Supplementary Appendix: Matching, Marriage, and Children

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

March 2016

---

[*] Corresponding author email: allen@sfu.ca, Tel: 778 782-3445, Fax: 778 782-4955. Department of Economics, Simon Fraser University, Burnaby BC, V5A 1S6. This research and analysis are based on CCHS and Census data from Statistics Canada, and the opinions expressed do not represent the views of Statistics Canada. Thanks to Statistics Canada for access to data through their RDC program, and to the Social Sciences and Research Council of Canada for funding.

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:

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

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)

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

---

[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:

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

Thus,

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

and therefore,

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

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 $U_{g',h}(\underline{x},\underline{y}) \leq \underline{U}^\sigma(g',h)$. 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.

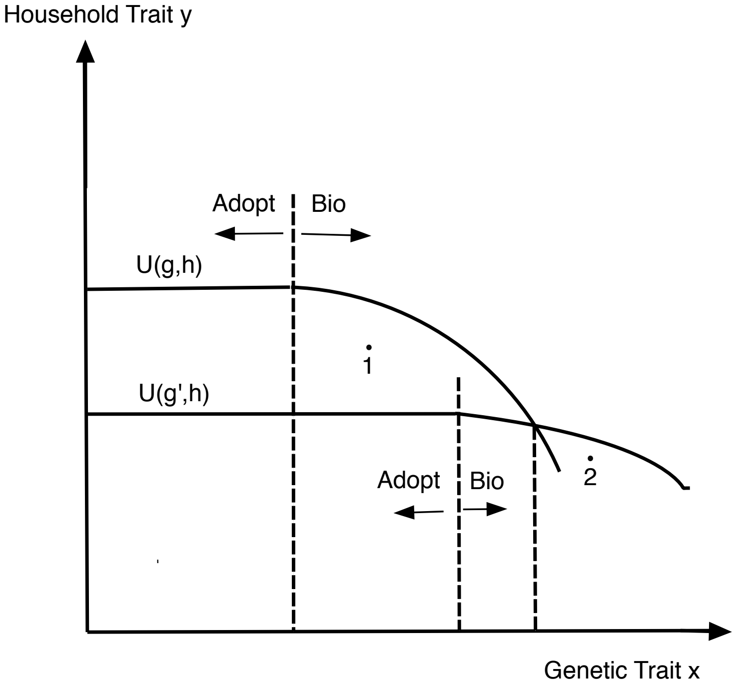

Figure 1: Example of Matching Boundary Conditions

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.

TABLE 4S: Logit Regressions for Marriage vs Cohabitation

	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.