What Makes You Click:
An Empirical Analysis of Online Dating*

Günter J. Hitsch

University of Chicago

Graduate School of Business

Ali Hortaçsu

University of Chicago

Department of Economics

Dan Ariely

MIT

Sloan School of Management

January 2005

Abstract

This paper uses a novel data set obtained from a major online dating service to draw inferences on mate preferences and the match outcomes of the site users. The data set contains detailed information on user attributes such as income, education, physique, and attractiveness, as well as information on the users' religion, political inclination, etc. The data set also contains a detailed record of all online activities of the users. In particular, we know whether a site member approaches a potential mate and receives a reply, and we have some limited information on the content of the exchanged e-mails. A drawback of the data set is that we do not observe any "offline" activities. We first compare the reported demographic characteristics of the site users to the characteristics of the population-at-large. We then discuss the conditions under which the user's observed behavior reveals their mate preferences. We estimate these preferences and relate them to own and partner attributes. Finally, we predict the equilibrium structure of matches based on the preference estimates and a simple matching protocol, and compare the resulting sorting along attributes such as income and education to observed online matches and actual marriages in the U.S.

1

1 Introduction

Economic models of marriage markets predict how marriages are formed, and make statements about the efficiency of the realized matches. These predictions are based on a specification of mate preferences, the matching protocol, i.e. the mechanism by which matches are made, the information structure of the game, and the strategic sophistication of the agents. The seminal work by Gale and Shapley (1962) and Becker (1973) is based on specific assumptions of these model primitives. Since then, the empirical literature on marriage markets has been concerned with the estimation of mate preferences and the relationship between preferences and the structure of observed matches, such as the correlation of men's and women's age or income in a marriage. Our paper contributes to this literature by exploiting a detailed data set of partner search from an online dating service. We provide a description of how men and women interact in the online dating market, and exploit the observed partner search behavior to relate mate preferences on both sides of the market to user attributes, in particular looks and socioeconomic factors such as income and education. Based on the preference estimates, we predict the structure of equilibrium matches, and compare the predictions to observed online matches and actual marriages in the U.S.

Our empirical analysis is based on a new data set that we obtained from a major online dating website. This data set records all activities of 23,000 users in Boston and San Diego during a three and a half months period in 2003. Anecdotal evidence and press coverage suggest that online dating is becoming a widespread means of finding a partner in both the U.S. and many other countries around the world.[1] For research purposes, online dating provides an unusual opportunity to measure mate attributes, and capture the users' search process and the interactions between potential partners. Users who join the dating service post a "profile" on the dating website, that provides their potential partners with information about their age, income, education level, ethnicity, political inclinations, marital status, etc. The users can also post one or more photographs of themselves on the website. Using a laboratory environment, we assigned a numeric looks rating to these users. Together with other information, such as the users' height and weight, this provides us with a measure of physical attractiveness that is otherwise hard to obtain from field data. Our data set lets

---

[1]According to a recent estimate based on ComScore Networks' analysis of Internet users' browsing behavior, 40 million Americans visited online dating sites in 2003, generating $214 million in revenues, making online dating the most important subscription-based business on the Internet. Match.com, which was founded in 1995 as one of the pioneering online dating sites, boasted 939,000 paying subscribers as of the fourth quarter of 2003. Although the sector is led by large and nationally advertised sites like Match.com, Matchmaker.com, and eHarmony.com, along with online dating services bundled by major online service providers (such as Yahoo!Singles), there are also numerous online dating sites that cater to more specialized audiences, such as JDate.com, which bills itself as the "The largest Jewish singles network," Gay.com, BlackSinglesConnection.com, and ChristianSingles.com.

us track the users' activities at a detailed level. At each moment in time, we know which profile they browse, whether they view a specific photograph, and whether they send or reply to a letter from another user. We also have some limited information on the contents of the e-mails exchanged; in particular, we know whether the users exchanged phone numbers or e-mail addresses. A drawback of our data set is that we do not observe whether an online exchange between two users finally results in a marriage, which is the ultimate object of our interest. Also, users may lie about their true attributes. However, when we compare the reported socioeconomic characteristics of the site users to local population characteristics surveyed by the U.S. Census, we do not find stark differences, especially after controlling for Internet use.[2]

The identification of preferences in matching models, and in marriage markets in particular, is complicated if only final matches are observed, or if agents behave strategically. For example, a man with a low attractiveness rating may not approach a highly attractive woman if the chance of forming a match with her is low, such that the *expected* utility from a match is lower than the cost of writing an e-mail or the disutility from a possible rejection. In that case, his choice of a less attractive woman does not reveal his true preference ordering. However, in section 4 we find evidence that the site users are more likely to approach a more attractive mate than a less attractive mate, regardless of their own attractiveness rating. I.e., even if strategic behavior has some impact on the users' choices, the effect is not strong enough to cloud the relationship between attractiveness and the probability of being approached. We then estimate preferences using the following simple identification strategy. Suppose that if user A is more attractive than user B, the probability of receiving an e-mail from any potential mate is higher for A than for B.[3] This assumption has empirical support in our data. Furthermore, suppose that men and women can be ranked according to a single dimensional "type". Under these assumptions, the number of unsolicited e-mails, i.e. the number of "first contacts" from a potential mate that a users receives (per unit of time) reveals her or his type. We can then use regression analysis to investigate how physical, socioeconomic, and other attributes enter into this single dimensional index of attractiveness.

The single dimensional type or index assumption means that user preferences are homogenous. We relax this assumption in two ways. First, we segment users into a priori chosen segments, for example low and high income users. If the chosen segmentation is

---

[2]There do appear certain patterns in our sample that are distinct. Men are overrepresented on the dating site, and minorities are largely underrepresented. Furthermore, the age profile appears more skewed towards the 20-30 year old range, which is of course as expected.

[3]Such behavior is consistent with a cut-off rule that arises in some models of mate search (Shimer and Smith 2000).

correct, the preferences of users within a specific segment can be estimated by regressing the first contacts from those users on mate attributes. Second, we estimate a discrete choice model of the decision to contact a potential mate after viewing his or her profile. We relate this decision to both the attributes of the user who makes the first contact decision, and the attributes of the potential mate. In particular, we consider both the level and the difference between the attribute levels of the two users, such as the difference in education. This approach allows for a more flexible way of assessing preference heterogeneity.

Our empirical analysis reveals the following findings: Many of the self-reported user attributes are strongly associated with online "success," in particular the number of introductory e-mails received. There are similarities, but also stark differences between the determinants of success of men and women. Online success is strongly increasing in men's and women's looks ratings, and the effect sizes are similar. Height and weight are strongly related to outcomes, but here the effects are qualitatively and quantitatively different for men and women. The most striking difference across genders is related to earnings and education. Both men and women prefer partners with higher incomes, but this preference is much more pronounced for women. While income preferences appear to be largely homogeneous, there is heterogeneity in the way men and women value their partner's education level. Generally, users prefer a partner who has a similar education level. However, while men have a particularly strong "distaste" for a better educated partner, women particularly try to avoid less educated men. The users of the dating service typically have strong preferences for a partner of their own ethnicity, and this effect is more pronounced for women than for men.

Since Becker's seminal work, the literature on marriage markets has focused on analyzing whether men and women sort along certain characteristics. Sorting along dimensions such as income and education has been argued to be among the determinants of long term trends in ability and the distribution of income. Sorting may arise in equilibrium due to people's preferences, or it may be due to search frictions, i.e. the time cost of meeting and getting to know a potential partner. In contrast to the traditional way of finding a mate, online dating provides an environment with much reduced search costs. Thus, we expect that online matches mostly reflect men's and women's preferences and the equilibrium mechanism by which matches are formed. In section 7, we present evidence on the structure of online matches[4] and actual marriages in the U.S., for example the correlation in age and income among matched partners. Based on our preference estimates, we then predict who matches with whom in a "stable" equilibrium, obtained by the Gale-Shapley algorithm.

---

[4]Our data set does not allow us to observe whether users who meet online eventually get married. However, utilizing information on the content of exchanged e-mails, such as a phone number or e-mail address, we can assess whether an online meeting resulted in an initial match.

4

Interestingly, while our predictions for the correlation in income is consistent with actual marriages and online matches, we vastly underpredict the correlation in education compared to marriages in the U.S. This suggests that the strong sorting along education as observed in actual marriages is only partially driven by preferences. Search frictions, and the resulting outcome in which people marry partners who they met in high school, college, or at work, seem to play an important role in the formation of marriages. A caveat to this suggestion concerns the interpretation of our preference estimates: If these preferences are over initial "dates," and differ from preferences over marriages, it is conceivable that people increase the weight placed on the education of their partner as the relationship progresses.

Our work relates to the economic literature on matching and marriage markets in several ways. A long literature in economics, sociology and demography has focused on reporting correlations between married couples' socioeconomic attributes. However, it is difficult to interpret these correlations in terms of underlying preferences without knowing the choice constraints faced by the matching parties. A long literature in psychology has thus taken the approach of measuring "stated" preferences through a wide variety of surveys in which participants are asked to rate hypothetical (or real) partners. Another approach is to assess "revealed" preferences by interpreting observed match outcomes through an explicit economic model which generates match outcomes as an equilibrium prediction (Wong 2001; Choo and Siow 2003). Our work, in contrast, tries to estimate preferences in an environment where the users' choices among potential mates can be more directly related to their preferences. Our work may thus be viewed as an attempt to measure "revealed" preferences using data in a setting where the matching protocol, information available to the agents (including choice alternatives), and the choices made by agents are observed. In this regard, the work that comes closest to ours is that of Fisman, Iyengar and Simonson (2004), who investigate revealed preference determinants of mate selection using an experimental speed dating market. In contrast to their work, we emphasize how preferences and match outcomes are related to socioeconomic characteristics such as income and education. Our large and diverse sample is more ideally suited to analyze this question than theirs, which is mostly composed of graduate students at one U.S. university. On the other hand, their approach is better suited to assessing the importance of factors that are hard to measure, such as shared interests between two potential partners.

The paper proceeds as follows. Section 2 provides a description of the workings of the dating site, and the characteristics and intentions of the site users. Section 3 outlines the modeling framework. Section 4 describes some aspects of the users' search behavior, and presents evidence that supports the monotonicity assumption that we use for the identification of user preferences. Section 5 relates online outcomes, in particular the number of

5

first contact e-mails received, to user attributes. Under our assumptions, these regressions reveal the users' preferences. In section 6, we take a discrete choice approach to estimating preferences, and account for preference heterogeneity in a more flexible way. Section 7 compares the predicted sorting based on our preference estimates with the structure of online matches and actual marriages. Section 8 concludes.

2   The Data and User Characteristics: Who Uses Online Dating?

Our data set contains socioeconomic and demographic information and a detailed account of the website activities of more than 23,000 users of a major online dating service. 11,390 users were located in the Boston area, and 11,691 users were located in San Diego. We observe the users' activities over a period of three and a half months in 2003. We first provide a brief description of online dating that also clarifies how the data were collected.

Upon joining the dating service, the users answer questions from a mandatory survey and create "profiles" of themselves.[5] Such a profile is a webpage that provides information about a user and can be viewed by the other members of the dating service. The users indicate various demographic, socioeconomic, and physical characteristics, such as their age, gender, education level, height, weight, eye and hair color, and income. The users also answer a question on why they joined the service, for example to find a partner for a long-term relationship, or, alternatively, a partner for a "casual" relationship. In addition, the users provide information that relates to their personality, life-style, or views. For example, the site members indicate what they expect on a first date, whether they have children, their religion, whether they attend church frequently or not, and their political views. All this information is either numeric (such as age and weight) or an answer to a multiple choice question, and hence easily storable and usable for our statistical analysis. The users can also answer essay questions that provide more detailed information about their attitudes and personalities. This information is too unstructured to be usable for our analysis. Many users also include one or more photos in their profile. We have access to these photos and, as we will explain in detail later, used the photos to constructed a measure of the users' physical attractiveness.

After registering, the users can browse, search, and interact with the other members of the dating service. Typically, users start their search by indicating an age range and geographic location for their partners in a database query form. The query returns a list

---

[5]Neither the names nor any contact information of the users were provided to us in order to protect the privacy of the users.

of "short profiles" indicating the user name, age, a brief description, and, if available, a thumbnail version of the photo of a potential mate. By clicking on one of the short profiles, the searcher can view the full user profile, which contains socioeconomic and demographic information, a larger version of the profile photo (and possibly additional photos), and answers to several essay questions. Upon reviewing this detailed profile, the searcher decides whether to send an e-mail (a "first contact") to the user. Our data contain a detailed, second by second account of all these user activities.[6] We know if and when a user browses another user, views his or her photo(s), sends an e-mail to another user, answers a received e-mail, etc. We also have additional information that indicates whether an e-mail contains a phone number, e-mail address, or keyword or phrase such as "let's meet", based on an automated search for special words and characters in the exchanged e-mails.[7]

In order to initiate a contact by e-mail a user has to become a paying member of the dating service. All users can reply to a received e-mail, independent of whether the are paying members or not.

In summary, our data provide detailed user descriptions, and we know how the users interact online. The keyword searches provide some information on the progress of the online relationships, possibly to an offline, "real world" meeting. We now give a detailed description of the users' characteristics.

Motivation for using the dating service   The registration survey asks users why they are joining the site. It is important to know the users' motivation when we estimate mate preferences, because we need to be clear whether these preferences are for a relationship that might end in a marriage, or only for casual sex. 39% of the users state that they are "hoping to start a long-term relationship," 26% state that they are "just looking/curious," and 9% declare that they are looking for a casual relationship. Perhaps not surprisingly, men seem to be more eager for a short term/casual relationship (14%) than women (4%).

Users who–according to their own stated preferences–joined the dating service to find a long-term relationship account for more than half of all observed activities. For example, men who are looking for a long-term relationship account for 57% of all e-mails sent by men; among women who are looking for a long-term relationship the percentage is 53%. The corresponding numbers for e-mails sent by users who are "just looking/curious" is 22% for men and 20% for women. Only a small percentage of user activities is accounted for by members who are seeking a casual relationship; the fraction of sent e-mails is 2.9% for men and 2.4% for women.

---

[6]We obtained this information in the form of a "weblog."

[7]We do not have access to the full content of the e-mail, or the e-mail address or phone number that was exchanged.

We conclude that at least half of all observed activities is accounted for by people who have a stated preference for a long-term relationship and thus possibly for an eventual marriage. In addition, many of the users who state that they are "just looking/curious" possibly choose this answer because it sounds less committal than "hoping to start a long-term relationship". Under this assumption, the activities of more than 75% of all users reveal attitudes towards a long-term partner.

Sexual preferences The registration also asks users about their sexual preferences. 93% of the users declare that they are heterosexual, while 9% of women and 5% of men are homosexual or bisexual. Most of our analysis focuses on the preferences and match formation among men and women in heterosexual relationships; therefore, we retain only the heterosexual users in our sample.[8] Among them, 2.5% of men and 6% of women state that they have had at least one homosexual experience or could be persuaded to have a homosexual experience. On the other hand, 8% of men and 5% of women declare that homosexuality offends them.

Demographic/socioeconomic characteristics We now investigate the reported characteristics of the site users, and contrast some of these characteristics to representative samplings of these geographic areas from the CPS Community Survey Profile (table 2.1). In particular, we contrast the site users with two sub-samples of the CPS. The first sub-sample is a representative sample of the Boston and San Diego MSA's (Metropolitan Statistical Area), and reflects information current to 2003. The second CPS sub-sample conditions on being an Internet user, as reported in the CPS Computer and Internet Use Supplement, which was administered in 2001.

A visible difference between the dating site and the population-at-large is the overrepresentation of men on the site. In San Diego, 55% of users, and in Boston, 54% of users are men.[9] Another visible difference is in the age profiles: Site users are more concentrated in the 26-35 year range than both CPS samples (the median user on the site is in the 26-35 age range, whereas the median person in both CPS samples is in the 36-45 age range). People above 56 years are underrepresented on the site compared to the general CPS sample; however, when we condition on internet use, this difference in older users attenuates somewhat.

---

[8]Unless noted otherwise, all sample statistics reported are with respect to our main sample of 23,000 heterosexual users.

[9]When we restrict attention to members who have posted photos online (29% of registered users in Boson and 35% of users in San Diego), the percentage difference between male and female participation decreases slightly: In Boston 52% of users is Boston and 53% of users in San Diege are men.

The ethnic profile of site users appears to roughly reflect the profile of the geographic regions covered by the site, especially when conditioning on Internet use, although Asians appear to be underrepresented on the San Diego site.[10]

The reported marital status of site users clearly represents the fact that most users are looking for a partner. 57% of the site users are single, 24% are divorced, and 4% are separated. The fraction of divorced women (27%) is higher than the fraction of divorced men (21%). A small number of the site users, especially among men, is "happily married" (1.9% of men and 0.6% of women) or "not-so-happily married" (3.2% of men and 1.1% of women). This suggests that people in a long term relationship may use the site as a search outlet. Of course, one may expect the *true* percentage of otherwise committed people on this site to be much higher than what is *reported*.[11]

The education profiles of the site users show that site users are in general more educated than the general CPS population. Above 54% of the site users have college degrees or above. However, the educational profiles appear quite close to that of the Internet using population, with only a slightly higher percentage of professional degree holders.

The income profiles also reflect similar patterns to the education profile. Site users are in general a higher income sample than the overall CPS population, but not compared to the "Internet using" population. One visible difference between income profiles is that about 3.5% of site users declare their annual income to be above $200,000, whereas the CPS samples contain 0.0% of the population in this cell.

These comparisons suggest that the online dating site attracts users that are typically single, somewhat younger, more educated, and have a higher income than the general population. Once we condition on household internet use, however, the remaining differences are not large. We believe this reflects that online dating has become an accepted and widespread means of partner search in the internet using population during recent years.

Reported physical characteristics of the users Our data set contains detailed (although self-reported) information regarding the physical attributes of the users. As mentioned before, 32% of the users have posted one or more photos online. For the rest of the users, the survey is the primary source of information about their appearance.

The survey asks the users to rate their looks on a subjective scale. 20% of men and 24% of women possess "very good looks," and 49% of men and women have "above average looks."

---

[10]We should note that the "Other" category in the site's ethnic classification includes several ethnicities grouped under "White" by the CPS. Once we reconcile these classifications, the differences in the "White" category disappear.

[11]Only 14 out of the 311 "happily married" users actually posted a picture, and only 17% of these users state that they are hoping to start a long term relationship, while 25% seek a casual affair.

Only a minority–29% of men and 27% of women–declare that they are "looking like anyone else walking down the street." That leaves less than 1% of users with "less than average looks", and a few members who avoid the question and joke that one should "bring your bag in case mine tears." Posting a photo online is a choice, and hence one might suspect that those users who post a photo are on average better looking. On the other hand, those users who do not post a photo might misrepresent their looks and give an inflated assessment of themselves. The data suggest that the former effect is more important. Among those users who have a photo online, the fraction of average looking members is smaller (24% of men and 19% of women) and correspondingly the fraction of above average looking members is larger compared to all site users.

The registration survey contains information on the users' height and weight. We compared these reported characteristics with information on the whole U.S. population, obtained from the National Health and Examination Survey Anthropometric Tables (the data are from the 1988-1994 survey and cover only Caucasians). Table 2.2 reports this comparison. Among women, we find that the average stated weight is less than the average weight in the U.S. population. The discrepancy is about 6 lbs among 20-29 year olds, 18 lbs among 30-39 year olds, and 20 lbs among 40-49 year olds. On the other hand, the reported weights of men are generally slightly above yet close to the national averages. The stated height of both men and women is somewhat above the U.S. average. This difference is more pronounced among men, although the numbers are small in size. For example, among 20-29 year olds, the difference is 1.3 inches for men and 1 inch for women. The weight and height differences translate into body mass indices (BMI) that are 2 to 4 points less than national averages among women, and about 1 point less than national averages among men.

The data also contain some interesting statistics on the site members' hair color, and indirectly on the usage of hair dye.[12] The majority of users (51% of men and 41% of women) have brown hair. Blond hair is quite prevalent among women, but not among men–29% of the female users, but only 12% of all male users state that they have blond hair. Also, 8% of all women have auburn, and 4% have red hair, while only 2.5% of men have one of these hair colors. Black hair, on the other hand, is more common among men–21% of men, but only 12% of women have black hair.

Measured Physical Characteristics of the Users   31% of men (3920 users) and 32% of women (3328 users) had photos available for browsing. To construct an attractiveness rating for these available photos, we recruited 100 subjects from the University of Chicago GSB

---

[12]A priori, we consider it less likely that selection is an important source of the gender differences in hair colors.

Decision Research Lab mailing list. The subjects were University of Chicago undergraduates and graduate students, falling in the 18-25 age group. Half of the subjects were male students and the other half were female students. Each subject rated 800 pictures: 400 male faces and 400 female faces. The study took approximately 1 hour to complete and subjects were paid $10 for their time. Each subject came in, read and signed the consent form (which told them the study was about rating the physical attractiveness of faces). Each subject was then brought to a computer and started the experiment, in which they were asked to rank the physical attractiveness of the pictures on a scale of 1 to 10. Once the experiment was done, the subjects were paid $10. There was a 1 minute obligatory break between the 200 first and second set of pictures within a gender, and a 3 minute obligatory break between the two sets of face genders.

The order in which photos were presented to the subjects was counterbalanced: half of the subjects rated 400 male faces followed by 400 female faces. The other half rated 400 female faces followed by 400 male faces. We assigned the same picture ordering for each male-female subject pair. Specifically, for each subject, we randomly took 400 male and 400 female pictures from the entire list of pictures (without replacement). Each particular ordering was assigned for both a male and a female subject. For the next male-female subject pair we took the next randomly chosen 400 male and female pictures and blocked them, etc. Since we had 100 subjects, we generated 50 such lists, which meant a given picture was used approximately 12 times across subjects.

Consistent with findings in a large literature in cognitive psychology, attractiveness ratings by independent observers appear to be positively correlated (for surveys of this literature, see Langlois et. al. 2000, Etcoff 2000, and Buss 2003). The average pairwise correlation across raters was found to be 0.24, and the Cronbach's alpha across 12 ratings per photo was calculated to be 0.80. As a comparison, in their analysis of the relationship between the looks and subsequent earnings of law school graduates, Biddle and Hamermesh (1998) found the average pairwise correlation of raters to be 0.40, with a Cronbach alpha of 0.75. A potential explanation for the lower degree of interrater correlation in our sample is the fact that our sample is more heterogeneous than Hamermesh and Liddle's, and that the photos we have are composed and formatted in a much less controlled fashion than the matriculation photos analyzed in their study. However, our Cronbach alpha measure is slightly higher and satisfies the reliability criterion (0.80) used in most social science applications of this measure.

To eliminate rater-specific mean and variance differences in rating choices, we followed Biddle and Hamermesh (1998) and standardized each photo rating by subtracting the mean rating given by the subject, and dividing by the standard deviation of the subject's ratings.

We then averaged this standardized rating across all subjects who rated a particular photo to calculate the "mean standardized rating" for the photo.

Table 2.3 reports the results of regressions of (reported) annual income on the attractiveness ratings. Our results largely replicate the findings of Hamermesh and Biddle (1994) and Biddle and Hamermesh (1998), although unlike Biddle and Hamermesh (1998), the cross-sectional rather than panel nature of our data set makes it difficult for us to argue in favor of a causal relationship between looks and earnings. Nevertheless, the estimated correlations between attractiveness ratings and reported income are economically significant. The coefficient estimates on the standardized attractiveness score imply that a one standard deviation increase in a man's attractiveness score is correlated with a 6% increase in his earnings, whereas for a woman, the attractiveness premium is 9%. Notice that these estimates are roughly commensurate with our estimates for the return to one additional year of schooling. Interestingly, there also appears to be a significant height premium for men – a one inch increase in a man's height is correlated with a 2% increase in earnings (a smaller 1% premium is found for women). Although point estimates of the coefficient on weight are negative, the coefficient estimates are not statistically significant.

3 A Modeling Framework for Analyzing User Behavior

The seminal economic models of marriage markets were developed by Gale and Shapley (1962) in the nontransferable utility (NTU) framework and by Becker (1973) and Shapley and Shubik (1972) in the transferable utility (TU) framework. These models rely on either centralized matching algorithms or the absence of search frictions. Recent authors, including Morgan (1995), Lu and McAfee (1996), Burdett and Coles (1997), Shimer and Smith (2000), Smith (2002), Adachi (2003), and Atakan (2004) have investigated the role played by search frictions. In particular, Adachi (2003) establishes the outcome of the Gale-Shapley procedure as the limiting outcome of a decentralized search and matching model where search costs are negligible (we review the Gale-Shapley procedure in Section 7). We now overview the Adachi (2003) model to motivate our data analysis.

Adachi (2003) considers a marriage market populated with ex-ante heterogeneous men and women. Each man (woman) is indexed by his type characterized by a real number $m \in M$ ($w \in W$). The distributions of types of single men and women are λ_M and λ_W. Time is discrete, with period discount factor ρ. Each period, a single man (woman) meets a woman (man) with probability δ, whose type is drawn according $\lambda_W(\lambda_M)$. A match with a type w woman gives a utility of $q_M(m, w)$ to type m man. Similarly, a type w woman gets utility $q_W(m, w)$ from this match. A match is made if both parties agree – if not, the

agents continue searching. To keep the distributions of types of men and women constant, Adachi (2003) assumes that if a type m man and a type w woman are matched, they are immediately replaced by a single type m man and a single type w woman.

Let $v_M(m)$ $(v_W(w))$ be the *reservation utility* of a type m man (type w woman) from staying single and continue searching. A type m man will agree to match with woman w if $q_M(m, w) \geq v_M(m)$. Similarly, a w woman will match with man m if $q_W(m, w) \geq v_W(w)$. Given these reservation utility search rules, we can define the *acceptance sets* (borrowing the terminology of Shimer and Smith (2000)):

$$A_M(m; v_M) = \{w \in W | q_M(m, w) \geq v_M(m)\}$$
$$A_W(w; v_W) = \{m \in M | q_W(m, w) \geq v_W(w)\}$$

where $A_M(m; v_M)$ $(A_W(w; v_W))$ is the set of women (men) that man m (woman w) would agree to match with.

Opposing the acceptance sets, we have the *opportunity sets*:

$$\Omega_M(m; v_W) = \{w \in W | q_W(m, w) \geq v_W(w)\}$$
$$\Omega_W(w; v_M) = \{m \in M | q_M(m, w) \geq v_M(m)\}$$

where $\Omega_M(m; v_M)$ $(\Omega_W(w; v_W))$ is the set of women (men) who would be willing to match with man m (woman w). Given these definitions, the Bellman equation characterizing the optimal search rule of man m is given by:

$$v_M(m) = \rho\delta \left\{ \int 1_{\Omega_M}(m, w; v_W) \max(q_M(m, w), v_M(m)) d\lambda_W(w) + v_M(m)\lambda_W(\Omega_M^C(m; v_W)) \right\}$$
$$+ \rho(1 - \delta)v_M(m)$$

where $1_{\Omega_M}(m, w; v_W)$ is the indicator for meeting woman w within man m's opportunity set $\Omega_M(m; v_W)$. $\Omega_M^C(m; v_W)$ is the complement of this opportunity set. The right hand side of this equation is the expected utility from entering into the next period single. With probability δ, man m will meet a woman. With probability $\lambda_W(\Omega_M^C(m; v_W))$, this woman will not be within $\Omega_M(m; v_W)$, hence the man will be rejected, and will stay single. For women within $\Omega_M(m; v_W)$, the man will choose to be matched with those who are within his acceptance set, and continue searching (and getting $v_M(m)$) if he meets women who are not acceptable.

The Bellman equation characterizing the optimal search rule of woman w is similarly

given by:

$$v_W(w) = \rho\delta \left\{ \int 1_{\Omega_W}(w,m;v_M)\max(q_W(m,w),v_W(w))d\lambda_M(m) + v_W(w)\lambda_M(\Omega_W^C(w;v_M)) \right\}$$
$$+ \rho(1-\delta)v_W(w)$$

Adachi (2003) characterizes the stationary equilibrium in this marriage market as the profile of reservation utilities $(v_M^*(m), v_W^*(w))$ that solve the (system of) Bellman equations. In particular, he shows that the set of equilibrium reservation utility profiles is nonempty, and forms a complete lattice under the partial ordering of male preferences. Most notably, Adachi (2003) also shows that as $\rho \to 1$, i.e. search costs become negligible, the set of equilibrium reservation utility profiles characterize matching profiles that are *stable* in the Gale-Shapley (1962) sense.

Empirical Implications of the Adachi (2003) Model The main assumption that allows us to use the Adachi (2003) framework to interpret our data is to assume that *searching members contact (with constant probability) every member they encounter that they would be willing to match with.* I.e. man m sends an e-mail to every woman $w \in A_M(m, v_M)$ (or a constant fraction of such acceptable women, where the fraction does not depend on the woman's type).

This assumption allows us to characterize each man (woman)'s behavior on the dating site as a series of binary decisions, where man m sends an e-mail to woman w if the utility of matching with this woman, $q_M(m,w)$, exceeds the utility of staying single, $v_M(m)$. We will utilize this binary decision rule as the basis of our empirical analysis in Section 7.

Note that this assumption relies on the time cost of e-mailing an otherwise acceptable man or woman to be negligible. If e-mailing is costly, however, agents may refrain from contacting acceptable partners who are not attainable, i.e. acceptable partners not within their opportunity set. In Section 4, we investigate whether such a "strategic" response by agents is an important empirical concern. Before we do that, however, we would like to argue why such a response might not be important within the online dating context. Unlike in conventional "marriage markets" where the costs of asking people out[13] may indeed be nontrivial, online dating is designed to provide an environment that minimizes this cost. Aside from any psychological cost of rejection, the main cost associated with sending an e-mail is the cost of composing it – however, the marginal cost of producing yet another witty e-mail should not be exaggerated since one can always personalize a polished form

---

[13]These costs may include the embarrassment of being rejected. In more traditional societies, the cost of "propositioning" the wrong woman (or man) may be the loss of life.

letter. Moreover, note that Adachi's model is one without uncertainty regarding the other side's preferences (i.e. the potential partner's type is perfectly observed) – in reality, one may expect the encountered person's preference ordering to have an unobservable random component. If the expected benefit from *any match* within one's acceptance set exceeds the marginal cost of sending e-mail, we should not expect the dating site users to strategically refrain from contacting people they find acceptable.

We now argue that additional assumptions imposed on the Adachi (2003) framework lead to an even simpler empirical strategy that we employ in section 5. These assumptions are:

1. All men (women) agree on women's (men's) rankings: i.e. if $q_M(m, w) \geq q_M(m, w')$, then $q_M(m', w) \geq q_M(m', w')$ for all m'. In particular, we can relabel a woman's type w as her "rank" among all women.

2. The meeting probability, δ, does not depend on types m and w. (It can be random across m and w, but its distribution can not depend on m or w.)

Under these assumptions, *higher ranked women (men) receive e-mails at a higher rate.* This allows us to interpret the observed rate of contacts received by an individual as his or her "rank." Since this "rank" can be interpreted as a utility index aggregating various attributes of a given woman that are observable to contacting men through her profile, we can regress this "rank" on attributes such as income, education, looks, and other socioeconomic and physical attributes to see how these factors make up the utility index.

4   Empirical Patterns in Browsing and E-Mailing Behavior

In this section we investigate how a user's propensity to send an e-mail is related to the attractiveness of the profile of a potential mate, and whether this propensity is different across attractive vs. unattractive searchers. As we mentioned in section 3, if the time cost of composing an e-mail or the psychological cost of rejection is significant compared to the expected benefit of eliciting a reply, one might expect searchers to avoid sending e-mails to potential mates that they deem unattainable. For instance, unattractive men may shy away from sending e-mails to very attractive women, and instead focus on e-mailing women near their own attractiveness level.

To investigate whether such a pattern is present in the data, we construct a choice set for each user consisting of all profiles that this user browsed, and of which he/she chooses

specific profiles to send an e-mail to.[14] We then construct a binary variable to indicate the choice of sending an e-mail (implicitly, this approach views this choice problem as sequential, rather than as a simultaneous problem over all browsed profiles). Our basic regression specification is a linear probability model of the form

$$\text{EMAIL}_{ij} = \beta \cdot \text{ATTRACTIVENESS}_j + u_i + \varepsilon_{ij}, \tag{1}$$

where EMAIL_{ij} equals 1 if browser i sends an e-mail to profile j. The term u_i indicates person-specific fixed effects (conditional logit estimates yielded very similar results). Within the context of a sequential search model, u_i can be interpreted as the (unobserved) optimal search threshold for sending an e-mail to profile j.

We first use our measure of physical attractiveness as a proxy for the overall attractiveness of a profile. We run regression (1) separately for users in different classes of physical attractiveness. I.e., we segment the users i who make the decision whether to send an e-mail or not according to their physical attractiveness, and allow for the possibility that the users in the different groups respond differently to the attractiveness of the profiles that they browse. Figure 4.1 shows the relationship between a browsed profile's photo rating and the estimated probability that the browser will send a first-contact e-mail. We see that regardless of the physical attractiveness of the browser, the probability of sending a first-contact e-mail in response to a profile is monotonically increasing in the attractiveness of the photo in that profile. Thus, even if unattractive men (or women) take the cost of rejection and composing an e-mail into account, this perceived cost is not large enough such that the net expected benefit of hearing back from an attractive mate would be less than the net expected benefit of hearing back from a less attractive mate.

Figure 4.2 provides some insight on the responder side of the market. This figure shows the relationship between the physical attractiveness of the person sending a first-contact e-mail and the probability that the receiver of that e-mail will respond to that e-mail. As expected, the relationship is monotonic in the attractiveness of the e-mail sender (there are no strategic concerns regarding rejection here, since the responder knows that the contacting person is interested). Note that men appear much more receptive to first-contact e-mails than women. The median man (in terms of photo attractiveness) can expect to hear back

---

[14]The median man browses 26 profiles (mean 75), and median woman browses 22 (mean 54). Of these browsed profiles, the median man sends an (unsolicited) email to only 1 (mean 7.2). Likewise, the median woman sends an unsolicited e-mail to only 1 (mean 3.8). We also investigated whether browsing intensity is different for people with different levels of physical attractiveness. When we regressed number of profiles browsed on the browser's attractiveness, we found the explanatory power of physical attractiveness to be extremely low (with R-squared never exceeding 0.01), and failed to find robust patterns between a person's level of attractiveness and the number of profiles this person browsed. Thus we concluded that the intensity of browsing is independent of the person's attractiveness.

16

from the median woman with a 40% chance, whereas the median woman can expect to get a reply with a 70% chance. We also see that men in the bottom 40 percentile of the attractiveness distribution can expect to hear back from the top 10th percentile of women with 14% probability. We believe that this is a pretty good return for spending 10-15 minutes on writing an introductory e-mail (or less than one minute on copying and pasting a previously prepared one), even for a busy economist!

Figures 4.1 and 4.2 also provide evidence that more attractive men and women are "pickier". The least attractive women are 2-4 times more likely to send a first-contact e-mail to a man than the most attractive women. The same difference in selectiveness is also evident in the reply probabilities. This finding is consistent with any search model in which more attractive types have higher outside options.

We also measure the "attractiveness" of a given profile by the number of first-contact e-mails it received during the observation period. Figure 4.3 shows the probability of sending a first-contact e-mail to a given profile by searchers of different attractiveness levels. Once again, these probabilities are monotonic in the attractiveness of the browsed profile.

These results provide direct support for the identification assumption discussed in section 3. If every men is more likely to respond to a more attractive woman than a less attractive woman, the number of first-contact e-mails received by a more attractive woman (in a given time period) will be larger than the number of first-contact e-mails received by a less attractive woman.

5 The Relationship Between User Attributes and Online Outcomes

In this section we explore how mate attributes, such as the stated goal for being on the dating site, looks, income, and education, are related to dating outcomes. We measure outcomes primarily by the number of e-mails a user receives as a first contact.[15] The number of first contact e-mails received indicates online success, at least if there is no systematic relationship between the number of first contacts and the average "type" of the users from who these e-mails originate. Furthermore, we have discussed in section 3 that under certain assumptions, the number of times a user is approached by others reveals his or her utility index, and thus the relationship between mate preferences and user attributes. The assumptions that have to hold are that *any* user, regardless of his or her type, is more likely to approach a high type mate than a low type mate, and that preferences over mates

---

[15]A "first contact" is a situation where user A e-mails user B, and no e-mail exchange has taken place between the two users before.

are homogeneous. Based on the results in section 4, we believe it is reasonable to rule out strategic behavior of a form that would violate the former assumption – this, in fact, is one of the major strengths of our data set.

We first present results that are based on the assumption that men and women have homogeneous preferences over their potential partners. We then relax this assumption, segment users into a priori chosen segments, and assess the importance of preference heterogeneity by examining how outcomes vary across these different segments of the dating population. We also discuss how the user attributes are related to two alternative outcome measures. One of these measures is the number of times a user's profile has been browsed. The other measure, obtained from a keyword search, is the number of e-mails containing a phone number or e-mail address received from other users. For both outcome variables, we only count the first browse and e-mail from each user. Compared to the first contact measure, the number of browses corresponds to an earlier stage and the number of keywords received corresponds to a later stage in the online relationship.

We estimate the relationship between first contacts and user attributes using a Poisson regression model. A count data model, such as a Poisson regression, is particularly appropriate for the integer outcomes in our application.[16] The conditional expectation of the outcome variable is specified as $\mathbb{E}(Y|x) = \exp(x'\beta)$, where x is a vector of user attributes. Under the Poisson assumption, this conditional expectation fully determines the distribution of the outcome variable. The Poisson assumption places strong restrictions on the data. In particular, the conditional variance of a Poisson distributed outcome variable equals the conditional expectation, $\mathrm{Var}(Y|x) = \mathbb{E}(Y|x)$. However, as long as the conditional expectation is correctly specified, the (quasi) maximum likelihood estimator associated with the Poisson regression model is consistent, even if the Poisson assumption is incorrect (Wooldridge 2001, pp. 648-649). We report robust (under distributional mis-specification) standard error estimates for the regressions (Wooldridge 2001, p. 651).

In our application, all regressors are categorical variables indicating the presence of a specific user attribute. If two users A and B differ only by one attribute that is unique to A, with the associated regression coefficient β_j, the ratio of expected outcomes is

$$\frac{\mathbb{E}(Y|x_A)}{\mathbb{E}(Y|x_B)} = \exp(\beta_j).$$

---

[16]Alternatively, a linear regression model has the obvious disadvantage of predicting negative outcome values for some user attributes. A logarithmic transformation of the outcome variable avoids this problem, but would force us to drop many observations for which the outcome measure is zero. Furthermore, it is not clear how the estimated conditional expectation $\mathbb{E}(\log(Y)|x)$ is related to the object of our interest, $\mathbb{E}(Y|x)$. The same problem pertains to the transformation $\log(1 + Y)$, which is defined for outcome values of zero.

The *incidence rate ratio,* $\exp(\beta_j)$, measures the premium (or penalty) from a specific attribute in terms of an outcome multiple. For example, using the number of e-mails received as outcome variable, the coefficient associated with "some college" education is 0.21 for men. Hence, holding all other attributes constant, men with some college education receive, on average, $\exp(0.27) = 1.31$ as many e-mails as the baseline group, men who have not finished high school yet. Alternatively, we can calculate the "college premium" for men as $100 \times (\exp(0.27) - 1) = 31\%$.

The results from the Poisson regressions are presented below for different types of user attributes, such as looks, income, and education. Separate regressions were estimated for men and women. As the outcome numbers are only meaningful if measured with respect to a unit period of time, we include the (log) number of days a user was active on the dating site as a covariate. Also, we include a dummy variable for users who were members of the website already before the start of the sampling period.

Table 5.1 presents summary statistics of the outcome measures. Women are browsed more often, and receive more first contact e-mails and e-mails containing a phone number or e-mail address than men. Hence, a first contact is more likely to be initiated by a man. While men receive an average of 2.6 first contact e-mails, women receive 12.6 e-mails. 54.5% of all men in the sample did not receive a first contact e-mail at all, whereas only 19.9% of all women were not approached by e-mail.

5.1 Outcome Regression Results: Homogeneous Preferences Over Mates

The detailed regression results for the Poisson models are reported in table A at the end of the paper. All 304 observed user attributes were used in the empirical analysis.

Goodness of fit A preliminary analysis shows what fraction of the variability in the outcome variable is explained by different user attributes. To that end, we present R^2 measures obtained from several OLS regressions using the transformed outcome measure $\log(1 + Y)$ as the dependent variable.[17] A similar, straightforward goodness of fit measure is not available for the Poisson regressions employed in the remainder of this section. We focus on the attributes looks, income, and education.

The results from several regressions are displayed in table 5.2. Focusing on first contact e-mails outcome variable, the full set of user attributes explains 29% of the outcome variability for men, and 44% of the outcome variability for women. "Looks" has the strongest explanatory power (31% for women and 19% for men), while income and education, if used as the only regressors, explain only a much smaller fraction of the outcome variance.

---

[17]The outcome Y is adjusted for the number of days a user was active during the sample period.

Effect of goals on outcomes   The members of the dating service can state in their profile why they joined the dating site. The majority of all users are "Hoping to start a long term relationship" (37% of men and 41% of women), or are "Just looking/curious" (26% of both men and women). An explicitly stated goal of finding a partner for casual sex ("Seeking an occasional lover/casual relationship") is more common among men (14%) than among women (4%).

The impact of these stated goals on online success differs across men and women (figure 5.1). Men who indicate a preference for a less than serious relationship or casual sex are contacted less often than men who state that they are "Hoping to start a long term relationship". Women, on the other hand, are not negatively affected by such indications. To the contrary, women who are "Seeking an occasional lover/casual relationship" receive 17% more first contact e-mails relative to the baseline, while men experience a 42% penalty. Men who are "Just looking/curious" receive 19% fewer first contact e-mails, and the statement "I'd like to make new friends. Nothing serious" is associated with a 24% outcome penalty. Either indication is mostly unrelated to women's' outcomes.

Looks and physical attributes   The users of the dating service describe many of their physical attributes, such as height and weight, in their profile. Also, about one third of all users post a photo online. We rated the looks of those members who posted one or more photos online in a laboratory environment, as previously described in section 2. We then classified the ratings into deciles, where the top decile was split again in two halves. This classification was performed separately for men and women. The looks of those member who did not post a photo online are measured using their self-description, such as "average looks", or "very good looks".

The relationship between the looks rating of the member who posted a profile and the number of first contact e-mails received is shown in table 5.2. Outcomes are strongly increasing in measured looks. In fact, the looks ratings variable has the largest impact on outcomes among all variables used in the Poisson regression analysis. Men and women in the lowest decile receive only about half as many e-mails as members whose rating is in the fourth decile, while the users in the top decile are contacted about twice as often. Overall, the relationship between outcomes and looks is similar for men and women. However, there is a surprising "superstar" effect for men. Those men in the top five percent of ratings receive almost twice as many first contacts as the next five percent; for women, on the other hand, the difference in outcomes is much smaller.

Having a photo online per se improves the members' outcomes. Women receive more than twice as many e-mails, and men receive about 50% more e-mails than those users who

did not post a photo and describe themselves as having "Average looks". Figure 5.3 also shows that outcomes are positively related to the user's self assessment, although the effect sizes are small compared to the impact of looks on outcomes for those users who include a photo in their profile.

Further evidence on the importance of physical attributes is provided by the members' description of their physique. Members who are "chiseled" and "toned" receive slightly more first contact e-mails than "height-weight proportionate" users, while "voluptuous/portly" and "large but shapely" members experience a sizable penalty.

Height matters for both men and women, but mostly in opposite directions. Women like tall men (figure 5.4). Men in the 6'3"-6'4" range, for example, receive about 60% more first contact e-mails than men in the 5'7"-5'8" range. In contrast, the ideal height for women seems to be in the 5'3"-5'8" range, while taller women experience increasingly worse outcomes. For example, the average 6'3" tall woman receives 40% fewer e-mails than a woman who is 5'5".

We examine the impact of a user's weight on his or her outcomes by means of the body mass index (BMI), which is a height adjusted measure of weight.[18] Figure 5.5 shows that for both men and women there is an "ideal" BMI at which success peaks, but the level of the ideal BMI differs strongly across genders. The optimal BMI for men is about 27. According to the American Heart Association, a man with such a BMI is slightly overweight. For women, on the other hand, the optimal BMI is about 17, which is considered underweight and corresponds to the figure of a supermodel. A woman with such a BMI receives about 77% more first contact e-mails than a woman with a BMI of 25.

Finally, as regards hair color (using brown hair as the baseline), we find that men with red hair suffer a moderate outcome penalty. Blonde women have a slight improvement in their online "success", while women with gray or "salt and pepper" hair suffer a sizable penalty. Men with curly hair receive about 22% fewer first contact e-mails than men in the baseline category, "medium straight hair". For women, "long straight hair" leads to a slight improvement in outcomes, while short hair styles are associated with a moderate decrease in outcomes.

Income   About 64% of men and 51% of women report their income. Figure 5.6 shows how these self-reported income measures are related to the members' dating outcomes. Income strongly affects the success of men, as measured by the number of first contact e-mails received. While there is no apparent effect below an annual income of $50,000, outcomes improve monotonically for income levels above $50,000. Relative to incomes below $50,000,

---

[18]The BMI is defined as $BMI = 703 \times w/h^2$, where w is weight in pounds and h is height in inches.

the increase in the expected number of first contacts is at least 32%, and as large as 156% for incomes in excess of $250,000. In contrast to the strong income effect for men, the online success of women is at most marginally related to their income. Women in the $35,000-$100,000 income range fare slightly better than women with lower incomes. Higher incomes, however, do not appear to improve outcomes, and are not associated with a statistically different effect relative to the $15,000-$25,000 income range.

Educational attainment   The relationship between online dating outcomes and education is less pronounced than the effect of income. However, we find some evidence that—similar to the income effect—higher levels of education increase the online success of men but not of women (figure 5.7). With respect to the number of first contact e-mails, there is a college and graduate education premium for men. Relative to high school graduates, a college degree is associated with a 35% increase in the number of first contacts. Graduate degrees are associated with a similar premium, but do not improve outcomes further relative to a college degrees. In contrast to these findings for men, the outcomes of women do not improve with their educational attainment. To the contrary, college juniors and seniors, women in a post-graduate program, and women with a master's degree incur a slight outcome penalty.

Occupation   Online success also varies across different occupational groups. Here, all outcomes are measured relative to the "success" of students, who are chosen as the baseline group. Holding everything else constant, the biggest improvement in outcomes is observed for men in legal professions (77% outcome premium), followed by the military (49%), fire fighters (45%), and health related professions (42%). Manufacturing jobs, on the other hand, are associated with an about 10% penalty. The occupation of women, on the other hand, has little influence on their outcomes; in fact, most professions are associated with a slightly lower number of first contacts relative to students.

5.2   Outcome Regression Results: Heterogeneous Preferences

We now relax the assumption that men and women rank their potential mates according to a single dimensional index, and provide some preliminary evidence on the extent of preference heterogeneity over mate attributes. We allow for preference heterogeneity by creating different user segments, which are chosen a priori according to observable user characteristics. We then create new outcome measures by separately counting the number of first contact e-mails from each segment. I.e., for each user, we record the number of first contacts from segment one, two, etc. Under the assumptions laid out in section **??**, the

number of first contact e-mails received from users in segment j is monotonically related to the preference ordering that the users in segment j have over their potential mates. In order to relate these preferences to user attributes we again estimate Poisson regression models, but now separately for each user segment.

We focus on preference heterogeneity that is related to looks, income, and education. First, we segment the site users into two halves according to their looks rating. Second, we create a low and high income segment. Site members in the low income group have annual incomes up to $50,000, while members in the high income group have incomes above $50,000. Third, we create three education segments of users who have or are working towards a high school, college, or graduate degree. For each of the three segmentation schemes, we focus on how preferences over looks, income, and education differ across segments.

We find the strongest evidence for preference heterogeneity with respect to education. In particular, women seem to have a strong preference for men with equivalent education levels (figure 5.8). For example, men with a master's degree receive 48% fewer first contact e-mails from high school educated women than high school educated men. From college educated women, on the other hand, they receive 23% more e-mails, and from women with or working towards a graduate degree they receive 84% more e-mails. We also find evidence that high school educated men have a preference for a woman with a similar education level, or alternatively, avoid women with college or graduate degrees. However, men with college or graduate degrees do not seem to base their choices on a woman's education.

As regards income, we find some evidence that high income women have a stronger preference for men's' income than low income women. Also, women with a looks rating in the lower 50% place slightly more emphasis on men's income than women in the top 50%.

Finally, we find no evidence that preferences over a potential mate's looks differ according to own looks, income, or education.

5.3 Does Race/Ethnicity Affect Outcomes?

The users of the online dating service can declare in their profile whether the ethnicity of a potential partner matters to them. We find a striking gender difference in this stated preference for ethnicity: 38% of all women, but only 18% of men say that they prefer to meet someone of the same ethnic background as themselves. This stated ethnicity preference also varies across users of different ethnic backgrounds (figure 5.9). For example, among Caucasians, 48% of all women and 22% of men declare a preference for Caucasian mates. On the other hand, only 25% of black women and 8% of black man declare that they want to meet only other blacks.[19]

---

[19]This, of course, could reflect self selection to a dating service with a majority of caucasian users.

The question is whether ethnicity preferences also influence the interaction between users, and whether the stated ethnicity preferences are reflected in these users' online behavior. We create four groups of users, based on whether they declare their ethnicity as Caucasian, Black, Hispanic, or Asian. We then construct first contact e-mail outcome measures for all users, separately from each segment, as we did before in the analysis of preference heterogeneity.

The regression results provide evidence that members of all four ethnic groups "discriminate" against users belonging to other ethnic groups (figure 5.10). For example, African American and Hispanic men receive only about half as many first contact e-mails from White women than White men, and Asian men are contacted only about one fourth as often. Note that these results fully control for all other observable user attributes, such as income and education. Also, note that these results are not due to a market size effect, as the outcomes reflect the relative success of the different ethnic groups with respect to the same population of potential mates. Overall, it appears that women discriminate more strongly against members of the different ethnicities than men. Also, Asian men and women seem to be least discriminating among the ethnicities, although the effects are not measured precisely.

Figure 5.11 shows the estimated ethnicity preferences separately for users who declare that they want to meet only users of their own race and users who do not have a declared preference. Due to sample size issues, we consider only first contact e-mails from Caucasians. It is evident that both members who declare a preference for their own ethnicity, and those who do not, discriminate against users who belong to different ethnic groups. However, the discrimination size is more pronounced for members of the former group, i.e. these users act consistent with their stated preferences. There is strong evidence, however, that also members of the latter group have ethnic preferences, which is in contradiction to their statement that ethnicity "doesn't matter" to them.

6   A Closer Look at Preference Heterogeneity

We now take an alternative, discrete choice based approach to estimating mate preferences. To an empirical microeconomist, such an approach may appear more straightforward than the outcome regressions of section 5, and in particular it allows us to control for preference heterogeneity in a more flexible way compared to the a priori segmentation approach pursued in section 5.2. However, this approach also comes at a cost, because the estimation approach is now computationally much more costly and hence forces us to limit the number of model parameters. Therefore, we need to make stronger functional form assumptions on the form

of preferences compared to the essentially non-parametric way by which variables such as income entered the utility index in section (5).

The estimation approach is based on a sequence of binary decisions as described in section 3, which we model using a random effects probit model. For each user, we observe the potential mates that he or she browses. The users needs to view the mate's profile in order to contact him or her. We then model the choice of sending or not sending an e-mail to the potential mate as a binary choice. The latent utility of man m from e-mailing woman w, which is related to the eventual utility from forming some sort of match, is specified as

$$U_{mw} = x'_w \alpha + |x_w - x_m|'_+ \beta + |x_w - x_m|'_- \gamma$$
$$+ \sum_{k,l=1}^{N} \{d_{mk} = v_k \text{ and } d_{wl} = v_l\} \cdot \delta_{kl} + c_m + \epsilon_{mw}. \tag{2}$$

The first component of utility is a simple linear valuation of the woman's attributes. The second and third components relate the preferences over woman w to man m's own characteristics. $|x_w - x_m|_+$ is the difference between the woman's and man's attributes if this difference is positive, and $|x_w - x_m|_-$ denotes the absolute value of this difference is the difference is negative.[20] For example, consider the difference in age between m and woman w. If the coefficient corresponding to the age difference in β and γ is negative, it means that users prefer someone of their own age. The fourth component in (2) consists of dummy variables that relate to categorical attributes of both m and w. For example, v_k could indicate that the ethnicity of m is "White," and v_l could indicate that the ethnicity of w is "Hispanic." Then δ_{kl} measures the relative preference of Whites for Hispanic mates. c_m is an unobserved user-specific component of preferences that measures the reservation utility of m relative to which the decision to contact a woman is made. Finally, $\epsilon_{mw} \sim N(0,1)$ represents the utility due to all unobserved factors. We assume that the random effects c_m are independent of all observed covariates and distributed $N(0, \sigma_c^2)$.

The assumption that the c_m's are independent of the covariates in (2) might be too strong; alternatively, following Chamberlain (1980), we could specify c_m to be conditionally normal with mean $\mu + x'_m \eta$. Note, however, that in this case our covariates are perfectly collinear, as $|x_w - x_m|_- - |x_w - x_m|_+ + x_w = x_w$. In other words, the effect of own characteristics on the reservation utility is not separately identified from the effect of own characteristics on the valuation of mate attributes, and care has to be taken in the interpretation of the parameter estimates in the latent utility specification. To assess whether the random effects are related to the observable user attributes, we conducted a preliminary analysis

---

[20]Formally, $|a - b|_+ = \max(a - b, 0)$ and $|a - b|_- = \max(b - a, 0)$.

of the relationship between user attributes and the rate at which users contact a potential mate. Specifically, for each user we calculated f_m, the fraction of browsed mates that m contacted, and regressed f_m on various user attributes in categorical form, as in section 5.[21] This regression had only low explanatory power with an R^2 of 0.047, most variables were not statistically significant, and for most variables we found no evidence for a systematic relationship with f_m. The only exception is the looks rating variable; better looking site members contact fewer of the potential partners that they browse (this finding is consistent with the analysis of e-mailing and browsing behavior in section 4). We take the findings of the preliminary analysis as evidence that most observable user attributes are not related to the propensity to contact other users, and that the random effects c_m are in fact independent of these covariates. As regards the users' looks, we allow for c_m to depend on the looks rating of user m, but do not include the difference between mate and own looks in the utility specification. We regard this restriction, which says that the utility from a partner's looks does not depend on own looks, as a priori plausible.

Table 6 presents the maximum likelihood estimates of the binary probit model. Because our final interest is in preferences over potential marriage partners, we only used observations on users who state that they are looking for a long term relationship in the estimation.[22] To be precise, we included only choices of those users who look for a long term relationship, but considered their full range of choices among users who joined the site for any reason. Overall, the results confirm the importance of the variables highlighted in section 5, but qualify some of the main findings.

First, as expected we find that the users of the dating service prefer a partner whose age is similar to their own. Men, in particular, try to avoid women who are older than them. Women, on the other hand, have a particularly strong distaste for younger men.

Women who are single tend to avoid divorced men, while divorced women have a relative preference for divorced men. The corresponding utility weights of men are of the same sign, but much smaller in size. Both men and women who have children prefer a partner who also has children. Members with children, however, are much less desirable to both men and women who themselves do not have children. Also, women, but not men, particularly prefer a partner who also indicates that he is seeking for a long term relationship.

Similar to the results from the outcome regressions, we find that looks matter to both

---

[21]We used categorical variables on the users' age, looks rating, weight, height, income, education, and ethnicity as covariates. A regression with $\log(f_m)$ as dependent variable yielded similar results.

[22]We also estimated the model with the choices of users who are "just looking/curious" included. The results were similar. For the full sample, where we also included the users who are seeking a short term relationship, many parameter estimates were smaller in absolute value. Hence, the choice behavior of these less "serious" users appears less focused than the behavior of the site members who try to find a long term partner.

men and women. The utility weight on the looks rating variable differs only little across men (0.277) and women (0.265). Also as in the case of the outcome regressions, men and women have a stronger preference for mates who describe their looks as "above average" than for average looking members, and they have an even stronger preference for members with self described "very good looks." Regarding height and weight, we find evidence of considerable preference heterogeneity. Men generally prefer shorter women, and they particularly try to avoid women who are taller than themselves. Women, on the other hand, prefer men who are taller than themselves, and they have a particularly strong aversion to shorter men. For example, our estimates imply that compared to a man who is five inches taller than a woman and earns $ 50,000 per year, a man who is five inches shorter than a woman would need to earn slightly more than half a million dollars per year to make up for his shortcoming. As regards weight, we find as before that generally men have a strong distaste for women with a larger BMI, while women tend to prefer somewhat heavier men. Both men and women also appear to have a preference for a partner that is closer to their own BMI, although the quantitative significance of this heterogeneity component is small compared to the preference over the BMI level.

We confirm that the partner's income matters to both men and women. Woman, however, place almost twice as much weight on income than men. Contrary to what we expected, men have a statistically significant distaste for women who are poorer than them, while women have a statistically significant distaste for men who are richer than them. The absolute value of these coefficients is small, however, and hence own income appears to matter only little in the evaluation of a partner's earnings.

The utility weight on the level of education is very small and statistically insignificant for both men and women. Quantitatively more important, however, is the heterogeneity component. Men in particular, have a distaste for a partner who is more educated than them. Women, on the other hand, try to avoid men who are less educated than them.

Finally, we find that both men and women have a preference for a partner of the same religion.

7 Who Matches With Whom? – Online Matching And Actual Marriages

Online dating lowers the search costs of finding a partner in a market characterized by large search frictions. In this section, we provide some evidence on the effect of this new technology on the resulting structure of matches. Traditionally, people find their marriage partners in the social and geographic environment they live in, such as the school, college, or church

27

they attend, at work, through friends or relatives, or in public places such as bars. Most people are therefore more frequently exposed to potential partners who are more similar to them in terms of their education, income, faith, or ethnicity than a randomly drawn partner from the general U.S. population. Therefore, the empirically observed correlations in marriages along certain attributes, such as income and education, may be purely due to the social institutions that bring partners together and only partially due to the preferences men and women have over their mates.[23] Compared to traditional marriage markets, online dating is characterized by only small search frictions, and the resulting matches are therefore largely driven by preferences and the equilibrium mechanism that brings partners together. In this section, we provide a comparison of observed and predicted online matches, and compare the associated correlation in traits, such as age, income, and education, with the correlation of traits observed in actual marriages. Such a comparison sheds some light on the role of institutional factors in the observed sorting of men and women.

We first provide some evidence on the observed "matches" from our dating service. A clarification of what we mean by a "match" is in order. A main limitation of our data is that we can only track the users' online behavior. We therefore do not know whether two partners who met online ever went on a date or eventually got married. However, our data provides some information on the contents of the exchanged e-mails. We observe whether users exchange a phone number or e-mail address, or whether an e-mail contains certain keywords or phrases such as "get together" or "let's meet." We therefore have some indirect information on whether the online meeting resulted in an initial match, i.e. a date between the users. We define such a match as a situation where both mates exchange such contact information (i.e., for a match it is not enough for a man to offer his phone number, we also require that the woman responds by sending her contact information).

Table 7 (II) shows the correlation of several user attributes in the observed online matches. Not surprisingly, age is strongly correlated across men and women ($\rho = 0.720$). Also looks are strongly correlated ($\rho = 0.329$), and is a somewhat smaller but still sizable positive correlation in height, BMI, income, and years of education.

We next examine whether these observed matches are as predicted from the preference estimates in section 6 and a specific assumption on the equilibrium mechanism by which matches are formed. We obtain an equilibrium of our dating market from the Gale-Shapley (1962) algorithm, which, as noted by Adachi (2003) and in Section 3, can be intepreted as the limiting outcome of a decentralized search and matching environment as the one considered here. This algorithm produces a stable matching, in the sense that no pair of

---

[23]Some of these institutions, such as "upscale" bars, may well have arisen endogenously to facilitate sorting along certain traits. Nonetheless, it is instructive to compare matching in environments with different degrees of search frictions.

men and women could leave their current partner and improve on the current outcome by forming a new match. The Gale-Shapley algorithm works as follows. Men make offers (proposals) to the women, and the women accept or decline these offers. The algorithm proceeds over several rounds. In the first round, each man makes an offer to his most preferred woman. The women then collect offers from the men, rank the men who made proposals to them, and keep the highest ranked men engaged. The offers from the other men are rejected. In the second round, those men who are not currently engaged make offers to the women who are next highest on their list. Again, women consider all men who made them proposals, including the currently engaged man, and keep the highest ranked man among these. In each subsequent round, those men who are not engaged make an offer to the highest ranked woman who they have not previously made an offer to, and women engage the highest ranked man among all currently available partners. The algorithm ends after a finite number of rounds. At this stage, men and women either have a partner or remain single. Clearly, the algorithm also produces a stable matching if the women make offers; generally, however, the equilibrium outcome of the algorithm will depend on which side makes offers.

Of course, the actual behavior in the online dating market that we study is not exactly as described by the Gale-Shapley algorithm. However, the algorithm captures some basic mechanisms that we consider to be central to the functioning and outcomes of any marriage market: First, the mates on each side of the market strive to attain a partner who, according to their own tastes, is as desirable as possible. Hence, people will not form a match if they can obtain a more preferred partner. Second, the sorting that arises in equilibrium incorporates the limitations imposed on the mates due to their own desirability, i.e. people cannot expect to be matched with a partner who can attain a more desirable mate.

Table 7 (III a) and (III b) shows the correlation in user attributes of the predicted matches, separately for the situation where either men or women make offers to their potential partners. The predicted correlations in user attributes are virtually identical for these two cases. The age correlation in observed and predicted matches is roughly similar. The predicted correlation in physical attributes, looks, height, and BMI, is somewhat smaller than the correlation in observed matches. Finally, while the predicted correlation in incomes is smaller than observed correlation, the Gale-Shapley algorithm predicts a higher correlation ($\rho = 0.196$) in years of education than observed ($\rho = 0.131$). Overall, it appears that the structure of online matches can be reasonably well explained from our preferences estimates and a simple yet straightforward mechanism that predicts how matches are formed.

Table 7 (I) displays the correlation of traits in actual marriages in Boston and San

Diego. The underlying data are from the 2000 IPUMS 5% sample; data from 1990 yield very similar results. We find that the correlation of age in marriages (0.944) is even stronger than the correlation in actual (0.720) and predicted (0.707) online matches. The correlation of income (0.127), on the other hand, is somewhat smaller than the correlation of observed matches (0.153) and somewhat larger than the correlation of predicted matches (0.136). We observe a particularly strong discrepancy between online and offline matches with respect to years of education. The correlation of education in actual marriages is 0.642, which is much larger than the correlation of actual (0.131) and predicted (0.196) online matches.

These findings suggest that the sorting in characteristics such as age and education in actual marriages is not only driven by preferences, but also by the search frictions that are inherent in the process of finding a partner. However, some caution needs to be exercised in the interpretation of the results. Due to the limitations of our data, we cannot be sure that the estimated preferences are over marriage partners. We observe only an early stage in the development of a relationship, and it is possible that preferences at a later stage of the relationship look differently, and possibly put more weight on age and education. Of course, the same argument would also apply to income, although we find that the correlation in actual marriages is quite consistent with the correlation in online matches. Even if we accept the argument that the revealed preferences differ from preferences over marriage partners, it is striking that the discrepancy between actual and online correlations differs so strongly across attributes. The ratio of the correlation in age between marriages and predicted online matches is 1.36, while the ratio for years of education is 3.28. We take the sheer size of the extent to which we underpredict the correlation in education as strong, although not ultimately conclusive evidence that the sorting in education in the actual marriages is partly due to social institutions such as schools and universities.

We previously presented evidence that income, looks, and weight are positively correlated in our sample of dating service members. Therefore, to some extent the match correlation in income and education could be driven by preferences over physical attributes, and not by preferences over income and education per se. To isolate the role of income and education preferences, we matched the users again under the assumption that each user had the same looks, height, and weight. The results from this matching are presented in table 7 (IV). The predicted age correlation is now somewhat larger than under the original matching, but–as conjectured–the correlation in income ($\rho = 0.097$) and education ($\rho = 0.167$) is somewhat lower, yet still sizable.

Many readers will find some of our results sobering. Our fate in love and marriage seems to be driven by factors such as looks, height, weight, and income, that are hard or impossible to change. We would like to stress, however, that these observable user attributes account

only for part of the preferences for a potential partner. Other attributes, such as personality traits, are unobserved or match-specific. In order to examine the role of these unobserved factors, we match the users again under the assumption that their preferences are only over observable attributes (i.e., we took the ϵ out of the latent utility specification (2)). Table 7 (V) shows the match correlations under this assumption. We observe that now the correlations in age, and in particular in income, education, and in all the physical attributes is much higher than under the original matching, where some attributes are unobserved or match-specific. For example, the correlation income is now 0.460 (previously 0.136), and the correlation in looks is 0.517 (previously 0.213). Hence, factors such as personality traits apparently allow us to partly make up for deficiencies in good looks or wealth.

8   Conclusions

This paper investigates mate preferences and matching in a dating or marriage market. Our analysis is based on unusually detailed data on the attributes and interactions of men and women, which are available to us due to the well-defined institutional rules of an online dating market. Our analysis of *revealed preferences*, and the relationship of these preferences to user attributes, confirms many findings obtained in psychology, anthropology, and sociology studies, which are based on *stated preference* data. For example, we find a stronger emphasis on a partner's income among women than among men. Our revealed preference data, however, are more ideally suited to consider preference heterogeneity. Also, revealed preference data allow us to investigate mate preferences that people might not truthfully reveal, in particular their behavior towards potential mates of different ethnicities.

We use the Gale-Shapley algorithm to predict the equilibrium sorting along attributes such as age, income, and education, based on our preference estimates. These predictions are made under the assumption of no search frictions, which we believe characterizes online dating well compared to the traditional "real world" way of finding a partner. We find that we can predict the correlation in income among men and women, but vastly underpredict the correlation in education levels in actual marriages. This suggests that the strong sorting along education observed in marriages is at least partly driven by search frictions, and not fully due to preferences over the partner's education level.

A drawback of our analysis is that we cannot observe whether an online meeting finally results in a marriage, which is the outcome that we are interested in. Therefore, we cannot fully exclude the possibility that some of our preference estimates are driven by preferences over a dating partner as opposed to a marriage partner. For future research, we hope to conduct exit/follow up surveys on a dating site in order to assess this issue.

References

[1] Adachi, H. (2003): "A search model of two-sided matching under non-transferable utility," Journal of Economic Theory, vol. 113, 182-198.

[2] Becker, G. S. (1973): "A Theory of Marriage: Part I," Journal of Political Economy, vol. 81 (4), 813-846.

[3] Becker, G. S. (1974): "A Theory of Marriage: Part II," Journal of Political Economy, vol. 82 (2), Part 2: Marriage, Family Human Capital, and Fertility, S11-S26.

[4] Biddle J. and D. Hamermesh (1998): "Beauty, Productivity, and Discrimination: Lawyers' Looks and Luchre," Journal of Labor Economics, vol. 16(1), 172-201.

[5] Buss, D. (2003): The Evolution of Desire: Strategies of Human Mating, Basic Books.

[6] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," Review of Economic Studies, 47, 225-238.

[7] Choo, E. and A. Siow (2003): "Who Marries Whom and Why," manuscript, University of Toronto.

[8] Etcoff, N (1999): Survival Of The Prettiest, Doubleday.

[9] Fisman, R., S. Iyengar, and I. Simonson (2004): "Revealed Preference Determinants of Mate Selection: Evidence from an Experimental Dating Market," manuscript, Columbia University.

[10] Hamermesh, D., J. Biddle (1994): "Beauty and the Labor Market," American Economic Review, vol 84(5), 1174-1194.

[11] Gale, D. and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," American Mathematical Monthly, 69, 9-14.

[12] Shimer, R. and L. Smith (2000): "Assortative Matching and Search," Econometrica, 68 (2), 343-370.

[13] Smith, L. (2002): "The Marriage Model with Search Frictions," manuscript, University of Michigan.

[14] Wong, L. Y. (2003): "An Empirical Study of Darwin's Theory of Mate Choice," manuscript.

Table 2.1 – Dating Service Members and County Profile of General Demographic Characteristics

Variable	San Diego			Boston		
	Dating Service	General Population	Internet User	Dating Service	General Population	Internet User
General Information						
Total Member and Population	11,691	2,026,020	1,180,020	11,390	2,555,874	1,581,711
Percentage of Males	55.5	49.9	49.4	54.2	49.0	50.6
Age Composition						
18 to 20 years	19.6	6.0	6.4	18.4	5.8	7.2
21 to 25 years	30.4	9.5	11.5	32.9	9.3	12.0
26 to 35 years	27.9	21.3	18.8	28.0	17.2	19.7
36 to 45 years	10.0	23.0	28.6	10.3	23.1	26.8
46 to 55 years	6.8	18.5	19.0	6.2	17.6	20.1
56 to 60 years	4.3	6.3	6.5	3.5	7.3	6.9
61 to 65 years	0.8	2.9	3.6	0.5	4.3	3.7
66 to 75 years	0.1	6.9	4.8	0.1	8.8	2.9
Over 76	0.2	5.7	0.8	0.2	6.8	0.7
Race Composition (1)						
Whites	65.4	61.9	71.3	73.7	84.2	89.1
Blacks	4.2	4.8	4.2	4.6	7.4	4.2
Hispanics	10.7	19.5	9.8	4.0	4.4	2.3
Asian	5.1	13.0	13.6	3.9	3.8	4.2
Other	14.7	0.9	1.1	13.8	0.3	0.2
Marital Status						
Males						
Never married	65.1	31.8	28.5	66.7	35.3	36.8
Married & not separated	6.1	57	62.0	6.9	54.1	56.7
Separated	4.0	1.2	0.7	4.8	1.1	0.3
Widowed	1.8	2.3	1.5	1.4	3.6	1.0
Divorced	23.1	8.1	7.4	20.3	6.0	5.2
Females						
Never married	61.3	20.2	23.9	65.6	28.0	32.7
Married & not separated	2.6	57	62.5	1.9	49.0	55.9
Separated	3.7	3.9	1.9	4.2	2.4	0.9
Widowed	3.3	6.3	2.0	3.0	13.4	3.5
Divorced	29.1	12.3	9.7	25.2	7.2	7.0
Educational Attainment (2)						
Have not finished high school	1.3	12.1	3.0	1.5	9.2	3.2
High school graduate	9.1	23.0	17.8	10.1	30.1	20.4
Technical training (2-year degree)	31.8	5.2	5.4	23.3	7.3	7.6
Some college	6.7	27.9	28.5	4.6	14.1	15.0

Bachelor's degree	29.0	22.7	31.5	34.4	22.2	29.7
Master's degree	11.5	6.0	9.0	16.5	11.7	16.3
Doctoral degree	3.4	1.5	2.6	3.8	3.3	5.2
Professional degree	7.3	1.7	2.3	5.9	2.0	2.6
Income (3)						
Total Individuals with	6,831	283,442	224,339	6,650	396,065	281,619
Income information						
Less than $12,000	7.7	12.5	12.4	8.4	7.6	4.6
$12,000 to $15,000	4.9	3.0	1.9	3.8	5.0	6.0
$15,001 to $25,000	8.6	13.8	10.1	6.0	21.4	16.2
$25,001 to $35,000	13.9	23.3	22.3	12.2	19.9	21.4
$35,001 to $50,000	20.6	12.4	10.6	22.1	16.5	18.5
$50,001 to $75,000	20.3	17.3	20.2	23.2	21.7	24.6
$75,001 to $100,000	10.5	7.2	9.1	12.4	4.8	4.5
$100,001 to $150,000	6.8	7.5	9.5	7.1	1.9	2.7
$150,001 to $200,000	2.7	3.2	4.0	2.1	1.1	1.6
$200,001 or more	3.9	0.0	0.0	2.9	0.0	0.0

Source. Estimates from CPS Internet and Computer use Supplement, September 2001. All the CPS estimates are weighted. All the individuals for the CPS and members are constrained to be 18 years of age or older. The percentages for the column "Internet user" is calculated conditioning the CPS sample to those individuals who declared to use the Internet.

Notes. Geographical information is based on Metropolitan Statistical Area. Boston PMSA includes a New Hampshire portion. San Diego geographic information corresponds to San Diego MSA. Member information as 2003.

(1) The figures for Whites, Blacks and Asian and Other race for the CPS data correspond to those with non-Hispanic ethnicity.

(2) Education excludes certain categories on member data that can not be translated into years of educational attainment.

(3) The income figures from the CPS data were adjusted to 2003 dollars.

Table 2.2 – Physical Characteristics of Dating Service Members vs. General Population

Variable	Men		Women	
	Dating Service	General Population	Dating Service	General Population
Weight (lbs)				
20-29 years	175.5	172.1	136.2	141.7
30-39 years	184.5	182.5	136.7	154.2
40-49 years	187.6	187.3	137.9	157.4
50-59 years	187.2	189.2	140.2	163.7
60-69 years	188.6	182.8	146.1	155.9
70-79 years	185.6	173.6	147.1	148.2
Height (inches)				
20-29 years	70.6	69.3	65.1	64.1
30-39 years	70.7	69.5	65.1	64.3
40-49 years	70.8	69.4	65.1	64.1
50-59 years	70.6	69.2	64.7	63.7
60-69 years	70.3	68.5	64.5	63.1
70-79 years	68.8	67.7	63.6	62.2
BMI**				
20-29 years	24.7	25.2	22.6	24.3
30-39 years	25.9	26.5	22.6	26.3
40-49 years	26.3	27.3	22.9	27.0
50-59 years	26.4	27.8	23.6	28.4
60-69 years	26.8	27.3	24.6	27.6
70-79 years	27.8	26.7	25.7	26.9

*General population statistics obtained from the National Health and Nutrition Examination Survey, 1988-1994 Anthropometric Reference Data Tables.
** BMI (body mass index) is calculated as weight (in kilograms) divided by height (in meters) squared.

Table 2.3 – Log Earnings and Photo Ratings

	Men (1)	Men (2)	Women (1)	Women (2)
Standardized photo rating	0.0612	0.0573	0.0922	0.0882
	(0.0129)***	(0.0132)***	(0.0152)***	(0.0163)***
Years of education	0.0832	0.0825	0.0776	0.0771
	(0.0053)***	(0.0053)***	(0.0062)***	(0.0062)***
Weight in lbs		-0.0008		-0.0004
		(0.0006)		(0.0006)
Height in inches		0.0226		0.0105
		(0.0050)***		(0.0057)*
Observations	1,781	1,781	1,240	1,240
R-squared	0.50	0.51	0.45	0.45

Note: The dependent variable in each regression is the log of reported annual income. Each regression also includes indicator variables controling for occupation, city (Boston or San Diego), ethnicity, and marital status. We also constructed we called "years on the workforce" which is the age of the respondent minus the total years of education minus five. We also included the square of this variable. Standard errors are reported in parentheses.

Table 5.1 – Description of Outcome Measures

	Browses[a]	First Contacts[b]	Keywords[c]
Men			
All Observations			
No. Observations	12,654	12,654	12,654
Median	11	0	0
Mean	45.9	2.6	1.5
SD	84.6	6.2	5.9
Min	1	0	0
Max	1,059	88	263
% Obs. Equal 0	0.0	54.5	67.1
Observations > 0			
No. Observations	12,654	5,763	4,158
Median	11	3	2
Mean	45.9	5.8	4.4
SD	84.6	8.2	9.6
Women			
All Observations			
No. Observations	10,427	10,427	10,427
Median	38	4	1
Mean	129.7	12.6	3.5
SD	195.7	21.8	7.2
Min	1	0	0
Max	1,649	202	372
% Obs. Equal 0	0.0	19.9	42.8
Observations > 0			
No. Observations	10,427	8,349	5,965
Median	38	7	3
Mean	129.7	15.8	6.1
SD	195.7	23.3	8.7

[a] Number of times user was browsed by unique users

[b] Number of first contact e-mails received

[c] Number of e-mails containing a phone number or e-mail address recieved

Table 5.2 – Explanatory Power of Several User Attributes

Attributes	Men Browse	Men First Contacts	Men Keywords	Women Browse	Women First Contacts	Women Keywords
Looks	0.38	0.19	0.04	0.49	0.31	0.13
Income	0.13	0.07	0.02	0.07	0.04	0.03
Education	0.10	0.06	0.02	0.03	0.02	0.02
Looks, Income	0.41	0.22	0.05	0.51	0.32	0.14
Looks, Income, Education	0.42	0.22	0.05	0.51	0.33	0.15
All	0.49	0.29	0.10	0.59	0.44	0.23

Note: The table reports R-squared measures from different OLS regressions, using $\log(1+Y)$ as dependent variable. Y is defined as the number of first contact e-mails received per day active. The regression in the last row of the table ("All") includes all user attributes that were used in the Poisson regression of section 5.

Table 6 – Binary Probit Estimates

	Men		Women	
	Estimate	SE	Estimate	SE
Age	**-0.0029**	0.0013	**-0.0040**	0.0013
Age Difference (+)	**-0.0415**	0.0020	**-0.0142**	0.0017
Age Difference (-)	**-0.0228**	0.0013	**-0.0458**	0.0021
Single, Mate: Divorced [a]	**-0.0200**	0.0169	**-0.0464**	0.0186
Divorced, Mate: Divorced	**0.0206**	0.0184	**0.1076**	0.0175
"Long Term", Mate: "Long Term"	**0.0003**	0.0110	**0.1160**	0.0131
Has Children, Mate: Has Children	**0.1080**	0.0181	**0.0989**	0.0175
No Children, Mate: No Children	**-0.1723**	0.0166	**-0.1773**	0.0193
Has Photo	**-0.0168**	0.0267	**0.0955**	0.0272
Looks Rating	**0.2773**	0.0108	**0.2652**	0.0136
Own Looks Rating	**-0.0754**	0.0194	**-0.1365**	0.0170
"Very Good" Looks	**0.3325**	0.0312	**0.2802**	0.0343
"Above Average" Looks	**0.1952**	0.0283	**0.0982**	0.0298
"Other Looks"	**0.0337**	0.1759	**0.0939**	0.1216
Height	**-0.0274**	0.0041	**-0.0044**	0.0039
Height Difference (+)	**-0.1140**	0.0124	**0.0308**	0.0034
Height Difference (-)	**-0.0150**	0.0039	**-0.2154**	0.0190
BMI	**-0.2454**	0.0155	**0.1169**	0.0263
BMI^2	**0.0031**	0.0003	**-0.0021**	0.0005
BMI Difference (+)	**0.0050**	0.0082	**-0.0034**	0.0033
BMI Difference (-)	**-0.0431**	0.0035	**-0.0234**	0.0061
Education (Years)	**0.0072**	0.0042	**-0.0003**	0.0043
Education Difference (+)	**-0.0284**	0.0040	**-0.0099**	0.0040
Education Difference (-)	**-0.0043**	0.0047	**-0.0390**	0.0051
Income ($ 1,000)	**0.0033**	0.0009	**0.0072**	0.0019
Income (>50) [b]	**-0.0015**	0.0014	**-0.0016**	0.0022
Income (>100) [b]	**-0.0027**	0.0016	**-0.0054**	0.0010
Income (>200) [b]	**-0.0013**	0.0025	**0.0027**	0.0011
Income Difference (+)	**0.0004**	0.0003	**-0.0007**	0.0003
Income Difference (-)	**-0.0006**	0.0002	**0.0000**	0.0003
Income "Only Accountant Knows"	**0.2225**	0.0356	**0.5243**	0.0834
Income "What, Me Work?"	**0.1489**	0.0423	**0.3384**	0.0929
White, Mate: Black	**-0.4326**	0.0637	**-0.3619**	0.0747
White, Mate: Hispanic	**-0.1697**	0.0297	**-0.2828**	0.0572
White, Mate: Asian	**-0.2433**	0.0335	**-1.0118**	0.1753
White, Mate: Other	**-0.0588**	0.0179	**-0.0452**	0.0199

Black, Mate: White	**0.4897**	0.0530	**0.1913**	0.1390
Black, Mate: Hispanic	**0.2342**	0.1905	**0.4966**	0.5013
Black, Mate: Asian	**-0.3916**	0.3521	**0.2351**	0.8187
Black, Mate: Other	**0.6179**	0.1197	**0.1299**	0.2287
Hispanic, Mate: White			**0.0063**	0.0482
Hispanic, Mate: Black	**0.4786**	0.3022	**0.2059**	0.2736
Hispanic, Mate: Asian	**0.0196**	0.1852		
Hispanic, Mate: Other	**0.1386**	0.1180	**0.0249**	0.1089
Asian, Mate: White	**-0.3546**	0.1085	**0.1974**	0.0599
Asian, Mate: Black			**-0.1784**	0.4215
Asian, Mate: Hispanic	**0.0733**	0.2256	**-0.1621**	0.2639
Asian, Mate: Other	**-0.0809**	0.1553	**-0.0142**	0.1238
Same Religion	**0.0934**	0.0157	**0.1673**	0.0157
Constant	**4.1590**	0.3609	**-3.4133**	0.4402
SD of Random Effect Component	0.6296	0.0079	0.6002	0.0091
Log-likelihood	-40,019.85		-35,189.30	
No. of Observations	143,533		143,184	
No. of Individuals	3,148		2,729	

[a] I.e., the user who makes the choice is single, and the potential mate is divorced.

[b] Income ($> x$) is the amount of income (in $ 1,000) above the income level x.

Note: The dependent variable is a 0/1 choice to contact a previously "browsed" user. The model includes random effects for each user. We used the full sample of 143,184 choices of women who state that they are "looking for a long-term relationship." There are many more choices observed by men who declare the same preference. In order to reduce the computation time, we took a random sample of these men's choices (i.e., we kept all such men, but randomly discarded some of their observed choices).

Table 7 – Attribute Correlations in Online Matches and Actual Marriages

	Marriages	Keywords	Gale-Shapley: Men Make Offers	Gale-Shapley: Women Make Offers	Gale-Shapley: No Physical Attributes	Gale-Shapley: No Unobserved Component
	(I)	(II)	(III a)	(III b)	(IV)	(V)
Age	0.944*	0.720	0.707	0.708	0.727	0.946
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
		3,631	7,857	7,857	7,857	7,857
Height	0.31-0.63**	0.156	0.131	0.128		0.517
		(0.000)	(0.000)	(0.000)		(0.000)
		3,631	7,857	7,857		7,857
Weight	0.08-0.32**	0.130	0.103	0.104		0.274
		(0.000)	(0.000)	(0.000)		(0.000)
		3,631	7,857	7,857		7,857
Looks Rating	0.54***	0.329	0.213	0.201		0.517
		(0.000)	(0.000)	(0.000)		(0.000)
		1,973	1,025	1,028		1,143
Income	0.127*	0.153	0.136	0.142	0.097	0.460
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
		832	2,443	2,443	2,428	2,428
Years of Education	0.642*	0.131	0.196	0.192	0.167	0.768
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
		3,631	7,857	7,857	7,857	7,857

Note: The table displays (Pearson) correlation coefficients between mate attributes. p-values are in parentheses, and the number of observations are displayed below. Entries marked with (*) in column (I) come from data on actual marriages (in Boston and San Diego), obtained from the 2000 IPUMS 5% sample. Entries marked as (**) report the range of results obtained by anthropometric studies (N=46 to 984) surveyed by Spuhler (1968). The entry for looks correlation in this column, marked as (***), comes from Hinsz (1989) who constructs the attractiveness (rated on scale of 1-9) correlation of photographs of 30 engaged and 30 married couples whose engagement and 25th anniversary announcements were published in an Upper-Midwestern newspaper. In column (II), we classify two users as "matched" if they exchanged e-mails containing contact information (a phone number or e-mail address) or if the e-mails contain certain phrases such as "let's meet." A description of the Gale-Shapley algorithm, utilized to construct the "matches" analyzed in columns (III)-(V), is given in section 7. In (IV), we assume that all users have the same looks rating, height, and weight, and in (V) we assume that there is no unobserved component in users' preferences.

Hinsz, Verlin B. (1989), "Facial Resemblance in Engaged and Married Couples," *Journal of Social and Personal Relationships*, 6, 223-229.
Spuhler, J.N. (1968), "Assortative Mating with Respect to Physical Characteristics," *Social Biology*, 15(2), 128-140.
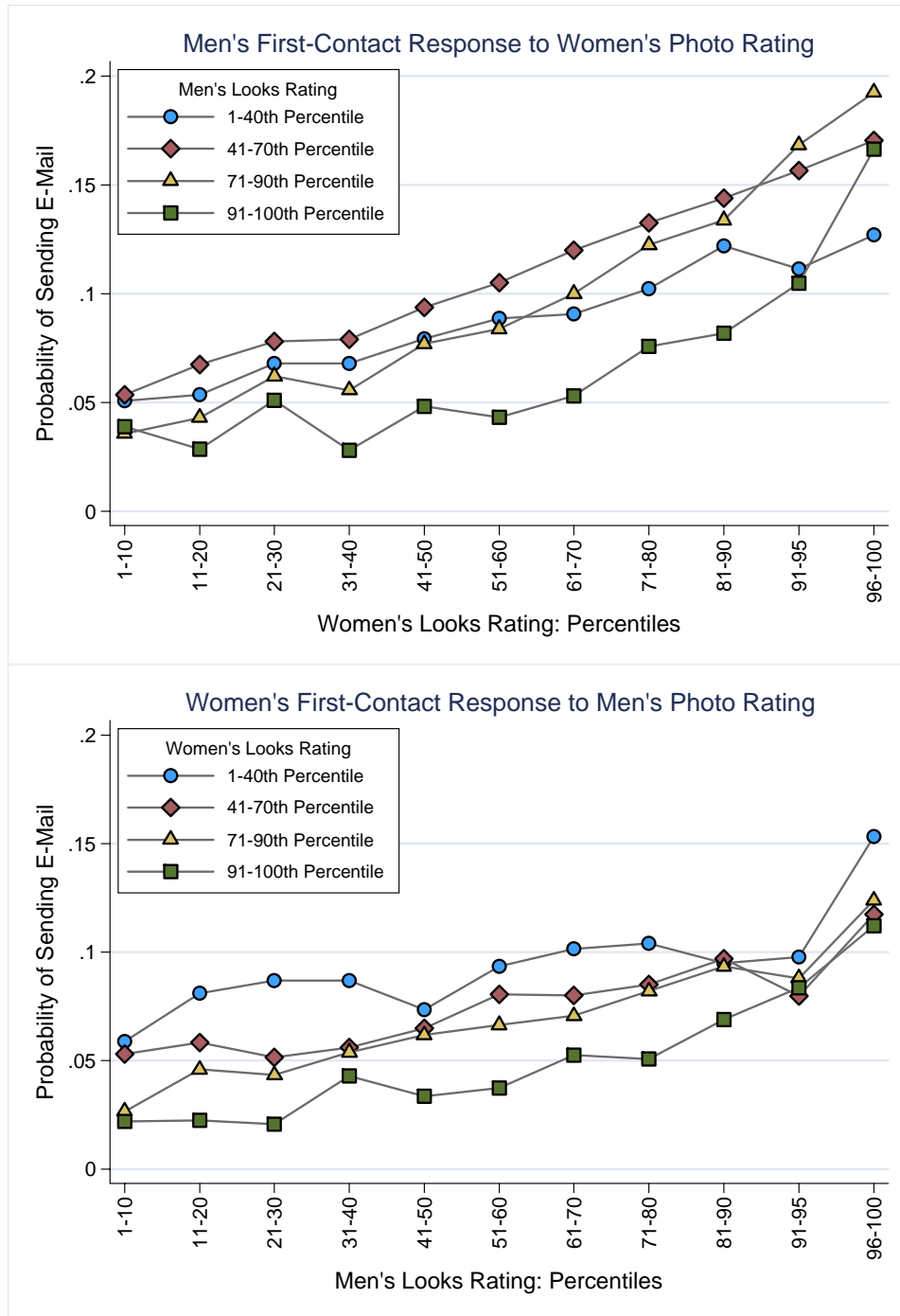
Figure 4.1 – Note: The figures report the result of an OLS regression where the dependent variable is an indicator variable for whether a user sends a first contact e-mail after browsing the profile of a potential mate. The independent variables are indicators for the photo rating of the user being browsed. The regressions also control for browser fixed effects. The vertical axis plots the estimated mean probability of sending a first-contact e-mail to a browsed profile. The horizontal axis indicates the photo rating of the browsed profile. The regressions were estimated separately for different groups of repliers. The first group comprises users who fall within the 40th percentile of the photo ratings distribution within their gender, etc.
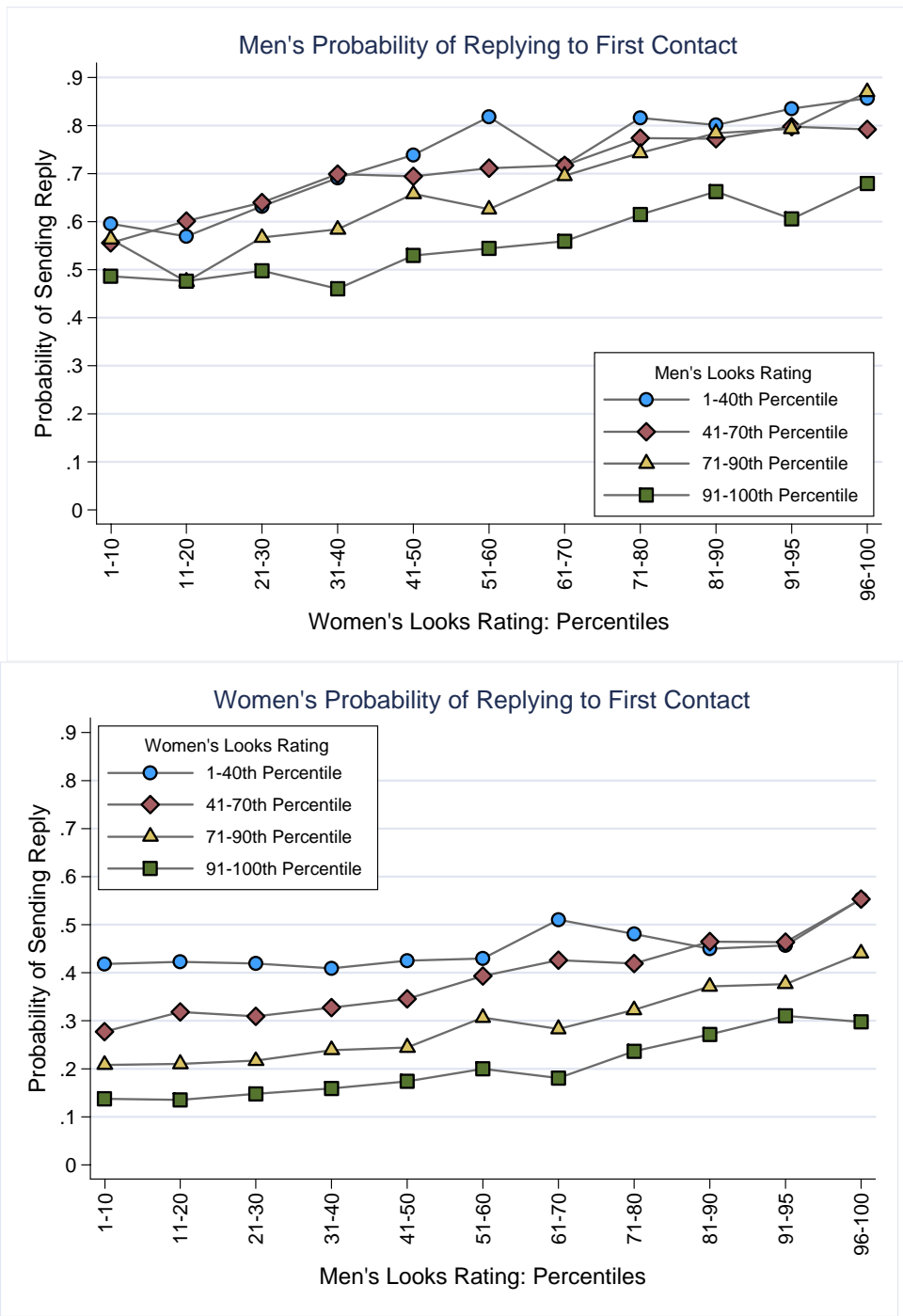
Figure 4.2 – Note: The figures report the result of an OLS regression where the dependent variable is an indicator variable for whether a user replied to a first-contact e-mail. The independent variables are indicators for the photo rating of the person sending the first-contact e-mail. The regressions also control for replier fixed effects. The vertical axis plots the estimated mean probability of sending a reply to a first-contact. The horizontal axis is the photo rating of the person sending the first-contact. The regressions were estimated separately for different groups of repliers. The first group comprises users who fall within the 40th percentile of the photo ratings distribution within their gender, etc.

First Contacts - Reason for Joining Site
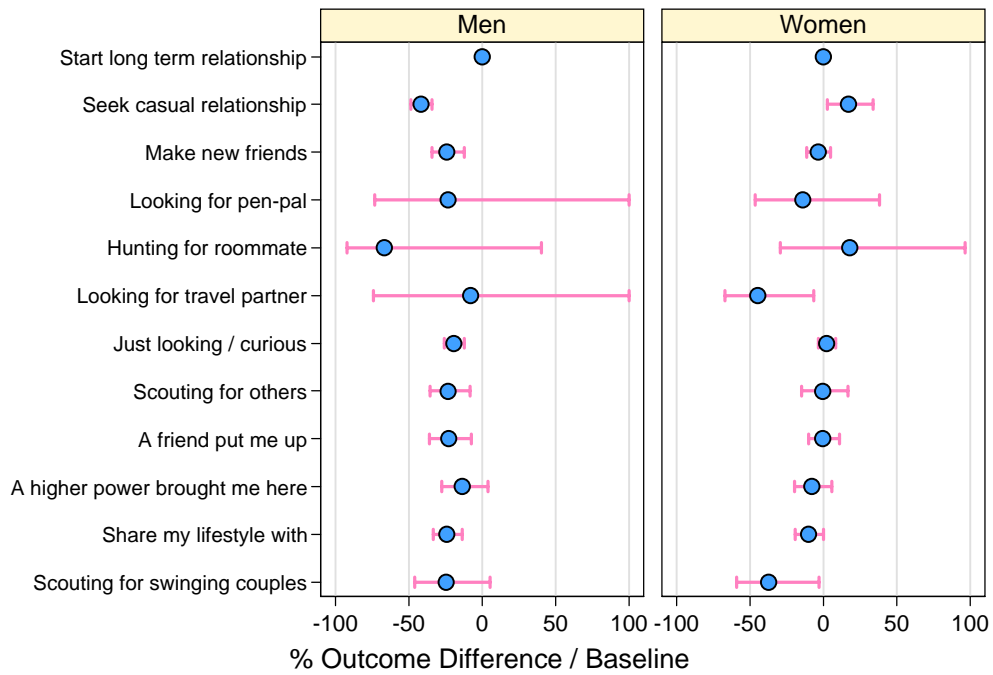
Figure 5.1

First Contacts - Looks Ratings
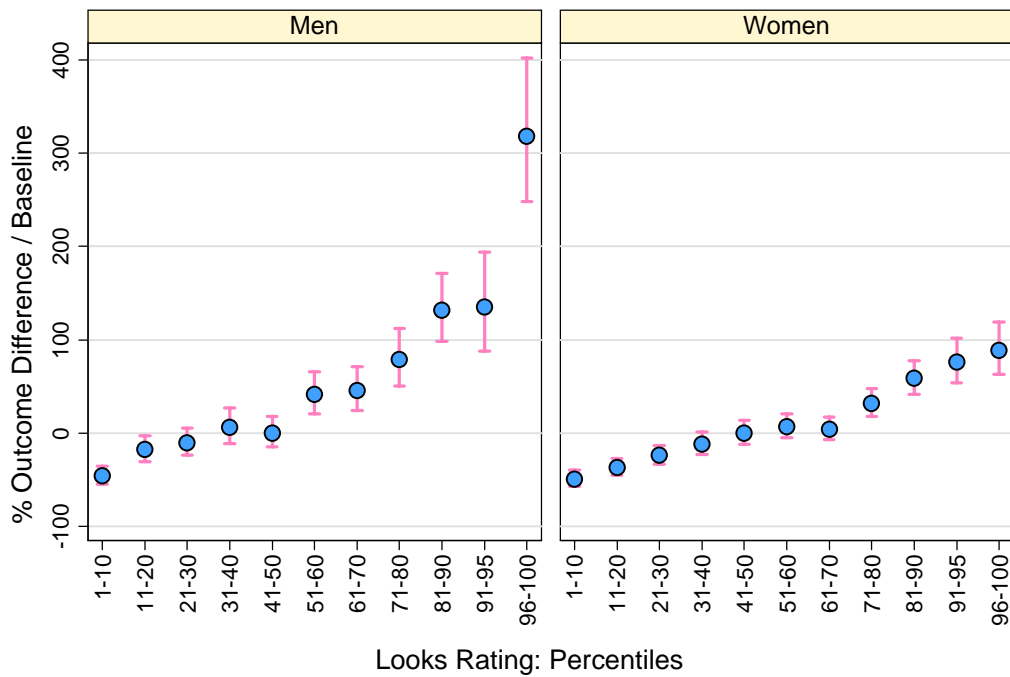
Figure 5.2

First Contacts - Self Described Looks

Note: Estimates for users who do not have a photo online
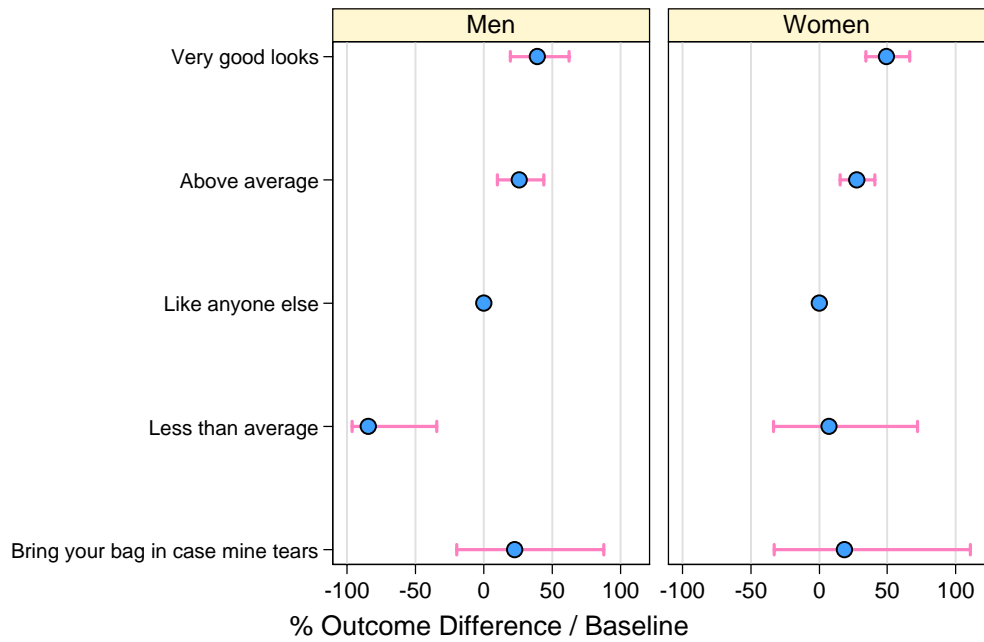
Figure 5.3

First Contacts - Height
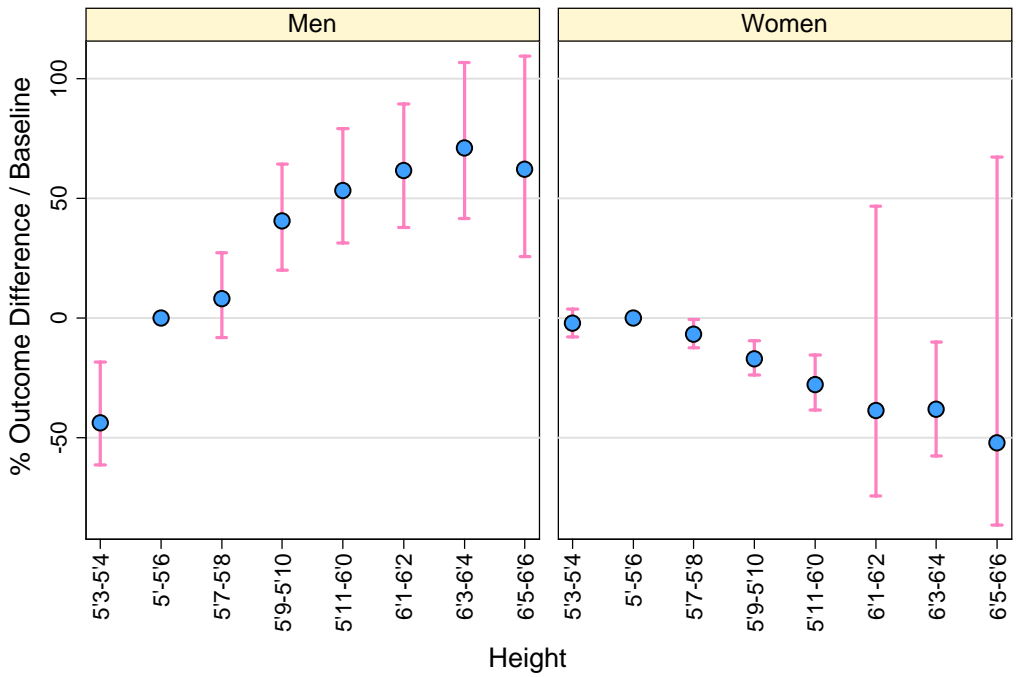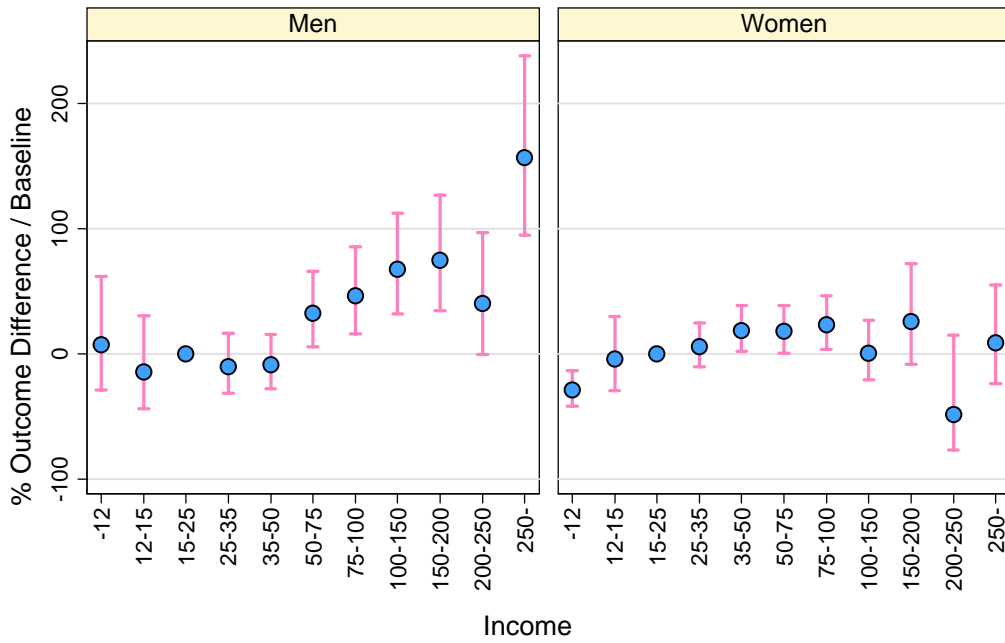

Figure 5.4

First Contacts - Body Mass Index
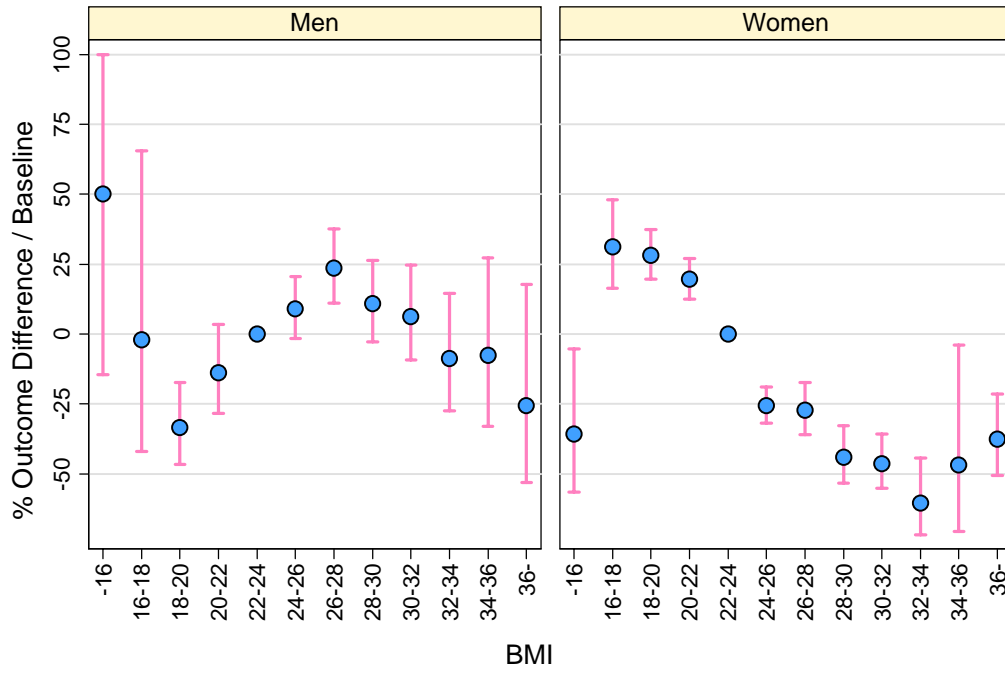
Figure 5.5

First Contacts - Income

Note: Income brackets in $1,000
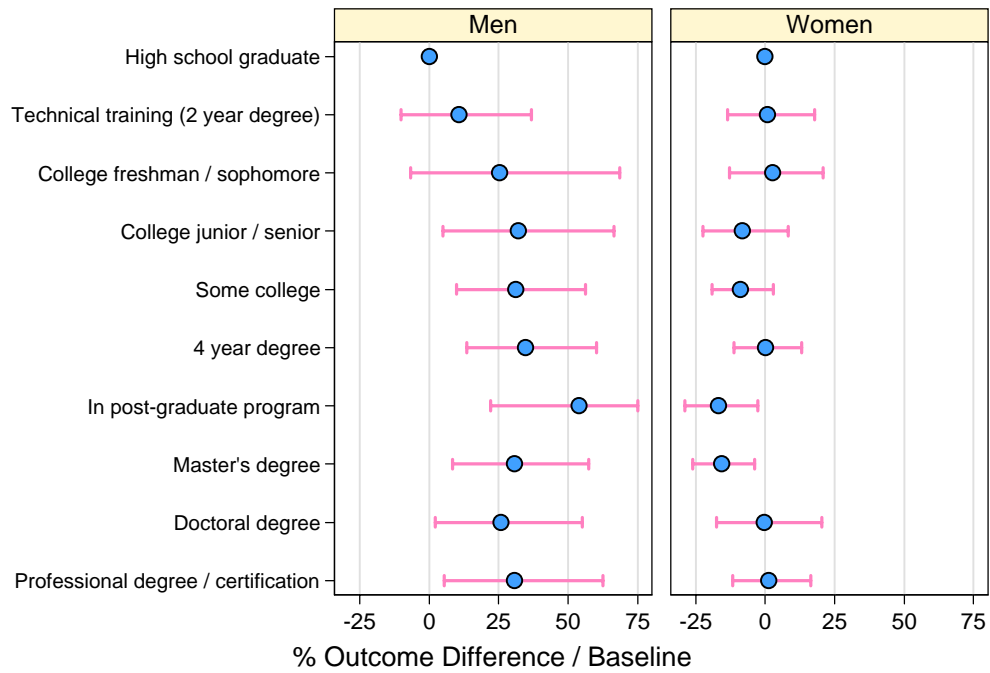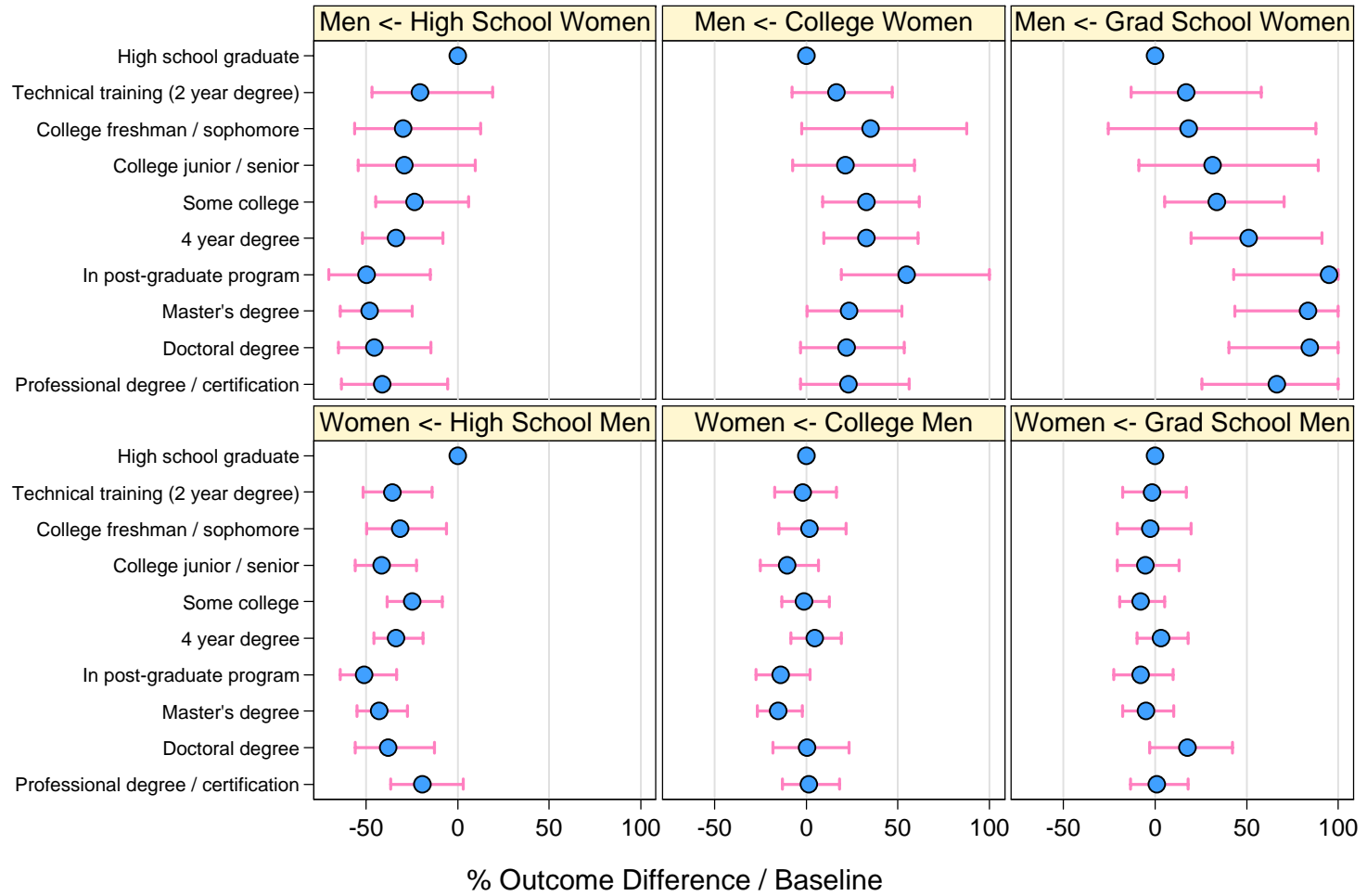
Figure 5.6

First Contacts - Educational Achievement

Figure 5.7

First Contacts - Educational Achievement
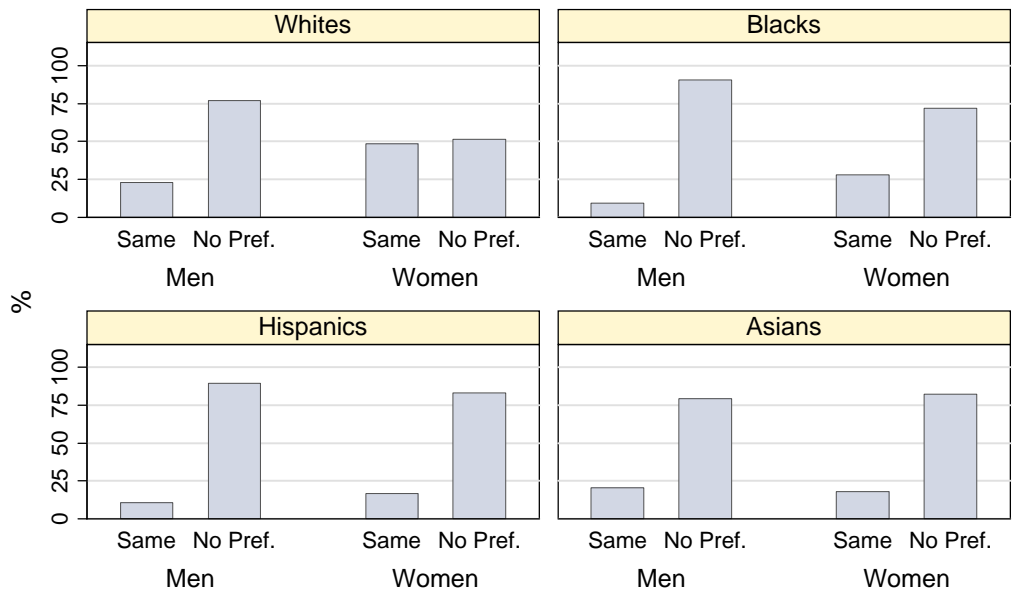
Segmentation: Education

Figure 5.8

Stated Preference for Ethnicity
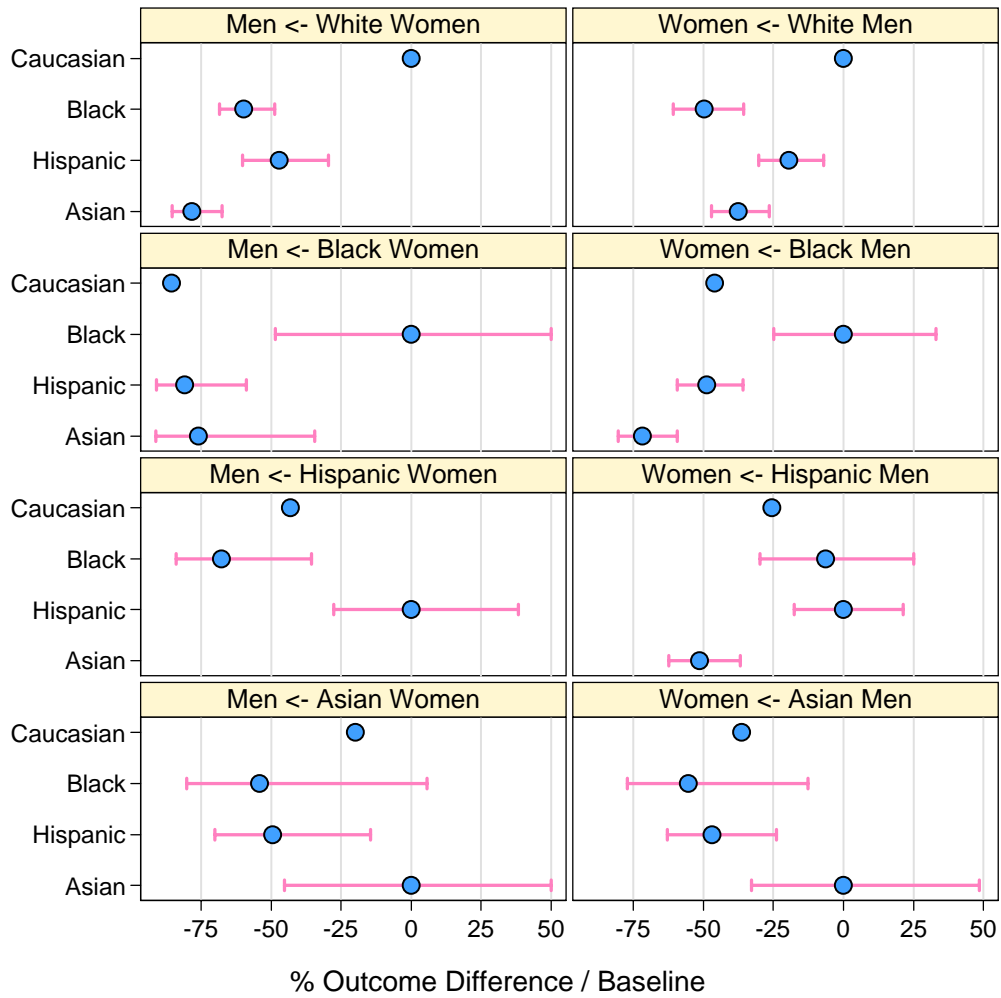### Which Ethnicity Do Users Prefer?

Figure 5.9

First Contacts - Ethnicity
Outcomes w.r.t. Ethnic Groups

% Outcome Difference / Baseline

Figure 5.10

First Contacts - Ethnicity

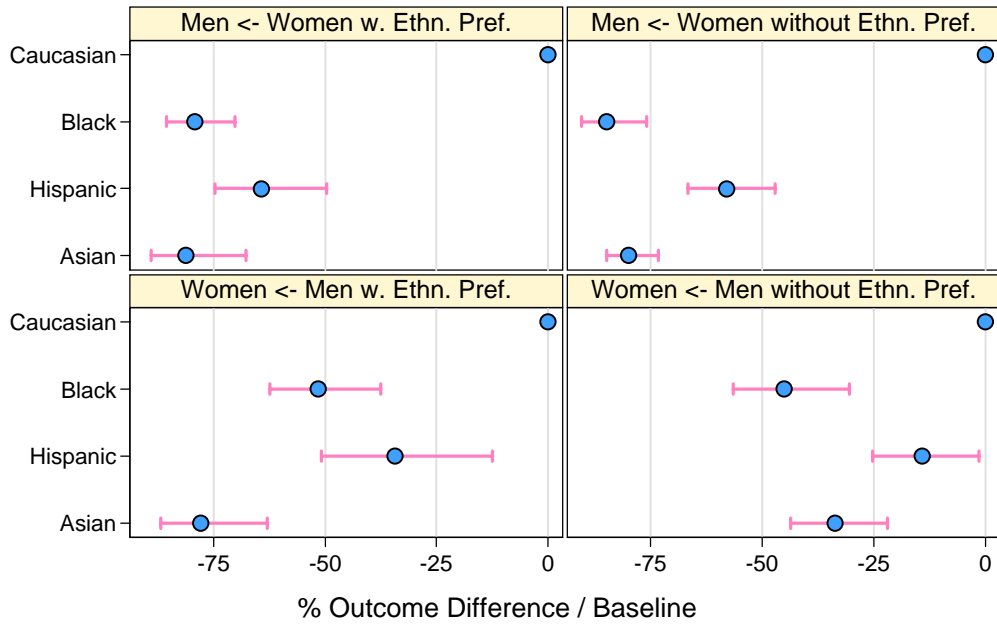## Outcomes w.r.t. Whites with or without a specific ethnicity preference

% Outcome Difference / Baseline

Figure 5.11