Matching, Marriage, and Children: Differences Across Sexual Orientations

By Douglas W. Allen and Shih En Lu[*]

May 30, 2016

ABSTRACT

There are many differences in behavior across couples of different sexual orientations — some well known, others not. We propose a model which explains differences in expected matching behavior, marriage rates, non-child-friendly activities, and fertility, based on different costs of procreation and complementarities between marriage and children. The model predicts that the biological traits of same-sex couples, unlike those of heterosexual couples, should not be correlated — holding constant other household production characteristics. In addition, the model predicts that heterosexuals have a higher probability of having children and getting married, and that childless heterosexuals are less likely to engage in behaviors not complementary with children than childless gays and lesbians. Using two nationally representative probability samples that self-identify sexual orientation, these predictions are confirmed.

Key Words: Matching, Same-Sex Couples.

JEL: J12, J15, J16

1. Introduction

Recent work on labor market choices has shown that same-sex couples often make different human capital decisions that reflect their lifestyle, gender composition, and procreative constraints.[1] This nascent economic work, however, seldom directly relates to family or marriage behavior.[2] Given the obvious difference in the ability to directly procreate between opposite and same-sex couples, and given the complementarities between marriage and children, it seems intuitive that there should be differences in matching, reasons to marry, frequency of marriage, behavior not complementary with children, and the presence of children.

We model different procreation and childbearing costs across sexual orientations, predict differences in the aforementioned behaviors, and confirm the predictions with high quality data. While some examined differences in behavior are *anecdotally* well-known or have been found in other small sample work, we make three major contributions. First, we show that a wide range of differences in behavior can hinge on differences in the cost of procreation and child raising. Other theories based on thin markets, preferences, or stigma can each explain some, but not all, of the behaviors that we examine. Second, we document that matching behavior differs across sexual orientations. To our knowledge, no other model predicts the specific matching behavior that we identify.[3] Finally, we test our model with two large Canadian probability samples.

Because same-sex couples are unable to procreate, they must engage in some type of more expensive procedure to acquire children when they want them. The current channels by which a same-sex couple must either conceive, adopt, or otherwise acquire children, are all considerably more costly than heterosexual sex. There is a large literature spread across several fields (law, tax, gender studies) supporting this cost difference. For example, cost differences arise over the need to find and contract with third parties, the expense of artificial reproductive technologies, discrimination over access to technologies, legal and social barriers to adoption, and differences in tax deductibility considerations for same-sex couples.[4] Once all of the legal and social hurdles have been met, the specific costs are still

---

[1] On differences, see Harris (2011), Negrusa and Oreffice (2011), Oreffice (2016), Black *et al.* (2002), or Black *et al.* (p. 54, 2007). On similarities, see Oreffice (2010), Jepsen and Jepsen (2002), or Allen (2015).

[2] The exceptions are Carpenter and Gates (2008), and Aldèn *et al.* (2015).

[3] Ciscato *et al.* (2015) examine same-sex matching on other dimensions.

[4] See, for example, Aldèn *et al.* (2015), Black *et al.* (2007), Daar (2008), Goldberg (2012), Ikemoto (1996), Morgan (2004), or Pratt (2009).

considerable. At the low end, artificial insemination for lesbians costs around $1000 per trial, with a 5-25% success rate — more if fertility drugs are used.[5] At the higher end, surrogacy for gay couples can cost in the tens of thousands of dollars.[6] Throughout the paper, we recognize and exploit the fact that these costs are greater for gays compared to lesbians.

There are also differences in the costs of raising children, due to some or all of the following: discrimination and stigma, bonding and affinity issues resulting from lack of biological connection, social disapproval, behavior problems arising over knowledge of having been donor inseminated, or the imperfect substitutability of "mothering" and "fathering." This cost claim is strongly supported by the literature, and the asymmetric list of issues faced by lesbian and gay couples is long.[7]

A higher cost of children means that: (i) fewer should be demanded by same-sex couples, even if they have the same desire for children as opposite-sex couples, and (ii) these couples have less incentive to avoid behaviors, lifestyles, and social capital investments that are not complementary with children, and thus should engage in them more frequently. Moreover, due to differences in the ways in which children are acquired, all else equal, heterosexuals should place greater value on the quality of their partner's inheritable traits than gays and lesbians, which should lead to (for heterosexuals relative to gays and lesbians): (iii) a stronger positive relation between these traits and the probability of marriage, and (iv) more pronounced assortative matching along these traits. In short, the marital behavior of same-sex couples is likely to be different from opposite-sex couples because the shadow prices they face are different, and these same differences should exist between gay and lesbian couples for the same reason.

Given that the "costs of children" are not directly observable in our data, one might object that our results are driven by differences in preferences for children, marriage, and other elements of relationships rather than by such costs. Our goal is to show that our

---

[5] See "Costs of fertility treatments in Canada" accessed online, January 18 2016 at http://www.babycenter.ca/a1028300/cost-of-fertility-treatments-in-canada.

[6] For example, Pratt (2009) argues that even the tax law for the deductibility of infertility treatments is biased against same-sex couples. Pratt analyzes a watershed case where a gay male's use of surrogates cost $95,903 over the course of two years. Gay couples face considerable legal challenges in forming families with surrogates that are not faced by lesbian couples dealing with sperm donation.

[7] See, for example, Dempsey (2012, 2013), Bergman (2010), or Jones (2005). Many children in same-sex households come from previous heterosexual unions (Gates 2011), which means prior spouses are involved in raising children.

model is consistent with a wide range of phenomena, even if one assumes that everyone has the same preferences.[8] Preferences *may not* be the same across orientations, but we simply argue that *even if* they were the same and even though both same-sex and opposite-sex couples are present in the marriage market, there should be differences in their family behaviors given the different costs of having and raising children.

We use two large data sets to analyze different types of households in terms of their potential marriage behavior. The first is the Canadian Community Health Survey (CCHS), which is a large, nationally representative, probability sample of Canadian households that self-identifies sexual orientation. These data have excellent measures of health, and allow for the direct identification of gay and lesbian *individuals* (single and married).[9] Despite its advantages, the CCHS is unable to test our matching hypothesis because the information it contains on the respondent's spouse is too limited. Hence, we also use the 2006 Canada Census, which contains a large 20% random sample of the population that self-identifies same-sex couples. The overall evidence of differences in behavior between gays, lesbians, and heterosexuals is strongly consistent with our model.[10]

We present our model in Section 2, and discuss our data and empirical results in Section 3.

2. A Model of Marriage, Children, and Matching

We present a matching model where the key feature is a difference in the cost of children across three different sexual orientations: heterosexual, gay, and lesbian.[11] We assume that members of each group only match with members from the same group, all pregnancies

---

[8] There are two other reasons for generating results without resorting to different preferences. First, differences in costs are observable, in principle. Explanations based on differences in unobservable preferences are *ad hoc* (Stigler and Becker, 1977). Second, there are widespread claims made in the academic literature, the popular press, and by professional bodies that preferences are similar (*e.g.*, American Psychological Association, Resolution on Marriage Equality, 2011; Badgett, p. 1105, 2010); or Patterson (p. 115, 1995).

[9] It also allows for the identification of bi-sexuals, but we drop bi-sexuals from the analysis. However, none of the general results of the paper change when bi-sexuals are included.

[10] The heterogeneity in behaviors between lesbians and gays suggests that differences in sexual orientation are more nuanced than the simple heterosexual versus homosexual split. Aldèn *et al.* (2015) also find difference between gay and lesbian couples in the reasons for marriage. They find that for gays, the main benefit is "resource pooling"; for lesbians, it is a "vehicle for family formation" (pp. 1265–1266).

[11] We abstract from the fact that some people may be able to choose between the same-sex and the heterosexual markets. Many of these individuals would self-identify as bi-sexual, which are excluded from our empirical analysis and a small part of the sample. Furthermore, these individuals face the same constraints on procreation as others in whichever market they choose.

are planned, and only couples have children. Individuals are initially randomly paired in a "date" and incur a search cost $k > 0$. Later they decide if they want to be a couple, and once a couple, they decide if they want to marry and/or to have children. Individuals can reject a date and go back to the dating pool, but once a person is coupled, they remain so. In addition, we assume that spouses have the same preferences over children and marriage and that there are no transfers, which allows for an abstraction of bargaining issues within a household and allows a focus on cost differences in conception and child rearing between the different couple types.

Every type of individual is described by two traits (g_i, h_i) distributed according to a positive density on $[0, G] \times [0, H]$. [12] The trait g_i is a quality index related to biological reproduction: it accounts for expected longevity, health, fertility and other features that *could be passed on to children,* and also accounts for an individual's reproductive fitness. The trait h_i is an index of characteristics such as education, talent, etc. that produce non-child household production. A component of every potential match payoff is $h_{ij} = m(h_i, h_j)$, where m is increasing in h_i and h_j, that measures the utility of household production *independent of children or marriage.* We *do not* assume that g is independent of h. Therefore, our model could easily accommodate characteristics that contribute to both traits: for example, intelligence may be passed on, and may also produce non-child related household production.

Stage Play

There are three stages of play to the matching game.

> **Stage 1:** Singles are randomly paired in a dating market at cost $k > 0$.
>
> **Stage 2:** Each person i decides if he wants to break up after observing the other person's g_j and their h_{ij}. Coupling is mutual, so either can break the date and return to stage 1.
>
> **Stage 3:** Each couple c_{ij} that remains together now observes $\epsilon_c \in \Re$, their suitability for marriage, and decides whether to marry, and whether to have children.[13]

---

[12] The closest one-dimensional counterpart to the matching portion of our model is Morgan (1998). Atakan (2006) studies a similar model, but with transferable utility. Both papers obtain assortative matching if surplus is super-modular in matched types. Lindenlaub (2014) studies multi-dimensional matching, but in a frictionless setting.

[13] To simplify notation, we drop the subscripts for couple c_{ij}. Couples are assumed to have the optimal number of children if they have any.

We assume that ϵ_c is independent of all other variables and distributed according to a continuous cumulative distribution function. We do not, however, constrain its sign: some couples may prefer the status and institutional protections of marriage, while others may prefer the flexibility of remaining unmarried.

Person i has the following separable utility function over children and household production when matched with person j:

$$U_{ij} = v_{ij} + h_{ij},$$
$$= \begin{cases} [\gamma(g_i, g_j) + M_c - s - f_c + \epsilon_c M_c] + h_{ij} & \text{if children} \\ \epsilon_c M_c + h_{ij} & \text{if no children} \end{cases} \quad (1)$$

Note that v_{ij} captures the utility related to marriage and children, and h_{ij} captures all other household utility. The sub-utility function v_{ij} has a number of components. First, $\gamma(g_i, g_j)$ is the expected utility of children, conditional on the biological attributes of the couple. We assume that $\gamma(g_i, g_j) = \max\{a, b(g_i, g_j)\}$, where a is the expected utility from adopting, and b is the expected utility from having own biological children.[14] For same-sex couples, option b is unavailable — this is not a critical assumption.[15] We assume b is increasing in both arguments; that is, g_i and g_j both improve child quality.

Second, M_c is an indicator variable for being married. Marriage is understood to be an institution complementary to children, and we normalize the value of this increase to 1.[16] The variable s is the value of activities foregone due to the presence of children. This is the value of behaviors that are not complementary with children and which are sacrificed when children arrive. Finally, our critical variable is f_c, the cost of having children. These costs are assumed to be zero for opposite-sex couples, for whom children are a by-product of sex. Importantly, these costs are positive for same-sex couples, and vary across gay and lesbian couples. Lesbian couples might have to engage in costly insemination procedures,

---

[14] Even assuming that all heterosexuals rear their own biological children, there can be a supply of children for adoption from various sources, such as parent deaths and foreign countries.

[15] Explicitly modelling surrogacy/insemination would *strengthen* predicted differences between gays/lesbians and heterosexuals — gays and lesbians with higher g would care *less* than gays and lesbians with lower g about their partner's g because only the higher g would be used — but adds complexity. Hence, we ignore these options.

[16] This is an important assumption in our model. It has been suggested that some benefits of marriage, such as companionship, may be more valuable to childless couples, so that marriage and children may be substitutes in certain ways. However, it appears reasonable to assume that, *overall*, marriage and children are complements (Aldén *et al.*, 2015). Therefore, one may interpret this "1" as the *net* amount by which marriage increases the value of children.

but these are less expensive than surrogacy. Therefore, we assume that the costs of children are related to sexual orientation; that is, $f_{Gay} > f_{Lesbian} > f_{Hetero} = 0$.[17]

Table 1 shows the four possible utility outcomes once a pairing decides to be a couple (subscripts have been suppressed).

<div align="center">

Table 1

Individual Marriage and Child Payoffs

</div>

	No Children	**Children**
Cohabiting	(A): h	(C): $h + \gamma - s - f_c$
Married	(B): $h + \epsilon_c$	(D): $h + \epsilon_c + \gamma + 1 - s - f_c$

No outcome dominates the others, and which outcome is chosen depends on the couple's specific values of the various utility components. The difference in utility between cohabitation with children (option (C)) and marriage with children (option (D)) is $1 + \epsilon_c$. The utility difference between cohabitation and marriage without children is just ϵ_c. These values may be greater or less than zero depending on the couples suitability for marriage. The difference in utility between married couples with children and married couples without children is $\gamma + 1 - s - f_c$, which can also be greater or less than zero. As a result, different couple combinations will choose different outcomes with respect to marriage and children.

In order to save space, below we present our propositions as intuitively as possible. The full model, formal results and proofs are available in an online appendix.[18]

Incentive to To Marry and Have Children

The model generates one proposition and four corollaries with respect to marriage and children that are quite intuitive.

---

[17] More broadly, f_c can also be understood as including the cost of raising children, which, as discussed in the introduction, may also be higher for same-sex couples than for opposite-sex ones. Moreover, one could assume f_c to be stochastic, and that the distributions by sexual orientation are ranked by first-order stochastic dominance. Doing so does not qualitatively impact any of our propositions and corollaries.

[18] See www.sfu.ca/~ allen/REHOAppendix.html.

Proposition 1. *Same-sex couples are no more likely to marry than heterosexual couples, and they are strictly less likely to do so than a heterosexual couple with biological traits (g_i, g_j) when $s > a - f_{Lesbian}$ and $s < \max\{a, b(g_i, g_j)\} + 1$.*[19]

Corollary 1. *Lesbian couples are at least as likely to marry as gay couples, and more so if $a - f_{Gay} < s < a + 1 - f_{Lesbian}$.*

Corollary 2. *Heterosexual couples are at least as likely to have children as lesbian couples, which are in turn at least as likely to do so as gay couples. These relations are strict as long as $a - f_{Lesbian} < s < a + 1 - f_{Lesbian}$, so that some, but not all lesbian couples adopt.*

Proposition 1 and the first two corollaries can be understood from Table 1. Consider an increase in f_c to f'_c, all else equal. Couples who would have chosen no children before the change do not change their behavior because f_c is not in their payoff function. Couples who would have chosen children with cohabitation reveal that $\gamma - s - f_c > 0$, and an increase in f_c means that those at the margin will now decide to have no children. Finally, couples who would have chosen marriage with children under the original cost will now continue with (D) or choose (A) or (B), depending on how close they are to indifference between having children or not.[20]

In other words, an increase in the cost of children leads some couples to forego having children, and some of these couples will also forego marriage as a result. No couple changes its decision in the opposite direction. Since $f_{Gay} > f_{Lesbian} > 0$, all other attributes equal, gay couples should be the least likely to have children and marry because f_c is greatest for them, followed by lesbian couples, and finally heterosexual couples. Two other intuitive results follow from the model:

Corollary 3. *Gay couples are at least as likely to engage in behaviors not complementary with children as lesbian couples, which in turn are at least as likely as heterosexual couples to engage in non-complementary behaviors.*[21]

Corollary 4. *Suppose couples A and B have biological traits (g_i, g_j) and (g'_i, g'_j), respectively, with $g_i > g'_i$, $g_j \geq g'_j$, and $b(g_i, g_j) > a$. Then if the couples are heterosexual, couple*

---

[19] These conditions are most likely to hold when a is low relative to $b(g_i, g_j)$, which implies a high cost of adoption.

[20] Note that they do not choose to cohabit with children (option C) because they have revealed $1 + \epsilon_c > 0$, and this does not depend on f_c.

[21] The empirical work examines three such behaviors: smoking, illegal drug use, and sexual activity with multiple partners.

A is more likely to marry than couple B, while if they are same-sex, they have the same probability of marriage.

Corollary 3 follows from Corollary 2, while Corollary 4 follows because, if the couples are heterosexual, couple A has a higher γ than couple B, which is equivalent to a lower f.

Matching Behavior

To close the matching model, we assume that individuals leaving the dating pool are replaced by individuals with the same characteristics.[22]

Same-Sex Matching

Given that same-sex couples cannot procreate together, their biological traits cannot be complementary, and are not considered in matching.

Proposition 2. *In any equilibrium for same-sex couples, for any $h \in [0, H]$, the set of partners that would accept type (g, h) as a mate, as well as as the set of partners that are acceptable to type (g, h) as mates, are independent of g.*

Therefore, conditional on h_i, the expected g of individual i's partner in a same-sex couple is independent of g_i. In other words, the biological fitness of same-sex partners should be uncorrelated, conditional on h. This will not be true in general for heterosexual couples because their biological fitness is passed on to their own offspring.

Heterosexual Matching

For heterosexual matching, we assume that $b(.,.)$ is super-modular; that is, if $g_i > g_i'$ and $g_j > g_j'$, then $b(g_i, g_j) + b(g_i', g_j') > b(g_i, g_j') + b(g_i', g_j)$.[23] For example, individuals with high g may place greater value on their children having high g because they do not want their children to face difficulties that they did not face. Alternatively, if parents are

---

[22] As discussed by Chiappori, McCann and Neishem (2010), there is no straightforward way to generalize the concept of assortative matching to multiple dimensions. We simply show that, in our setting, equilibria exhibit characteristics suggesting that heterosexuals with higher g (holding h fixed) tend to have partners with higher g. This is the hypothesis that we will take to the data.

[23] In the online appendix, we show that if the expected utility from biological children $b(.,.)$ is super-modular, then the expected utility of marriage and children $v(.,.)$ is weakly super-modular. The argument is simple: if $b(.,.)$ is super-modular, then the expected utility from children $\gamma(.,.)$ must be weakly super-modular. Adding marriage into the mix does not change this fact because marriage and children are complementary.

risk-averse with respect to the quality of traits passed on to the child, then it is more important for an individual with high g to have a partner with high g (so that good traits are passed on for sure) than for an individual with a low g.

Under the above assumption, heterosexuals with higher g are more selective than individuals with lower g when considering partners with low g. This points to assorting along the g dimension for heterosexual matching, and leads to the result below. To avoid confusion between specific own type and a generic partner's type, we denote the partner's type as (x, y), where x is the biological trait, and y is the household trait. Moreover, we assume that for every g_i, $v(g_i, G) - v(g_i, 0) > 0$. That is, every heterosexual cares about the biological trait of their partner at least to some extent.

Proposition 3. *There is weakly positive assortative matching in g for heterosexual couples, in the sense that, in equilibrium, for any h and whenever $g > g'$, when considering partners with sufficiently low x, type (g, h) is no less selective than (that is, require y as least as high as) type (g', h) is, and is strictly more selective if the probability of having biological children is positive.*

To understand the intuition for this result, consider an indifference curve for types (g, h) and (g', h), where $g > g'$, plotted on a plane with the partner's biological trait on the horizontal axis and the partner's household trait on the vertical axis in Figure 1.

[Figure 1 Here]

Consider the indifference curve for either "high" type (g, h) or "low" type (g', h). For low values of x, the benefits from adopting children are greater than from procreation, and the indifference curves remain flat: the biological trait x has no value in this region. Once $b > a$, the indifference curves start to fall because one type is willing to substitute x and y in a partner. This occurs sooner for high type (g, h) than for low type (g', h) since $g > g'$. Moreover, whenever the marginal utility of x for high type (g, h) is nonzero, it is greater than the marginal utility of x for low type (g', h). Hence, whenever it is not flat, high type (g, h)'s indifference curve at a given (x, y) is strictly steeper than low type (g', h)'s at the same (x, y).[24]

Consider a potential mate given by point 1 in the graph. High type (g, h) would reject this person as a match because they fall below their indifference curve boundary. On the other hand, low type (g', h) would accept this person as a mate. Now consider another

---

[24] The downward sloping part of the indifference curves need not be concave as depicted.

potential mate given by point 2. This person has a higher x and lower y than the person at point 1. Now high type (g, h) finds this person acceptable, while low type (g', h) does not. The reason is the supermodularity of expected utility in the biological traits: high biological traits matter more to high type (g, h) than to low type (g', h). Figure 1 is one example of the how the boundary conditions might look for types (g, h) and (g', h). There are actually three cases, and each is considered in the online appendix. All cases generate the result.

Our model thus predicts differences in behavior without positing any difference in preferences, marriage market conditions (thinness), costs of marriage, or type distributions across sexual orientations. Instead, they all occur due to simple variations across orientations in the availability of means of conception and/or the cost of having children. More complicated models are possible, but our goal is to examine if differences in the costs of children can explain a wide range of differences in behavior between couples of different orientations.[25]

3. Empirical Results

3.1. The Data

Data for most of our tests come from years 2005, 2007-2010, 2013, and 2014 of the restricted CCHS master files — a probability sample survey with a cross-section design.[26] The target population of the CCHS is all Canadians aged 12 and over, it covers 98% of the provincial populations, and data is collected through computer assisted interviews. Given the cross-section structure, our data is suitable for finding the correlations predicted by our model, but not for establishing causal linkages.

The CCHS has extensive information on the respondent, but only limited information on all other members of the household. What makes it particularly unique for a large probability sample is that it self-identifies sexual orientation — heterosexual, gay, lesbian,

---

[25] These predictions would hold for other types of couples who, *ex ante*, would be predicted to have high costs of procreation. For example, these predictions would apply to couples that are elderly or infertile at the time of matching. Unfortunately, the data set used here does not allow the identification of such couples.

[26] There was no survey in 2006, and the critical self-reported health variable was dropped from the survey in 2011 and 2012. The full CCHS is not a public use data set. Results are screened by Statistics Canada, and as a result, no maximums, minimums, or sample counts for variables are reported in this paper, and the data are not available from the authors. The data and procedures are available from Statistics Canada for researchers approved for access by Statistics Canada.

and bi-sexuality — for all individuals, including singles. Some might critique direct self-reporting of sexual orientation on the grounds that some individuals are unwilling to reveal such sensitive information; however, self-reporting is better than the alternatives, and the CCHS has some additional advantages. First, it does *not* indirectly identify same-sex couples through responses to a series of questions. Such methods fail to identify gays or lesbians who are single, fail to distinguish bi-sexual individuals, are subject to the same under-reporting problem, and have the added problem of capturing large numbers of heterosexual couples who incorrectly record the wrong sex.[27] Second, the CCHS's refined identification of bi-sexual individuals is helpful for reducing measurement error in identifying gays and lesbians.

Finally, the CCHS data are from Canada, where one could argue there has been little official discrimination against same-sex couples for some time: same-sex couples have had all taxation and government benefits since 1997, and same-sex marriage has been legal since 2001–2005.[28] Other social scientists have noted that legalization has reduced the stress and stigma of homosexuality in Canada, which makes it more likely that respondents would answer questions honestly.[29] All things considered, the CCHS is an excellent large, random sample data set available to study non-heterosexuals.[30]

3.2. Basic Demographics

---

[27] The Williams Institute 2010 census study claims that the total national error rate is approximately 0.25%. The problem is that small errors in the large heterosexual response rate leads to large errors in the small same-sex sample.

[28] Regarding his 1967 Omnibus bill that legalized homosexuality, Justice Minister Pierre Trudeau famously stated that "There's no place for the state in the bedrooms of the nation ... what's done in private between adults doesn't concern the Criminal Code." (http://www.cbc.ca/archives/categories /politics/rights-freedoms/trudeau). Many Canadians consider this the watershed moment for the acceptance of homosexuality in Canada. In Canada, same-sex couples were allowed to adopt before they were allowed to marry. Ontario became the first province to allow adoption in 1997. Others quickly followed. The first Canadian same-sex marriages took place on January 14, 2001 at the Toronto Metropolitan Community Church. These became the basis of a successful legal challenge that ended at the court of appeal on June 10, 2003. On July 20, 2005, the federal government passed the Civil Marriage Act that made Canada the fourth country in the world to legalize same-sex marriage. Thus, different people date the arrival of same-sex marriage in Canada as 2001, 2003, or 2005.

[29] See, for example, Biblarz and Savci (p. 490, 2010).

[30] It is possible that those who self-identify as gay or lesbian may also be more likely to self-identify the use of drugs, multiple sex partners, and the like. This, however, would not explain other systematic differences in behaviors that carry less stigma. For example, though not reported, there are differences in alcoholic consumption rates across the sexual orientations that one would not expect if the results were purely selection driven.

Table 2 shows some estimated population characteristics for the three household types in Canada.[31] The numbers are consistent with other findings based on same-sex couples from nonrandom samples. The facts are also consistent with our model predictions.

Table 2 shows that gays and lesbians make up tiny fraction of the population, and, as in several other data sets, there are more gays than lesbians.[32] Overall, 43.2% of lesbians are single, and only 14.5% of them are married. Gay men are much more likely to be single (59.7%), and less likely to be married (only 9.8%); the opposite is true for heterosexuals (29.2% and 48.4% respectively). The percentage of households with children under 18 is 10.3% for lesbians, only 2.7% for gays, and 24.7% for heterosexuals.

In terms of income, the CCHS confirms findings from other studies: gay and lesbian households do not appear to suffer any household income penalty.[33] Heterosexuals, despite their average incomes, are more likely to own their home compared to all other groups, and heterosexuals are more likely to report no health problems. Lesbians are more likely to be white, and on average both gays and lesbians are considerably more educated than heterosexuals. Heterosexuals are less likely to be smokers, on average. However, perhaps the most striking difference is with respect to sexual behavior. In this regard, lesbians and heterosexuals appear quite similar on average: 85.29% and 83.22% had only one sexual partner in the past twelve months, and around 3% of them had more than four. In contrast, gays are much less likely to have one sexual partner in the past twelve months, and much more likely to have had more than four (20.6%). All of these unconditional averages are qualitatively consistent with our model.

3.3. Children, Behavior, and Marriage

Our model makes a series of predictions regarding family behaviors for heterosexuals, gays, and lesbians. In Table 3, we present the results of five logit regressions, estimated together as seemingly unrelated regressions, to test these predictions.[34]

---

[31] All of the results of this paper are weighted estimates from the CCHS sample. As a result, we will normally drop the adjective "estimated" unless the context calls for clarification.

[32] These estimates are not that different from fractions found in other random samples. For example, Wainright *et al.* (2004), using the National Longitudinal Survey of Adolescence Health, find that lesbians make up about 1/3 of one percent of the sample. Golombok *et al.* (2003), using the Avon Longitudinal Study of Parents and Children, find that .22% of the mothers are lesbians.

[33] See Ahmed, Andersson, & Hammarstedt (2011) or Carpenter (2004).

[34] Many other combinations of controls, weighting, and robust errors were tried, but unreported to keep the table sizes manageable. The results from these unreported regressions are always consistent with those reported here; that is, the results are robust. Running the logit regressions independently made virtually no difference. The definitions of the variables used are in Table 1A in the appendix.

Presence of Children

Our model predicts that children are least likely in gay households and most likely in heterosexual households. The summary statistics in Table 2 confirm this: children are rare among gay and lesbian households without controlling for household characteristics. Table 3, column (1) confirms the findings from Table 2. These logit regression results are based on the full sample, using full controls, robust standard errors, and regression weights, where the dependent variable is whether or not a child under 18 is present in the household.[35] Although both types of households are less likely to have such children present, there is a considerable difference between gay and lesbian households. Looking at the odds ratio, the coefficient for gays means that the odds of having children present in the home are almost 20 times smaller compared to heterosexual homes. On the other hand, the odds of lesbians having children are only 42% as large as those for heterosexuals. This difference is consistent with our prediction that non-heterosexual households are less likely to include children, and that this effect is stronger among gay households compared to lesbian households.

Behaviors Non-Complementary With Children

Table 3 investigates a series of behaviors that most would consider non-complementary with the presence of children: smoking, illegal drug use, and sexual activity with more than four partners in the past year. Columns (2)–(4) contain selected coefficients from three regressions, depending on the dependant variable. Each column reports the regression results for the full sample when all controls, weights, and robust standard errors are used.

Our model predicts that the presence of children should reduce the frequency of these behaviors, and we find that the presence of children is associated with less smoking, less illegal drug use, and a lower likelihood of having many sex partners in the past year, for all three sexual orientations. Furthermore, if gays and lesbians without children are less likely than childless heterosexuals to *expect* having children in the future, then our model also predicts that: (i) on average, childless gays and lesbians should engage in these behaviors more often than childless heterosexuals; and (ii) the presence of children should be associated with a larger reduction in the incidence of these behaviors for gays and lesbians than for heterosexuals, because the difference in expectation between gays

---

[35] Where an individual reports no income, an income is imputed by running a regression of actual income on age, education (categorical), white, marital status, and presence of children.

and lesbians that have children and those that do not is larger than the difference for heterosexuals. Columns (2)–(4) show strong support for these predictions as well.

Taken together, the results from columns (2)–(4) show that childless gays and lesbians more frequently engage in the three examined types of "non-family-friendly" behavior, relative to childless heterosexuals.[36] However, the results show that while the presence of children is associated with a lower prevalence of these behaviors for all sexual orientations, this "child effect" is generally larger for gays and lesbians.

Probability of Marriage Given Health

Our model predicts that heterosexuals with high g's should be more likely to have children, and therefore, more likely to marry. Fortunately, the CCHS contains excellent information on an individual's self-reported health status. It provides information on many health problems, but also calculates a health utility index based on vision, speech, hearing, dexterity, cognition, mobility, or emotional disorders.[37] We use this health index as a measure of biological fitness. Although the health index ranges from negative values to one, we create a health dummy variable that equals zero if the health index is less than one, and equals one otherwise.[38]

Column (5) reports a logit regression on a couple's choice to marry or cohabit, based on the full sample, robust standard errors, weighted observations, and controls. The reported coefficients are the sexual orientation variables, these variables interacted with the health index fixed effect, and these variables interacted with the child fixed effect. The variables of interest for matching are the interactive terms of sexual orientation and the health index. The sorting hypothesis predicts that the interactive terms should only matter for heterosexuals.[39]

---

[36] Marginal effects are somewhat ambiguous with nonlinear models with interactions. Evaluated at the means, the marginal effects of being gay on smoking, drug use, and having more than four sexual partners are 0.106, 0.038, and 0.021, respectively. The marginal effects of being lesbian on the same dimensions are 0.043, 0.019, and 0.191, respectively.

[37] It is reasonable to assume that all of these are exogenous to the institutional decision, with the possible exception of emotional disorders. Excluding this within a different index makes no effective difference in the results.

[38] We do this to avoid imposing cardinality on our biological measure and introducing measurement error, since the index does not measure positive health attributes that would lead to higher g's. However, this makes almost no difference to the estimates.

[39] We use a health fixed effect to avoid measurement error; however, when we use the actual cardinal index, the results are similar. That is, we find the following odds ratios (t-statistics): Gay $\times$ Healthy, 0.44 (-0.59); Lesbian $\times$ Healthy, 1.10 (0.11); and Hetero $\times$ Healthy, 1.24 (3.03).

Column (5) confirms the summary statistic findings of the first four tables: lesbians and gay men are significantly less likely to be married relative to cohabitation. In terms of marriage and biological fitness, the health status of gays appears unrelated to marriage, and for lesbians is negative — although both estimates are insignificant . On the contrary, health status matters for heterosexuals. Healthy heterosexuals have odds of marrying that are 47% higher than the odds for unhealthy heterosexuals, a statistically significant effect that is much larger than the point estimates for gays and lesbians.[40]

Assortative Matching

Our model predicts that heterosexual couples should match along biological and re-productive fitness lines (g), holding constant household traits (h). At the same time, same-sex couples should not match along g because these couples cannot procreate. The one weakness of the CCHS data set is that it does not contain health information on the respondent's spouse. To resolve this we turn to another data set: the 2006 Canada census.

The 2006 Canada census only identifies same-sex couples (both married and cohabitating), but it does contain the same information on each spouse. This information includes a crude measure of health status. In particular, the census asks if the individual has home, leisure, education, work, or other activities limited due to poor health. The question goes on to define poor health as a condition resulting from injury, illness, mental illness, and hereditary diseases. Respondents only have the three options of "never," "sometimes," or "often." As such, the census health measure is a noisy measure of our g parameter. Moreover, spousal health also contributes to h, as it generates well-being independently from reproduction. Therefore, the model does not rule out sorting along the health dimension for same-sex couples. Rather, the prediction is that such sorting should be more pronounced for heterosexual couples due to the importance of biology for the g parameter.

We use the 20% restricted census master file.[41] From this file, all couples, either married or cohabitating, were selected. Statistics Canada does not allow the sample sizes

---

[40] Since marriage may impact health, we have run the regression with only individuals between 18–35, an age group where marriage is less likely to have yet had an effect on health. The odds ratios (t-statistics) then become: Gay × Healthy, 0.39 (-1.18); Lesbian × Healthy, 1.08 (0.15); and Hetero × Healthy, 1.42 (8.50). We conduct three other falsification tests using age, income, and education in lieu of health and interact these variables with sexual orientation. In these cases, the interaction results are often reversed (and always insignificant for heterosexuals), which suggests that our results are picking up differences in the costs of having children rather than a cohort or other spurious effects.

[41] Like the CCHS, this is not a public use data set, and the separate procedures for access are identical to those of the CCHS. Empirical work was conducted at the SFU Research Data Center, and all results were screened by Statistics Canada before release. Statistics Canada does not allow any unweighted observations or descriptives to be released, nor any maximums or minimums of weighted estimates.

to be released; however, the weighted estimates of the population based on this sample are: 19,575 lesbian couples; 23,125 gay couples; 1,296,250 cohabitating heterosexual couples; and 5,920,270 married heterosexual couples.[42]

Table 4 shows our simple test of assortative matching. Columns (1) – (4) run our regression selecting samples based on sexual orientation. Column (5) pools the data. The dependent variable is the ordinal health status of the spouse identified as person 1 in the census. This is regressed on the health status of this person's spouse and a host of controls. In the case of the pooled regression, the health status of person 2 is interacted with sexual orientation. We report the coefficient on the health status of the spouse and one of the controls: the presence of children. The results confirm our hypothesis. Gay and lesbian couples sort less along the health dimension than cohabitating heterosexuals. Heterosexual married couples sort the most strongly along the health dimension than same-sex couples: the differences between the coefficients on the spouse's health are statistically significant.[43] Pooling produces identical results. In addition, when children are present in the marriage, the health status of each spouse is higher for heterosexual couples, but not for same-sex couples — again, a finding consistent with our model.[44]

4. Conclusion and Discussion

This paper has exploited two data sets that allow for reliable estimates of demographic characteristics of different sexual orientations, and for some investigation of lifestyle choices and mate matching behavior. The model presented ties a series of different behaviors together and shows how they result from a simple and observable cost difference, which leads to heterosexuals having a stronger expectation of children than lesbians and gays do. Other studies, mostly based on small nonrandom samples, have found similar results for

---

[42] As required, these numbers are rounded to the nearest 5. About 85% of the lesbian and gay couples are cohabitating, and since the results of interest from Table 4 were the same whether same-sex married couples were separated out or not, we have combined them to make the table simpler.

[43] If the census health measure only identified biological traits that could be passed on to children, then we might expect the health coefficient for same sex couples to be zero. However, it measures general health status, which is likely to be correlated with our household characteristic. Hence, it is not too surprising that even same sex couples have a positive health correlation.

[44] An alternative explanation for the higher health correlation in heterosexual couples is that it might be difficult to care simultaneously for children and an ill partner. This would increase the importance of health for couples with or wanting children for all orientations; the coefficient on partner's health would still be higher for heterosexuals since they more frequently have children. However, as observed here, gay and lesbian couples with children do not exhibit better health than those without, which does not support this alternative reasoning.

fertility and non-family behaviors, but none to our knowledge have tied these together with the matching problem, nor have they used a data set with the qualities of the CCHS.[45]

An alternative explanation for lower marriage rates among same-sex couples (and, if marriage and children are complementary as we hypothesize, the lower prevalence of children in same-sex households) is that, while same-sex marriage is legal in Canada, it may still be stigmatized, or may simply not have reached a steady state. However, our model additionally explains differences in matching behavior. Thus, while a higher cost of marriage for same-sex couples may contribute to some of this paper's findings, it is neither necessary nor sufficient for explaining all of them. Our contribution suggests that these differences are likely to persist in the long run, even after transitory factors vanish, because they are based on a biological constraint.

One may also worry that our results about behavior non-complementary with children may be caused by a selection effect: i) people that are comfortable reporting their homosexuality may also be more comfortable admitting to sensitive behavior, or ii) our results stem from older gays who were unable to marry. While we cannot completely rule out either effect, we make several observations that alleviate these concerns. First, same-sex households with children are not more likely than heterosexual households with children in reporting such behavior. Second, our matching results show that differences in non-sensitive behavior occur, so it appears plausible that differences in sensitive behavior would occur as well. Finally, we ran our regressions without the older gays and found similar results.

In the end, we have provided a simple and plausible model of household behavior that explains a wide range of behavior differences across couples of different sexual orientation, which are documented by our empirical work. We do not claim that our model is the only explanation for these correlations, as we have not established causality. However, we have demonstrated that a parsimonious model based on a single fundamental difference — in procreation and childrearing costs — can generate the many differences in behavior that are observed. Other factors may well contribute to the magnitude of these effects, but do not tie all these phenomena together. We leave it to future work to compare the importance of various explanations.

---

[45] An exception would be Aldèn *et al.* (2015). They use an excellent data set and exploit differences in the ability to reproduce, the explore fertility and income differences across sexual orientation. They do not examine the matching issue, however.

TABLE 2: Population Estimates of Household Characteristics
Weighted Observations

Characteristic	Heterosexual Household	Gay Household	Lesbian Household
% of Population	97.79	.57	.32
% HH with children <18	24.7	2.37	10.3
% Married	48.4	9.8	14.5
% Single	29.2	59.7	43.2
Avg. HH Income	80,748	86,301	76,805
% Homeowners	74.7	55.7	63.0
% White	84.0	88.7	92.2
% Smokers	15.7	25.8	20.2
% Health	53.7	47.8	47.6
% High School	76.4	96.2	95.1
% Graduate Work	9.9	17.5	14.1
# of Sex Partners in past 12 months, if at least one			
% One	85.29	55.52	83.22
% Two	7.91	14.67	11.41
% Three	3.38	8.95	2.35
% Four +	3.42	20.86	3.02

TABLE 3 Children, Non-Complementary Behaviors, & Marriage
SUR Logit Regressions

Variable	Children	Smoking	Illegal Drug Use	More Than 4 Sex Partners	Marriage
	(1)	(2)	(3)	(4)	(5)
Gay coefficient	−3.08	0.68	0.58	1.10	−1.97
Gay odds ratio	0.05	1.99	1.79	3.02	0.14
	(−5.88)*	(4.65)*	(2.61)*	(4.00)*	(−6.58)*
Lesbian coefficient	−0.86	0.32	0.33	3.02	−2.04
Lesbian odds ratio	0.42	1.37	1.39	22.81	0.13
	(−2.73)*	(1.54)	(1.57)	(2.10)*	(−8.59)*
Gay× Children		−1.35	−2.21	−0.91	2.05
Gay× Children odds ratio		0.26	0.11	0.40	7.77
		(−2.18)*	(−2.13)*	(−1.01)	(3.35)*
Lesbian× Children		−0.94	0.15	−2.32	0.51
Lesbian× Children odds ratio		0.39	1.16	0.17	1.67
		(−2.36)*	(0.30)	(−3.71)	(0.97)
Children		−0.39	−0.05	−0.66	0.66
Children odds ratio		0.97	0.95	0.51	1.95
		(1.15)	(1.37)	(−8.54)*	(24.58)*
Lesbian × Healthy					0.04
odds ratio					1.03
					(0.11)
Gay × Healthy					−0.31
odds ratio					0.73
					(−0.80)
Heterosexual × Healthy					0.38
odds ratio					1.47
					(14.83)*
N	351,890	351,890	351,890	106,788	199,891
Log Likelihood	−7,416,475	−7,446.865	−4,208,411	−872,623	−5,102,258
Pseudo R2	0.27	.04	.05	0.18	0.16

* Significant at the 5% level. t-statistics in parentheses.

Controls: Age, Year, White, Income, Urban, Graduate Work, Obesity, Children (except in (1)), Marriage, and the sexual orientation health interactions.

SUR uses weighted observations and robust standard errors.

TABLE 4: OLS Regressions for Assortative Matching

Dependent Variable: Spouse 1 Health

Variable	Gay Couples	Lesbian Couples	Heterosexual Common Law	Heterosexual Married	Pooled Sample
	(1)	(2)	(3)	(4)	(5)
Spouse 2 Health	0.29	0.33	0.37	0.42	
	(12.08)*	(13.31)*	(98.71)*	(487.36)*	
Gay× Health					0.28
					(12.06)*
Lesbian × Health					0.33
					(13.29)*
Common Law × Health					0.37
					(102.61)*
Married × Health					0.42
					(301.59)*
Children	−0.04	−0.02	0.01	0.005	−0.0009
	(−0.98)	(−0.81)	(7.00)*	(4.46)*	(−0.89)
Controls	Yes	Yes	Yes	Yes	Yes
Weighted obs.	Yes	Yes	Yes	Yes	Yes
N	23,125	19,575	1,296,250	5,920,270	7,259,220
R^2	0.19	0.18	0.22	0.29	0.96

* Significant at the 5% level. t-statistics in parentheses.

Controls: Age, White, Rooms, Education, Ethnicity, Province, Citizenship, Value, Urban Size, Work. For the pooled regression the CONSTANT is interacted with the family type fixed effect as well.
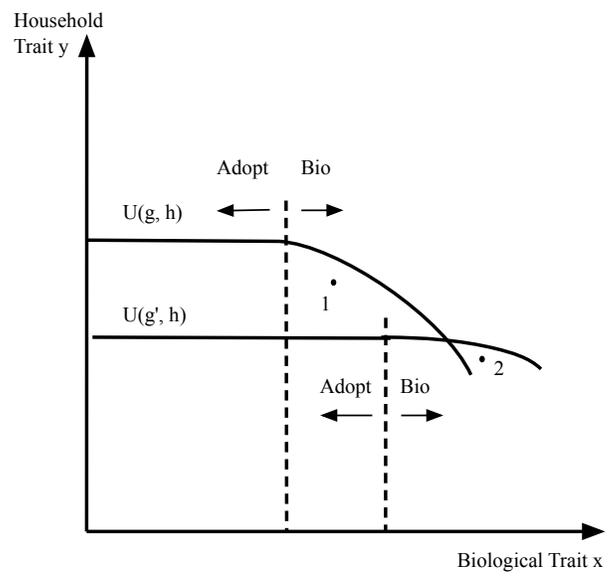
Figure 1: Example of Matching Boundary Conditions

References

Ahmed, A., L. Andersson and M. Hammarstedt. "Inter- and Intra-Household Earnings Differentials among Homosexual and Heterosexual Couples." *British Journal of Industrial Relations* 2011.

Aldèn, L., L. Edlund, M. Hammarstedt, and M. Mueller-Smith. "Effect of Registered Partnership on Labor Earnings and Fertility for Same-Sex Couples: Evidence From Swedish Register Data" *Demography* 52(4), August 2015: 1243–1268.

Allen, D.W. "An Economic Assessment of Same-Sex marriage" *Harvard Journal of Law and Public Policy* Vol. 29(3), Summer 2006, 949-980.

———. "Household Production and Sexual Orientation" *Economic Inquiry* 53(1) January 2015: 406–418.

———. "What Does She See In Him: The Effect of Sharing on the Choice of Spouse." *Economic Inquiry* 30, January 1992, 57–67.

Atakan, A.E. "Assortative Matching with Explicit Search Costs" *Econometrica* 74(3), May 2006, 667-680.

Badgett, M.V.L. "The Economic Value of Marriage For Same Sex Couples" *Drake Law Review* 58(4), Summer 2010, 1081–1120.

Bergman, K., R. Rubio, R.J. Green, and E. Padron. "Gay Men Who Become Fathers via Surrogacy: the Transition to Parenthood." *Journal of GLBT Family Studies,* 6(2), 2010: 111–141.

Biblarz, T., and E. Savci. "Lesbian, Gay, bi-sexual, and Transgender Families" *Journal of Marriage and Family* 72, June 2010.

Black, D., S. Sanders, & L. Taylor. "The economics of lesbian and gay families", *Journal of Economic Perspectives* 21(2), 2007.

———., G. Gates, S. Sanders, and L. Taylor. "Why Do Gay Men Live in San Francisco?" *Journal of Urban Economics* 51(1) 2002: 54–76.

Blankenhorn, D. *The Future of Marriage* (New York: Encounter Books, 2007).

Carpenter, C. "New Evidence on Gay and Lesbian Household Income" *Contemporary Economic Policy* 22(1) 2004.

———. and G. Gates. "Gay and Lesbian Partnership: Evidence from California" *Demography* 45(3) 2008.

Chiappori, P.A., R. McCann, and L. Nesheim. "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness" *Economic Theory* 42(2), 2010, 317-354.

Chiappori, P.A., S. Oreffice, and C. Quintana-Domeque. "Fatter Attraction: Anthropometric and Socioeconomic Characteristics in the Marriage Market" *Journal of Political Economy* (forthcoming).

Ciscato, E., A. Galichon, and M. Gousse. "Like Attract Like? A Structural Comparison of Homogamy Across Same-Sex and Different-Sex Households" (Sciences Po working paper, 2015)

Cohen, Lloyd. "Rhetoric, The Unnatural Family, and Women's Work" *Virginia Law Review* 81, November 1995.

———. "Marriage, Divorce, and Quasi Rents: or, 'I Gave Him the Best Years of My Life'" *Journal of Legal Studies* XVI June 1987.

Daar, Judith F. "Accessing Reproductive Technologies: Invisible Barriers, Indelible Harms" *Berkeley Journal of Gender, Law, and Justice* 23, 2008: 18–82.

Dempsey, Deborah. "Surrogacy, gay male couples and the significance of biogenetic paternity" *New Genetics and Society* 32(1), 2013: 37–53.

———. "More Like a Donor or More Like a Father? Gay Men?s Concepts of Relatedness to Children." *Sexualities,* 15 (2), 2012: 156–174.

Dnes, A., and R. Rowthorn. "The Law and Economics of Marriage & Divorce" (Cambridge: Cambridge University Press, 2002).

Gallagher, Maggie. "(How) Will Gay Marriage Weaken Marriage as a Social Institution: A Reply to Andrew Koppelman" *University of St. Thomas Law Journal* 2 Fall 2004.

Gates, Gary. "Family Formation and Raising Children Among Same-Sex Couples" *Family Focus On ... LGBT Families* Issue FF51, Winter 2011.

Geoghegan, Tom. "The gay people against gay marriage." http://www.bbc.com/news/magazine-22758434 (2013).

Goldberg, A.E. *Gay Dads: Transitions to Adoptive Fatherhood* (New York: NY University Press, 2012).

Golombok, S., B. Perry, A. Burston, C. Murray, J. Mooney-Somers, M. Stevens , and J. Golding. "Children With Lesbian Parents: A Community Study" *Developmental Psychology* 39(1), 2003.

Greif, A. *Institutions and the Path to the Modern Economy* (Cambridge: Cambridge University Press, 2006).

Harris, B. "An explanation for the sexual orientation wage gap" (Simon Fraser working paper, 2011)

Ikemoto, L. C. "The In/Fertile, the Too Fertile, and the Dysfertile," *Hastings Law Journal* 47(4) 1996: 1007–1062.

Jepsen, L., and C. Jepsen. "An Empirical Analysis of the Matching Patterns of Same-sex and Opposite-sex Couples" *Demography* 39(3), 2002.

Jones, C. "Looking Like a Family: Negotiating Biogenetic Continuity in British Lesbian Families Using Licensed Donor Insemination." *Sexualities*, 8(2), 2005: 221–237.

Lindenlaub, I. "Sorting Multidimensional Types: Theory and Application" (Yale working paper, 2014)

Matouschek, N., and I. Rasul. "The Economics of the Marriage Contract: Theories and Evidence" *Journal of Law and Economics* 51, February 2008.

Morgan, Benjamin C. "Adopting Lawrence: Lawrence v. Texas and Discriminatory Adoption Laws", *Emory Law Journal* 53(3), 2004: 1491–1532.

Morgan, P.B. "A Model of Search, Coordination and Market Segmentation" (SUNY Buffalo working paper, 1998)

Negrusa, B., and S. Oreffice. "Sexual orientation and household financial decisions: evidence from couples in the United States" *Review of Economics of the Household*, 18(2), 2011.

Oreffice, S. "Sexual orientation and household decision making. Same-sex couples balance of power and labor supply choices" *Labour Economics* 18(2), 2010: 145–158.

———. "Sexual Orientation and Marriage" *Estudios de Economia Aplicada* 34(1), 2016: 7–34.

Patterson, C. "Families of the Lesbian Baby Boom: Parents' Division of Labor and Children's Adjustment" *Developmental Psychology* 31(1), 1995, 115–123.

Pratt, Katherine. "Deducting the Costsof Fertility Treatment: Implications of Magdalin v. Commissioner for Opposite-se Couples, Gay and Lesbian Same-Sex Couples, and Single Women and Men," *Wisconsin Law Review* Issue 6, 2009: 1283–1348.

Stigler, G. and G. Becker. "De Gustibus Non Est Disputandum" *American Economic Review* 67(2), 1977: 76–90.

Vonk, M. "One, Two, or Three Parents? Lesbian Co-mothers and a Known Donor with 'Family Life' Under Dutch Law." *International Journal of Law, Policy and the Family* 18, 2004, 103–117.

Wainright, J., S. Russell, and C. Patterson. "Psychosocial Adjustment, School Outcomes, and Romantic Relationships of Adolescents with Same sex Parents " *Child Development* November 2004.

Williams Institute. "Census Snapshot: 2010 Methodology adjustment procedures for same-sex couple data." At http://services.law.ucla.edu/williamsinstitute/ pdf/AdjustmentProcedures.pdf, accessed August 29, 2011.

Zaidel, D.; Aarde, S.; Baig, K. "Appearance of symmetry, beauty, and health in human faces". *Brain and Cognition* 57(3), 2005, 261–263

TABLE 1A: Definitions of Variables
Canadian Community Health Survey

Variable Name		Definition
Gay	=	1 if respondent self-identified as gay.
Lesbian	=	1 if respondent self-identified as lesbian
Bi-sexual male	=	1if respondent self-identified as male bi-sexual
Bi-sexual female	=	1if respondent self-identified as male bi-sexual
Age	=	age in years.
Year	=	year of survey, either 2005, 2007, 2008, or 2009.
White	=	1 if respondent was white.
Smoking	=	1 if respondent was a daily smoker, and had smoked more than 100 cigs. in life.
Income	=	self reported income of respondent.
Urban	=	1 if respondent lived in urban area.
Obesity	=	1 if body mass index was greater than 30.
Children	=	1 if any child in household was less than 18.
Health	=	1 if respondent had no serious health problems.
Graduate Work	=	1 if respondent had completed graduate degree.
High School	=	1 if respondent graduated from high school.
Drug Use	=	1 if the respondent has used marijuana, cocaine, speed, ecstasy, hallucinogens, glue, or heroin
Health Index	=	1 if person suffered from vision, speech, hearing, dexterity, cognition, mobility, or emotional disorders.

TABLE 2A: Definitions of Variables
2006 Canada Census

Variable Name		Definition
Spouse 1 Health	=	1 if poor health often prohibits job, school, or other activities.
	=	2 if poor health sometimes prohibits job, school, or other activities.
	=	3 if poor health never prohibits job, school, or other activities.
Spouse 2 Health	=	same as Spouse 1 Health.
Children	=	1 if children are present in home.
Age1	=	age of spouse 1.
Age2	=	age of spouse 2.
Education1	=	highest grade achieved by spouse 1.
Education2	=	highest grade achieved by spouse 2.
Ethnicity1	=	ordinal ethnic category for spouse 1.
Ethnicity2	=	ordinal ethnic category for spouse 2.
Citizen1	=	1 if spouse 1 is Canadian citizen.
Citizen2	=	1 if spouse 2 is Canadian citizen.
White1	=	1 if spouse 1 is white.
White2	=	1 if spouse 2 is white.
Work1	=	number of weeks worked in 2005 for spouse 1.
Work2	=	number of weeks worked in 2005 for spouse 2.
Rooms	=	number of rooms in residence.
Value	=	value of residence.
Province	=	ordinal value for province.
Urban Size	=	population of rural/urban district.

Supplementary Appendix:
Matching, Marriage, and Children

By Douglas W. Allen and Shih En Lu*

March 2016

1. Introduction

In order to meet the page limit restrictions of the journal, the full theoretical model has been moved to this online appendix. After the full model, we present some robustness tests from empirical results based on restricted samples.

2. The Full Model of Marriage, Children, and Matching

We present a stylized model of matching in which the key feature is a difference in the cost of children across three different sexual orientations: heterosexual, gay, and lesbian.[1] We assume that members of each group only match with members from the same group, all pregnancies are planned, and only couples (married or not) have children (no single parents). Individuals are initially randomly paired in a "date" and incur a search cost $k > 0$. Later they decide if they want to be a couple, and once a couple, they decide if they want to marry and/or to have children. Individuals can reject a date and go back to the dating pool, but once a person is coupled, they remain so. In addition, we assume that spouses have the same preferences over children and marriage and that there are no transfers, which allows for an abstraction of bargaining issues within a household and allows a focus on cost differences in conception, pregnancy, and child rearing between the different couple types.

Every type of individual is described by two traits (g_i, h_i) distributed according to a positive density on $[0, G] \times [0, H]$. The trait g_i is a quality index related to biological reproduction: it accounts for genetic features such as expected longevity, health, fertility and other features that could be passed on to children, and also accounts for an individual's reproductive fitness. The trait h_i is an index of characteristics such as education, talent, etc. that produce non-child household production. A component of every potential match payoff is $h_{ij} = m(h_i, h_j)$, where m is increasing in h_i and h_j, that measures the utility of household production *independent of children or marriage*. We do not assume that g is independent of h. Therefore, our model could easily accommodate characterisics that

---

[1] We abstract from the fact that some agents may be able to choose between the same-sex and the heterosexual markets. Many of these individuals would self-identify as bi-sexual, which are excluded from our empirical analysis and a small part of the sample. Furthermore, these individuals face the same constraints on procreation as others in whichever market they choose.

contribute to both traits: for example, intelligence may be genetically passed on, and may also produce non-child related household production.[2]

Stage Play

There are three stages of play to the matching game.

Stage 1: Singles are randomly paired in a dating market at cost $k > 0$.

Stage 2: Each person i decides if he wants to break up after observing the other person's g_j and their h_{ij}. Coupling is mutual, so either can break the date and return to stage 1.

Stage 3: Each couple c_{ij} that remains together now observes $\epsilon_c \in \Re$, their suitability for marriage, and decides whether to marry, and how many children to have.[3] We assume that ϵ_c is independent of all other variables and distributed according to a continuous cumulative distribution function. We do not, however, constrain its sign: some couples may prefer the status and institutional protections of marriage, while others may prefer the flexibility of remaining unmarried.

Person i has the following separable utility function over children and household production when matched with person j:

$$U_{ij} = v_{ij} + h_{ij},$$
$$= \begin{cases} [\gamma(g_i, g_j) + M_c - s - f_c + \epsilon_c M_c] + h_{ij} & \text{if children} \\ \epsilon_c M_c + h_{ij} & \text{if no children} \end{cases} \tag{1}$$

Note that v_{ij} captures the utility related to marriage and children, and h_{ij} captures all other household utility. The sub-utility function v_{ij} has a number of components. First, $\gamma(g_i, g_j)$ is the expected utility of children, conditional on the genetic attributes of the couple. We assume that $\gamma(g_i, g_j) = \max\{a, b(g_i, g_j)\}$, where a is the expected utility from adopting, and b is the expected utility from having own biological children.[4] For same-sex couples, option b is unavailable because we ignore surrogacy and alternative insemination

---

[2] The closest one-dimensional counterpart to the matching portion of our model is Morgan (1998). Atakan (2006) studies a similar model, but with transferable utility. Both papers obtain assortative matching if surplus is super-modular in matched types.

[3] To simplify notation, we drop the subscripts for couple c_{ij}. Couples are assumed to have the optimal number of children if they have any.

[4] Even assuming that all heterosexuals rear their own biological children, there can be a supply of children for adoption from various sources, such as parent deaths and foreign countries.

options for all couples — this is not a critical assumption.[5] We assume b is increasing in both arguments; that is g_i and g_j both improve child quality.

Second, M_c is an indicator variable for being married. Marriage is understood to be an institution that increases the value of children — either by raising the quality of children, lowering the costs of raising them, or maintaining the marriage in stressful times. We normalize the value of this increase to 1.[6] The variable s is the value of foregone consumption that results from having children. This is the value of consumption activities that are not complementary with children and which are sacrificed when children arrive. Finally, our critical variable is f_c, the extra cost of having children for same-sex couples. These are the extra costs of parenting that might arise in a same-sex relationship.[7] We assume that these costs are greater for gays than for lesbians; that is, $f_{Gay} > f_{Lesbian} > f_{Hetero} = 0$. [8] [9]

Table 1 shows the four possible utility outcomes once a pairing decides to be a couple (subscripts have been suppressed).

Table 1

Individual Marriage and Child Payoffs

	No Children	**Children**
Cohabiting	(A): h	(C): $h + \gamma - s - f_c$
Married	(B): $h + \epsilon_c$	(D): $h + \epsilon_c + \gamma + 1 - s - f_c$

---

[5] Introducing surrogacy and other options to the model strengthens our results – gays and lesbians with higher g would care *less* about their partner's g because only the higher g would be used – but adds complexity. Hence, we ignore them.

[6] It has been suggested that some benefits of marriage, such as companionship, may be more valuable to childless couples, so that marriage and children may be substitutes in certain ways. However, it appears reasonable to assume that, *overall*, marriage and children are complements (Aldèn *et al.*, 2015). Therefore, one may interpret this "1" as the *net* amount by which marriage increases the value of children.

[7] As discussed in the introduction, these costs of parenting might arise from the inability (or reduced ability) to have a sexual division of labor; that is, "mothering" and "fathering" might be imperfect substitutes. They might also arise from stigma and discrimination against same-sex couples and their children.

[8] Surrogacy is also significantly more costly than insemination, so including these options would strengthen our $f_{Gay} > f_{Lesbian}$ assumption.

[9] More realistically, one could assume f_c to be stochastic, and that the distributions by sexual orientation are ranked by first-order stochastic dominance. Doing so does not qualitatively impact any of our propositions and corollaries.

No outcome dominates the others, and which outcome is chosen depends on the couple's specific values of the various utility components. The difference in utility between cohabitation with children (option (C)) and marriage with children (option (D)) is $1 + \epsilon_c$. The utility difference between cohabitation and marriage without children is just ϵ_c. These values may be greater or less than zero depending on the couples suitability for marriage. The difference in utility between married couples with children and married couples without children is $\gamma + 1 - s - f_c$, which can also be greater or less than zero. As a result, different couple combinations will choose different outcomes with respect to marriage and children.

Incentive to To Marry and Have Children

Because we assume that ϵ_c is observed after a dating pair becomes a couple in stage 3 and is independent of g and h, every *same-sex* couple has the same probability of marriage.[10] We show that the probability of marriage is weakly greater for every heterosexual couple, regardless of the couple's g_i and g_j. It is strictly greater when not all same-sex couples adopt and at least some heterosexual couples have children. This is stated in Proposition 1.

Proposition 1. *Same-sex couples are no more likely to marry than heterosexual couples, and they are strictly less likely to do so than a heterosexual couple with genetic traits (g_i, g_j) when $s > a - f_{Lesbian}$ and $s < \max\{a, b(g_i, g_j)\} + 1$.*[11]

Proof of Proposition 1

All couples who do not have biological children, both same-sex and opposite-sex, will marry if and only if (up to measure zero indifference cases)

$$\epsilon_c > \begin{cases} 0 & \text{if } f_c + s \geq a + 1, \ (\textit{i.e. couple never adopts}) \\ f_c + s - a - 1 & \text{if } a \leq f_c + s < a + 1, \ (\textit{i.e. couple adopts if married}) \\ -1 & \text{if } f_c + s < a, \ (\textit{i.e. couple always adopts}). \end{cases} \quad (2)$$

---

[10] This is not true for heterosexual couples because their decision whether to have children, which impacts the value of marriage, may depend on the genetic attributes g_i, g_j. Hence, different heterosexual couples will have a different threshold for ϵ_c.

[11] These conditions are most likely to hold when a is low relative to $b(g_i, g_j)$, which implies a high cost of adoption.

Because the right-hand side is independent of g and h, all couples of a given sexual orientation will have the same probability of marrying. That is, all gay couples have the same probability of marriage, all lesbian couples have the same probability, and all heterosexual couples that do not have biological children have the same probability. However, the probability of marriage is always weakly greater for heterosexual couples than gay and lesbian couples because the right hand side, the net cost of marriage, is weakly smaller for heterosexual couples.

For the marriage probability to be strictly greater for heterosexual couples, we need $s < a + 1$ so that some heterosexual couples adopt, and $f_{Lesbian} + s > a$, so that some same-sex couples do not adopt. If no heterosexual couples adopt or if all same-sex couples do adopt, then right-hand side of equation (2) is the same regardless of sexual orientation.

Couples who choose to have biological children have an even greater probability of marriage since their value of having children is $b(g_i, g_j) > a$. ∎

Corollary 1. *Lesbian couples are at least as likely to marry as gay couples, and more so if $a - f_{Gay} < s < a + 1 - f_{Lesbian}$.*

Proof of Corollary 1

Use the same reasoning as in the first half of the proof of Proposition 1, and the assumption that $f_{Gay} > f_{Lesbian}$. ∎

Corollary 2. *Heterosexual couples are at least as likely to have children as lesbian couples, which are in turn at least as likely to do so as gay couples. These relations are strict as long as $a - f_{Lesbian} < s < a + 1 - f_{Lesbian}$, so that some, but not all lesbian couples adopt.*

Proof of Corollary 2

This is a direct consequence of Proposition 1, Corollary 1, the fact that marriage increases the utility from having children, and $f_{Gay} > f_{Lesbian} > 0$. ∎

Proposition 1 and its corollaries are rather intuitive. Consider an increase in f_c to f_c', all else equal. Couples who would have chosen no children before the change do not change their behavior because f_c is not in their payoff function. Couples who would have chosen children with cohabitation reveal that $\gamma - s - f_c > 0$, and an increase in f_c means that those at the margin will now decide to have no children. Finally, couples who would have chosen marriage with children under the original cost will now continue with (D) or choose

(A) or (B), depending on how close they are to indifference between having children or not.[12]

In other words, an increase in the cost of children will lead some couples to forego having children, and some of these couples will also forego marriage as a result. No couple changes its decision in the opposite direction. Since $f_{Gay} > f_{Lesbian} > 0$, all other attributes equal, gay couples should be the least likely to have children because f_c is greatest for them, followed by lesbian couples, and finally heterosexual couples.[13]

Corollary 3. *Gay couples are at least as likely to engage in behaviors not complementary with children as lesbian couples, which in turn are at least as likely as heterosexual couples to engage in non-complementary behaviors.*

Proof of Corollary 3

This is a direct consequence of Corollary 2. ∎

Another intuitive result follows from the model:

Corollary 4. *Suppose couples A and B have genetic traits (g_i, g_j) and (g'_i, g'_j), respectively, with $g_i > g'_i$, $g_j \geq g'_j$, and $b(g_i, g_j) > a$. Then if the couples are heterosexual, couple A is more likely to marry than couple B, while if they are same-sex, they have the same probability of marriage.*

Proof of Corollary 4

Given the assumptions, for heterosexuals, we have $\gamma_g > 0$, so couple A has a higher γ than couple B. This is equivalent to facing a lower f, so by the same logic as Proposition 1, couple 1 is more likely to marry. For gays and lesbians, $\gamma = a$ does not depend on g. ∎

Matching Behavior

We now examine matching behavior for different sexual orientations. We focus on stationary situations (which we call *threshold equilibria*) where each type's reservation utility remains constant. We assume that individuals leaving the dating pool are replaced by individuals with the same characteristics. First, we show that same-sex matching

---

[12] Note that they do not choose to cohabitate with children (option C) because they have revealed $1 + \epsilon_c > 0$, and this does not depend on f_c.

[13] If f_c is not driven completely by biology, and depends in part on social stigma and discrimination, then this effect would be reduced over time as such stigma is reduced.

occurs independently from g. This implies that any correlation in g between same-sex partners should disappear once we control for h. Then, we show that this is not the case for heterosexuals. Indeed, under the assumption that $b(.)$ is super-modular in its arguments, heterosexual sorting should have assortative characteristics (defined later) in the g dimension, even controlling for h.[14]

We use the following definitions:

Let $U_i(g_j, h_j)$ be the expected utility of person i, with characteristics (g_i, h_i), coupling with a partner of type (g_j, h_j), prior to observing ϵ_c. That is, $U_i(g_j, h_j)$ is the expectation of U_{ij} maximized over the child and marriage decisions.

A stationary strategy profile σ is a *threshold equilibrium* if there exists a collection of utility thresholds $\underline{U}^\sigma(g_i, h_i)$ such that each type (g_i, h_i) accepts couple with type (g_j, h_j) if and only if $U_i(g_j, h_j) \geq \underline{U}^\sigma(g_i, h_i)$, and in doing so, maximizes their expected utility.

In addition, in a threshold equilibrium σ, let:

$A^\sigma(g_i, h_i) = \{(g_j, h_j) : U_i(g_j, h_j) \geq \underline{U}^\sigma(g_i, h_i)\}$ be the set of types that (g_i, h_i) accepts,

$B^\sigma(g_i, h_i) = \{(g_j, h_j) : U_j(g_i, h_i) \geq \underline{U}^\sigma(g_j, h_j)\}$ be the set of types that accept (g_i, h_i).[15]

Note that due to the search cost k, type (g_i, h_i) may "settle" (accept a type even though better ones exist in $B^\sigma(g_i, h_i)$), even in the absence of transfers, as is the case here.

Same-Sex Matching

Proposition 2. *In any threshold equilibrium σ for same-sex couples, for all $g, g' \in [0, G]$ and all $h \in [0, H]$, $B^\sigma(g, h) = B^\sigma(g', h)$ and $A^\sigma(g, h) = A^\sigma(g', h)$.*

Proof of Proposition 2:

Since gays and lesbians only care about h, $B^\sigma(g, h) = B^\sigma(g', h)$ for all h. Therefore, (g, h) and (g', h), which have the same preferences, must have the same threshold, so $A^\sigma(g, h) = A^\sigma(g', h)$. ∎

---

[14] As discussed by Chiappori, McCann and Neishem (2010), there is no straightforward way to generalize the concept of assortative matching to multiple dimensions. We will simply show that, in our setting, equilibria exhibit characteristics suggesting that heterosexuals with higher g (holding h fixed) tend to have partners with higher g. This is the hypothesis that we will take to the data.

[15] As usual, $A^\sigma(g_i, h_i) = B^\sigma(g_i, h_i) = \emptyset$ is a trivial equilibrium. We will restrict our attention to other equilibria.

Therefore, conditional on h_i, the expected g of individual i's partner in a same-sex couple is independent of g_i. In other words, the genetic fitness of same-sex partners should be uncorrelated, conditional on h. This will not be true in general for heterosexual couples because their genetic fitness is passed on to their own offspring.

Heterosexual Matching

For our study of heterosexual matching, we assume that $b(.)$ is super-modular; that is, if $g_i > g_i'$ and $g_j > g_j'$, then $b(g_i, g_j) + b(g_i', g_j') > b(g_i, g_j') + b(g_i', g_j)$. For example, individuals with high g may place greater value on their children having high g because they do not want their children to face difficulties that they did not face. Alternatively, if parents are risk-averse with respect to the quality of genes passed on to the child, then it is more important for an individual with high g to have a partner with high g (so that good genes are passed on for sure) than for an individual with a low g.

Under the above assumption, we show that heterosexuals with higher g will be more selective than individuals with lower g when considering partners with low g. This points to assortativeness along the g dimension for heterosexual matching.

Recall from equation (1) that the utility function is separable in genetic and household characteristics. We denote the expectation of v_{ij} (before observing ϵ_c) as $v(g_i, g_j)$. Taking the expectation of equation (1) gives: $U_i(g_j, h_j) = h_{ij} + v(g_i, g_j)$. Therefore, $v(g_i, g_j)$ is the expected utility from marriage and/or children, over and above the utility from coupling.

Lemma 1. $v(g_i, g_j)$ is super-modular: if $g_i > g_i'$ and $g_j > g_j'$, then $v(g_i, g_j) - v(g_i, g_j') \geq v(g_i', g_j) - v(g_i', g_j')$. Moreover, if $v(g_i, g_j) - v(g_i, g_j') > 0$, then $v(g_i, g_j) - v(g_i, g_j') > v(g_i', g_j) - v(g_i', g_j')$.

Proof of Lemma 1:

Because $\gamma(g_i, g_j) = \max\{a, b(g_i, g_j)\}$, we have: $\gamma(g_i, g_j) - \gamma(g_i, g_j') \in \{b(g_i, g_j) - b(g_i, g_j'), b(g_i, g_j) - a, 0\}$. Thus, $\gamma(g_i, g_j) - \gamma(g_i, g_j') \leq b(g_i, g_j) - b(g_i, g_j')$. Therefore:

If $\gamma(g_i, g_j) - \gamma(g_i, g_j') = b(g_i, g_j) - b(g_i, g_j')$, then $\gamma(g_i', g_j) - \gamma(g_i', g_j') \leq b(g_i', g_j) - b(g_i', g_j') < b(g_i, g_j) - b(g_i, g_j') = \gamma(g_i, g_j) - \gamma(g_i, g_j')$.

If $\gamma(g_i, g_j) - \gamma(g_i, g_j') = b(g_i, g_j) - a > 0$, then, because $b_1 > 0$, $\gamma(g_i', g_j) - \gamma(g_i', g_j') = \max\{b(g_i', g_j) - a, 0\} < b(g_i, g_j) - a = \gamma(g_i, g_j) - \gamma(g_i, g_j')$.

If $\gamma(g_i, g_j) - \gamma(g_i, g_j') = 0$, then clearly $\gamma(g_i', g_j) - \gamma(g_i', g_j') = 0$.

Thus, $\gamma(g_i, g_j) - \gamma(g_i, g'_j) \geq \gamma(g'_i, g_j) - \gamma(g'_i, g'_j)$.

Therefore, even with the same probability of child rearing, type g_i would benefit weakly more than type g'_i from increasing the genetic type of her partner. Furthermore, for any given type of the partner, type g_i is weakly more likely than type g'_i to have children. Thus, $v(g_i, g_j) - v(g'_i, g_j) \geq v(g_i, g'_j) - v(g'_i, g'_j)$.

If $v(g_i, g_j) - v(g_i, g'_j) > 0$, then it must be that $\gamma(g_i, g_j) - \gamma(g_i, g'_j) \neq 0$. The argument above shows that $\gamma(g_i, g_j) - \gamma(g_i, g'_j) > \gamma(g'_i, g_j) - \gamma(g'_i, g'_j)$. Moreover, if $v(g_i, g_j) - v(g_i, g'_j) > 0$, then it must be that a couple with genetic traits g_i, g_j has a strictly positive probability of having children. Therefore, $v(g_i, g_j) - v(g'_i, g_j) > v(g_i, g'_j) - v(g'_i, g'_j)$. ∎

Lemma 1 shows that if, as we assume, the expected utility from biological children $b(.,.)$ is super-modular, then the expected utility of marriage and children $v(.,.)$ is weakly super-modular. The argument is simple: if $b(.,.)$ is super-modular, then the expected utility from children $\gamma(.,.)$ must be weakly super-modular. Adding marriage into the mix does not change this fact because marriage and children are complementary.

The super-modularity of v leads to result below. To avoid confusion between specific own type and a generic partner's type, we denote the partner's type as (x, y), where x is the genetic trait, and y is the household trait. Moreover, we assume that for every g_i, $v(g_i, G) - v(g_i, 0) > 0$. That is, every heterosexual cares about the genetic trait of their partner at least to some extent.

Proposition 3. *If $g > g'$, then in any threshold equilibrium σ for heterosexual couples, for all h, we have $B^\sigma(g, h) \supseteq B^\sigma(g', h)$, and there exists $g^* > 0$ such that:*

– if $x < g^*$ and $(x, y) \in A^\sigma(g, h)$, then $(x, y) \in A^\sigma(g', h)$; and

– as long as type (g, h) has biological children with positive probability, there exists a positive measure of types (x, y) with $x < g^*$ such that $(x, y) \notin A^\sigma(g, h)$, but $(x, y) \in A^\sigma(g', h)$.

Proof of Proposition 3:

Denote type (g_j, h_j)'s utility from being matched with type (g, h) as $U_{g_j, h_j}(g, h)$. Since $g > g'$, we have $U_{g_j, h_j}(g, h) \geq U_{g_j, h_j}(g', h)$ for all (g_j, h_j). It follows that $B^\sigma(g, h) \supseteq B^\sigma(g', h)$.

Case 1: The boundaries of $A^\sigma(g, h)$ and $A^\sigma(g', h)$ cross. It follows from the discussion about indifference curves in the main text that they must cross at only one type. Letting g^* be this type's genetic characteristic yields the desired result.

Case 2: $A^\sigma(g,h) \subseteq A^\sigma(g',h)$. Here, we must simply show that $A^\sigma(g,h)$ and $A^\sigma(g',h)$ must differ by a set of positive measure. Due to the assumption that every g_i, $v(g_i,G) - v(g_i,0) > 0$, Lemma 1 implies that $v(g,G) - v(g,0) > v(g',G) - v(g',0)$. Thus, no indifference curve for any type can be flat across the entire type space. Again, by the discussion about indifference curves in the main text, the boundary of $A^\sigma(g,h)$ must be strictly steeper than the boundary of $A^\sigma(g',h)$ in some range.

Case 3: $A^\sigma(g',h) \subset A^\sigma(g,h)$. The rest of the proof shows that this case cannot arise if type (g,h) has a positive probability of having biological children.

In a threshold equilibrium, we must have:

$$\int_{A^\sigma(g,h) \cap B^\sigma(g,h)} [U_{g,h}(x,y) - \underline{U}^\sigma(g,h)] dF(x,y) = k$$

$$= \int_{A^\sigma(g',h) \cap B^\sigma(g',h)} [U_{g',h}(x,y) - \underline{U}^\sigma(g',h)] dF(x,y).$$

Define $L = A^\sigma(g,h) \cap B^\sigma(g,h)$ and $L' = A^\sigma(g',h) \cap B^\sigma(g',h)$. Recall that $B^\sigma(g,h) \supseteq B^\sigma(g',h)$. Thus, if $A^\sigma(g',h) \subset A^\sigma(g,h)$, it must be that $L' \subseteq L$. Since $U_{g,h}(x,y) - \underline{U}^\sigma(g,h) > 0$ for all $(x,y) \in A^\sigma(g,h)$, it follows that:

$$\int_{L'} [U_{g,h}(x,y) - \underline{U}^\sigma(g,h)] dF(x,y) \le \int_L [U_{g,h}(x,y) - \underline{U}^\sigma(g,h)] dF(x,y) = k$$

Thus,

$$\int_{L'} [U_{g,h}(x,y) - \underline{U}^\sigma(g,h)] dF(x,y) \le k = \int_{L'} [U_{g',h}(x,y) - \underline{U}^\sigma(g',h)] dF(x,y),$$

and therefore,

$$\frac{\int_{L'} [v(g,x) - v(g',x)] dF(x,y)}{\int_{L'} dF(x,y)} \le \underline{U}^\sigma(g,h) - \underline{U}^\sigma(g',h). \tag{B1}$$

Let $\underline{x}$ be the lowest x such that there exists y for which $(x,y) \in A^\sigma(g,h)$. Let $\underline{y}$ be this y, so that $U_{g,h}(\underline{x},\underline{y}) = \underline{U}^\sigma(g,h)$. Since $A^\sigma(g',h) \subset A^\sigma(g,h)$, it must be that $U_{g',h}(\underline{x},\underline{y}) \le \underline{U}^\sigma(g',h)$. Then we have:

$$\underline{U}^{\sigma}(g,h) - \underline{U}^{\sigma}(g',h) \leq U_{g,h}(\underline{x},\underline{y}) - U_{g',h}(\underline{x},\underline{y}) = v(g,\underline{x}) - v(g',\underline{x}). \qquad (B2)$$

Combining equations B1 and B2 gives:

$$\frac{\int_{L'}[v(g,x) - v(g',x)]dF(x,y)}{\int_{L'} dF(x,y)} \leq v(g,\underline{x}) - v(g',\underline{x}). \qquad (B3)$$

Since $g > g'$ and all $(x,y) \in A^{\sigma}(g',h)$ must have $x \geq \underline{x}$, it follows from Lemma 1 that $v(g,x) - v(g',x) \geq v(g,\underline{x}) - v(g',\underline{x})$. Therefore, for B3 to hold, it must be that $v(g,x) - v(g',x) = v(g,\underline{x}) - v(g',\underline{x})$ for almost all types $(x,y) \in A^{\sigma}(g',h) \cap B^{\sigma}(g',h)$. By the second half of Lemma 1, this means that $v(g,x) - v(g,\underline{x}) = 0$. This implies that type (g,h) does not care about the genetic trait of its partner, which means that it will not have biological children. ∎

This proposition states that for heterosexuals, if individual 1 has a higher g and the same h as individual 2, then individual 1 is acceptable to (weakly) more types, and is weakly more selective (requires equal or higher y) among partners with a low x. Furthermore, if individual 1 has biological children with positive probability, individual 1 is strictly more selective than individual 2 among partners that have a low x, and the additional low x types that individual 1 rejects as a result form a nonzero fraction of the population.

To understand the intuition for this result, note that the boundary of $A^{\sigma}(g,h)$ must be one of type (g,h)'s indifference curves. Figure 1 shows an example of indifference curves for types (g,h) and (g',h), where $g > g'$, plotted on a plane with the partner's genetic trait on the horizontal axis and the partner's household trait on the vertical axis.

Consider the indifference curve for either "high" type (g,h) or "low" type (g',h). For low values of x, the benefits from adopting children are greater than from procreation, and the indifference curves remain flat: the genetic trait x has no value in this region. Once $b > a$, the indifference curves start to fall because one type is willing to substitute x and y in a partner. This occurs sooner for high type (g,h) than for low type (g',h) since $g > g'$. Moreover, by Lemma 1, whenever the marginal utility of x for high type (g,h) is nonzero, it is greater than the marginal utility of x for low type (g',h). Hence, whenever it is not flat, high type (g,h)'s indifference curve at a given (x,y) is strictly steeper than low type (g',h)'s at the same (x,y).[16]

---

[16] The downward sloping part of the indifference curves need not be concave as depicted.
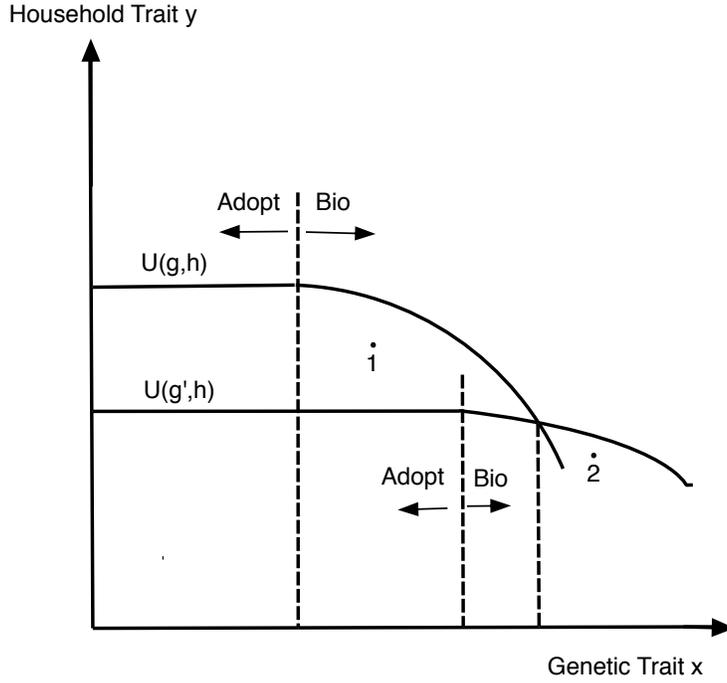
Figure 1: Example of Matching Boundary Conditions

Consider a potential mate given by point 1 in the graph. High type (g, h) would reject this person as a match because they fall below their indifference curve boundary. On the other hand, low type (g', h) would accept this person as a mate. Now consider another potential mate given by point 2. This person has a higher x and lower y than the person at point 1. Now high type (g, h) finds this person acceptable, while low type (g', h) does not. The reason is found in Lemma 1: the supermodularity of expected utility in the genetic trait means that point 2's high genetic trait matters more to high type (g, h) than to low type (g', h).

Figure 1 is one example of the how the boundary conditions might look for types (g, h) and (g', h). There are actually three cases to consider:

1. The indifference curves corresponding to the lowest acceptable utility cross, as depicted in Figure 1. In this case, type (g, h) is less selective than type (g', h) among partners

with high x. Since type (g, h) is more selective among partners with low x, matching along the genetic dimension has assortative properties.

2. Type (g, h) is more selective than type (g', h) among partners of all x. If we assume that h_{ij} is log-separable in h_i and h_j, then even with different partners, types (g, h) and (g', h) have the same marginal utility for y. In this case, type (g, h) has a steeper boundary of acceptable types, which again suggests assortative matching.

3. Type (g, h) is less selective than type (g', h) among partners of all x. The proof of Proposition 3 shows that this is not possible because type (g, h) is more willing to wait for a partner of higher genetic quality than type (g', h), and because type (g, h) is acceptable to more types than type (g', h).

Summary of the Model's Predictions

Our model delivers the following testable implications.[17]

1. Non-heterosexuals are less likely to marry and less likely to have children.

2. Among non-heterosexuals, gays are less likely to marry and less likely to have children compared to lesbians.

3. Non-heterosexual couples are more likely to engage in the consumption of goods that are non-complementary with children, not controlling for the presence of children.

4. Heterosexuals should be more likely to marry and less likely to cohabitate as their g increases. No such relationship should exist for gays and lesbians.

5. Conditional on h, non-heterosexuals should not sort for mates along genetic lines, but there should be positive assortative matching for heterosexuals on genetic lines.

These differences in predicted behavior arise without positing any difference in preferences, marriage market conditions, costs of marriage, or type distributions across sexual orientations. Instead, they all occur due to simple variations across orientations in the availability of means of conception and/or the cost of having children. More complicated models are possible, but our goal is to examine if differences in the costs of children can explain differences in behavior between couples of different orientations.[18]

---

[17] These predictions would hold for other types of couples who, *ex ante*, would be predicted to have high costs of procreation. For example, these predictions would apply to couples that are elderly at the time of matching or infertile. Unfortunately, the data set used here does not allow the identification of such couples.

[18] A model based on differences in preferences (ie., gays and lesbians have a smaller demand for children) will lead to the same predictions as this model because the effect on the shadow price of children would be the same (see Pollak and Wachter (1975)).

3. Robustness Results

3.1. The Data

3.2. Behaviors Non-Complementary With Children

Tables 1S to 4S replicate the tests conducted in Table 3 of the paper. Here we restrict the sample to men and women separately, and reproduce the results with differing controls. The last column of each set of regressions reproduces the results from the paper. The definitions of the variables are the same as in the paper.

TABLE 1S Smoking Behavior
Logit Regression

Variable	Males		Females		Full Sample
	(1)	(2)	(3)	(4)	(5)
Gay coefficient	0.47	0.55			0.68
Gay odds ratio	1.60	1.74			1.99
	$(5.86)^*$	$(3.76)^*$			$(4.65)^*$
Lesbian coefficient			0.39	0.43	0.32
Lesbian odds ratio			1.48	1.54	1.37
			$(3.38)^*$	$(2.06)^*$	(1.54)
Gay× Children		-1.29			-1.35
Gay× Children odds ratio		0.27			0.26
		$(-2.10)^*$			$(-2.18)^*$
Lesbian× Children				-0.98	-0.94
Lesbian× Children odds ratio				0.37	0.39
				$(-2.43)^*$	$(-2.36)^*$
Children		-0.04		0.03	-0.39
Children odds ratio		0.96		1.03	0.97
		(-1.26)		(1.00)	(-1.15)
Exogenous Controls	Yes	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes	Yes
Weighted observations		Yes		Yes	Yes
Robust Std. Errors		Yes		Yes	Yes
N	216,455	165,779	259,941	186,111	351,890
Pseudo R2	0.01	0.03	0.03	0.06	0.03

* Significant at the 5% level. t-statistics in parentheses.

Exogenous Controls: Age, Year, White. Other Controls: Income, Urban, Graduate Work, Obesity, Children, Marriage, and the sexual orientation health interactions.

TABLE 2S Illegal Drug Use
Logit Regression

Variable	Males		Females		Full Sample
	(1)	(2)	(3)	(4)	(5)
Gay coefficient	0.51	0.46			0.58
Gay Odds Ratio	1.66	1.59			1.79
	$(4.16)^*$	$(2.08)^*$			$(2.61)^*$
Lesbian coefficient			0.74	0.45	0.33
Lesbian Odds Ratio			2.10	1.57	1.39
			$(5.99)^*$	$(2.11)^*$	(1.57)
Gay× Children		−2.22			−2.21
Gay× Children odds ratio		0.10			0.11
		$(-2.14)^*$			$(-2.13)^*$
Lesbian× Children				−0.18	0.15
Lesbian× Children odds ratio				0.86	1.16
				(−0.06)	(0.30)
Children		−0.23		−0.21	−0.05
Children odds ratio		0.79		0.81	0.95
		$(-4.30)^*$		$(-3.62)^*$	$(-1.37)^*$
Exogenous Controls	Yes	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes	Yes
Weighted observations		Yes		Yes	Yes
Robust Std. Errors		Yes		Yes	Yes
N	137,918	102,378	164,185	112,266	214,644
Log Likelihood	−2,527,473	−1,927,439	−2,019,022	−1,494,923	−3,433,477
Pseudo R2	0.12	0.14	0.13	0.14	0.14

* Significant at the 5% level. t-statistics in parentheses.

Exogenous Controls: Age, Year, White. Other Controls: Income, Urban, Graduate Work, Obesity, Children, Married, and the sexual orientation health interactions.

TABLE 3S More Than Four Sex Partners
Logit Regression

Variable	Males		Females		Full Sample
	(1)	(2)	(3)	(4)	(5)
Gay coefficient	1.86	0.93			1.10
Gay Odds Ratio	6.44	2.52			3.02
	(12.88)*	(3.28)*			(4.00)*
Lesbian coefficient			1.18	3.48	3.02
Lesbian Odds Ratio			3.22	32.6	22.81
			(2.07)*	(2.30)*	(2.10)*
Gay× Children		−0.90			−0.91
Gay× Children odds ratio		0.40			0.40
		(−1.00)			(−1.01)
Lesbian× Children				−1.61	−2.32
Lesbian× Children odds ratio				0.29	0.17
				(−1.37)	(−3.71)*
Children		−0.55		−0.54	−0.66
Children odds ratio		0.57		0.58	0.51
		(−5.95)*		(−4.53)*	(−8.54)*
Exogenous Controls	Yes	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes	Yes
Weighted observations		Yes		Yes	Yes
Robust Std. Errors		Yes		Yes	Yes
N	64,626	54,639	73,648	59,397	106,788
Log Likelihood	−804,617	−720,213	−342,582	−289,271	−1,044,123
Pseudo R2	0.08	0.13	0.08	0.16	0.18

* Significant at the 5% level. t-statistics in parentheses.

Exogenous Controls: Age, Year, White. Other Controls: Income, Urban, Graduate Work, Obesity, Children, Married, and the sexual orientation health interactions.

	Males		Females		Full Sample
Variable	(1)	(2)	(3)	(4)	(5)
Now Married					
Gay coefficient	−1.72	−1.76			−1.97
Gay odds ratio	0.17	0.17			0.14
	(−7.59)*	(−7.17)*			(−6.58)*
Lesbian coefficient			−1.69	−1.71	−2.04
Lesbian odds ratio			.18	−0.17	0.13
			(−9.57)*	(−8.53)*	(−8.59)*
Gay× Children		1.00			2.05
Gay× Children odds ratio		2.73			7.77
		(1.81)			(3.35)*
Lesbian× Children				0.14	0.51
Lesbian× Children odds ratio				1.23	1.67
				(0.05)	(0.97)
Children		0.81		0.38	0.66
Children odds ratio		2.24		1.46	1.95
		(23.67)*		(12.40)*	(24.58)*
Lesbian × Healthy			0.02	0.01	0.04
odds ratio			1.01	1.01	1.03
			(0.07)	(0.05)	(0.11)
Gay × Healthy	−0.31	−0.28			−031
odds ratio	0.73	0.75			0.73
	(−1.07)	(−0.86)			(−0.80)
Heterosexual × Healthy	0.28	0.19	0.14	0.05	0.38
odds ratio	1.32	1.22	1.14	1.04	1.47
	(11.30)*	(6.68)*	(6.45)*	(1.80)	(14.83)*
Exogenous Controls	Yes	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes	Yes
Weighted observations		Yes		Yes	Yes
Robust Std. Errors		Yes		Yes	Yes
N	146,189	118,600	196,123	144,749	199,891
Log Likelihood	−3,031,551	−2,469,837	−2,874,101	−2,190,262	−4,687,039
Pseudo R2	0.09	0.07	0.02	0.03	0.16

* Significant at the 5% level. t-statistics in parentheses.

Exogenous Controls: Age, Year, White. Other Controls: Smoking, Income, Spouse Income, Urban, Graduate Work, Obesity.