = x' U' \Lambda V y = \tilde{x} \Lambda \tilde{y}$.

Here, we need to slightly adapt this idea to take advantage of the symmetry of A and of the requirement that the matrix of loadings U and V should be identical. The natural solution is the eigenvalue decomposition of A , which leads to the existence of an orthogonal loading matrix U and a diagonal $\Lambda = \text{diag}(\lambda_i)$ with nonincreasing (but not necessarily positive) coefficients on the diagonal such that

$$A = U' \Lambda U.$$

This allows us to introduce a new vector of indices $\tilde{x} = Ux$ and $\tilde{y} = Uy$, which are orthogonal transforms of the previous indices. That way, the joint surplus between individuals x and y is given by

$$\Phi(x, y) = x'U'\Lambda Uy = \tilde{x}'\Lambda\tilde{y} = \sum_{p=1}^K \lambda_p \tilde{x}^p \tilde{y}^p$$

hence this term only reflects pairwise interactions of dimension p of \tilde{x} and \tilde{y} , which are either complements (if $\lambda_p > 0$) or substitute (if $\lambda_p < 0$), and there are no complementarities across different dimensions.

The following statement formalizes this finding:

Theorem B. *Assume that $\mathbb{E}_{\hat{\pi}}[X] = 0$ and that $\text{var}_{\hat{\pi}}(X^i) = 1$ for all i . Then there exists an orthogonal loading matrix \hat{U} and a diagonal $\hat{\Lambda} = \text{diag}(\lambda_i)$ with nonincreasing coefficients on the diagonal such that*

$$\hat{A} = \hat{U}'\hat{\Lambda}\hat{U}$$

and, denoting $\tilde{x} = \hat{U}x$ and $\tilde{y} = \hat{U}y$, the estimator of the surplus function is given by

$$\hat{\Phi}(x, y) = \tilde{x}'\hat{\Lambda}\tilde{y} = \sum_{p=1}^K \lambda_p \tilde{x}^p \tilde{y}^p.$$

Proof. Because \hat{A} is symmetric, it has the following eigenvalue decomposition

$$\hat{A} = \hat{U}'\hat{\Lambda}\hat{U}$$

where \hat{U} is orthogonal, and $\hat{\Lambda} = \text{diag}(\lambda_i)$ is diagonal with nonincreasing coefficients. Denoting $\tilde{x} = \hat{U}x$ and $\tilde{y} = \hat{U}y$,

$$x' Ay = x' \hat{U}' \hat{\Lambda} \hat{U} y = \tilde{x}' \hat{\Lambda} \tilde{y} = \sum_{p=1}^K \lambda_p \tilde{x}^p \tilde{y}^p.$$

□

In the presence of categorical variables, the presence of a block λI_r reflecting assortativeness on the categorical variable in the expression (3.5) implies that the singular values of A in (3.5) will be the singular values of A in addition to λ with multiplicity r . Therefore, it is recommended to perform Saliency Analysis simply on the upper left block A .

3.4. Comparison across markets. Affinity matrices are a useful tool to analyze the marital surplus. It is very tempting to use them to compare different markets. In the present context, we would like to compare the structure of marital surplus across the homosexual and heterosexual marriage markets. A caveat is, however, in order. Indeed, recall from the above discussion that A and σ are not identified individually, but A/σ is identified; the expression of the total surplus in a given market, namely

$$\mathcal{W}(A, \sigma) = \mathbb{E}_\pi [X'AY] - \sigma \mathbb{E}_\pi [\ln \pi(X, Y)]$$

is scale invariant. However, for cross-market comparison purposes, the normalization $\sigma = 1$ which we have adopted so far can be misleading, as it assumes that the standard deviation of the heterogeneity in preferences is the same across all markets considered. This is not a very satisfying assumption. For comparison purposes, it seems plausible to work under the different assumption that the total quantity of surplus $\mathcal{W}(A, \sigma)$ is the same across markets. That is:

Assumption 3. *Assume that the average surplus on each market is the same, that is $\mathcal{W}(A, \sigma)$ is the same across every market.*

Assumption 3 allows for comparison of affinity matrices across markets. Other assumptions would be possible; for instance, one could assume that σ is the same across markets; or that $\mathbb{E}[X'AY]$ is the same across markets. However, the observable characteristics may weight more or less in the sum of social welfare across markets, and it seems more plausible to assume that the overall quantity of welfare is the same, even though the breakdown observable/unobservable may vary. Assumption 3 then allows us to impose the normalization $\mathcal{W}(A, \sigma) = \mathcal{W}^{ref}$ on each market. Therefore, in each market, we shall determine the scale k such that $\mathcal{W}(kA, k) = \mathcal{W}^{ref}$, that is $k = \mathcal{W}^{ref}/\mathcal{W}(A, 1)$. Thus, instead of the estimator \hat{A} given by (3.4), we shall use $\hat{A}/\mathcal{W}(\hat{A}, 1)$.

4. DATA

4.1. Data on same-sex couples. Empirical studies on homosexuality have traditionally needed to cope with poor data, due to the late legal recognition of their partnerships – still

unachieved in several countries – and with misreporting issues, due to social pressure on respondents⁷. Social scientists have largely relied on the data collected by the US Census Bureau for large-scale analysis of homosexuality issues (Jepsen and Jepsen, 2002; Black, Sanders, and Taylor, 2007; Schwartz and Graf, 2009). Starting from the 1990 decennial census, individuals could report themselves as “unmarried partner” within the household, regardless of their sex, so that homosexual couples could be identified. In more recent databases from the US Census Bureau, homosexual couples are still identifiable as out-of-marriage cohabiting partners. Indeed, although same-sex marriages have been officiated in some American states since 2004, they were recognized at federal level only in 2013, and currently available surveys conducted by the Census Bureau do not allow reporting marriage bonds other than the heterosexual one.

Several researchers soon realized the inaccuracy of the US Census Bureau data on homosexual couples. In particular, Black, Gates, Sanders, and Taylor (2007); DeMaio, Bates, and O’Connell (2013) and O’Connell and Gooding (2006) point out the unreliability of the 2000 decennial Census. Going back to the 1990 Census, if respondents declared themselves of the same sex and a married couple at one time, the answer was tagged as illogical and the sex of the householder’s partner was automatically *allocated* by the Bureau. However, it was argued that some homosexual couples voluntarily declared themselves as being married if their partnership concretely resembled a marriage bond. Therefore, the 1990 Census would underestimate the number of homosexual couple households.

In the 2000 decennial Census, the Bureau adopted a different *allocation* strategy in order to improve the accuracy in measuring of homosexual households. In case of a questionnaire reporting a same-sex married couple, while sex variables were not touched any more, the marital status variable was now switched to “unmarried couple”⁸. However, a relatively

⁷Underreporting is a serious issue in the study of same-sex couple (Coffman, Coffman, and Ericson, 2013) and could be a potential problem for our analysis if it causes the available sample to be biased with respect to the homosexual population. While presenting our work, we were suggested that an analysis of sorting patterns of households with only two roommates of the same sex could shed some light on the persistence of underreporting.

⁸In 2000, homosexual marriage had not been introduced in any state yet. Later on, the same strategy has been kept, since homosexual marriage was not recognized at Federal level. This provision has specifically

small measurement error in the whole population for one variable can significantly generate misclassification issues for small subgroups. In the 2000 Census case, a large share⁹ of the same-sex married couples turned out to be different-sex ones that wrongly compiled the questionnaire. As a result, estimates on the number of homosexual couple households turned out to be inflated. Since in the database it is possible to identify through an allocation flag variable those individuals whose marital status has been reallocated by the Bureau, it has been advised to exclude such observations from samples for studies on homosexuality, sometimes significantly reducing the sample size (e.g. in Schwartz and Graf (2009)). Besides, some studies argue that similar flaws affect other US Census Bureau datasets, notably the 2005-2007 American Community Survey (ACS) (Gates and Steinberger, 2009) and the 2010 decennial Census (DeMaio, Bates, and O’Connell, 2013).

To tackle this issue, the US Census Bureau implemented some improvements in the questionnaire layout and in the data editing tools, in order to minimize measurement errors on the sex of heterosexual couple partners (see US Census Bureau (2013) for further explanations). Such changes resulted in a sharp decline in the estimates on homosexual couples between 2007 and 2008, a consequence of an increase in accuracy (Gates, 2009; DeMaio, Bates, and O’Connell, 2013)¹⁰. Therefore, though not flawless, the ACS data gathered since 2008 represent the best available database among those provided by the Census Bureau to study homosexuality issues.

Accordingly, the present work relies on the five-year Public Use Microdata Sample (PUMS) for 2008-2012 coming from the ACS, conducted by the US Census Bureau. We restricted our sample to the state of California, which first legalized same-sex marriage on June 16, 2008 following a Supreme Court of California decision, and then - after some judicial and

been imposed since 1996 by the Federal Defense of Marriage Act (DOMA) (DeMaio, Bates, and O’Connell, 2013).

⁹About three quarters of the same-sex married couples were actually different-sex married couples (Black, Gates, Sanders, and Taylor, 2007; Gates and Steinberger, 2009).

¹⁰DeMaio, Bates, and O’Connell (2013) use 2010 ACS data as a benchmark to show the inaccuracy of 2010 decennial Census data, which are collected with the older methodology.

political controversies that impeded the officialization of same-sex weddings from November 5, 2008 to June 27, 2013¹¹- another decision of the Supreme Court finally accomplished full legalization. Restricting the sample to one state allows focusing on a marriage market undergoing a uniform judicial framework. Moreover, in states where same-sex marriage is recognized, estimates on the number of married same-sex couple households are more reliable, i.e. the incidence of the measurement error is smaller (Gates, 2010; Virgile, 2011).

The sample is limited to those individuals involved in a cohabiting partnership, both married and unmarried, thus excluding singles but also couples whose partners live far from each other. Each couple is identified as a householder with his/her partner, where both share the same ID household number.

Furthermore, we restrict the number of couples to those where both partners are between 25 and 50 in age and exclude those who have not completed their studies yet. The patterns of observed couple characteristics are subject to attribute changes over time and to a selection effect through partnership dissolution (Schwartz and Graf, 2009). The PUMS cross-sectional data allows describing a static situation in a fixed point in time, without following couples over time. It is therefore appropriate to restrict the sample to those couples that formed recently to limit the effects of time variations. In addition, Alexander, Davern, and Stevenson (2010) call attention to the correctness of US Census Bureau gender data for individuals aged 65 or older: since the gender dummy is crucial in the construction of the sample, excluding the elderly should boost the reliability of the data.

4.2. Descriptive statistics. The main database is composed of 681,060 individuals in couples who have completed their schooling. The restriction to prime age couples (25-50 year old) reduces our sample to 285,546 individuals. 3,654 couples (1.28% of the sample) live in same-sex couples, of which 2,034 live in male couples (0.71 %) and 1,620 live in female couples (0.57 %). 87.39% of the sample are married heterosexuals and 11.33 % are cohabiting heterosexuals. For estimation purposes, after randomly selecting a subsample of

¹¹In this period, marriage licenses issued to same-sex couples held their validity.

the different-sex couples set¹², a total of 9,820 couples are considered, of which 4,959 are married and 4,799 are not.

To compare different marriage markets, following Jepsen and Jepsen (2002), the main sample is divided into four subsamples: same-sex male couples, same-sex female couples, different-sex unmarried couples and different-sex married couples. This repartition is based on the assumption that individuals enter into separate markets according to their sexuality. However, another criterion is used to differentiate two of the subgroups: married and unmarried heterosexual couples are treated as two separate subpopulations¹³, since empirical evidence has reported significant differences in patterns between these two kinds of partnership (Jepsen and Jepsen, 2002; Schwartz and Graf, 2009). Although it is impossible to know *a priori* if a person is interested in a marital union rather than in a less binding relationship, this repartition can be of great interest and deepen the analysis. Nevertheless, even if California represents the larger state-level ACS sample in the US, further splitting the male and female homosexual groups into two parts unfortunately implies working with relatively small samples. Moreover, although same-sex marriage is permitted, it has been recognized only recently and at the end of many legal struggles, which may have prevented a part of those same-sex couples that wished to marry from doing so. Whenever the number of registered homosexual partnerships increased and the share of married couples were known with more certainty, then considering married and unmarried homosexual couples separately would be extremely interesting, as proved by recent research of Verbakel and Kalmijn (2014).

This study takes into consideration several variables, some related to the labor market and some others to the general background. Non-labor market traits include age, education and race. Age and education are treated as continuous variables, with the latter defined as the highest schooling level attained by the individual. Thanks to the detailed data of the ACS, the variable has been built in order to reflect as many distinguished educational stages

¹²We randomly select 4% of married couples and 30% of unmarried couples.

¹³See Mourifié and Siow (2014) for a very interesting discussion of the endogenous choice of the form of marital relationship.

as possible. As considers race and ethnicity, we consider five large racial/ethnic groups: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian, Hispanic and Others¹⁴.

Finally, among labor market variables, we compute and include hourly wage¹⁵. Note that yearly wage is top-coded for very high values (over \$999,999). Moreover, the usual amount of hours worked per week is included.

Table 1 presents some descriptive statistics of our sample. Individuals in same-sex couples are on average more educated than individuals in opposite-sex couples. As observed by Black, Sanders, and Taylor (2007), young homosexual women are much more likely to be part of the labor force than heterosexual ones and also have higher wages. We observe that unmarried heterosexual couples are much younger than married couples and same-sex couples. Among the 25-50 year-old, unmarried heterosexual men and women are on average four year younger than others. Cohabitation is often (but not always) a “trial” period before marriage, which explains the age difference. Table 2 presents the distribution of ethnics among couples: same-sex couples are much more present among White people, whereas there is a remarkably high share of Black individuals among lesbian couples. On the contrary, Asians and Hispanics are under-represented in the homosexual population.

Table 3 presents the correlation rates among traits. It shows that age and education attainment are much more correlated among heterosexual couples than among homosexual ones. Moreover, the correlation is stronger for lesbian couples than for gay ones. Education is also more correlated among young couples than among older ones. Correlations on labor market outcomes are particularly interesting: hours worked are negatively correlated only for different-sex couples, a possible clue of stronger household specialization, whereas the

¹⁴American demographic institutions do not include a Hispanic category in variables on race, furnishing a separate variable for Hispanic origins, which is why there is some overlapping and the other categories bear the specification “Non-Hispanic”. The issue concerns the conceptual differences of “race” and “ethnicity”. See for instance Rodriguez (2000) for clarifications.

¹⁵The variable is computed as follows: we divide yearly wage by 52 in order to have the average weekly wage for last year and then we divide it again by the usual number of hours worked per week, which is available in the dataset. The hourly wage is partly approximated because the exact number of weeks worked in the last 12 months is not available.

correlation is low and positive for same-sex couples. On the other hand, wages display a positive correlation in every market, with male same-sex couples exhibiting the lowest correlation.

Table 4, 5 and 6 present the homogamy rates of couples with respect to race for different types of couples. The homogamy rate is the ratio between the observed number of couples of a certain type and the counterfactual number which should be observed if individuals formed couples randomly¹⁶. For instance, Table 4 shows that lesbian couples among black women form 10 times much more than if they were formed randomly among the lesbian population.

5. RESULTS

Homogamy rates and correlations presented in section 4 are interesting measures of assortative mating and provide a good starting point for our analysis. However, no conclusion on preferences can be drawn from their observation. In fact, as it was already noted by Becker (1973), we need to know how specific traits interact when all the others are held constant. Our structural approach allows to analyze the degree of assortativeness for each couple of traits *ceteris paribus*.

Moreover, bear in mind that, since we directly estimate the parameters of the systematic surplus function contained in our matching model, it is implied that differences in assortative mating reveal a different structure of preferences. We interpret assortativeness as a consequence of the structure of the household production function - as meant by Becker (1973) - rather than as a consequence of search dynamics (notably, geographic factors and search frictions).

We report in appendix the estimates of the affinity matrix for male homosexuals in Table 7, for female homosexuals in Table 12, for cohabiting heterosexuals in Table 17 and for married heterosexuals in Table 21.

¹⁶If couples formed randomly across the population, the number of couples between a man of type i and a woman of type j would be equal to $n_i n_j / N$. The homogamy rate is then the ratio of the observed number of couples on this theoretical number of couples. See Vanderschelden (2006).

5.1. Age, education and race/ethnicity. First of all, our estimates of the diagonal elements of the affinity matrices are highly positive and significant for age, education and ethnicity, which strongly confirms the positive assortative mating observed in the literature. The complementarity in these non-labor market traits is once again empirically assessed. In line with the results by Jepsen and Jepsen (2002) and Schwartz and Graf (2009), we find that heterosexuals have a stronger preference for homogamy with respect to age and ethnicity. Nonetheless, our structural estimation reveals that sorting on education is stronger on the same-sex marriage market.

Going into detail, we find that the intensity of both affinity by age and preference for ethnic homogamy are the lowest for male same-sex couples (both equal to 0.64). We observe intermediate values for female same-sex couples (0.81 for age, 1.30 for ethnicity) and unmarried different-sex couples (1.19, 1.97), whereas married different-sex couples exhibit the strongest homogamous preferences (2.25, 2.42). Contrarily, education plays a more relevant role on homosexual marriage markets: complementarity of schooling levels is the strongest for lesbian couples (1.23), followed by gay couples (0.87). Affinity by education is weaker for married heterosexuals (0.85) and is the lowest for unmarried heterosexuals (0.66).

Our estimates on the level of educational sorting not only contradicts previous findings, but seem at odds with the theoretical predictions formulated by the social sciences literature¹⁷ that homosexuals have a weaker preference for socio-economic homogamy. Nevertheless, one possible explanation to our unexpected result may be found in childrearing patterns: heterosexuals, who are more likely to have children¹⁸, may be more specialized also with respect to education. As we discuss in the next section, the household specialization process should mainly concern labor-market variables. Nonetheless, it is possible that, in different-sex couples, one spouse - typically the wife - may be more likely to give up on her studies in anticipation of her role of child-bearer.

¹⁷The main reference works about mating among homosexuals are listed in our introduction. We refer to Schwartz and Graf (2009) and Verbakel and Kalmijn (2014) who, drawing from literatures from different social sciences, both provide a complete and updated review on this topic

¹⁸In our sample, among the 25-50 years-old, 14.5% of gays have children, 37.8% of lesbians, 58.04% of cohabiting different-sex couples and 83.5% of married different sex couples.

In order to ease the interpretation of our findings, we also estimate the affinity matrix considering only childless couples as a robustness check¹⁹ (tables 11, 16, 24 and 20). Interestingly, we find that, with respect to age and ethnicity, both childless homosexuals and heterosexuals have weaker homogamous preferences than their respective counterparts with children. It seems therefore that those individuals who plan to have children look for a more similar partner than those who do not. On the contrary, sorting on education is stronger for childless couples and the former “ranking” on complementarity in education is reversed. Childless married heterosexuals exhibit the strongest taste for educational homogamy (1.11), followed by lesbians (0.94), unmarried heterosexuals (0.92) and gays (0.67). In addition, complementarity in education drops to 0.89 for married same-sex couples with only one child and to 0.69 for those with three children (tables 25 and 26). These additional findings support the previous observation that complementarity in education is weaker on marriage markets where childrearing plays a more relevant role.

5.2. Labor market traits. To describe labor market traits, we must be very cautious as these outcomes are potentially endogenous. Since we do not observe these traits at the moment of the match formation but possibly much later, the specialization process at work in couples may have already begun. In particular, we expect that this specialization effect is strong in heterosexual couples, who are more likely to have children²⁰. Raising children takes time and most of mothers leave the labor force or reduce their working hours. Consequently, because of interrupted careers and less paid part-time jobs, their hourly wage does not rise as much as that of their male counterparts and we observe many associations between low-wage women and high-wage men. This phenomenon could bias our estimates and we cannot interpret them directly as preference estimates. However, the differences we observe

¹⁹We are aware that this is an “artificial” marriage market, since it does not make much sense to state that couples with children formed on a different market. However, our model does not have a specification that explicitly accounts for choices related to childbearing.

²⁰Antecol and Steinberger (2013) and Jepsen and Jepsen (2015) showed that to a lesser extent some household specialization also exists within same-sex households. Moreover, Antecol and Steinberger (2013) stress that childless different-sex couples are less specialized and thus more similar to same-sex couples

between the estimates for our four types of couples help us to shed light on the specialization effect, as we now argue.

First, we observe a significant positive assortative mating on hourly wages for all types of couples although the coefficient is higher for same-sex couples (0.06) and for unmarried couples (0.12) than for married heterosexual couples (0.03). Furthermore, we do not observe positive assortative mating on working hours for married heterosexual couples, whereas we observe much higher and significant positive estimates for same-sex couples (0.21 for female same-sex couples and 0.12 for male same-sex couples). The coefficient for unmarried couples is also positive and significant (0.10). As unmarried couples are more often young couples, these estimates must be much closer to the true preference estimates as the specialization process has not had time to happen yet²¹. Estimates must also identify preference parameters for same-sex couples as homosexuals are less likely to have children and consequently have weaker incentives to specialize. When estimating our model on childless couples, we find that married childless couples have a much higher and significant coefficient for both wage (0.14) and working hours (0.19). Our results are in line with those of Jepsen and Jepsen (2015). It is worth noting that the estimate for working hours is much higher for lesbians than for gays. Homosexual women may prefer a partner with similar time schedule.

The cross-estimate between the wage of one partner and working hours of the other partner is also very interesting to analyze, although we may not be able to interpret it as a preference parameter. Instead, it must represent the well-known income effect at work in couples: when the wage of an individual rises, the household's income rises and the partner of this individual is free to work less. The estimate is highly negative for homosexual women (-0.20) and heterosexual married women (-0.14). It is also negative and significant for homosexual men (-0.07) and heterosexual married men (-0.18). It seems that same-sex couples and married couples coordinate their labor supply: they pool a part of their income

²¹Unmarried couples can also be long dating couples who do not want to marry ever. Still, in that case, the specialization is not as strong as in married couples, since they might not commit to the community as strongly as married individuals.

and adjust their work in reaction to variations of the labor market traits of their partner²². However, the coefficient is non-significant for unmarried couples. As unmarried couples are on average younger, their income effect is weaker since they do not pool their income yet and are more likely to stay financially independent from their respective partners.

5.3. Cross-interactions and symmetry. Other significant positive cross-effects have been found for some elements of the affinity matrix that lie off the diagonal. The parameter capturing the interaction between wage and education is persistently high and positive. This suggests that higher wage individuals have a preference for more educated partners, keeping constant their wage and all other characteristics. Once again, same-sex couples exhibit the weakest affinity between these two variables (0.13 for gays and 0.20 for lesbians). Complementarity between the two inputs is stronger for unmarried and married heterosexuals, but both markets exhibit asymmetry in this cross-interaction: for married couples, a joint increase of husband's wage and wife's education generates higher surplus than the other way around (0.47 against 0.38), whereas the opposite holds for unmarried couples (0.27 against 0.52). The complementarity between the two inputs can be explained in two ways. First of all, as a matter of preferences, in that high-income individuals - independently of their educational level - may enjoy the company of cultured partners, whereas the latter benefit from matching with a partner with a high wage. A second possible explanation is related to the impact of endogenous labor market choices on these estimates. A positive association of wage and education might simply resemble the complementarity between partners' wages, especially when one of them is younger and has just entered the job market. Similarly, we might have that one partner quits the labor market (for instance, after having a child) despite his/her earning potential ensured by a relatively high schooling level.

Another cross-interaction that arises from the estimation is the substitutability between age and hours worked on same-sex marriage markets (the estimates are equal to -0.14 for gays and -0.10 for lesbians). This interaction might be looked at as the equilibrium outcome of the family's bargaining dynamics, as explained by Oreffice (2011): younger

²²Other similarities between same-sex couples and heterosexual couples household dynamic has also been found in Oreffice (2011), in which the author shows that individuals in same-sex couples bargain over the income as heterosexuals and respond to bargaining power shifts.

partners enjoy higher bargaining power and thus can afford reducing their labor supply. Interestingly, unmarried different-sex couples exhibit similar patterns, although the effect is weaker and present only in one direction: women can afford reducing their labor supply when cohabiting with older partners, while men cannot. We observe exactly the same patterns for married different-sex couples only when we consider the subsample of childless couples, whereas on the whole different-sex marriage market we observe an opposite trend: older married women are optimally matched with men that are ready to increase their labor supply, a pattern that may be also linked with childrearing.

Finally, as a last robustness check for our main results, we estimate a bipartite matching model of the same-sex marriage market (tables 10 and 15). The purpose of this exercise is to understand whether the resulting affinity matrix is symmetric when symmetry is not required by the model. However, note that, to run a bipartite estimation, we need to define two separate subpopulations. On one side of the market, we group all those homosexuals that are registered as “householders”, whereas on the other we group their “cohabiting partners”. This repartition is highly artificial, since it implies that two homosexuals that are householders before finding a partner can never match: in general, it seems implausible to divide the same-sex population in two separate subgroups with the data that we have at hand. Nonetheless, it is interesting to check - under the strict assumption of predetermined roles - whether some asymmetry in cross-interactions occurs. We observe that the affinity matrices for both male and female homosexuals (respectively Tables 10 and 15) are not much different than in the unipartite case²³. The main exception is the cross-interaction between age and hours: for gays, the interaction appears only in one direction, with relatively young cohabiting partners that can reduce their labor supply when matching with an older householder. The effect is only present in the opposite direction in lesbian couples: it is now relatively young householders that benefit from a shift in bargaining power.

5.4. Matching on unobservables. Thanks to our identification assumption, we can evaluate the parameter σ for each market. As anticipated in section 2, this parameter has a simple interpretation: the higher σ , the more the matching appears as random to the

²³We do not run a formal test, although this would be possible and possibly very useful in other situations.

econometrician. Since our observable characteristics are meant to capture the main socio-economic traits, we expect that a higher σ implies that mating is less “deterministic”: indeed, we expect that, for higher σ , the socio-economic background of an individual matters relatively less, whereas other unobservable traits (e.g. personality or physical appearance) may matter relatively more.

We find that σ is higher on same-sex marriage markets (1.22 for gays and 1.19 for lesbians), whereas it is lower on different-sex marriage markets (1.03 for unmarried couples and 0.98). Hence, it seems that socio-economic background matters more among heterosexuals, and even more among those who are married.

5.5. Saliency analysis. A way to bring further insights to the main drivers of preferences of individuals over different characteristics is to decompose the affinity matrix in orthogonal dimensions. As detailed in the estimation part we will do the decomposition analysis on all variables but race. We present such decomposition in the appendix in Tables 8, 9, 13, 14, 18, 19, 22 and 23. In the four markets, we show that more than 90% of the joint surplus left could be explained in two orthogonal dimensions that we could name “indices of attractiveness” as in DG. These indices load on different characteristics of individuals. For all markets, the first index is almost only composed of age. It explains by itself around 50-60% of the surplus. Then the second dimension of sorting relies mostly on education for all markets and explains around 30-40%.

When we consider heterosexual couples, the indices of mutual attractiveness could differ between genders. For married heterosexual individuals, the education/wage index (second dimension) loads positively twice as much on wages for men than for women, whereas there is a penalty on working hours for women that does not appear for men.

Separately, we compute the share of the systematic part of the total surplus that is attributable to race (table 27). We find this share reaches more than 45% for heterosexual couples, 42% for lesbian couples and only 33% for gays.

6. DISCUSSION AND PERSPECTIVES

The contributions of the present paper are twofold. From a methodological point of view, this paper is the first to propose a tractable theoretical framework for same-sex marriage. Our methodology could be applied to many other markets (e.g. roommates, teammates, co-workers). In addition, we apply the model in order to provide an empirical analysis of mating preferences in the same-sex marriage market in California. We conduct a cross-market comparison: we analyze the heterogeneity in preferences between homosexual and heterosexual couples. First, we find that, as concerns age and ethnicity, the heterosexual population has a stronger “preference for homogamy” than the homosexual one, whereas the inverse is true as concerns sorting on education. Second, we discuss the differences in complementarity and substitutability of inputs in the household production function as defined in the theory of the family put forward by Gary Becker. Our findings seem to suggest that labor market traits are substitutes for married heterosexual couples but complementary for other types of couples. This result challenges the traditional concept of the marriage gain based on specialization within the couple.

The need for effective analytical frameworks to study and describe relatively modern forms of families has recently emerged in the economic literature, both as concerns same-sex couples (Black, Sanders, and Taylor, 2007; Orefice, 2011) and cohabiting partners (Steven-son and Wolfers, 2007; Gemici and Laufer, 2011). In this paper, we first identified three separate subpopulations according to sexual preferences, and then we separately analyzed the matching of married couples and of unmarried ones. We found preferences disparities between the four markets. However, can we state with certainty that these markets are mutually exclusive? In fact, individuals may first endogenously choose into which market they are willing to match. Moreover, there could be spillovers between markets.

In this paper, we show that simple cohabitation and marriage correspond to different preferences for mating and household organization, which is what has also been observed in the literature (Schwartz and Graf, 2009; Gemici and Laufer, 2011; Verbakel and Kalmijn, 2014). Cohabitation is a developing phenomenon and is associated with a lower degree of specialization and a lower degree of positive assortative mating. A promising area of

research would be to understand the preferences for marriage or cohabitation jointly with sorting preferences. Mourifié and Siow (2014) set a first model in that direction for heterosexual couples. An empirical paper of Verbakel and Kalmijn (2014) separately analyzes the marriage and cohabitation markets in Netherlands also for homosexuals. In the same vein, Aldén, Edlund, Hammarstedt, and Mueller-Smith (2013) study the implications of the legal recognition of same-sex unions on fertility and labor market outcomes of homosexuals. With the consolidation of the same-sex marriage and the availability of more and more accurate data, it will soon be possible to expand our understanding of differences and similarities across markets. Families and household arrangements are evolving quickly and we need to understand the underlying forces of these changes.

One issue that needs to be investigated is the possible presence of spillovers between markets. Many opponents of same-sex marriage fear that it will cause the marriage institution to lose its value and favor alternative forms of families, typically more flexible/less stable, such as cohabitation. For now, researchers have found no effect of same-sex marriage on the number of different-sex marriages or on the number of divorces (Trandafir, 2014, 2015). However, we wonder whether the legal recognition of same-sex marriage could somehow impact the preferences observed on the different markets. What changes should we expect in the behavior of heterosexuals? And could it be that same-sex couples become more homogamous as homosexual marriage is institutionalized?

Finally, in Becker’s theory, a rationale for marriage is the home production complementarities between men and women skills. However, the traditional gains from marriage have diminished for two main reasons. First, the progress in home technology has decreased the value of domestic production; second, as women took control over their fertility and have been getting more and more educated, their opportunity cost to stay at home has increased (Stevenson and Wolfers, 2007; Greenwood, Guner, Kocharkov, and Santos, 2012). Despite the decrease in the traditional marriage gains, the institution of marriage has not disappeared. On the contrary, there has been a high demand for same-sex legal marriage in many developed countries. Stevenson and Wolfers (2007) argue that individuals now look for a mate with whom they “share passions” and the new rationale for marriage is now “consumption complementarities” instead of “production complementarities”. It is

also possible that the act of marriage itself is still considered as intrinsically valuable for cultural and social reasons. In any case, this evolution may lead to even higher correlation of traits. Time will tell how these changes will impact macroeconomic outcomes, life quality and social distance among individuals.

^b *Economics Department, Sciences Po, Paris.*

Address: Sciences Po, Department of Economics, 27 rue Saint-Guillaume, 75007 Paris, France. Email: edoardo.ciscato@sciencespo.fr.

[†] *New York University, Economics Department and CIMS, and Economics Department, Sciences Po, Paris.*

Address: NYU Economics, 19W 4th Street, New York, NY 10012, USA. Email: ag133@nyu.edu.

[§] *Economics Department, Laval University.*

Address: Université Laval, Département d'Economie, Québec, (QC) G1V 0A6, Canada. Email: marion.gousse@ecn.ulaval.ca.

REFERENCES

- ALDÉN, L., L. EDLUND, M. HAMMARSTEDT, AND M. MUELLER-SMITH (2013): “Domestic Partnership What a Difference a Yay Makes,” *mimeo*.
- ALEXANDER, J. T., M. DAVERN, AND B. STEVENSON (2010): “The Polls—Review Inaccurate Age and Sex Data in the Census Pums Files: Evidence and Implications,” *Public opinion quarterly*, 74(3), 551–569.
- ANDERSSON, G., T. NOACK, A. SEIERSTAD, AND H. WEEDON-FEKJÆR (2006): “The demographics of same-sex marriages in Norway and Sweden,” *Demography*, 43(1), 79–98.
- ANTECOL, H., AND M. D. STEINBERGER (2013): “Labor Supply Differences between Married Heterosexual Women and Partnered Lesbians: A Semi-Parametric Decomposition Approach,” *Economic Inquiry*, 51(1), 783–805.
- BECKER, G. (1973): “A theory of marriage: Part I,” *Journal of Political economy*, 81(4).
- (1981): “A Treatise on the Family, 1991,” *Enlarged Edition (Cambridge, Mass.: Harvard University Print)*.

- BLACK, D., G. GATES, S. SANDERS, AND L. TAYLOR (2002): “Why do gay men live in San Francisco?,” *Journal of Urban Economics*, 51(1), 54–76.
- (2007): “The measurement of same-sex unmarried partner couples in the 2000 US Census,” *California Center for Population Research*.
- BLACK, D. A., S. G. SANDERS, AND L. J. TAYLOR (2007): “The economics of lesbian and gay families,” *The Journal of Economic Perspectives*, 21(2), 53–70.
- BLUMSTEIN, P., P. SCHWARTZ, ET AL. (1983): *American couples: Money, work, sex*. Morrow New York.
- CANDELON, B., AND A. DUPUY (2014): “Hierarchical Organization and Performance Inequality: Evidence from Professional Cycling,” *International Economic Review*, Forthcoming.
- CHIAPPORI, P. A., A. GALICHON, AND B. SALANIÉ (2012): “The Roommate Problem Is More Stable Than You Think,” .
- 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.
- CHIAPPORI, P.-A., B. SALANIÉ, AND Y. WEISS (2011): “Partner Choice and the Marital College Premium,” *Discussion Papers*.
- CHOO, E., AND A. SIOW (2006): “Who Marries Whom and Why,” *Journal of Political Economy*, 114, 175–201.
- COFFMAN, K. B., L. C. COFFMAN, AND K. M. M. ERICSON (2013): “The Size of the LGBT Population and the Magnitude of Anti-Gay Sentiment are Substantially Underestimated,” Discussion paper, National Bureau of Economic Research.
- DAGSVIK, J. K. (1994): “Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes,” *Econometrica: Journal of the Econometric Society*, pp. 1179–1205.
- DEMAIO, T. J., N. BATES, AND M. O’CONNELL (2013): “Exploring measurement error issues in reporting of same-sex couples,” *Public Opinion Quarterly*, 77(S1), 145–158.
- DUPUY, A., AND A. GALICHON (2014): “Personality Traits and the Marriage Market,” Discussion paper, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.

- EIKA, L., M. MOGSTAD, AND B. ZAFAR (2014): “Educational Assortative Mating and Household Income Inequality,” Discussion paper, National Bureau of Economic Research.
- ESKRIDGE JR, W. N. (1993): “A history of same-sex marriage,” *Virginia Law Review*, pp. 1419–1513.
- FOX, J. T. (2010): “Identification in matching games,” *Quantitative Economics*, 1(2), 203–254.
- GALICHON, A., AND B. SALANIÉ (2014): “Cupid’s invisible hand: Social surplus and identification in matching models,” .
- GATES, G. (2009): “Same-sex spouses and unmarried partners in the American Community Survey, 2008,” *The Williams Institute*.
- GATES, G. J. (2010): “Same-sex couples in US Census Bureau Data: Who gets counted and why,” .
- GATES, G. J., AND M. STEINBERGER (2009): “Same-sex unmarried partner couples in the American Community Survey: the role of misreporting, miscoding and misallocation,” in *Annual Meetings of the Population Association of America, Detroit, MI*.
- GEMICI, A., AND S. LAUFER (2011): “Marriage and cohabitation,” *New York University, mimeo*.
- GOLDIN, C., L. F. KATZ, AND I. KUZIEMKO (2006): “The Homecoming of American College Women: The Reversal of the College Gender Gap,” *The Journal of Economic Perspectives*, pp. 133–156.
- GREENWOOD, J., N. GUNER, G. KOCHARKOV, AND C. SANTOS (2012): “Technology and the changing family: A unified model of marriage, divorce, educational attainment and married female labor-force participation,” Discussion paper, National Bureau of Economic Research.
- (2014): “Marry your like: Assortative mating and income inequality,” Discussion paper, National Bureau of Economic Research.
- HARRY, J. (1984): *Gay couples*. Praeger.
- HAYES, A. F. (1995): “Age preferences for same-and opposite-sex partners,” *The Journal of social psychology*, 135(2), 125–133.

- HECKMAN, J. (2007): “Notes on Koopmans and Beckmann’s ”Assignment Problems and the Location”,” Lecture notes, University of Chicago.
- JEPSEN, C., AND L. K. JEPSEN (2015): “Labor-Market Specialization within Same-Sex and Difference-Sex Couples,” *Industrial Relations: A Journal of Economy and Society*, 54(1), 109–130.
- JEPSEN, L. K., AND C. A. JEPSEN (2002): “An Empirical Analysis of the Matching Patterns of Same-Sex and Opposite-Sex Couples,” *Demography*, 39(3), 435–453.
- KENRICK, D. T., AND R. C. KEEFE (1992): “Age preferences in mates reflect sex differences in human reproductive strategies,” *Behavioral and Brain Sciences*, 15(01), 75–91.
- KURDEK, L. A., AND J. P. SCHMITT (1987): “Partner homogamy in married, heterosexual cohabiting, gay, and lesbian couples,” *Journal of Sex Research*, 23(2), 212–232.
- LOW, C. (2013): “Pricing the Biological Clock: Reproductive Capital on the US Marriage Market,” Discussion paper, November 2013. Mimeo.
- MEIER, A., K. E. HULL, AND T. A. ORTYL (2009): “Young adult relationship values at the intersection of gender and sexuality,” *Journal of Marriage and Family*, 71(3), 510–525.
- MENZEL, K. (2013): “Large Matching Markets as Two-sided Demand Systems,” Discussion paper, working paper.
- MOURIFIÉ, I. Y., AND A. SIOW (2014): “Cohabitation vs Marriage: Marriage matching with peer effects,” Discussion paper, University of Toronto.
- O’CONNELL, M., AND G. GOODING (2006): “The use of first names to evaluate reports of gender and its effect on the distribution of married and unmarried couple households,” *Population Association of America, Los Angeles, March*.
- OREFFICE, S. (2011): “Sexual orientation and household decision making.: Same-sex couples’ balance of power and labor supply choices,” *Labour Economics*, 18(2), 145–158.
- RODRIGUEZ, C. E. (2000): *Changing race: Latinos, the census, and the history of ethnicity in the United States*. NYU Press.
- ROSENFELD, M. J., AND B.-S. KIM (2005): “The independence of young adults and the rise of interracial and same-sex unions,” *American Sociological Review*, 70(4), 541–562.
- SCHWARTZ, C. R., AND N. L. GRAF (2009): “Assortative Matching Among Same-Sex and Different-Sex Couples in the United States, 1990–2000,” *Demographic research*, 21, 843.

- STEVENSON, B. (2008): “The Evolution of the American Family: An Economic Interpretation,” *American Journal of Family Law*, 22(3), 141–147.
- STEVENSON, B., AND J. WOLFERS (2007): “Marriage and Divorce: Changes and Their Driving Forces,” *The Journal of Economic Perspectives*, pp. 27–52.
- SULLIVAN, A. (2009): *Same-sex marriage: Pro and con*. Random House LLC.
- TRANDAFIR, M. (2014): “The effect of same-sex marriage laws on different-sex marriage: Evidence from the Netherlands,” *Demography*, 51(1), 317–340.
- (2015): “Legal recognition of same-sex couples and family formation.,” *Demography*, Forthcoming.
- US CENSUS BUREAU (2013): “Frequently Asked Questions, About Same-Sex Couple Households,” Discussion paper, US Census Bureau, Fertility and Family Statistics Branch.
- VANDERSCHelden, M. (2006): “Homogamie socioprofessionnelle et ressemblance en termes de niveau d’études: constat et évolution au fil des cohortes d’unions,” *Économie et statistique*, 398(1), 33–58.
- VERBAKEL, E., AND M. KALMIJN (2014): “Assortative Mating Among Dutch Married and Cohabiting Same-Sex and Different-Sex Couples,” *Journal of Marriage and Family*, 76(1), 1–12.
- VIRGILE, M. (2011): “Measurement error in the relationship status of same-sex couples in the 2009 American Community Survey,” Discussion paper, ACS11-RER-10, US Government Printing Office, Washington, DC.
- WEISS, Y., J. YI, AND J. ZHANG (2013): “Hypergammy, Cross-Boundary Marriages, and Family Behavior,” .

APPENDIX A. DESCRIPTIVE STATISTICS

Type of couples	Age	Education	Wage*	Hours	Sample size	Share
All						
<i>Married Heterosexuals</i>						
Men	53.14	12.42	31.24	42.26	306798	90.09 %
Women	50.49	12.28	22.56	35.76	306798	
<i>Unmarried Heterosexuals</i>						
Men	41.54	11.39	19.48	40.74	29058	8.53 %
Women	39.87	11.63	18.02	37.57	29058	
<i>Homosexuals</i>						
Men	49.99	13.82	33.45	40.94	5197	0.76 %
Women	50.00	13.69	28.35	39.92	4150	0.61 %
25-50 year old						
<i>Married Heterosexuals</i>						
Men	40.22	12.33	31.34	43.60	124772	87.39 %
Women	38.37	12.47	22.84	36.16	124772	
<i>Unmarried Heterosexuals</i>						
Men	36.31	11.17	19.79	41.33	16174	11.33 %
Women	34.84	11.55	18.45	38.30	16174	
<i>Homosexuals</i>						
Men	40.00	13.93	35.19	42.71	2034	0.71 %
Women	39.35	13.78	28.44	40.74	1620	0.57 %

Table 1: Sample means

Ethnic	Heterosexual	Gay	Lesbian	All
White	42.7	67.3	63.5	43.0
Black	2.9	2.5	5.2	2.9
Others	0.6	0.8	1.2	0.6
Asian	16.7	8.8	5.8	16.6
Hispanic	37.1	20.7	24.3	36.9
Total	100.0	100.0	100.0	100.0

Table 2: Race (25-50 year old)

Type of couples	Age	Education	Wage	Hours
All				
Heterosexual couples	0.82	0.55	0.11	-0.02
Gay couples	0.58	0.31	0.05	0.07
Lesbian couples	0.66	0.34	0.20	0.07
25-50 year old				
Heterosexual couples	0.76	0.70	0.17	-0.10
Gay couples	0.56	0.56	0.15	0.03
Lesbian couples	0.66	0.65	0.20	0.08

Table 3: Couple correlations

	White	Black	Others	Asian	Hispanic
White	1,28	0,39	0,74	0,70	0,47
Black	0,39	10,67	1,00	0,82	0,53
Others	0,74	1,00	20,00	0,91	0,87
Asian	0,70	0,82	0,91	8,00	0,13
Hispanic	0,47	0,53	0,87	0,13	2,69

Table 4: Homogamy rates of lesbians (25-50 year old)

	White	Black	Others	Asian	Hispanic
White	1,12	0,79	0,46	0,94	0,67
Black	0,79	12,31	0,00	0,44	0,57
Others	0,46	0,00	40,00	2,14	1,21
Asian	0,94	0,44	2,14	3,08	0,33
Hispanic	0,67	0,57	1,21	0,33	2,39

Table 5: Homogamy rates of gays (25-50 year old)

Men	Women				
	White	Black	Others	Asian	Hispanic
White	1,96	0,32	0,87	0,37	0,28
Black	0,49	24,02	1,38	0,34	0,36
Others	0,84	0,62	60,91	0,40	0,46
Asian	0,15	0,08	0,27	5,08	0,07
Hispanic	0,26	0,16	0,44	0,09	2,32

Table 6: Homogamy rates of heterosexuals (25-50 year old)

APPENDIX B. GAYS

	Age	Education	Wage	Hours	Race
Age	0.64	-0.06	-0.02	-0.14	0.00
Education	-0.06	0.87	0.13	-0.07	0.00
Wage	-0.02	0.13	0.05	-0.07	0.00
Hours	-0.14	-0.07	-0.07	0.12	0.00
Race	0.00	0.00	0.00	0.00	0.64
σ	1.22				

Table 7: Affinity matrix for gays (25-50 year old, 1,017 couples)

	I1	I2	I3	I4
Share of surplus	54.07	37.31	7.88	0.74
(Standard deviation)	(3.39)	(3.65)	(1.84)	(2.12)

Table 8: Shares of systematic surplus for gays (25-50 year old, 1,017 couples)

	I1	I2	I3
Age	0.97	-0.11	0.16
Education	-0.15	-0.94	0.30
Wage	-0.04	-0.27	-0.59
Hours	-0.19	0.19	0.74

Table 9: Indices of attractiveness for gays (25-50 year old, 1,017 couples)

	Partner				
Head	Age	Education	Wage	Hours	Race
Age	0.62	-0.06	-0.00	-0.17	0.00
Education	0.00	0.83	0.12	-0.02	0.00
Wage	-0.01	0.16	0.06	-0.07	0.00
Hours	-0.05	-0.09	-0.04	0.11	0.00
Race	0.00	0.00	0.00	0.00	0.67
σ	1.32				

Table 10: Affinity matrix from bipartite estimation for gays (25-50 year old, 1,017 couples)

	Age	Education	Wage	Hours	Race
Age	0.58	-0.08	-0.04	-0.11	0.00
Education	-0.08	0.67	0.11	-0.02	0.00
Wage	-0.04	0.11	0.06	-0.07	0.00
Hours	-0.11	-0.02	-0.07	0.18	0.00
Race	0.00	0.00	0.00	0.00	0.47
σ	1.26				

Table 11: Affinity matrix for childless gays (25-50 year old, 870 couples)

APPENDIX C. LESBIANS

	Age	Education	Wage	Hours	Race
Age	0.81	0.04	0.05	-0.10	0.00
Education	0.04	1.23	0.20	-0.01	0.00
Wage	0.05	0.20	0.06	-0.19	0.00
Hours	-0.10	-0.01	-0.19	0.21	0.00
Race	0.00	0.00	0.00	0.00	1.30
σ	1.19				

Table 12: Affinity matrix for lesbians (25-50 year old, 810 couples)

	I1	I2	I3	I4
Share of surplus	49.30	35.09	12.10	3.51
(Standard deviation)	(3.20)	(3.68)	(1.21)	(1.66)

Table 13: Shares of systematic surplus for lesbians (25-50 year old, 810 couples)

	I1	I2	I3
Age	0.98	0.17	0.14
Education	0.15	-0.97	0.15
Wage	0.09	-0.19	-0.50
Hours	-0.13	0.03	0.84

Table 14: Indices of attractiveness for lesbians (25-50 year old, 810 couples)

	Partner				
Head	Age	Education	Wage	Hours	Race
Age	0.76	0.02	0.06	-0.06	0.00
Education	0.06	1.22	0.18	0.05	0.00
Wage	0.03	0.18	0.08	-0.25	0.00
Hours	-0.11	-0.07	-0.16	0.22	0.00
Race	0.00	0.00	0.00	0.00	1.22
σ	1.29				

Table 15: Affinity matrix from bipartite estimation for lesbians (25-50 year old, 810 couples)

	Age	Education	Wage	Hours	Race
Age	0.78	0.08	0.03	-0.09	0.00
Education	0.08	0.94	0.24	-0.02	0.00
Wage	0.03	0.24	0.11	-0.20	0.00
Hours	-0.09	-0.02	-0.20	0.35	0.00
Race	0.00	0.00	0.00	0.00	1.16
σ	1.27				

Table 16: Affinity matrix for childless lesbians (25-50 year old, 504 couples)

APPENDIX D. UNMARRIED HETEROSEXUALS

	Women				
Men	Age	Education	Wage	Hours	Race
Age	1.19	-0.02	-0.06	-0.04	0.00
Education	-0.11	0.66	0.52	0.02	0.00
Wage	-0.04	0.27	0.12	0.01	0.00
Hours	-0.02	-0.04	0.06	0.10	0.00
Race	0.00	0.00	0.00	0.00	1.97
σ	1.03				

Table 17: Affinity matrix for unmarried heterosexuals (25-50 year old, 4,799 couples)

	I1	I2	I3	I4
Share of surplus	59.55	34.49	4.50	1.45
(Standard deviation)	(2.08)	(1.55)	(0.49)	(0.82)

Table 18: Shares of systematic surplus for unmarried heterosexuals (25-50 year old, 4,799 couples)

	I1 M	I1 W	I2 M	I2 W	I3 M	I3 W
Age	0.99	0.99	-0.16	-0.13	0.02	0.04
Education	-0.16	-0.11	-0.96	-0.90	0.06	-0.22
Wage	-0.04	-0.07	-0.24	-0.42	-0.13	0.42
Hours	-0.02	-0.04	0.03	-0.02	0.99	0.88

Table 19: Indices of attractiveness for unmarried heterosexuals (25-50 year old, 4,799 couples)

	Women				
Men	Age	Education	Wage	Hours	Race
Age	1.13	0.04	-0.04	-0.05	0.00
Education	-0.08	0.92	0.35	0.09	0.00
Wage	-0.01	0.09	0.02	0.02	0.00
Hours	-0.06	0.04	0.02	0.24	0.00
Race	0.00	0.00	0.00	0.00	1.51
σ	1.05				

Table 20: Affinity matrix for childless unmarried heterosexuals (25-50 year old, 4,335 couples)

APPENDIX E. MARRIED HETEROSEXUALS

	Women				
Men	Age	Education	Wage	Hours	Race
Age	2.25	-0.21	-0.01	0.00	0.00
Education	-0.01	0.85	0.38	-0.11	0.00
Wage	0.05	0.47	0.03	-0.14	0.00
Hours	0.06	0.02	-0.18	-0.01	0.00
Race	0.00	0.00	0.00	0.00	2.42
σ	0.98				

Table 21: Affinity matrix for married heterosexuals (25-50 year old, 4,959 couples)

	I1	I2	I3	I4
Share of surplus	63.44	29.61	4.88	2.07
(Standard deviation)	(1.89)	(1.22)	(0.20)	(0.49)

Table 22: Shares of systematic surplus for married heterosexuals (25-50 year old, 4,959 couples)

	I1 M	I1 W	I2 M	I2 W	I3 M	I3 W
Age	1.00	0.99	-0.06	-0.10	0.05	0.03
Education	-0.07	-0.10	-0.92	-0.96	0.38	0.04
Wage	0.02	-0.00	-0.37	-0.20	-0.90	0.41
Hours	0.00	-0.03	-0.06	0.14	-0.22	0.91

Table 23: Indices of attractiveness for married heterosexuals (25-50 year old, 4,959 couples)

	Women				
Men	Age	Education	Wage	Hours	Race
Age	1.91	-0.08	-0.02	-0.07	0.00
Education	-0.08	1.11	0.19	-0.03	0.00
Wage	0.04	0.35	0.14	-0.18	0.00
Hours	-0.01	0.02	-0.21	0.19	0.00
Race	0.00	0.00	0.00	0.00	2.10
σ	0.98				

Table 24: Affinity matrix for childless married heterosexuals (25-50 year old, 5,223 couples)

	Women				
Men	Age	Education	Wage	Hours	Race
Age	2.12	-0.18	0.01	-0.04	0.00
Education	-0.10	0.89	0.42	-0.15	0.00
Wage	0.07	0.37	0.06	-0.15	0.00
Hours	0.12	0.01	-0.15	0.06	0.00
Race	0.00	0.00	0.00	0.00	2.42
σ	0.98				

Table 25: Affinity matrix for married heterosexuals with one child (25-50 year old, 5,318 couples)

	Women				
Men	Age	Education	Wage	Hours	Race
Age	2.35	-0.19	0.02	0.01	0.00
Education	0.04	0.69	0.11	-0.07	0.00
Wage	0.04	0.33	0.01	-0.15	0.00
Hours	0.03	-0.00	-0.06	-0.03	0.00
Race	0.00	0.00	0.00	0.00	2.65
σ	1.00				

Table 26: Affinity matrix for married heterosexuals with three children (25-50 year old, 4,322 couples)

APPENDIX F. SHARE OF TOTAL SYSTEMATIC SURPLUS EXPLAINED BY ETHNICITY

Marriage market	Share
Gays	32.92%
Lesbians	42.11%
Unmarried heterosexuals	45.31%
Married heterosexuals	46.41%

Table 27: Shares of total systematic surplus explained by ethnicity on each marriage market