INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

423

2012

Marry for What? Caste and Mate Selection in Modern India

Abhijit Banerjee, Esther Duflo, Maitreesh Ghatak, Jeanne Lafortune.

www.economia.puc.cl • ISSN (edición impresa) 0716-7334 • ISSN (edición electrónica) 0717-7593

Versión impresa ISSN: 0716-7334
Versión electrónica ISSN: 0717-7593

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA

Oficina de Publicaciones
Casilla 76, Correo 17, Santiago
www.economia.puc.cl

MARRY FOR WHAT? CASTE AND MATE SELECTION IN MODERN INDIA

Abhijit Banerjee
Esther Duflo
Maitreesh Ghatak
Jeanne Lafortune*

Documento de Trabajo Nº 423

Santiago, Marzo 2012

* jlafortune@uc.cl

INDEX

Marry for What? Caste and Mate Selection in Modern India*

Abhijit Banerjee, Esther Duflo, Maitreesh Ghatak and Jeanne Lafortune†

March 2012

Abstract

This paper analyzes how preferences for a non-economic characteristic, such as caste, can affect equilibrium patterns of matching in the marriage market, and empirically evaluates this in the context of arranged marriages among middle-class Indians. We develop a model that demonstrates how the equilibrium consequences of caste depend on whether we observe a bias towards one's own group or if there is a preference for "marrying up". We then estimate actual preferences for caste, education, beauty, and other attributes using a unique data set on individuals who placed matrimonial advertisements in a major newspaper, the responses they received, and how they ranked them. Our key empirical finding is the presence of a strong preference for in-caste marriage. We find that in equilibrium, as predicted by our theoretical framework, these preferences do little to alter the matching patterns on non-caste attributes, and so people do not have to sacrifice much to marry within caste. This suggests a reason why caste remains a persistent feature of the Indian marriage market.

JEL classification: D10, J12, O12
Key words: Caste, Marriage, Stable matching

*We thank the *Anandabazar Patrika* for their cooperation for this project, and Prasid Chakrabarty and the team of SRG investigators for conducting the survey. We thank Raquel Fernandez, Ali Hortascu, Patrick Bajari, George Mailath, Whitney Newey, Parag Pathak, Andrew Postlewaite, Debraj Ray, Alvin Roth, Ken Wolpin, many seminar participants and three anonymous referees for very helpful comments. Finally, we also thank Sanchari Roy and Tommy Wang for outstanding research assistance.

†The authors are from the departments of economics at MIT, MIT, LSE, and University of Maryland, College Park respectively.

1 Introduction

Marriage is, among other things, an important economic decision. Sorting in families has an impact on child outcomes and the accumulation of human capital, and consequently, on long term economic development and inequality (Fernandez and Rogerson 2001, Fernandez 2003). In developing countries, where many women do not work outside their homes, marriage is arguably the single most important determinant of a woman's economic future.[1] In India, the setting for this paper, several studies have shown that marriage is indeed taken as an economic decision, managed by parents more often than by the prospective spouses. For example, Rosenzweig and Stark (1989) show that parents marry their daughters in villages where incomes co-vary less with respect to their own village. Foster and Rosenzweig (2001) show that demand for healthy women in the marriage market influences investments in girls.

Yet, "status"-like attributes, such as caste, continue to play a seemingly crucial role in determining marriage outcomes in India. In a recent opinion poll in India, 74 percent of respondents declared to be opposed to inter-caste marriage.[2] The institution is so prevalent that matrimonial advertisements (henceforth, ads) in Indian newspapers are classified under caste headings, making it immediately obvious where prospective brides or grooms can find someone from their own caste.

It is well known that these types of non-meritocratic social preferences can impede economic efficiency – a point that is often made in the literature on discrimination (Becker 1957). At the same time there is also the view that economic forces will tend to undermine institutions or preferences that generate impose large economic costs on people.[3] Indeed we do see the role of caste changing with economic growth and the diversification of earnings opportunities in India: the correlation between caste and income in India is significantly lower now, and caste plays much less of a role in determining the job someone has (Munshi and Rosenzweig 2006).

This paper is an attempt to understand why the stated role of caste in marriage remains so strong.[4] One possibility is that this is just something that people say, but they do not actually act

[1]Even in our sample of highly educated females and males, fewer than 25 % of matched brides were working after marriage.

[2]We use the word caste in the sense of *jati* (community) as opposed to *varna*. The latter is a broad theoretical system of grouping by occupation (priests, nobility, merchants, and workers). The jati is the community within which one is required to be married, and which forms ones social identity.

[3]In the context of the marriage market, for example, Cole et al. (1992) characterize an "aristocratic equilibrium" which is characterized by low levels of productivity because of the weight people put on status. They go on to show that the aristocratic equilibrium may be broken by increased economic mobility because it leads to the emergence of low status men who are nevertheless high wealth, who may be in a position to attract a high status, low wealth woman.

[4]This is related to the literature in the United States which has looked at how religious homogamy can persist despite the fact that some groups are clearly in minority. Bisin et al. (2004), Bisin and Verdier (2001) and Bisin and Verdier (2000) argue that in this context, there is a strong preference for horizontal matching in order to socialize children within one's faith. In addition, this homogamy may depend on the availability of partners of one's own religion and members of a minority such as Jews may actually exhibit higher rates of endogamy. However, these

on it–perhaps if we were to observe their actual marital choices we would see that caste is much less important than it is claimed. The fact that many people do end up marrying in caste is not enough to reject this view since it is well-known that caste is correlated with many other attributes and those could be driving the observed choices. To answer the question of whether caste actually matters in the choice of a spouse, we follow the methodology developed in Hitsch et al. (2009), and Fisman et al. (2008) for studying partner choice in the United States. Hitsch et al. (2009) use on-line dating data to estimate racial preferences in the US: they observe the set of partner profiles that a potential dater faces as well as which profiles they actually click on, which is what they interpret as the first act of choice. Since they observe all the attributes that the decision-maker observes, this allows them to identify the decision-maker's actual preferences. Fisman et al. (2008), do something very similar, using the random assignment of people to partners in speed dating. Both find strong evidence of same-race preferences, the equivalent of same-caste preferences in our context.

To look at the strength of caste preferences in marriage, we apply this methodology to a data-set that we collected based on interviews with 783 families who placed newspaper matrimonial ads in a major Bengali newspaper. We asked ad-placers to rank the letters they have received in response to their ad, and list the letters they are planning to follow up with, and use these responses to estimate the "marginal" rate of substitution between caste and other attributes. We find, evidence for very strong own caste preferences: for example our estimates suggest that the bride's side would be willing to trade off the difference between no education and a master's degree in the prospective husband to avoid marrying outside their caste. For men seeking brides, the own caste effect is twice the effect of the difference between a self-described "very beautiful" woman and a self-described "decent-looking" one. This is despite the fact that the population in our sample is urban, relatively well off, and highly educated (for example, 85% have a college degree). Interestingly, this preference for caste seems much more *horizontal* than *vertical*: we see little interest in "marrying up" in the caste hierarchy among both men and women, but a strong preference for in-caste matches. This is similar to the strong preference for same-race matches that the literature in the United States finds, though our context makes it even more striking: these are ads for arranged marriage in a relatively conservative society where the goal is clearly marriage and as a result, the motives of the decision-makers are likely to be much more classically economic than those involved in online dating or speed dating. Dugar et al. (fort), who study newspaper partner search in India using people's choices among nine randomly manipulated profiles, finds very similar results on the strength of own-caste preferences.[5]

papers do not include any characteristics of spouses other than religion, and thus do not address our notion of "cost" of marrying within one's religion. Bisin et al. (2004) note that this is an important question left for future research in their conclusions.

[5]Another paper on this general area that is more or less contemporaneous with ours is Lee (2007) who uses data from Korea data on online dating to study partner choice.

The central contribution of this paper is to suggest an explanation for the persistence of such a strong preference for caste-based matching. Specifically, we propose a model that explains why such a strong in-group preference might have survived despite the changing economic incentives, which should have made other characteristics (such as education, income, etc.) increasingly attractive. Our basic hypotheses is that this is because, as we saw, preferences for caste are primarily "horizontal" in the sense that people prefer to marry their own caste over marrying into any other caste. This goes against the traditional story about the caste system, which emphasizes its hierarchical structure, but is consistent with the sociological evidence on the nature of caste today (Fuller 1996). When caste preferences are horizontal and a particular condition we call *balance* holds, the theoretical section of the paper demonstrates, the matching patterns along non-caste dimensions are actually very similar to the ones that would obtain in the absence of any caste preferences. As a result, the equilibrium price of caste, which is the opportunity cost of the marriage option that one has to give up to marry in caste tends to be quite low. One possible reason why caste persists, therefore, is that it actually does not cost very much to marry within caste.[6]

To check that this line of reasoning actually works in the data, we use our estimated preferences and the assumption of stable matching to predict matching patterns, both in the current scenario and in various counterfactual settings. We surveyed our original respondents after one year to obtain information on their outcome on the marriage market: whether they married, and whom. We thus directly look at whether the actual matching pattern is similar to that predicted by the estimated preferences. Specifically, we use the Gale-Shapley (Gale and Shapley 1962) algorithm to generate the stable matches predicted by these preferences and compare them with the actual matches. Hitsch et al. (2009) perform the same exercise in their data, using an exchange of emails as the final outcome. Using metrics similar to the ones they use we find that the predicted and observed matches more or less line up, and therefore conclude that that stable matching is reasonable way to model marriage market equilibrium.

This brings us to the central empirical exercise in our paper: we compute the set of stable matches that would arise in our population if preferences were exactly as estimated above except that all caste variables were ignored. Our results indicate that the percentage of intra-caste marriages drops dramatically. This implies that caste is not just a proxy for other characteristics households also care about and that there are several potential matches for each individual, both inside and outside his or her caste. At the same time, we also find that individuals are matched with spouses who are very similar on all non-caste characteristics to the mate they would have selected when caste was included within one's preferences.

[6]This is in contrast with the results by Abramitzky et al. (2011) who estimate the impact of changing sex ratios in France after WWI on the propensity to marry across social class, clearly a vertical characteristic. In that setting, changing the supply of males has a large effect on the matching patterns because the price of marrying within one's social class has been greatly increased for women.

Second, we estimate the "equilibrium price" of caste in terms of a variety of attributes, defined as the difference between the spouses of two observationally identical individuals, one who is from the same caste and the other who is not. This is done by regressing a spousal characteristic, such as education, on all observable characteristics of the individuals and a dummy for whether the match is "within caste" among the set of simulated matches. There is no characteristic for which this measure of price is significantly positive.

To complete the argument we also estimate the equilibrium price for a vertical attribute, beauty, in terms of education. As our theory would predict, we see a non-zero price in this case.

A number of conclusions follow from our findings. First, there is no reason to expect that economic growth by itself will undermine caste-based preferences in marriage. Second, caste-based preferences in marriage are unlikely to be a major constraint on growth. Finally, one might worry that if caste becomes less important, inequality might increase along other dimensions as we will see more assortative matching. Given that the matching is already close to being assortative, this is probably not an important concern. [7]

While these conclusions are particularly important in the context of India, they are also more broadly relevant for any setting where we may observe strong in-group preferences in a matching context. Our theoretical conclusions, in particular, suggest that these preferences will have more impact for matching patterns in equilibrium whenever they display a "vertical" nature. Racial preferences for spouses, for example, may not have large equilibrium consequences if groups have a preference for marrying someone of their own race rather than hoping to marry a particularly favored racial group. On the other hand, preferences for social status (e.g., marrying into aristocracy) might be more vertical (Abramitzky et al. (2011) and Almenberg and Dreber (2009)).

While this is the main point of the paper, our data allows us to perform a number of other innovative exercises, which help bolstering the claim that we are identifying true preferences, rather than strategic behavior. Since we observe all the information seen by the ad-placer at the time they make a decision to reply or not to a particular letter means we do not have to worry about unobserved variables seen by the ad-placer and not seen by the econometrician.[8] However, there are still a number of possible alternative explanations for the choices we observe other than a pure preference for caste. These are not unique to our paper: all the papers that use this kind of methodology for estimating preferences face the same problem. However, our data and setting allow us to explore these alternatives.

First, we may be concerned about signaling. Perhaps there is no real preference for marrying in caste; because no one actually does it in equilibrium, however, those who make proposals to

[7] An important caveat to these conclusions is that they were obtained in a particular sample of highly educated West Bengali: while this population is anticipated to have weaker preferences for caste, thus potentially implying that our results are a lower bound for what one could expect in a different setting, this population is also potentially more "balanced" which implies that assortative matching may be easier to achieve here than in other contexts.

[8] This is why, unlike Dugar et al. (fort) we do not conduct an experiment: there is no econometric problem that an experiment can solve here.

non-caste members are treated with suspicion. We examine this by looking at the actual matches of those who make proposals out of caste and find that they are no different from that of others, suggesting that their underlying unobserved quality must be similar. Second, we need to deal with the possibility of strategic responses, i.e. the fact that some candidates may choose their responses based on who they expect to respond back positively rather than their true preferences. To get at this we compute an index of the quality of each ad and each letter, and show that the relative likelihood of responding to a "high quality" letter and a "low quality" letter are the same for "low quality" ad placers and "high quality" ad placers.[9] The fact that the ad-placer's ranking of the letters, and the decision to reply to the letters gives us very similar results also suggests that the respondents are not strategic in deciding whom to reply to.

The remainder of the paper proceeds as follows: Section 2 first sketches a model where caste and other attributes interact on the marriage market. Section 3 presents the data while Section 4 elaborates on the methodology and the results of preference estimation. Section 5 highlights the results of the stable matches and Section 6 uses these results to derive conclusions regarding the equilibrium. Finally, Section 7 concludes.

2 Model

In this section we develop a simple model of marriage. The model introduces caste-based preferences (see for example Anderson (2003))into an otherwise standard model of marriage (Becker 1973). The novelty is that we allow two-sided matching, and both horizontal and vertical caste-based preferences (as opposed to just vertical preferences). Our goal is to derive how marriage market outcomes are affected by going from vertical to horizontal caste preferences, under the assumption that there is another "vertical" attribute (beauty, education, earnings, etc.) that the decision-makers also care about. We characterize conditions under which non-assortative matching will take place, and when it does, characterize the price of marrying within caste or outside caste (in terms of what the decision makers gives up along the second, vertical dimension). These results will motivate our empirical analysis and help us interpret the main results.

2.1 Set up

Assume a population of men and women differentiated by "caste" where the caste of an individual is $i \in \{1, 2\}$. They are ranked in descending order: $i = 1$ is the higher caste, followed by $i = 2$.

Men and women are also differentiated according to a "vertical" characteristic that we will refer to as quality, that affects their attractiveness to a potential partner. The quality of men will

[9]We find more evidence of strategic behavior at other steps of the process, in particular in deciding which ad to send a letter to.

be denoted by $x \in [H, L]$ and the quality of women will be denoted by $y \in [H, L]$. We can think of these as education levels of men and women, or, income and beauty.

We denote the total number of women of type y who belong to caste i by $\omega_{yi} > 0$ and the number of men of type x who belong to caste i by $\mu_{xi} > 0$, where $x, y = H, L$ and $i = 1, 2$. We assume the following condition regarding the distribution of men and women:

Condition 1. *A population is said to satisfy balance (**B**) if $\omega_{t1} + \omega_{t2} = \mu_{t1} + \mu_{t2}$ for each $t = H, L$, and $\omega_{Hi} + \omega_{Li} = \mu_{Hi} + \mu_{Li}$ for each $i = 1, 2$.*

In other words, there is a balanced sex ratio within each type and within each caste. [10] If populations do not follow this assumption, non-assortative matching will follow trivially.

The payoffs of men and women are both governed by the quality of the match. We assume that in a union where the man's quality is given by x and the woman's by y the payoff function has two (multiplicatively) separable elements, one governed by the vertical characteristics, $f(x, y)$, and the other by caste, $A(i, j)$ where the latter is the payoff of someone who is of caste i and who is matched with someone of caste j.

We assume that the function $f(x, y) > 0$ is increasing with respect to both arguments. Thus, other things constant, everyone prefers a higher attribute partner. Also, for ease of exposition, we assume $f(x, y)$ is symmetric, i.e., $f(H, L) = f(L, H)$.

In order to generate conditions that are easy to interpret we give the function $A(i, j)$ a specific form:

$$A(i, j) = 1 + \alpha\{\beta(2 - j) - \gamma(i - j)^2\}$$

where $\alpha \geq 0, \beta \geq 0, \gamma \geq 0$. It is readily verified that $A(i, j) > 0$ as long as $\alpha\gamma < 1$ (which we assume) and as long as $\gamma > 0$ the function displays strict complementarity with respect to caste: $\frac{\partial^2 A(i,j)}{\partial i \partial j} > 0$.

This caste-based match quality function is flexible. It allows a vertical as well as a horizontal component to caste. For example, if $\beta = 0$ then caste is purely horizontal: people want to match within their caste. Otherwise, the higher the caste of the partner (lower is j) the higher the match specific gain to an individual of caste i. On the other hand, if $\gamma = 0$ then caste is purely vertical with everyone preferring a higher caste partner, as in Anderson (2003).

We also assume that a number ν_{yi}, $y = H, L, i = 1, 2$, $0 < \nu_{yi} < \omega_{yi}$ of women and a corresponding number $0 < \kappa_{xi} < \mu_{xi}$ of men have caste-neutral (**CN**) preferences, $\alpha = 0$. These individuals put no weight on the caste of a potential partner, i.e., for them $A(i, j) = 1$ for all $i = 1, 2$ and $j = 1, 2$. Those who are caste-conscious value a caste-neutral individual of caste i ($i = 1, 2$) in the same way as they would a caste-conscious (**CC**) individual of caste i ($i = 1, 2$). As we will see, the data clearly supports the idea that a fraction of individuals are caste-neutral.

[10] As is well-known, the sex-ratio in South Asia tends to make men more abundant. However, as Rao (1993) has shown, the gap between the normal age at marriage between men and women, combined with the fact that the population is growing, counteracts this effect and almost all men do manage to find spouses.

Given these two elements governing the quality of a match, we assume that the payoff of an individual of gender G, of caste i who is matched with someone of caste j in a union where the man's quality is given by x and that of the woman's by y is given by:

$$u^G(i,j,x,y) = A(i,j)f(x,y) \text{ for } G = M, W.$$

We have imposed a lot of symmetry here: For example, a man of type 1 of caste 1 marrying a woman of type 2 of caste 2 gets the same payoff that a woman of caste 1 of type 1 would get from marrying a man of caste 2 of type 2. This is convenient for stating the results in a more compact form, but is by no means essential.

We also assume that the utility of not being matched is zero. Since because both $f(x,y)$ and $A(i,j)$ are positive, the utility of being matched with anyone is always better than that of remaining single. Since the total number of men and women are the same, everyone should match in equilibrium.

Finally we assume that matching is governed by these preferences—in particular there are no transfers, so that we have what in the literature is called non-transferable utility (NTU) matching (as in recent studies of the United States matching market by Hitsch et al. 2009, Fisman et al. 2006 and Fisman et al. 2008). This assumption is less common in the development economics literature on marriage than the alternative transferrable utility (TU) assumption (e.g., Becker 1973, Lam 1988), where dowries are interpreted as the instrument of transfer. Demanding a dowry is both illegal and considered unethical in middle–class urban Bengali culture,[11] and as a result, no one mentions dowries in the ads or the letters, unless it is to announce that they do not want a dowry. Our presumption is that some fraction of the population will eventually ask for a dowry, but a substantial fraction will not (given that they spend money to say that in their ads). Therefore we cannot assume that we are in either of the pure cases. Our strategy therefore is to go ahead as if we are in a pure NTU world but argue that we would get very similar results if we made the TU assumption. However the presence of dowry can potentially affect the interpretation of our estimated preferences from the NTU model: the next sub-section, while a slight detour in the exposition of the theory, deals with this important concern. We return briefly to the stable matching patterns under TU at the end of section 2.3.

[11] We have so far failed to locate a study on dowry in this population that would throw light on its extent. However, we note that while Kolkata has 12 percent of the population of the largest metropolitan cities in India, it has only 1.9 percent of the so-called "dowry deaths" in these cities (about 6,000 in a year, India-wide), which are episodes where a bride is killed or driven to commit suicide by her in-laws following negotiation failure about the dowry. To the extent that the prevalence of dowry death partly reflects the prevalence of dowry, it suggests that they are less prevalent in Kolkata than in other major cities in India.

2.2 An important caveat: preferences estimation with unobserved attributes

Our empirical strategy relies on the fact that the econometrician observes everything that the decision-maker observes. However unobserved attributes may still play a role–exactly as they would if the observed characteristics were randomly assigned–if the decision-makers take into account the correlation between observables and what they do not observe. A key example of such an unobservable is the expected "ask"–some people will demand dowry and others will not.

However, note that dowries, like many other unobservables, will get revealed in a future round of the marriage negotiations. Given that many people will not ask for a dowry, and you can always reject the ones who ask for too much later (or offer too little), it makes sense to first short-list every prospect worth exploring *ignoring the possibility of their asking for a dowry or offering one*, and to actually find out whether or not they want a dowry (or want to offer one) by contacting them. They can then discard the ones who ask for too much or offer too little based on better information. Obviously this logic only works if the cost of contacting an additional person is small which, given the large numbers of contacts that are made by people, seems plausible. It is straightforward to formalize this argument, and we do so in a separate online appendix.[12] Assuming that the conditions of this proposition hold (namely, the exploration costs are not too high), it tells us what we observe in the data is people's true ordering between those whom they consider and those whom they reject, even if dowry and other still to be revealed attributes will eventually be an important consideration in the decision. Based on this ranking we infer people's preferences over a range of attributes. We will, however, come back to discuss some direct evidence that the estimated preferences are consistent with the assumption that people ignore dowry at this stage.

None of this helps us with the possibility that there are unobserved attributes that will never be observed, but may yet be driving the decisions because of their correlations (actual or hypothesized) with the observables. We do try to test of some specific hypotheses of this class (e.g., is caste really a proxy for "culture') using ancillary data, but at one level this is obviously an impossible quest.

2.3 Stable matching patterns

To start with, observe that if everyone were **CN** all H types would want to match with H types and since there are the same number of H type men and women, this is indeed what would happen – people would match **assortatively**. There may be out of caste matches, but

[12]The assumption here is that the unobserved attribute has a fixed value. It is more like something like attractiveness than like a demand for dowry, which is something that might adjust to exactly compensate for differences in other attributes. Nevertheless, as long as each set of candidates with the same observable characteristics contains a sufficiently large subset which is on average identical to the rest of the group in everything except for the fact it will not accept a dowry, and as long as it is not possible to predict this in advance (dowry demands or offers are not made in writing), it makes sense to rank everyone as if no one wants a dowry, as long as the cost of search is not too large.

those who match out of caste will have the same quality of matches as those who marry in caste. We formalize this idea by introducing the concept of an average price of caste.

Definition The average price of caste (**APC**) for women (men) is the difference, in terms of average quality of the matches, of women (men) of the same quality who marry in or below caste, relative to the average quality of those who marry above or in caste averaged over all types of women.

The **APC** is zero if everyone matches assortatively as in the case where everyone is **CN**. With caste preferences, there is a potential trade-off between marrying assortatively and marrying based on caste preferences and therefore **APC** need not be zero. For example, consider a configuration where the only out of caste match is between high types of caste 2 and low types of caste 1 for both men and women, and all other matches are assortative. The price of caste will be positive because those H types who match in caste get a higher quality match relative to those who match with a higher caste.

Define xic to be a x-type individual ($x = H, L$) from caste i ($i = 1, 2$) who has caste preference $c \in \{C, N\}$ where $\alpha(C) = \alpha > 0 = \alpha(N)$, that is, people can be either caste-conscious (C) or caste-neutral (N). Therefore, we have eight types of individuals for each gender: $H1C$, $H1N$, $H2C$, $H2N$, $L1C$, $L1N$, $L2C$, and $L2N$. Sometimes we will refer to just the type and caste of an individual (and not his/her caste-preference): in that case we will refer them to as a xi type (where $x = H, L$ and $i = 1, 2$). Furthermore, if X-Y are a match, X is the type and caste of the female and Y is that of the male.

Proposition 1 establishes that, if an additional condition which limits the fraction of **CN** people in the population holds, then pure assortative matching cannot be an equilibrium when the vertical dimension of preferences is strong enough.

Condition 2. ***Limited Caste Neutrality (LCN)****: The number of* ***CN*** *H1 men is less than the number of caste conscious H2 women, and the number of* ***CN*** *H1 women is less than the number of caste conscious H2 men.*

Clearly this cannot hold unless **CN** people are a sufficiently small fraction of the population.

Let

$$\beta_0 \equiv \frac{1}{\alpha}\left(\frac{f(H,H)}{f(H,L)} + \alpha\gamma - 1\right)$$

and

$$\beta_1 \equiv \frac{1}{\alpha}\left(\frac{f(H,H)}{f(H,L)}(1-\alpha\gamma) - 1\right).$$

Below, we show that assortative matching is an equilibrium as long as the attraction of matching with the high caste (the vertical dimension) is not too strong:

Proposition 1. *All equilibria will include only assortative matches if* $\beta \leq \beta_1$. *On the other hand, if* $\beta > \beta_0$ *the following properties must hold: (i) all equilibria must have some non-assortative matching as long as condition* ***LCN*** *holds; (ii) if there is at least one non-assortative match there must be at least one out-of-caste non-assortative match; (iii) all out-of-caste non-assortative matches must involve an H type of caste 2 matching with an L type of caste 1.*

Proof. Suppose an equilibrium with one non-assortative match exist. By our assumption of a balanced sex ratio by type, there must be ate least another such match. Given that $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$ are the four possible matchings in terms of caste, if we treat identical matches with the gender roles reversed as the same match, then there are ten logical possibilities for pairs of non-assortative matches: (i) H1-L1 and L1-H1; (ii) H1-L1 and L1-H2; (iii) H1-L1 and L2-H1; (iv) H2-L2 and L1-H1; (v) H1-L2 and L2-H1; (vi) H1-L2 and L1-H2; (vii) H1-L2 and L2-H2; (viii) H2-L1 and L1-H2; (ix) H2-L1 and L2-H2; (x) H2-L2 and L2-H2. Of these (i), (iii), (v), (vi), and (x) are clearly unstable since there is a rematch from these two pairs of matches that would make both parties better off in at least one match. Case (vii) is also unstable because the H1 man would always prefer at least matching with an L1 woman who would accept such a match unless already matched with an H1 male. Thus, to be stable, this case would have to be combined with case (ii) or (iv). If it is combined with case (ii), the H1 man would be able to attract away the H2 woman paired with an L1 man. If combined with case (iv), the H2 man and woman would form a match. Thus, case (vii) is unstable.

For case (viii) and (ix) to be stable, the H2-types must prefer being matched with an L1 than with each other. This will occur when

$$(1 + \alpha\beta - \alpha\gamma) f(H, L) > f(H, H)$$

which translates into $\beta > \beta_0$. When $\beta < \beta_0$ (which holds when $\beta < \beta_1$ since $\beta_1 < \beta_0$), those matches are unstable.

Both cases (ii) and (iv) will be unstable if H1 prefers matching with H2 than with L1. This occurs when

$$(1 + \alpha\beta) f(H, L) < (1 - \alpha\gamma) f(H, H)$$

or when $\beta < \beta_1$. Thus, if $\beta < \beta_1$, all matches will be assortative.

On the other hand, if $\beta > \beta_0$, then, starting from an assortative match (which exists given **B**), a H2C will always want to match with a L1 unless she is already matched with a H1, and any L1N would accept her offer. Therefore the only way there can be an assortative equilibrium is if all H2Cs are matched with H1s. But if there are two H1Cs who are each matched to a H2, they would want to deviate and match with each other. Therefore it must be the case that either the number of H1N men (women) is at least as large as the number of H2C women (men). This cannot be true if condition **LCN** holds.

The next step is to observe that if there is at least one non-assortative match then there must be an out-of-caste non-assortative match. Suppose on the contrary that the population only contains assortative matches, and, non-assortative matches of the form H2-L2 or H1-L1 (or the reverse). Since we have a balanced sex ratio within type, we must have another non-assortative match. If those are of the same form (H2-L2 or H1-L1), then both H-types would match together and this is clearly unstable. Let us consider the possibility that the equilibrium includes only groups of pairs of the form $H2$-$L2$ and $L1$-$H1$. While H2 would clearly be willing to match with H1, H1 will be unwilling to do so as $\beta > \beta_1$. On the other hand, there is at least one L1N in the population (by our assumption) and this individual will be willing to match with H2. Now, L1N will be willing to do so as long as she is not already matched with an H-type. She cannot be matched with an H2 type since all non-assortative matches are within caste by assumption. Also, she cannot be matched with an H1 since in that case, the two unassortatively matched H1 would pair together. Thus, there will always be at least one non-assortative out-of-caste match, which is a contradiction. To complete the proof, note that there are two possible out-of-caste non-assortative matches: H2-L1 or H1-L2. The latter is ruled out since all the cases involving it were ruled at the beginning as not stable. ■

Of course, Proposition 1 does not guarantee that assortative matching is the only possible configuration when $\beta < \beta_0$. Moreover $\beta < \beta_1$ is a fairly stringent condition. This multiplicity of equilibria and the corresponding need to impose strong conditions to be able to limit the set of possible equilibrium patterns is a direct result of introducing caste neutrality. As is well-known, indifference introduces significant complications in matching problems.(See, for example, Abdulkadiroğlu et al. 2007 and Erdil and Ergin 2006). However indifference in our framework cannot be dismissed as a non-generic phenomenon. As we shall see, when we estimate preferences person by person, about 30% of the population show no caste preference of the type modeled here– which reflects the fact that their caste preferences are sufficiently weak so that other factors dominate their decisions and therefore given realistic choice environments (say 50 letters to chose from), we will never see them acting on their caste preferences. This is what indifference is meant to capture.

The next proposition provides a stronger characterization by adding the requirement that within a caste-type, the fraction of caste-neutral types is the same among men and women (see A.2 for the proof). For this we need to make a stronger assumption about the population distribution. We define

Condition 3. *A population is said to satisfy strong balance (**SB**) if $\omega_{ri} = \mu_{ri}$, and $\nu_{ri} = \kappa_{ri}$ where $r = H, L$ and $i = 1, 2$.*

Proposition 2. *Suppose the population satisfies **SB**. If $\beta < \beta_0$ then only assortative matchings are stable. Conversely, when $\beta > \beta_0$, all equilibria must have some non-assortative out-of-caste matching as long* *as condition* **LCN** *holds. Moreover if there is non-assortative out-of-caste matching it must involve, in addition to assortative matches, combinations of* $m \geq 0$ *L1-H2 and H2-L1 pairs and* $n \geq 0$ *either H2-L2 and L1-H2 pairs or L2-H2.and H2-L1 pairs. Finally the APC is zero when* $\beta < \beta_0$ *and positive if* $\beta > \beta_0$.

It is useful to ask whether we would get very different matching patterns if we took the same population (i.e one that satisfies **SB**), but used a TU framework. It should be clear that with TU matching, all **CC** H1 will match with each other and so will all **CC** L2 (the **SB** assumption makes this feasible) and all **CN.**

Under TU, it is sufficient to look at the total surplus under a given match and compare it with the total surplus under alternative matches. Let $v(xic, yjc)$ denote the total surplus when a man of type xic is matched with a woman of type yjc. Under our assumptions

$$v(xic, yjc) = \left[2 + \alpha(c)\{\beta(4 - (i + j) - 2\gamma(i - j)^2\}\right] f(x, y).$$

Given **SB**, we can show that if $\beta \geq 2\beta_0 - \gamma$ then an equilibrium with non-assortative matches is possible. Suppose not, and therefore start without loss of generality with an assortative matching equilibrium where individuals are matched to someone of the opposite sex identical to them in terms of type, caste and caste-preference. Consider a match between a $H2C$ and a $L1N$:

$$v(H2C, L1N) = [2 + \alpha\,(\beta - \gamma)]\, f(H, L).$$

Since $v(H2C, H2C) = 2f(H, H)$, so long as $\beta \geq 2\beta_0 - \gamma$ a $H2C$ type is better off matching with a $L1N$ type. Also, as $f(H, H) > f(L, L)$, and $v(L1N, L1N) = 2f(L, L)$, by a similar argument a $L1N$ type is better off matching with a $H2C$ type. When $\beta < 2\beta_0 - \gamma$, we can show that all pairs of non-assortative matches listed in the proof of Proposition 1 would not exist in equilibrium as they generate surpluses that are smaller than the ones obtained by assortatively matching the H-types. Thus, if $\beta < 2\beta_0 - \gamma$ and **SB** holds, all pairings would be assortative.

To sum up, our model suggests that the impact of caste preferences on equilibrium outcomes depends crucially on whether these preferences are vertical or horizontal. When preferences are mostly horizontal, out-of-caste matches will look like in-caste matches on non-caste attributes, i.e. they will be assortative, to the extent that demographics allow it. Furthermore, little would change in matching patterns on non-caste attributes if caste preferences were to be ignored and the "price of caste" will be zero. On the other hand, when preferences are strongly vertical, some fraction of out-of-caste matches would be non-assortative and we will see a positive “price of caste” in equilibrium.

Given these theoretical predictions, the empirical sections that follow will focus on estimating the magnitude of the caste preferences in our sample and determining whether they are mostly horizontal or vertical. Then, using these estimates, we will explore empirically the equilibrium consequences that these caste preferences generate for marital pairing and highlight their resemblances to the theoretical predictions generated here.

3 Setting and data

3.1 Setting: the search process

Our starting point is the set of all matrimonial ads placed in the Sunday edition of the main Bengali newspaper, the *Anandabazar Patrika* (ABP) from October 2002 to March 2003. With a circulation of 1.2 million, ABP is the largest single edition newspaper in India and it runs a popular special matrimonial section every Sunday. The search process works as follows.

First, the parents or relatives of a prospective bride or groom place an ad in the newspaper. Each ad indicates a PO box (provided by the newspaper), and sometimes a phone number, for interested parties to reply. They then get responses over the next few months (by mail or by phone), and elect whether or not to follow up with a particular response. While ads are placed by both sides of the market, "groom wanted" ads represent almost 63 percent of all ads placed. One can both post an ad and reply to one.

When both parties are interested, the set of parents meet, then the prospective brides and grooms meet. The process takes time: in our sample, within a year of placing an ad, 44 percent of our sample of ad-placers whom we interviewed were married or engaged although most had placed only a single ad. Of those who got married, 65 percent met through an ad, the rest met through relatives or, in 20 percent of the cases, on their own (which are referred to as "love marriages ").

3.2 Sample and data collection

We first coded the information in all the ads published in the Sunday edition over this time period. We excluded ads placed under the heading "Christian" or "Muslims" in the newspaper given our focus on caste, which is primarily (though not exclusively) a phenomenon among Hindus. The details on the information provided and the way it was coded are provided below. We refer to this data set of 22,210 ads as the "ad-placer sample."

We further restricted our attention to ads that did not mention a phone number, and requested all responses to be sent at the newspaper PO Box or to a personal mailing address.[13] This restriction was necessary to make sure that what the ad-placer knows about his/her respondents is fully captured by the letters. About 43 percent of the ad-placer sample included a phone number (sometimes in addition to a PO Box and sometimes as the only way to contact the ad-placer). We find little differences between the characteristics of the ads that included a phone

[13] Only a small fraction of ads included only a personal mailing address (namely, 4 percent of our interview-sample, and 8 percent of the ad placer sample).

number and those that did not, except in terms of geographical location: fewer ad placers with phone numbers were from Kolkata.

After excluding these ads from the ad-placer sample, we randomly sampled 784 ads. With ABP's authorization, respondents were approached and asked whether they would agree to be interviewed when they came to collect the answers to their ads at the newspaper PO Box. Only one sampled respondent refused to be interviewed. The ads placed by the 783 individuals who completed the survey form the "interview sample."

The interview was conducted in the ad-placer's home after a few days with the person in charge of the search, usually the parent, uncle or older brother of the prospective groom or bride. Detailed information was collected on the prospective groom or bride, his family and the search process for a marriage partner.[14] In particular, ad-placers were asked whether they also replied to other ads and, when they did, to identify the ads they had responded to among the ads published in the past few weeks. Ad placers were also asked how many letters they received in response to their ad (on average 83 for bride-wanted and 23 for groom-wanted ad placers), and to identify the letters they were planning to follow up with (the "considered" letters). We then randomly sampled five letters from the set of "considered" letters (or took the entire set if they had less than five in this category), and ten (or all of them if they had less than ten in this category) from the set of the "non-considered" letters, and requested authorization to photocopy them. The information in these letters was subsequently coded, using the procedure outlined below. We refer to this data set as the "letter data set."

Finally, a year after the first visit, this original interview-sample was re-interviewed, and we collected information regarding their current marital status and their partner's choice. Only 33 ad-placers out of the entire sample could not be contacted. Out of those we reached, 346 were married or engaged, and 289 of those agreed to a follow-up interview and gave us detailed information regarding their selected spouse, the date of the marriage and their overall search process including the number of ads posted and the way the match was made. Appendix Tables C.1 and C.2 compare ad-placers found and not found and those who agreed or refused to answer the follow up questions. There appears to be little systematic differences between the two groups.

3.3 Variable construction

Ads and letters provide very rich and mostly qualitative information. A data appendix describes the coding process. In this subsection, we mainly discuss the coding process for the caste information.

If caste was explicitly mentioned in the ad or letter, we used that information as the caste of the person. Caste is often not explicitly mentioned in the ad because the ad is usually placed

[14]The questionnaire is available online at `https://sites.google.com/site/jeannelafortune/research`.

underneath a particular heading in the newspaper corresponding to a caste. If caste is not directly mentioned in the ad, the heading is used for this classification. The information on caste is readily available, directly or indirectly, in the overwhelming majority of ads (98 percent). In the letters, caste is explicitly mentioned in about 70 percent of the cases.

As already mentioned, Hindu society is divided into a number of broad castes (*varnas)* but each of these castes, in turn, is divided into a number of sub-castes (*jati*s).Ad-placers or letters can be more or less specific in identifying themselves. Historically, there was a more or less clear hierarchy among the broad caste groups, but within each broad group, there was no clear ordering. We therefore grouped castes into eight ordered broad-caste groups, based on the classifications in Risley (1981) and Bose (1958), with Brahmin at the top (with the rank of 8, and various schedule castes at the bottom, with the rank of 1). Appendix Table C.3 presents the classification.

To determine whether a letter writer and an ad-placer are from the same caste, we attributed to each letter or ad the specific sub-caste mentioned in the ad. If the ad-placer or letter writer only mentioned a broad group, he or she is assumed to be from any of the specific sub-castes. For example, a self-identified Kulin Brahmin is considered to be from a different caste as a self-identified Nath Brahmin (though the vertical distance between them is set to zero), but is considered to be of the same caste as someone who simply identified himself as a Brahmin.

Another relevant piece of information is the stated preference regarding caste. Among the sampled ads, more than 30 percent of individuals specify their preference for marrying within their caste (using phrases such as "Brahmin bride wanted"). Another 20 to 30 percent

explicitly specify their willingness to unions outside their own caste by the use of phrases such as "caste no bar." The remaining 40 to 50 percent do not make any mention of preferences regarding caste.

The remaining variables coded were: education (in 7 categories), earnings and occupation for men (we construct an occupational score, referred to as "wage" in what follows), family origin, physical characteristics, and some more rarely mentioned traits (astrological signs, blood types, etc.). The data appendix provides more details on the coding and appendix table C.4 shows the fraction of ads in which each characteristic is not mentioned.

3.4 Summary statistics

Table 1 presents summary statistics for both our interview sample and the full set of ads. The two samples look quite similar, except that the interview sample is more likely to live in Kolkata (the Kolkata sample was less likely to provide a phone number).

It is important to emphasize that our sample is not at all representative of India, or even West Bengal. It is drawn mostly from the Bengali (upper) middle class, as evidenced both by the prevalence of higher caste individuals (a quarter of the sample are Brahmin), and educational achievement. Education levels are mentioned in the ad by 90 percent of women and 80 percent of men. Almost all men and women (90 percent) have at least a bachelor's degree. Both men and women have occupational scores significantly higher than the median urban formal sector occupational score (from Bargain et al. 2007 and Glinskaya and Lokshin 2005). This group enters the marriage market after they have completed their education and (at least for men) found a job: the average age is 27 for women, and 32 for men. Around 50 percent of the sample lives or works in Kolkata and slightly less than half consider their family as originating from West Bengal. This paper is not meant to be a characterization of the marriage market in India, but a description of how this particular market works; it is quite striking that even in this very well educated and quite well off sample, caste remains so important.

Physical characteristics clearly play an important role in the marriage market. Height is mentioned in the ad by 96 percent of the women and 90 percent of the men. A prospective bride's skin tone and beauty are mentioned in 75 percent and 70 percent of the groom wanted ad, respectively beauty. There does not appear to be much boasting about physical appearance, however. More ads describe the bride as being "decent-looking" than either "beautiful" or "very beautiful."

Table 2 shows summary statistics for this sample, comparisons between the ad-placers and the letters they have received, as well as with their eventual spouses. In this table, as well as in the remainder of the paper, all differences are presented in terms of the difference between the characteristics of the man and the characteristics of the woman.[15]

Two-thirds of the letters that mention caste are from someone from the same caste as the ad-placer. The fraction of within-caste marriages among actual matches is a little higher than the fraction of letters that come from within one's caste: 72 percent of the prospective grooms and 68 percent of the prospective brides who are married after a year have married within their own narrow caste. This fraction increases to 76 percent and 72 percent respectively if we use the broad classification in terms of caste. Men who marry outside of caste tend to marry women from a lower caste while women who marry outside of caste tend to marry someone from a higher caste. Women tend to marry grooms who have either the same education (42 percent) or who are more educated than them (45 percent). Men are more likely to marry similarly or more educated women than themselves and 72 percent to 75 percent of the brides and grooms are from the same family origin (i.e., West or East Bengal).

[15]Since the sampling was stratified with unequal weights, each letter is weighted by the inverse of its probability of selection.

4 Estimating preferences

Using this data, we now estimate the preferences over various characteristics, exploiting the choices made by ad-placers and people who replied to their ads. We first discuss our basic empirical strategy and present the results. We then empirically examine various concerns about why the coefficients we observe may not actually represent households' preferences.

4.1 Basic empirical strategy

The first goal of this section is to estimate relative preferences for various attributes in a prospective spouse.

We assume that the value of a spouse j to a particular individual i can be described by the following function:

$$U(X_j, X_i) = \alpha X_j + \beta f\left(X_i, X_j\right) + \mu_i + \varepsilon_{ij} \tag{1}$$

where α captures the effect of the characteristics of person j, β specifies how this effect might be different depending on person's i own characteristics and μ_i represents ad-placer fixed effects.

We have in our data several indications of individual's revealed preference for one potential spouse over another that can allow us to estimate the parameters of equation (1).

First, we know whether an ad-placer is following up with a particular letter writer or not. We thus have information that he preferred this letter to the letters he did not consider. Second, the ad-placers also provided us with their ranking of each letter we sampled. For this exercise, we asked them to give us their true preference ordering, regardless of whether they were considering responding to the letter. In addition, for ad-placers who have themselves replied to ads, we know which ads they decided to reply to (and we also know the universe of ads they could have replied to on that particular date). Furthermore, we know that a letter writer decided to reply to an ad. Finally, we also know how many responses an ad received.

We focus in what follows on the decision of the ad-placer to respond to a particular letter. The results using the ranking of letters provided by the respondent (provided in the appendix) are extremely similar. We prefer to consider the ad-places responses to the letters he has received over the other choices we observe in the data for three reasons. First, we can be sure that the ad-placers have read all the letters they have received, so the set over which choices are made is well-defined. Second, strategic behavior is *a priori* less likely in this sample since the letter writer has already expressed interest in the ad-placer. The results from these other strategies are presented in the appendix, and the relevant differences are discussed below.

The regressions we estimate thus take takes the following form:

$$y_{ij} = \alpha X_j + \beta f\left(X_i, X_j\right) + v_i + \epsilon_{ij}, \tag{2}$$

where y_{ij} is a dummy equal to 1 if ad-placer i replied to letter j.[16] In the empirical analysis, we specify $f(X_i, X_j)$ to include dummies for whether the value of some elements of the X vector are equal for i and j (for education, caste, location), the difference between the value of the elements of the vector for some attributes (always normalized such that we take out the average difference between men and women), and its square.[17] We estimate equation (2) using a conditional logit with fixed-effects for each person i, and OLS with fixed effects. Note that for characteristics of ad-placers, we could use either the information provided in their ad or their response to our interview questions. In order to use these estimates in the stable matching exercises that follow, the former was employed. However, very similar results were obtained when using the interview data.

4.2 Results

Table 3 presents the results of fixed-effects and conditional logit regressions, where the binary decision of whether or not an ad-placer i responded to a letter j is regressed on a set of characteristics of the letter, and its interactions with those of the ad.

Columns 1 to 5 present the specifications for groom-wanted ads (ads placed by females), and columns 6 to 10 present the specifications for bride-wanted ads. Recall that in both cases, differences are presented in terms of the difference between the characteristics of the man and the characteristics of the woman. A positive difference in education, for example, means that the prospective groom is more educated than the prospective bride. Also, given that we code the higher castes with a higher number, a positive difference between the man's and woman's caste indicates that the man is of a higher caste. A variable is set to zero if the letter did not mention that characteristic, and we include a dummy variable to indicate a missing characteristic.[18]

Many variables are not individually significant, but overall, the regression has an R square of around 32%. Most attributes have the expected signs in the utility function: both women and men prefer more educated spouses; science and commerce are the preferred fields. Women prefer men with higher incomes. Men prefer younger women, and women prefer men their own age. Both dislike large differences in age. As Hitsch et al. (2009), we find that looks matter: men prefer women who describe themselves as beautiful or very beautiful, and seem to have a strong preference for lighter-skinned brides. For example, the OLS estimate suggests that the probability of getting a call back would be higher for a very light-skinned woman without any

[16] This is similar to the regression framework of Hitsch et al. (2009).

[17] For linear variables, such as age or height, we include only the difference between the value of the variable for the man and the woman and its square, not the level of age or height for the letter writer: this is because once we include a fixed effect for the ad-placer, the age of the letter writer and the difference in age are co-linear.

[18] All models were estimated with and without including a series of additional covariates (for example, how "cultured" the family is, its wealth level, astrological sign). To save space we focus on the more parsimonious specification in the tables; the results are extremely similar when these additional controls are included.

education than for a dark-skinned woman with a college degree. Both men and women prefer a spouse who lives in Kolkata (recall that a majority of our families are from Kolkata), and with similar family origin (i.e., East or West Bengal).

Caste plays a very prominent role. In particular, both men and women seem to have a very strong preference for marrying within the same caste. The OLS estimates indicate that a woman is 13 percentage points more likely to call back a prospective groom if he is from the same caste, controlling for all other attributes. A man is 17 percentage points more likely to call back a woman from his caste. These are large differences, considering that the average call back rate is about 28 percent. These results also indicate a high preference for caste relative to other attributes. For example, in the bride-wanted ads, the probability to be called back is the same for a man from the same caste and no education as that for a man from a different caste with a master's degree. Men are willing to sacrifice three shades of skin tone to marry someone within their caste (column 6). The coefficient of the logit specification imply similarly high marginal rates of substitution between castes and other characteristics. Overall, about 30 percent of the explanatory power of our regressors (excluding adplacers' fixed effects) come from caste-related variable, the remainder coming from all other characteristics included in the regressions.

Given our theoretical framework, an important issue is whether preference for caste is horizontal or vertical. It appears to be purely horizontal for women: Women prefer men who are as close to their caste as possible. Among men, conditional on marrying out of caste, those from relatively low castes prefer women from the highest available caste. The magnitudes of the coefficient on the difference in caste, however, are much smaller than those for being of the same caste.

Several of the variables in these regressions may be co-linear proxies for the same underlying attribute. Specifically, the basic specification includes income (when reported), education, type of degree, and occupational score (when reported). This may artificially depress the coefficient of these variables relative to the caste variable. To investigate this possibility, we estimate in columns (4) and (9) a more parsimonious specification. We first regressed the log income of the letter writer (when reported) on all the education variables and the occupational score (including dummies when not reported). We then constructed for each ad-placer and letter writer a "predicted income" measure using the coefficients of that regression, and included this variable instead of all the education, income, and wage variables, adjusting the standard errors for the fact that this regressor is generated by using the method suggested by Murphy and Topel (1985). Predicted income has a strong and significant impact on the probability of call back, but the preference for marrying in the same caste is high relative to that for income. A woman from a given caste would be as likely to contact a man from her own caste with a given predicted income level than a man from a different caste who is predicted to earn 50 percent more.

Appendix Table C.6 presents similar regressions, using the ranking of the ad provided by the ad-placers as the dependent variable.[19] The results from these regressions are virtually identical to the ones presented in the previous table, as displayed graphically in Appendix Figures C.1 and C.2.

Appendix Tables C.7 and C.8 present similar regressions, this time exploring the determinants of which ad is selected by a letter writer or by another ad-placer, or of the number of letters received by an ad-placer. In all these specifications, the importance of caste in the choice is at least as important as in the main specification. There are nevertheless interesting differences between these specifications and the ones presented here as far as the other variables are concerned, which we discuss in greater detail below.

4.3 Heterogeneity in preferences

The previous analysis suggests a strong horizontal preference for caste. This seem to hold across castes, and is not driven by the highest or lowest castes (results omitted).

To further explore whether there is a lot of heterogeneity among ad-placers, we allow for heterogeneity in the coefficient of horizontal preferences for castes in two ways. First, we estimate a hierarchical binary logit model, as suggested by Rossi et al. (2006). This estimation method allows for the coefficients of our binary choice model equation to differ across individuals but imposes that they (the coefficients) are drawn from a normal distribution. We allow the heterogeneity in responses to depend on a few characteristics of the ad-placer, namely his or her caste, age, height and predicted income and the default prior suggested by Rossi et al. (2006). Figure 1 presents the results of this estimation for the preference for marrying within caste obtained using 20000 Markov Chain Monte Carlo draws.[20]

The mean horizontal preference for caste is similar to what was estimated before (being of the same caste increases the probability of being called back by 15%), but the results suggest a considerable degree of heterogeneity in this coefficient. Around one-third of the sample appears to have no preference for marrying within caste, a figure comparable to the fraction of actual out-of-caste matches. The fraction is larger among women (40%).

Second, we estimated the parsimonious regression using a OLS model but letting every single ad-placer have his or her own coefficient for the variable "same caste". The distribution of coefficients was very similar to what the hierarchical binary logit model suggested (results omitted to save space): about 30% to 40% don't put a positive weight on marrying within caste.

[19] The sample size is a bit smaller due to missing observations (e.g., some ad-placers refused to provide ranking).
[20] The remaining estimates are available from the authors upon request.

4.4 Do these coefficients really reflect preferences?

We want to argue that these estimates provide us with information on the relative preferences for different attributes. There are however two other potential interpretations of these coefficients which we examine in detail here.

4.4.1 Strategic behavior

A first concern is that ad-placers may behave strategically when they choose to which letters they will respond. For example, they may prefer not to reply to a letter that appears to be "too good" because they think there is little chance of success. As we mentioned above, this is unlikely to be happening in this setting since the fact that the respondent has sent a letter to the ad-placer already signals his potential interest. Moreover, while the decision to reply or not to a letter may be strategic, the rank we ask them to give to the letter has no reason to be, since we ask them to judge the letter by how much they like it. The fact that the results using the rank closely mirror those using the decision to consider a letter or not is thus a good indication that their behavior is probably not strategic.

Nevertheless, we further investigate the issue here. Specifically, we study whether ad placers with certain characteristics are relatively less likely to reply to "good" ads. To do so, we first compute an absolute measure of "quality" of the letter, by regressing the probability that a letter in our sample is considered, without any interactions with characteristics of the ad-placer who received the letter. In other words, for P_{ij}, a dummy indicating whether letter j is considered by ad-placer i, we estimate the equation $P_{ij} = X_j\beta + \epsilon_{ij}$ without any fixed effect for the ad-placer. The quality indicator is then given by $Q_j = X_j\hat{\beta}$. We also predict the quality of the ad-placer, using the same coefficients $Q_i = X_i\hat{\beta}$. [21]

Figure 2 plots the probability of considering a letter based on the quality of the ad-placer and that of the letter for males and females. If the responses displayed strategic behavior or if the characteristics in the quality index were mostly "horizontal", we would expect that low quality ad-placers would be less likely to consider high quality letters. In fact, the figures show little difference in the relative probability of considering letters of different quality by the quintiles of quality of the ad-placer (particularly for females), although higher quality ad-placers appear to consider on average a smaller fraction of letters of all quality levels.[22]

Interestingly, the decision to respond to an ad (reported in the appendix tables C.8) seems to reflect more strategic behavior than the choice of whether to respond to a letter an ad-placer

[21] We exclude from this indicator measures of caste.

[22] To confirm that the exercise has power to detect behavior of "like write to like", we repeated the exercise using mostly "horizontal" characteristics in the quality index, although they may have some horizontal component as well, namely age, height and family origin. In this case, as expected, the slope increases with the "quality" (which is somewhat arbitrarily defined) of the add placer.

received: for example, more educated letter-writers do not receive more call-backs. Furthermore, when we regress the number of responses received on a polynomial function of our measure quality Q_i (computed as before), we find that the best fit between quality of an ad and the overall number of responses is an inverse-U shaped curve. This may indicate that, at the ad stage, higher quality ads are only replied to by people who stand a chance.

Thus, there is evidence that families behave strategically when they respond to ads, but not obviously at subsequent stages. This is perhaps not surprising, as they have to choose between a very large number of ads. While the average person sees more than 800 ads every Sunday over the 12 months they spend on the market before getting married, they only respond to on average 16 of these for females and 35 for males. In contrast, it appears that each ad-placer considers each of the 40 letters received during their search as a potential prospect, and therefore does not behave strategically about whom to respond to (ad-placers respond to about 30 percent of the letters they receive).[23]

4.4.2 What does caste signal?

One of our main empirical results is the fact that families (ad-placers as well as people who write to them) are much more likely to write to, and to follow up with, people from their own caste. Caste preferences thus display a strong horizontal component. Does this reflect a preference for caste in itself, or does caste signal something else?

We first explore the possibility that caste is a shortcut for the prospective spouse's background and culture.

Starting with background, while it is true that, in general, lower ranked castes have "worse" characteristics, there is a large amount of overlap. About 40 percent of individuals of the lowest ranked caste are more educated than the median Brahmin (among those reporting their education level). Similar statistics are obtained when looking at income, occupational scores and skin tones. There is thus little evidence, in this population, that caste is a perfect proxy for other attractive attributes of individuals.

Caste does not appear to proxy for culture either, as far as we can measure it: the strong preference for caste does not seem to be affected by controlling for a host of variables including cultural variables (e.g., ability to sing) and it remains very strong in regressions restricted to the four highest castes that are supposed to be culturally and economically more homogenous than the rest (Appendix Table C.9). It therefore does not appear that caste is just a proxy for cultural similarity. Furthermore, columns (3) and (8) of Table 3 also include a dummy variable for being from the same broad caste group. The results suggest that it is the narrow caste that matters for preference. If caste was a proxy for cultural identity, broad caste groupings should be

[23]This is less costly than an equilibrium where letter writers would send a message to most ads and would leave the ad-placers to strategically consider (or not) the letters received.

stronger than smaller groups. However, caste could still capture unobservables that are common to smaller caste groups but our setting does not allow us to decompose what makes matching within caste particularly attractive but we can exclude that it proxies for observable cultural characteristics mentioned in the letters.

A second possibility is the preference of ad-placers for letter writers who are from the same caste as themselves reflects the fact that, in equilibrium, only people with some bad unobservable characteristics write to people who are not in their castes (or who are above them or below them). Writing "out of caste" would then be a signal of bad quality.

We first look at whether people who write to, or receive letters from, people belonging to other castes are observationally different from those who do not. In columns 1 and 3 of Panel A in Table 4, we show the average quality index Q for ad placers who told us that they have responded to at least one letter from a caste that is below or above them, compared to the quality of those who only responded to people from their caste. Each cell is the difference in mean quality between those who satisfy the condition and those who do not. This table indicates that there does not seem to be significant observable differences between people who write to someone from a different caste and people who do not. There is also no difference between the people who receive letters from other castes and those who don't (panel B).

This still leaves open the possibility that these individuals are different along unobservable dimensions. However, we have an excellent measure of the unobservable (at the time of ad placing or letter writing) quality of people: we know their eventual outcome. We compute our quality index for each ad-placer's future spouse, and we contrast the eventual marriage outcomes of those who have written to at least one person from another caste to those of people who have only written to people within their caste. In an alternative specification, we also regress the quality of the eventual mate of an ad-placer on the share of ads they replied to that were not from the same caste. The results (presented in Columns 2 and 4 of Table 4) suggest that the ultimate marriage outcomes of those who write out of caste are no different than those who do not (panel A). Likewise, those who get letters from other castes eventually marry people of the same observable quality (panel B). This is a strong indication that writing out of caste does not send the signal that something is "wrong" with the ad-placer.

These results therefore suggest that the fact that ad-placers are more likely to follow up with people from their own caste reflects a true preference for eventually marrying within the same caste. This preference seems to be related to caste itself, rather than characteristics caste could be a proxy for. Compared to the other attributes, this preference also appears to be extremely strong: it appears that the parents of prospective grooms or brides would be willing to give up a lot to ensure that their child marries within their caste. Furthermore, the preference for caste appears to be strongly "horizontal" rather than "vertical,"as defined above in the theoretical section.

4.5 Do these preferences reflect dowry?

So far we have ignored dowries. We argued in the theory section that even if some people do eventually ask for dowries, the decision of who to write back to will be based on people's true (i.e, not dowry-based) preferences, as long as the cost of pursuing the option until the information on dowry – or other unobservable variables – is revealed, is not too high. One way to check the validity of this argument is to test one of its implications: those who either say that they do not want a dowry should be treated the same as others. To verify this conjecture in the data, we re-estimate the preferences in the sample of letters that explicitly mention not wanting a dowry. In Appendix Table C.10, we interact not wanting a dowry with each characteristic of the letter. The full specification is presented in columns (1) and (2), and the parsimonious specification is presented in columns (3) and (4).[24] The even columns correspond to the interaction terms and the odd columns to the main effect. The results are noisier for the interactions than for the main effects given the sample size, but, overall, we cannot reject the hypothesis that the interaction terms are jointly equal to zero. Interestingly, caste plays an even bigger role for this sample (the coefficient of the interaction between not wanting a dowry and being of the same caste is positive, although it is not significant), while the role of predicted income does not change. This suggests an even larger marginal rate of substitution between caste and income, which is the opposite of what would have been predicted if rich grooms were also thought to require higher dowries.

In addition, we find that ad-placers who either announce that they will not offer a dowry or state that they will not demand one do not receive systematically different numbers of letters, and their attributes as mentioned in the letter are valued similarly.[25]

5 Predicting observed matching patterns

Having established that strong horizontal caste preferences in our sample exist, we compute the set of stable matches implied by the preferences estimated to study the role of caste in equilibrium. A stable match is defined, following Gale and Shapley (1962), as a pairing where nobody who is matched would rather be with another partner who would also prefer being with them (see Hitsch et al. 2009 and Lee 2007 for other applications of this method to the marriage market). These simulated matches will then be used to answer questions regarding the equilibrium role of caste.

[24] We present these results only for the "bride-wanted" sample since only prospective grooms specify whether or not they will accept a dowry. No prospective bride is advertised as refusing to pay a dowry in the letters and a very small proportion do so in the ads.

[25] All these results are clearly context-specific and may not generalize to other settings or populations.

5.1 Empirical strategy

The pool of men and women attempting to match within this market is defined as the entire set of ads posted during the period of the survey, from October 2002 to March 2003 (most individuals on the market usually place one and only one ad, which makes this approximation acceptable).

We want to construct ordinal preferences over the entire set of bride (groom) wanted ads for each man (woman), in the sample. To do so we use the estimated parameters in equation (1) to construct the predicted "utility" that each man i in the sample (the set of ads) would get from matching with woman j (and vice versa for women) using the following equations.[26]

$$U_{ij}^{k} = \hat{\alpha}_k X_i + \hat{\beta}_k f\left(X_i, X_j\right) \text{ for } k = m, f \tag{3}$$

Functions U^m and U^f are then transformed into ordinal ranking such that

$$R_{ij}^{k} = n \quad if \quad \left\{ \begin{array}{c} U_{ij'}^{k} > U_{ij}^{k} > U_{i\tilde{j}}^{k} \\ and \quad R_{ij'}^{k} = n-1 \quad and \quad R_{i\tilde{j}}^{k} = n+1 \end{array} \right\} \text{ for } k = m, f.$$

The preference estimates for the results presented below were all obtained from the linear specification as presented in columns (1) and (6) of Table 3. However, very similar results were obtained using the logit specification or the ranking estimates as presented in Appendix Table C.6.[27]

Applying this methodology for all males and females in the sample generates a full set of ordinal preferences for each ad-placer with respect to all ad placers of the opposite gender. We continue to assume, as we did in the model, that remaining single is a worse option than being married to any spouse.

The Gale-Shapley algorithm can be computed in many ways. In most of the results presented in this section, we assume that men make an offer to women. We later explore how the results change when women propose to men instead. When men propose to women, the algorithm works as follows. All men first propose to their most highly-ranked women. Women consider all the offers they receive and select the best one (staying single is considered to be a worse option than any marriage). All men who haven't been retained then select their second choice. If a woman receives a new offer that is preferable to the one she is currently holding, she releases the old offer and this man must then propose to the next woman on his list. This continues until all

[26]The input required by the stable matching algorithm is a measure of ordinal and not cardinal utility, so fixed effects can be ignored. This is because the fixed-effect of male i, for example, simply affects the overall preference of person i towards all potential mates and not the relative ranking of each mate within his set of preferences.

[27]Note that this ignores the heterogeneity in preferences that we estimated. In appendix we also present the results using a draw from the random coefficients estimates instead of the OLS estimate. The results are qualitatively similar, though noisier, and seem to give us less good a fit of the actual matching patterns, probably reflecting the fact that the estimation demands a lot from the data.

men have been matched. Since they are the long side of the market, some women will remain single. Ties are broken randomly, without loss of generality in this setting (unlike the example discussed by Erdil and Ergin 2008).

In order to obtain confidence intervals for the results of the matching algorithm, 250 estimates of the parameter estimates of equation (2), α and β were obtained by bootstrapping the above estimation procedure.[28] Then, using each of the 250 sets of parameters, the matching algorithm was separately run. This resulted in 250 stable matches that define the range of outcomes that could stem from the distribution of preference parameters. All the stable matching results will present the 2.5th and 97.5th percentiles of each characteristic of interest to bound the range of results obtained.

In performing this exercise, we are thus assuming that the noise in the regression comes from measurement error, and is not "structural noise", i.e. does not represent genuine random utility taste shock in our sample. This is obviously an assumption, which is probably not entirely correct. In their main specification, Hitsch et al. (2009) take the opposite approach and assume that all the noise is structural noise. The difference in assumptions seems warranted by the different settings: in this case, we observe essentially everything the prospective groom or bride observes: there may thus be less scope for them to like a profile more than another. On the other hand, we ask people to recall who they actually responding to, or two tell us who they consider: this process may introduce a fair amount of measurement error. In Hitsch et al. (2009), some important dimensions of the choice may be unobserved by the econometrician (for example the picture, when provided, is ranked, but one can imagine that the person likes it differently than the graders that were used), but on the other hand, what is observed is user's decision to click, which has presumably much less error. To investigate this, in the last 3 columns of Appendix Table C.13, we present results of estimating the Gale-Shapley algorithm assuming the preference are estimated without error, but the noise is structural: in other words, for each simulation, we use the coefficients from the OLS specification, and we then add a random error term for each possible pair, drawn from the distribution of estimated residuals.

One may worry that the assumption of frictionless matching, implied by the Gale-Shapley algorithm, is inappropriate. To explore this issue, we introduce search frictions in the following way. First, we constrain males to contact individuals close to their unconstrained optimal choice (within 1000 ranks). This is a proxy for the value of their outside option as we now allow individuals to prefer remaining single than to marry a choice that is much below their reference point. Second, at every offer period, a man may be unable to offer to a particular woman with 75 percent probability and may thus be constrained to skip this woman and offer to the next preferred candidate. With search frictions, some males remain unmatched but without all find

[28]This was done using a "block bootstrap" by ad-placer, that is, either all letters in response to an ad are randomly selected into the sample or they are all excluded.

a spouse because they are on the short-side of the market. While this may appear ad-hoc, other versions were explored and none significantly changed our results.

Finally, to compare the results of the algorithm to those observed in the data, the summary statistics for the algorithm results are computed only for the individuals in our original interview-sample, though using the ad-placer samples gives very similar results.

5.2 Results

This section presents the stable matches estimated with the algorithm as described above. We then compare the simulated outcomes to the actual ones.

5.2.1 Who stays single?

The algorithm predicts who stays single and who gets married. While in this paper, we are more interested in who marries whom than in who stay single, in Appendix Table C.11, we show the mean differences in the value of key attributes between single and married females in the simulations and in the observed data, that is, the difference between the characteristics of single women and those who are married.

For women, the algorithm does an acceptable, but not stellar, job in predicting who stays single: In most cases, the differences between those who get married and those who stay single observed in the stable matching have the same signs as the actual differences. For seven out of the sixteen variables, the actual difference between single and married in our data lies within the confidence interval of the stables matches. In six more cases, the confidence intervals overlap. There are three variables for which the stable matching algorithm gets the sign wrong: whether a woman has a science degree, whether she has a degree from another field and her log wage. Overall a chi-square test of equivalence of the moments of the algorithm with the mean values observed in the actual match data rejected their equivalence. Introducing search friction does not change the results much.

Men are predicted to all marry without search frictions. With search friction, the algorithm performs somewhat better for men than for women: the signs are now congruent for all the variables, and the observed mean differences between those who stay single and those who get married fits within the 95 percent predicted by the stable matching algorithm in nine out of thirteen characteristics.

In most cases where the point estimate of the difference in the actual data does not lie within the bounds of the stable matches estimate, the stable matches overestimate the differences between the variable. This probably reflects the fact that factors other than these attributes eventually determine whether or not people decide to marry; this will thus dampen the role of the variable in the case of actual matches. The way we introduce search friction does not seem to fully capture this phenomenon.

5.2.2 Who marries whom?

We now compare the characteristics of the couples in the stable matches and in our actual sample. Table 5 displays the main results. Columns 1, 2 and 3 present mean, 2.5^{th} percentiles and the 97.5^{th} percentile for the stable matches, using the "considered" response to estimate the preferences while columns 4 to 6 present the actual comparison between ad-placers and the letters they consider. Columns 7 to 9 compare the ad-placers and their actual matches. All the differences are expressed in terms of the difference between the husband and the wife.

The stable matching algorithm predicts the characteristics of the couples reasonably well. For all the statistics we look at, the sample equivalent in the actual marriages fits within the range of the stable matches estimate in 14 cases out of 21, and the confidence intervals overlap in 15 cases, even though for many variables, the bounds on the stable matches are quite tight.[29]

Not surprisingly, a dominant feature is the tendency to marry within one's caste. The stable matching predicts that 83 to 99 percent of the couples will have the same caste. In practice, a lower share (almost 70 percent) of the couples are from the same caste. The simulations thus over-predict the fraction of same caste marriages. The main reason, as we will see below, appears to be that a good fraction of the out of caste marriages are "love" marriages (i.e. the groom and bride married on their own), which are not a direct reflection of the preference of the parents.

To the extent they do not marry within castes, in the simulations, the tendency is for men to marry down and women to marry up, as it is the case in actual matches but not as much in the considered pairings.

Turning to other characteristics, the predictions regarding age are roughly similar in the simulations and in the data. Husbands are almost six years older than their wives on average. Height differences are slightly underestimated but we predict too much assortative matching by height as given by the spousal heights correlation. Both the data and the simulations suggest that husbands are 10 to 12 centimeters taller than their wives.

For education, we correctly predict the fraction of couples with the same education level and the correlation between the education of the spouses, although we tend to predict that husbands will be less educated than their wives, and the opposite is true in the data. This is surprising. This might be due to the fact that men from the top of the educational distribution may be less likely to report their education than females as they can signal that quality using their wage/occupation.

Comparing our indices of quality, we find that males have higher indices than their spouses

[29] However, because the stable matching differs greatly from the actual matches in a few instances, a chi-square test of the algorithm moments and the mean values for either considered or match individual rejected the hypothesis of their equality.

though this measure is slightly overestimated compared to the observed data.[30] These indices are also positively correlated according to the algorithm and in reality.

The algorithm does not have much to say on predicted wage and income differences. This appears to stem from the fact that few women report their wage and income and that these variables are not part of the estimated preferences for males. Finally, we seem to severely overestimate the correlation in family origins.

Introducing search frictions slightly improves the fit of the algorithm result, but does not substantially change the results. The education and wage differences get smaller. Family origin matching is still overestimated when compared to the observed matches.

The last three columns of Appendix Table C.13 presents the results of the simulation where the noise is assumed to be structural. Not surprisingly, compared to our main specification, we find lower correlations between the horizontal characteristics: this is because the noise slightly scrambles the strong preference for castes (similarly, setting the random utility terms to zero in their paper increase the chance that black mates with black from 29% to 74% in Hitsch et al. (2009)). Some characteristics that are predicted to be "too similar" in our simulation are now within the range of observed matches; this is the case for family origins and education for example. For the caste similarity, while ignoring the noise leads to an overprediction of the number of same caste matches (presumably because some noise occurs at the time of matching), assuming all the noise is structural leads to an underprediction of the number of same caste matches: This probably reflects that, not surprisingly, the noise is partly structural and partly due to measurement error. Nevertheless, even with all the noise assumed to be structural, we still predict that 49% of the marriage would happen within the same caste, which once again reflects the strong preference for same caste marriage, which is not swamped by noise. [31]

We also computed the equilibrium under two variants, presented in Table C.12. First, we computed the equilibrium under the assumption that women propose rather than men. The equilibrium we obtain is very similar in terms of who marries whom: less than 2 percent of the matches differ between the two algorithms.[32] Furthermore, the characteristics of who remains single and who finds a match are almost identical when women propose and a very small number of women (less than 0.025 percent) are single when they propose and find a spouse when men propose (results omitted to save space). This suggests an almost unique stable matching. Finally, we also imposed a balanced sex ratio by randomly selecting a subset of females equal to the

[30]This is driven by two elements. First, male letter writers have higher response rates and thus the indices are larger for males than for females in general. Second, since women with lower quality indices are remaining single, the matches are such that there is an even larger difference between spouses.

[31]The first 3 columns of the same table shows the results of the Gale Shapley algorithm estimated where, in each draw, we pick a random draw in the preferences from the random coefficient estimation. The results are qualitatively similar, though the fit with the real matching pattern is generally less good, probably reflecting the fact that this estimation demands more from our data. Here again, we find 42% of in caste matches in equilibrium.

[32]This is similar to findings by Roth and Peranson (1999) in the context of medical residency matching and by Pathak and Sönmez (2008) in the context of Boston public school matching.

number of male ads in the sample. The results are again similar to the ones presented in the main tables. Finally, we also re-weighted the summary statistics using Kolkota's distribution of scheduled caste versus non-scheduled caste by educational category and obtained congruent results with the ones presented here (results omitted).

6 The role of caste preferences in equilibrium

6.1 Model Predictions

In Section 2, the theoretical model emphasized that the equilibrium role of caste crucially depends on whether preferences for caste are horizontal or vertical. Section 4 has then argued that the estimation of preferences suggests that the preference for caste is horizontal rather than vertical.

The theoretical model discussed above suggests that the impact of caste preferences also depends on whether the distribution of male and female "quality" is balanced across castes. In our sample, we know that there is a surplus of females given that more ad-placers are looking for a groom. However, is there evidence of a difference in the quality distribution across castes that differ by gender? To evaluate this question, we compared the distribution of the quality index Q measure defined above by caste for males and females among the interview and the letter samples. We find that the distributions are fairly balanced for all major caste groups (results omitted). Another indication that characteristics appear to be relatively balanced within caste is the fact that the share of couples that are caste-matched varies little when we introduce search frictions. Given the evidence of horizontal preferences and balance, we should expect a relatively limited impact of caste on the pattern of matches along other dimensions.

6.2 Simulations

What do the simulation results tell us about the actual role of caste in the matching equilibrium?

Table 6 takes one cut at this issue. The first columns reproduce columns 1 through 3 of the first panel of Table 5. The second set of columns constrains all marriages to take place within one's caste while the last ignores caste when computing the preference of each ad-placer for each prospective bride or groom.

The striking result in this table is that neither of these manipulations greatly affects how matches look along non-caste dimensions. As expected, the correlations in age, height and education increase as the preferences for caste diminish (they are the highest when matches are restricted to be within caste, and the lowest when preferences for caste are "shut down"), but the gradient is fairly low, and very few of the other variables are affected. [33] Moreover, the proportion of within-caste marriage falls by a large fraction when preferences are caste-blind. This suggests that caste does not proxy for other attributes. There are many potential matches for each person, both within and outside his or her caste.[34]

Overall, when we impose caste-blindness, the individuals marry almost identical individuals but from another caste. This would suggest that the equilibrium price of caste ought to be low. Indeed, in the data, there is no evidence that men or women who marry outside their caste sacrifice "quality" measured in a variety of ways. However, this could be due to selection: individuals who have less of a preference for caste could select to marry outside their caste, and not get any compensation in equilibrium, since they do not require one. We therefore turn to the matches generated by the algorithm to shed light on this question. The exercise consists in comparing the spouses of two observationally equivalent individuals where one is matched within (and above) his or her caste and the other is matched to an individual of a lower caste. To do this, for each iteration of the algorithm, we run a regression of various measures of the quality of the match on two indicators of whether the match is within and above one's caste (the omitted category is then being married to someone of a lower caste). Table 7 presents the mean and the 2.5 and 97.5 percentile of the distribution of the coefficients on whether one's partner is of a caste above one's own, in the first two columns, or of the same caste, in columns (3) and (4). The coefficients are small, insignificant, and often have the wrongly sign. For example, females who marry above their own caste are (although not statistically) more likely to marry more educated individuals than those marrying down in terms of caste and males who marry females of a caste above or equal to theirs are also more likely to marry a more beautiful woman than those marrying "down". This suggest that the price of keeping caste is zero, consistent with the model in the case where preferences are horizontal, and there is balance in terms of quality.

In contrast, our theory would make us expect that vertical characteristics, such as education, should have a positive price. We compute the equilibrium price of education in a similar fashion. The last columns of Table 7 consider the case where as opposed to caste, individuals are forced to choose between, for example, beauty and the educational level of a woman. A man who marries a woman who has more education also marries one who is older, less beautiful and darker-skinned. This suggests that this test has sufficient power to pick up the "price" of a vertical attribute.

[33]We repeated this exercise in the version of the algorithm where we also include draws from residuals and find the impact of these manipulations to be very similar to the ones presented here.

[34]While not reported, we repeated the exercise for each caste and found that despite the fact that smaller and lower ranked castes are less balanced than the major ones, the gradient is also fairly low for these sub-groups, indicating that our results are not entirely driven by the balancedness in our population.

7 Conclusion

Our results indicate that while caste is highly valued in terms of preferences, it does not require a very high price in equilibrium. This is consistent with assuming that preferences are relatively horizontal and that the populations are close to being balanced. Both these conditions appear to hold in the data we collected for arranged marriages in West Bengal. However, there are trends that suggest that caste-based preferences might be changing. Despite the value placed on caste and its low equilibrium price, 30 percent of people in our sample *do not* marry within their caste. In part, this is due to heterogeneity in caste preferences, with some people having caste-neutral preferences. But there is something else. About 40 percent of the sons and daughters of our respondents eventually marry through a channel other than the ads (e.g., through friends and family networks), and 20 percent enter into a "love marriage." This suggests that while economic forces have not been able to undermine the role of caste-based preferences on marriage market outcomes, these preferences themselves might be undergoing changes. What drives this is an interesting topic of future research.

References

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2007). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. Mimeo Harvard Business School.

Abramitzky, R., A. Delavande, and L. Vasconcelos (2011). Marrying up: The role of sex ratio in assortative matching. *American Economic Journal: Applied Economics 3*(3), 124–57.

Almenberg, J. and A. Dreber (2009). Lady and the trump: Status and wealth in the marriage market. *Kyklos 62*(2), 161–181.

Anderson, S. (2003). Why Dowry Payments Declined with Modernization in Europe but Are Rising in India. *The Journal of Political Economy 111*(2), 269–310.

Bargain, O., S. K. Bhaumik, M. Chakrabarty, and Z. Zhao (2007). Returns to education and earnings differences between Chinese and Indian wage earners. Unpublished manuscript.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Becker, G. S. (1973). A theory of marriage: Part I. *The Journal of Political Economy 81*(4), 813–46.

Bisin, A., G. Topa, and T. Verdier (2004). Religious intermarriage and socialization in the united states. *Journal of Political Economy 112*(3), 615–664.

Bisin, A. and T. Verdier (2000). "beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics 115*(3), 955–988.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97*(2), 298–319.

Bose, N. K. (1958). Some aspects of caste in Bengal. *The Journal of American Folklore 71*, 397–412.

Cole, H. L., G. J. Mailath, and A. Postlewaite (1992). Social norms, savings behavior, and growth. *The Journal of Political Economy 100*(6), 1092–1125.

Dugar, S., H. Bhattacharya, and D. Lucking-Reiley (fort). Can't Buy Me Love? A Field Experiment Exploring the Trade-Off Between Income and Caste-Status in an Indian Matrimonial Market. *Economic Inquiry*.

Erdil, A. and H. Ergin (2006). Two-sided matching with indifferences. Working Paper, Harvard Business School.

Erdil, A. and H. Ergin (2008). What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review 98*(3), 669–89.

Fernandez, R. (2003). Household formation, inequality, and the macroeconomy. *Journal of the European Economic Association 1*(2-3), 683–697.

Fernandez, R. and R. Rogerson (2001). Sorting and long-run inequality. *The Quarterly Journal of Economics 116*(4), 1305–1341.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics 121*(2), 673–697.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2008). Racial preferences in dating. *Review of Economic Studies 75*(1), 117–132.

Foster, A. and M. Rosenzweig (2001). Missing women, the marriage market and economic growth. Unpublished manuscript.

Fuller, C. J. (1996). *Caste today*. Oxford University Press, Delhi ; New York :.

Gale, D. and L. S. Shapley (1962). College admissions and the stability of marriage. *The American Mathematical Monthly 69*(1), 9–15.

Glinskaya, E. and M. Lokshin (2005). Wage differentials between the public and the private sectors in India. World Bank Policy Research Working Paper No. 3574.

Hitsch, G. J., A. Hortacsu, and D. Ariely (2009). Matching and Sorting in Online Dating Markets. *American Economic Review (Forthcoming)*.

Lam, D. (1988). Marriage markets and assortative mating with household public goods: Theoretical results and empirical implications. *The Journal of Human Resources 23*(4), 462–487.

Lee, S. (2007). Preferences and choice constraints in marital sorting: Evidence from Korea. Unpublished manuscript.

Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review 96*(4), 1225–1252.

Murphy, K. M. and R. H. Topel (1985). Estimation and Inference in Two-Step Econometric Models. *Journal of Business and Economic Statistics 3*(4), 370–79.

Pathak, P. A. and T. Sönmez (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review 98*(4).

Rao, V. (1993). The rising price of husbands: A hedonic analysis of dowry increases in rural india. *The Journal of Political Economy 101*(4), 666–677.

Risley, H. H. (1981). *The Tribes and Castes of Bengal*, Volume II. Firma Mukhopadhyay.

Rosenzweig, M. R. and O. Stark (1989). Consumption smoothing, migration, and marriage: Evidence from rural India. *The Journal of Political Economy 97*(4), 905–926.

Rossi, P., G. Allenby, and R. McCulloch (2006). *Bayesian Statistics and Marketing*. Wiley.

Roth, A. E. and E. Peranson (1999). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review 89*(4), 748–780.

8 Tables and figures

Table 1: **Summary statistics: Ad-placers**

Variable	**Ads placed by females**				**Ads placed by males**			
	Full set (N=14172)		**Interviewed (N=506)**		**Full set (N=8038)**		**Interviewed (N=277)**	
	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.
Number of responses			22.67	19.84			82.71	76.10
Caste								
Brahmin	0.26	0.44	0.26	0.44	0.27	0.44	0.25	0.44
Baidya	0.04	0.20	0.04	0.20	0.03	0.18	0.05	0.21
Kshatriya	0.02	0.13	0.02	0.13	0.02	0.13	0.01	0.12
Kayastha	0.30	0.46	0.35	0.48	0.29	0.45	0.32	0.47
Baisya and others	0.18	0.39	0.19	0.39	0.20	0.40	0.18	0.38
Sagdope and others	0.13	0.34	0.10	0.30	0.13	0.34	0.12	0.33
Other castes	0.02	0.14	0.02	0.13	0.02	0.12	0.03	0.16
Scheduled castes	0.06	0.23	0.03	0.16	0.05	0.21	0.04	0.20
Physical characteristics								
Age	26.68	3.90	26.59	3.65	31.58	4.31	32.14	4.45
Height (meters)	1.56	0.04	1.58	0.04	1.68	0.06	1.70	0.06
Skin tone	2.36	0.84	2.30	0.80				
Very beautiful	0.06	0.24	0.08	0.27				
Beautiful	0.56	0.50	0.44	0.50				
Education and Income								
Less than high school	0.03	0.16	0.02	0.15	0.01	0.12	0.01	0.08
High school	0.06	0.23	0.08	0.28	0.07	0.25	0.08	0.27
Post-secondary	0.01	0.10	0.00	0.04	0.03	0.18	0.04	0.20
College	0.46	0.50	0.49	0.50	0.36	0.48	0.35	0.48
Master's	0.29	0.45	0.26	0.44	0.17	0.37	0.15	0.36
PhD	0.06	0.24	0.05	0.22	0.13	0.34	0.18	0.39
Log wage	5.55	0.36	5.54	0.35	5.20	0.79	5.61	0.53
Log income	9.22	0.83	8.75	0.77	9.46	0.75	9.44	0.67
Geography								
Living in Calcutta	0.51	0.50	0.80	0.40	0.50	0.50	0.76	0.43
Family from West Bengal	0.44	0.50	0.39	0.49	0.45	0.50	0.39	0.49
Demands mentioned								
Only within caste	0.09	0.29	0.10	0.30	0.10	0.30	0.08	0.28
Caste no bar	0.31	0.46	0.33	0.47	0.26	0.44	0.24	0.43
No dowry demanded	0.03	0.16	0.02	0.12	0.12	0.32	0.10	0.31

Statistics are computed only among individuals reporting a given characteristic. Statistics on the number of ads which omitted given characteristics can be found in Appendix Table C.4

Table 2: **Summary statistics: Letters and matches**

Variables	**Ads placed by females**				**Ads placed by males**			
	Letters (N=5630)		**Matches (N=158)**		**Letters (N=3944)**		**Matches (N=131)**	
	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.
Considered	0.34	0.47			0.28	0.45		
Caste								
Brahmin	0.23	0.42	0.27	0.45	0.21	0.41	0.24	0.42
Baidya	0.03	0.17	0.04	0.19	0.04	0.19	0.05	0.23
Kshatriya	0.01	0.10	0.01	0.08	0.02	0.14	0.03	0.17
Kayastha	0.38	0.48	0.43	0.50	0.36	0.48	0.37	0.49
Baisya and others	0.20	0.40	0.15	0.36	0.20	0.40	0.16	0.37
Sagdope and others	0.12	0.32	0.07	0.26	0.11	0.32	0.11	0.31
Other castes	0.01	0.08	0.01	0.11	0.02	0.14	0.01	0.09
Scheduled castes	0.04	0.19	0.02	0.14	0.04	0.19	0.03	0.17
Same caste	0.66	0.47	0.68	0.47	0.64	0.48	0.72	0.45
Difference in caste	-0.17	1.37	0.10	1.43	-0.04	1.23	-0.11	1.08
Physical Characteristics								
Age	32.60	4.37	32.49	3.67	26.34	3.96	27.33	3.67
Age difference	6.25	2.92	6.61	2.95	5.93	2.65	4.60	2.84
Height (meters)	1.70	0.06	1.71	0.08	1.58	0.04	1.59	0.05
Height difference (m)	0.12	0.06	0.13	0.08	0.12	0.07	0.12	0.06
Skin tone					1.41	0.77		
Very beautiful					0.10	0.31		
Beautiful					0.51	0.50		
Education and Income								
Less than high school	0.00	0.06	0.00	0.00	0.02	0.12	0.01	0.09
High school	0.08	0.27	0.06	0.22	0.16	0.37	0.08	0.28
Post-secondary	0.04	0.19	0.03	0.16	0.00	0.06	0.02	0.12
College	0.51	0.50	0.35	0.48	0.58	0.49	0.44	0.50
Master's	0.21	0.41	0.25	0.44	0.18	0.39	0.34	0.48
PhD	0.13	0.33	0.32	0.47	0.02	0.13	0.11	0.32
Same education level	0.44	0.50	0.42	0.49	0.37	0.48	0.46	0.50
Male is more educated	0.28	0.45	0.45	0.50	0.44	0.50	0.23	0.42
Log wage	5.47	0.59	5.53	0.57	5.50	0.35	5.46	0.36
Log income	9.31	0.73	9.47	0.79	8.85	0.68	1.75	3.54
Geography								
Living in Calcutta	0.55	0.50	0.59	0.50	0.54	0.50	0.53	0.50
Same residence	0.50	0.50	0.64	0.49	0.44	0.50	0.42	0.50
Family from West Bengal	0.39	0.49	0.46	0.50	0.41	0.49	0.42	0.50
Same family origin	0.75	0.43	0.75	0.43	0.71	0.46	0.72	0.45
Demands mentioned								
No dowry demanded	0.07	0.26	0.00	0.00				

Statistics are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Statistics are computed only among individuals reporting a given characteristic (Statistics on the number of individuals who omitted given characteristics can be found in Appendix Table C.5). Ads placed by females (males) received letters by males (females): the first four columns refer to prospective and actual grooms, the last four to prospective and actual brides.

Table 3: **Probability of considering a letter**

	Ads placed by females					Ads placed by males				
	Basic (1)	No caste (2)	Main caste (3)	Limited (4)	Logit (5)	Basic (6)	No caste (7)	Main caste (8)	Limited (9)	Logit (10)
Same caste	0.1317***		0.1347**	0.1395***	0.8604***	0.1707***		0.1769***	0.1800***	1.0454***
	(0.0329)		(0.0425)	(0.0330)	(0.2068)	(0.0351)		(0.0442)	(0.0352)	(0.2052)
Same main caste			0.0273					-0.0331		
			(0.0485)					(0.0554)		
Diff. in caste*Higher caste male	-0.0119		-0.0276	-0.0108	-0.0788	-0.0175		-0.0099	-0.0138	-0.1990
	(0.0151)		(0.0197)	(0.0152)	(0.0928)	(0.0170)		(0.0232)	(0.0171)	(0.1081)
Diff. in caste*Lower caste male	0.0145		0.0056	0.0103	0.1393	-0.0399*		-0.0301	-0.0428*	-0.2958**
	(0.0133)		(0.0160)	(0.0134)	(0.0903)	(0.0172)		(0.0220)	(0.0173)	(0.0990)
Same caste*only within	0.0954		0.0918	0.0968	35.1982	0.1234		0.1217	0.1162	1.5756
	(0.1093)		(0.1093)	(0.1097)	(1288.88)	(0.1409)		(0.1410)	(0.1418)	(1.7103)
Diff. in caste*only within	-0.0163		-0.0158	-0.0188	-11.6502	0.0024		0.0010	-0.0056	0.0674
	(0.0400)		(0.0400)	(0.0402)	(429.6274)	(0.0596)		(0.0596)	(0.0597)	(0.6857)
Same caste*no bar	-0.0560		-0.0549	-0.0563	-0.4950*	-0.0565		-0.0574	-0.0629	-0.2599
	(0.0366)		(0.0366)	(0.0367)	(0.2187)	(0.0428)		(0.0429)	(0.0430)	(0.2424)
Diff. in caste*no bar	-0.0084		-0.0098	-0.0052	-0.0528	0.0121		0.0118	0.0115	0.1194
	(0.0121)		(0.0121)	(0.0121)	(0.0786)	(0.0151)		(0.0152)	(0.0152)	(0.0880)
Diff. in age	-0.0019	-0.0035	-0.0019	-0.0032	0.1647***	0.0443***	0.0471***	0.0443***	0.0394***	0.2933***
	(0.0047)	(0.0047)	(0.0047)	(0.0047)	(0.0458)	(0.0083)	(0.0083)	(0.0083)	(0.0082)	(0.0545)
Squared diff. in age	-0.0008**	-0.0008**	-0.0008**	-0.0008**	-0.0203***	-0.0023***	-0.0025***	-0.0023***	-0.0023***	-0.0150***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0035)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0038)
Diff. in height	1.2508***	1.3455***	1.2490***	1.3028***	8.1805***	0.7228*	0.6829*	0.7153*	0.7585*	10.2634***
	(0.2745)	(0.2754)	(0.2745)	(0.2752)	(1.7128)	(0.3329)	(0.3348)	(0.3331)	(0.3339)	(2.6758)
Squared diff. in height	-3.4695***	-3.8398***	-3.4465***	-3.5684***	-22.4174***	-6.2532***	-6.1518***	-6.2375***	-6.3265***	-60.1849***
	(0.9692)	(0.9718)	(0.9694)	(0.9709)	(5.9882)	(1.2451)	(1.2522)	(1.2455)	(1.2491)	(10.2198)
High school	0.0732	0.0907	0.0751		0.0770	0.1043	0.1133	0.1038		0.6122
	(0.1097)	(0.1102)	(0.1097)		(0.6478)	(0.0623)	(0.0628)	(0.0624)		(0.3896)
Post-secondary	0.1216	0.1413	0.1238		0.3391	0.0832	0.0701	0.0808		0.5283
	(0.1187)	(0.1192)	(0.1188)		(0.6995)	(0.1403)	(0.1409)	(0.1403)		(0.8193)
Bachelor's	0.1019	0.1132	0.1024		0.2708	0.0966	0.1224	0.0965		0.3744
	(0.1183)	(0.1188)	(0.1183)		(0.6942)	(0.0879)	(0.0884)	(0.0880)		(0.5294)
Master's	0.2242	0.2330	0.2245		0.9356	0.1679	0.1928*	0.1678		0.8527
	(0.1219)	(0.1224)	(0.1219)		(0.7154)	(0.0913)	(0.0918)	(0.0914)		(0.5464)
PhD	0.2589*	0.2636*	0.2595*		1.1708	0.2626*	0.2835**	0.2624*		1.6229**
	(0.1248)	(0.1254)	(0.1248)		(0.7319)	(0.1031)	(0.1035)	(0.1031)		(0.6068)
Same education	0.0412	0.0435	0.0413		0.2482	0.0174	0.0084	0.0173		0.0296
	(0.0239)	(0.0240)	(0.0239)		(0.1393)	(0.0307)	(0.0309)	(0.0307)		(0.1636)
Male more educated	0.0571	0.0646	0.0571		0.3556	-0.0057	-0.0098	-0.0057		-0.1400
	(0.0379)	(0.0381)	(0.0379)		(0.2166)	(0.0419)	(0.0422)	(0.0419)		(0.2352)
Non-rankable degree	0.2126	0.2371*	0.2140		0.8966	0.2125**	0.2201**	0.2123**		1.2286*
	(0.1143)	(0.1148)	(0.1143)		(0.6698)	(0.0822)	(0.0828)	(0.0823)		(0.4877)
Science	0.1002***	0.0951***	0.0999***		0.5945***	0.0456*	0.0423*	0.0457*		0.3074**
	(0.0214)	(0.0215)	(0.0214)		(0.1252)	(0.0192)	(0.0192)	(0.0192)		(0.1026)

Continued on next page

	Ads placed by females					Ads placed by males				
	Basic	No caste	Main caste	Limited	Logit	Basic	No caste	Main caste	Limited	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Commerce	0.0529*	0.0525*	0.0526*		0.3096*	0.0781**	0.0819**	0.0785**		0.4895***
	(0.0222)	(0.0223)	(0.0222)		(0.1312)	(0.0259)	(0.0260)	(0.0259)		(0.1379)
Other field	0.0332	0.0321	0.0326		0.2229	0.0154	0.0162	0.0153		-0.2174
	(0.0518)	(0.0521)	(0.0518)		(0.2774)	(0.0742)	(0.0741)	(0.0742)		(0.4218)
Calcutta	0.0734***	0.0771***	0.0735***	0.0757***	0.4089***	0.0620**	0.0588**	0.0621**	0.0591**	0.3915***
	(0.0137)	(0.0138)	(0.0138)	(0.0138)	(0.0777)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.1064)
Same location	0.0469	0.0445	0.0463	0.0412	0.2988	-0.0437	-0.0455	-0.0438	-0.0442	-0.1492
	(0.0352)	(0.0353)	(0.0352)	(0.0352)	(0.2060)	(0.0289)	(0.0290)	(0.0289)	(0.0290)	(0.1593)
Same family origin	0.0348	0.0513**	0.0351	0.0363	0.2641*	0.0926***	0.1067***	0.0932***	0.0977***	0.6472***
	(0.0194)	(0.0194)	(0.0194)	(0.0194)	(0.1127)	(0.0214)	(0.0214)	(0.0214)	(0.0215)	(0.1246)
Log income	0.0995***	0.0953***	0.0992***		0.6010***					
	(0.0148)	(0.0148)	(0.0148)		(0.0853)					
Log wage	0.1046***	0.1093***	0.1050***		0.5581***					
	(0.0144)	(0.0145)	(0.0144)		(0.0837)					
Skin tone						-0.0506***	-0.0518***	-0.0508***	-0.0534***	-0.3004***
						(0.0101)	(0.0102)	(0.0101)	(0.0101)	(0.0595)
Beautiful						0.0071	0.0100	0.0071	0.0043	0.0920
						(0.0190)	(0.0191)	(0.0190)	(0.0191)	(0.1035)
Very beautiful						0.0532	0.0575	0.0533	0.0465	0.3279*
						(0.0300)	(0.0301)	(0.0300)	(0.0301)	(0.1569)
Predicted income				0.3478***					0.0817***	
				(0.0194)					(0.0229)	
N	5628	5628	5628	5628	5628	3944	3944	3944	3944	3944
R-square/Log likelihood	0.3204	0.3084	0.3206		-2214.9462	0.2675	0.2522	0.2677		-1783.215
F-test/Chi2 of caste variables	4.96***		4.56***		87.80***	4.20***		3.81***		82.93***

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristic, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Ads placed by females (males) received letters by males (females): the first five columns refer to decisions made by females regarding prospective grooms, the last five to decisions made by males regarding prospective brides. Standard errors in parentheses.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 4: **Quality indices by caste categories**

	Ads placed by females			Ads placed by males		
	Own	Match	Share	Own	Match	Share
	(1)	(2)		(3)	(4)	
Panel A: By letters written by ad placers						
Any letter to caste above	0.0067	-0.0118	0.2558	-0.0360	-0.0122	0.3673
	(0.0147)	(0.0413)		(0.0365)	(0.0139)	
Any letter to caste below	-0.0072	-0.0526	0.3101	-0.0110	-0.0049	0.3673
	(0.0155)	(0.0382)		(0.0369)	(0.0207)	
N	123	37		41	23	
Panel B: By letters received by ad placers						
Any letter from caste above	-0.0101	0.0073	0.3981	0.0160	0.0255	0.5158
	(0.0066)	(0.0191)		(0.0111)	(0.0197)	
Any letter from caste below	0.0001	-0.0138*	0.5771	0.0163	0.0029	0.5860
	(0.0065)	(0.0066)		(0.0113)	(0.0067)	
N	285	158		526	131	

All cells correspond to a univariate regression of quality on a dummy variable indicating caste relationship. Standard errors in parentheses. Columns (1) and (3) refer to the quality of the ad-placer and columns (2) and (4) to the quality of the eventual match. Males (females) who place ads eventually marry females (males). Columns (2) and (3) are thus referring to quality of males while columns (1), (4) to quality of females.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 5: **Couples' characteristics, simulated and observed**

	Simulated			Observed-considered			Observed-matched		
	Mean	2.5 ptile	97.5 ptile	Mean	2.5 ptile	97.5 ptile	Mean	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Without search frictions									
Age diff.	**5.90**	**5.55**	**6.34**	5.90	5.82	5.99	5.70	5.35	6.05
Age corr.	**0.88**	**0.80**	**0.92**	0.83	0.81	0.85	0.65	0.57	0.73
Height diff.	**0.11**	**0.11**	**0.12**	0.12	0.12	0.12	0.12	0.11	0.13
Height corr.	0.86	0.81	0.90	0.38	0.35	0.42	0.39	0.29	0.49
Same caste	0.93	0.83	0.99	0.75	0.73	0.77	0.69	0.64	0.75
Caste diff.	**0.21**	**-0.02**	**0.64**	-0.09	-0.13	-0.05	0.01	-0.14	0.16
Caste corr.	**0.85**	**0.50**	**1.00**	0.85	0.82	0.87	0.76	0.69	0.83
Same education	**0.56**	**0.23**	**0.79**	0.45	0.43	0.47	0.44	0.38	0.50
Education diff.	-0.25	-0.50	0.00	0.34	0.31	0.38	0.29	0.14	0.44
Education corr.	**0.39**	**0.18**	**0.57**	0.42	0.38	0.46	0.36	0.24	0.47
Same family origin	1.00	0.99	1.00	0.78	0.77	0.80	0.76	0.71	0.82
Family origin diff.	**0.00**	**0.00**	**0.01**	0.01	-0.02	0.03	0.04	-0.02	0.11
Family origin corr.	1.00	0.98	1.00	0.54	0.50	0.58	0.51	0.39	0.64
Same residence	**0.77**	**0.30**	**1.00**	0.45	0.43	0.50	0.48	0.38	0.58
Location corr.	**0.24**	**-0.25**	**0.98**	0.04	-0.04	0.12	-0.06	-0.22	0.21
Log wage diff.	-0.33	-0.55	-0.12	0.14	0.08	0.19	0.25	0.13	0.36
Log wage corr.	**0.13**	**-0.21**	**0.43**	0.07	-0.07	0.20	0.19	-0.13	0.50
Income diff.	**20855**	**-10000**	**115501**	9277	-3842	22397	28374	-16	56764
Income corr.	**0.29**	**-1.00**	**1.00**	0.58	0.49	0.81	0.45	0.08	0.81
Quality diff.	*0.14*	*0.13*	*0.15*	0.10	0.10	0.11	0.12	0.11	0.13
Quality corr.	**0.27**	**0.10**	**0.44**	0.04	-0.24	0.34	0.20	0.07	0.32
Panel B: With search frictions									
Age diff.	**5.89**	**5.41**	**6.35**	5.90	5.82	5.99	5.70	5.35	6.05
Age corr.	**0.86**	**0.79**	**0.91**	0.83	0.81	0.85	0.65	0.57	0.73
Height diff.	**0.11**	**0.11**	**0.12**	0.12	0.12	0.12	0.12	0.11	0.13
Height corr.	0.83	0.78	0.88	0.38	0.35	0.42	0.39	0.29	0.49
Same caste	0.94	0.83	0.99	0.75	0.73	0.77	0.69	0.64	0.75
Caste diff.	**0.19**	**-0.01**	**0.60**	-0.09	-0.13	0.05	0.01	-0.14	0.16
Caste corr.	**0.86**	**0.52**	**1.00**	0.85	0.82	0.87	0.76	0.69	0.83
Same education	**0.55**	**0.21**	**0.78**	0.45	0.43	0.47	0.44	0.38	0.50
Education diff.	*-0.17*	*-0.45*	*0.15*	0.34	0.31	0.38	0.29	0.14	0.44
Education corr.	**0.40**	**0.17**	**0.60**	0.42	0.38	0.46	0.36	0.24	0.47
Same family origin	1.00	0.99	1.00	0.78	0.77	0.80	0.76	0.71	0.82
Family origin diff.	**0.00**	**-0.01**	**0.01**	0.01	-0.02	0.03	0.04	-0.02	0.11
Family origin corr.	0.99	0.97	1.00	0.54	0.50	0.58	0.51	0.39	0.64
Same residence	**0.72**	**0.00**	**1.00**	0.47	0.43	0.50	0.48	0.38	0.58
Location corr.	**0.17**	**-0.50**	**1.00**	0.04	-0.04	0.12	-0.06	-0.22	0.21
Log wage diff.	-0.25	-0.52	-0.03	0.14	0.08	0.19	0.25	0.13	0.36
Log wage corr.	**0.12**	**-0.24**	**0.48**	0.07	-0.07	0.20	0.19	-0.13	0.50
Income diff.	**19722**	**-11300**	**188000**	9277	-3842	22397	28374	-16	56764
Income corr.	**0.25**	**-1.00**	**1.00**	0.58	0.49	0.81	0.45	0.08	0.81
Quality diff.	*0.15*	*0.13*	*0.16*	0.10	0.10	0.11	0.12	0.11	0.13
Quality corr.	**0.26**	**0.02**	**0.45**	0.04	-0.24	0.34	0.20	0.07	0.32

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval.
Entries in italic have overlapping confidence intervals with the observed distribution.

Table 6: **Couples' characteristics and the impact of caste**

	Without restrictions			With forced caste matching			Caste-blinded		
	2.5 ptile (1)	Mean (2)	97.5 ptile (3)	2.5 ptile (4)	Mean (5)	97.5 ptile (6)	2.5 ptile (7)	Mean (8)	97.5 ptile (9)
Age diff.	5.21	5.77	6.30	5.47	5.92	6.45	5.57	5.97	6.46
Age corr.	0.77	0.86	0.92	0.79	0.87	0.92	0.89	0.93	0.96
Height diff.	0.10	0.11	0.12	0.11	0.11	0.12	0.10	0.11	0.12
Height corr.	0.78	0.84	0.89	0.81	0.85	0.89	0.89	0.93	0.95
Same caste	0.89	0.95	0.99	1.00	1.00	1.00	0.16	0.20	0.24
Same education	0.23	0.55	0.79	0.26	0.57	0.79	0.19	0.60	0.85
Education diff.	-0.44	-0.12	0.16	-0.53	-0.24	0.02	-0.60	-0.25	0.08
Education corr.	0.22	0.44	0.63	0.21	0.41	0.58	0.24	0.47	0.69
Log wage diff.	-0.39	-0.17	0.04	-0.54	-0.32	-0.10	-0.48	-0.29	-0.10
Log wage corr.	-0.17	0.14	0.48	-0.13	0.16	0.43	-0.08	0.19	0.43
Quality diff.	0.13	0.15	0.16	0.11	0.18	0.24	0.13	0.14	0.15
Quality corr.	0.05	0.24	0.42	0.00	0.51	0.89	0.06	0.29	0.48

Table 7: **Distribution of costs of...**

	Marrying a higher caste		**Marrying within caste**		**Education**	
	Male	Female	Male	Female	Male	Female
Education	0.0221	0.0601	0.0011	0.2718		
	[-2.0860, 2.0264]	[-1.6911, 1.7130]	[-1.8724, 1.8661]	[-0.8832, 0.9835]		
Height difference	0.0046	-0.0011	0.0018	-0.0047	0.0004	0.0020
	[-0.0635, 0.0685]	[-0.0500, 0.0450]	[-0.0581, 0.0588]	[-0.0284, 0.0192]	[-0.0047, 0.0050]	[-0.0021, 0.0071]
Age difference	-0.2376	-0.2140	-0.3602	0.3806	-0.2034	**-0.7158**
	[-3.1020, 2.3743]	[-2.9389, 1.9959]	[-3.1448, 2.2427]	[-0.8830, 2.0624]	[-0.5026, 0.0918]	**[-1.2118, -0.3242]**
Income	-4369.36		1755.83		-2264.11	
	[-83866.24, 51988.03]		[-75943.67, 54369.01]		[-20318.10, 11768.52]	
Wage	-0.0136	-0.0398	-0.0643	0.1620	**0.2081**	
	[-1.2465, 1.1745]	[-1.0386, 0.5833]	[-0.9269, 0.9749]	[-0.5564, 0.8919]	**[0.0874, 0.3176]**	
Very beautiful		0.0536		-0.0261		**-0.0565**
		[-0.3368, 0.9606]		[-0.3498, 0.2429]		**[-0.1199, -0.0093]**
Beautiful		0.0353		0.0499		-0.0241
		[-0.7927, 1.0335]		[-0.3736, 0.5953]		[-0.1198, 0.0716]
Skin tone		0.0430		-0.1472		0.0871
		[-1.0612, 1.9401]		[-0.7197, 0.5070]		[-0.0256, 0.2457]

The 2.5 and 97.5 percentile of the distribution of coefficients is presented in brackets. Bold entries mark significance at 5% or more. The reference category for the first 4 columns is marrying someone of a lower caste.

Figure 1: **Distribution of preferences for own caste**

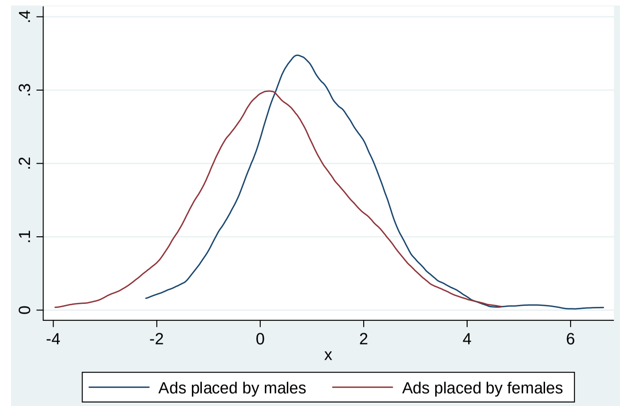
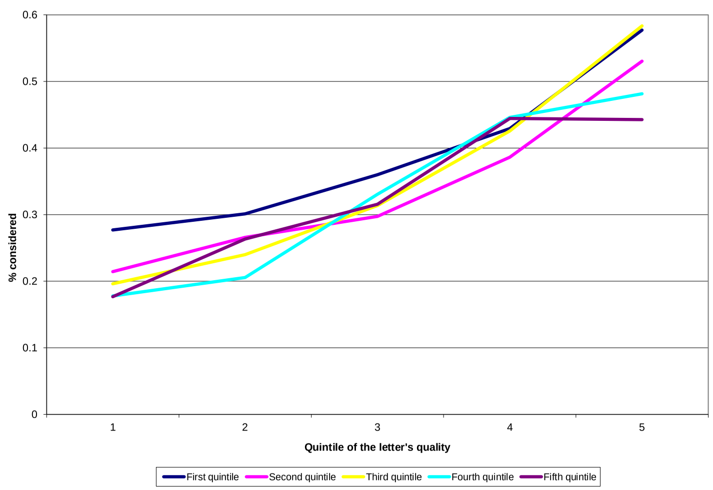
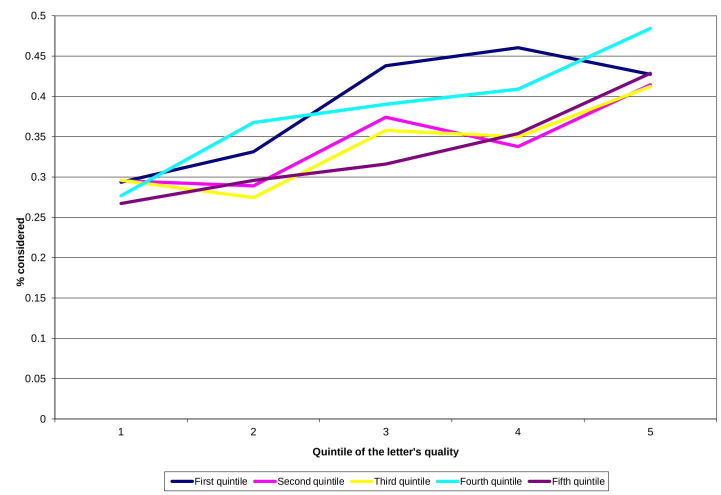


Figure 2: **Proportion of considered letters by quality of the letter and ad placer**



Ads placed by females



Ads placed by males

A Theoretical Appendix

A.1 Adding unobserved characteristics

This section proves that if exploration is not too costly, what individuals choose to be the set of options they explore reflects their true ordering over observables, even in the presence of an unobservable characteristic they may also care about.

Formally, we assume that in addition to the two characteristics already in our model, x and y, there is another (payoff-relevant) characteristic z (such as demand for dowry) not observed by the respondent that may be correlated with x. Is it a problem for our empirical analysis that the decision-maker can make inferences about z from their observation of x? The short answer, which this section briefly explains, is no, as long as the cost of exploration (upon which z is revealed) is low enough.

Suppose $z \in \{H, L\}$ with $H > L$ (say, the man is attractive or not). Let us modify the payoff of a woman of caste j and type y who is matched with a man of caste i and type (x, z) to $u^W(i, j, x, y) = A(j, i)f(x, y)z$. Let the conditional probability of z upon observing x, is denoted by $p(z|x)$. Given z is binary, $p(H|x) + p(L|x) = 1$. In that case, the expected payoff of this woman is:

$$A(j, i)f(x, y)p(H|x)H + A(j, i)f(x, y)p(L|x)L.$$

Suppose the choice is between two men of caste i whose characteristics are x' and x'' with $x'' > x'$. If x and z are independent (i.e., $p(z|x) = p(z)$ for $z = H, L$ for all x), or, x and z are positively correlated, then clearly the choice will be x''. Similarly, if it is costless to contact someone with type x'' and find out about z (both in terms of any direct cost, as well as indirect cost of losing out on the option x') the choice, once again, will be x'' independent of how (negatively) correlated x and z are.

More formally, for this simple case, suppose we allow x and z to be correlated in the following way: $p(H|x'') = p\mu, p(L|x'') = 1 - p\mu$, $p(H|x') = p$, and $p(L|x') = 1 - p$. If $\mu > 1$ we have positive correlation between z and x, if $\mu < 1$ we have negative correlation, and if $\mu = 1$, x and z are independent. Suppose exploring a single option costs c. Let us assume that $Hf(x', y) > Lf(x'', y)$ – otherwise, it is a dominant strategy to explore x'' only.

We consider two strategies. One is to explore only one of the two options and stick with the choice independent of the realization of z. The other is to explore both the options at first, and discard one of them later.

If the decision-maker explores both options, the choice will be x'' if either the z associated with it is H or if both x'' and x' have $z = L$ associated with them. Otherwise, the choice will be x'. The *ex ante* expected payoff from this strategy is

$$p\mu Hf(x'', y) + (1 - p\mu)[(1 - p)Lf(x'', y) + pHf(x', y)] - 2c.$$

This is obviously more than what he gets by exploring either one alone (namely, $f(x', y)\{pH + (1-p)L\} - c$ or $f(x'', y)\{p\mu H + (1-p\mu)L\} - c$) as long as c is small enough for any fixed value of $\mu > 0$.

Proposition 3. *For any fixed value of $\mu > 0$, so long as the exploration cost c is small enough, x'' will be chosen at the exploration stage whenever x' is chosen.*

In other words, as long as exploration is not too costly, what people choose to be the set of options to explore reflects their true ordering over the observables. In other words the indifference curve we infer from the "up or out" choices reflects their true preferences over the set of observables.

A.2 Omitted Proofs

A.2.1 Proof of Proposition 2

The fact that when $\beta \geq \beta_0$, all equilibria must have some non-assortative out-of-caste matching as long as condition **LCN** holds, follows from the previous proposition by virtue of the fact that **SB** was a possibility in our previous distributional assumption.

We now show that when $\beta < \beta_0$ and **SB** holds, cases (ii) and (iv) will be unstable and thus all equilibria will be assortative.

(ii): Clearly H1 must be **CC** in this case, otherwise he would deviated and matched with H2. But by **SB**, there must be another H1C type of the opposite sex who is in a X-H1 pair, where $X \neq H1$. But then the two H1 types should deviate and match with each other. This pair cannot be a part of a stable match.

(iv): For the pair H2-L2 and L1-H1 to be a stable match, one among H1 and H2 must be **CC**. Say H1 is **CC**. Then by **SB** there must exist another pair where a H1C who is in a H1-X pair where $X \neq H1$. This is not possible since the H1Cs would deviate and match. Now say the H2 is **CC** and H1 is not. Then H2 must prefer matching with a L2 to matching with a H1 (who would be willing to match with her). But there must be another H2C who is in a H2-X match where $X \neq H2$. Suppose $X = L2$. Then the two H2Cs should deviate and match. We know that X cannot be $H2$ by assumption. It cannot be $H1$ since from the two initial pairs, there is a H1N available and is not chosen. Then $X = L1$ but that is dominated by $H1$. Therefore the two H2Cs should deviate and match.

The final step of this part of the proof is to observe that H2-L2 and L2-H2 cannot co-exist since the H2s would immediately deviate. Hence all non-assortative matches must involve some H2-L1 and L1-H2 pairs and some either H2-L2 and L1-H2 pairs or L2-H2 and H2-L1 pairs.

To characterize the **APC** the fact that it is zero as long as $\beta < \beta_0$, follows from the fact that with only assortative matches everyone of a particular type matches the same type irrespective of whether they marry in caste or out of caste.

When $\beta \geq \beta_0$ there are non-assortative matches, but the type of possible non-assortative matches is quite restricted, as we saw above. Suppose there are $m \geq 0$ H2-L1 and L1-H2 pairs and $n \geq 0$ H2-L2 and L1-H2 pairs plus some number of assortative pairs. Since each pair contains two H2s, the total number of H2 females in assortative pairs is equal to the number of males. Since no H1 participates in a non-assortative pair, this is also true of H1s. By **SB** if there are $s \geq 0$ H1-H2 matches, there must also be exactly s H2-H1 matches.

However since we have an H2-L2 paired with an L1-H2, for each such pair there must be exactly one L2-L1 pair (therefore the number of L2 females in assortative matches exceeds the number of L2 males). Given that there are n H2-L2 and L1-H2 pairs this tell us that there must be at least n L2-L1 pairs. However if there are $n + t$ L2-L1 pairs there must be exactly t L1-L2 pairs.

So let the population consist of k H1-H1 matches, l H2-H2 matches, s H1-H2 matches, s H2-H1 matches.m H2-L1 and L1-H2 matches each, n H2-L2 and L1-H2 matches, p L1-L1 matches, q L2-L2 matches, $n + t$ L2-L1 matches and t L1-L2 matches. The H type woman who matches in or below caste matches with someone of average type $\frac{(k+l+s)H+mL}{k+l+s+n}$ as compared to $\frac{(k+l+s)H+(m+n)L}{k+l+s+m+n}$, for those who marry above or in caste. Since the former is larger the contribution of H types to the **APC** is positive.

Turning L type women, the average match of someone who matches in or below caste is $\frac{(m+n)H+(p+q+t)L}{m+n+p+q+t}$ while those who match above or in caste is L. Hence the L types also contribute positively to the **APC.** The **APC** for women is therefore positive. Similar (tedious) calculations show the same result for men.

B Data Appendix

Ads and letters provided very rich qualitative information that had to be coded to make the data analysis possible. We first coded caste, using the process described in the text.

Second, we coded information provided on education levels. Educational attainment was classified into seven categories: less than high school, high school completion, non-university post-secondary, bachelor's, master's, PhD or professional degree and non-classifiable degree.[35] In addition, we also coded, when available, the field in which the degree was obtained. We sorted these into four groups: humanities and social sciences (B.A, B.Ed, M.A, etc.), commerce (B.Com, MBA), science (B.Sc., B.Eng, M.Sc., etc.) and other fields (law, religion, etc.).

Third, we coded the available information on earning levels. When provided in the ad, self-reported earnings were converted into a monthly figure. This value will be referred to as "income." In addition, when the ad-placer or the letter writer provided his or her occupation,

[35] This last group mostly includes degrees in computer science from private institutions that were difficult to place within the existing ranking.

we used the National Sample Survey of India to construct an occupational score for the occupation (we refer to this below as "wage"). Note that prospective brides almost never report this information, and it will therefore be used only for the letters and ads from prospective grooms.

Fourth, we coded information on the origin of the family (East or West Bengal) and the current location of the prospective bride or groom under the following categories: Kolkata, Mumbai, other West Bengal, or other (mainly, abroad).[36]

Fifth, a very large fraction of ads from prospective brides specify physical characteristics of the women, using fairly uniform language and the same broad characteristics. Skin color was coded into four categories (from "extremely fair" to "dark") and we associate each category with a number from 1 to 4, with higher numbers representing darker skins. General beauty was divided into three categories ("very beautiful," "beautiful" and "decent-looking").

Finally, ads occasionally mention a multitude of other characteristics, such as "gotras" (a sub-group within one's caste based on lineage such that inter-marriages are ruled out under exogamy), astrological signs, blood type, family characteristics, personality traits, previous marital history, and specific demands. These were coded as well. However, each of these is rarely mentioned and so including or excluding them does not affect our results.

[36]At the time of Independence, the state of Bengal was partitioned into two states, one that remained in India, West Bengal, and the other that joined Pakistan, East Pakistan (which later became Bangladesh). Many Hindus migrated from East to West Bengal. There are some variations in terms of dialect, cultural and social norms among Bengalis depending on their family origin. This has some relevance in the arranged marriage market.

C Appendix tables

Table C.1: **Characteristics of ads by attrition status in second interviews**

Variable	Ads placed by females				Ads placed by males			
	Means		Difference		Means		Difference	
	Found	Not found	Mean	Sd. Error	Found	Not found	Mean	Sd. Error
Number of responses	23.004	18.000	5.00	4.65	79.874	89.071	-9.20	19.88
Caste								
Brahmin	0.27	0.21	0.06	0.10	0.25	0.29	-0.03	0.12
Baidya	0.04	0.16	**-0.12**	**0.05**	0.05	0.00	0.05	0.06
Kshatriya	0.02	0.00	0.02	0.03	0.02	0.00	0.02	0.03
Kayastha	0.35	0.21	0.14	0.11	0.31	0.36	-0.04	0.13
Baisya and others	0.19	0.21	-0.03	0.09	0.18	0.14	0.04	0.11
Sagdope and others	0.10	0.16	-0.06	0.07	0.12	0.14	-0.02	0.09
Other castes	0.02	0.00	0.02	0.03	0.02	0.07	-0.05	0.04
Scheduled castes	0.02	0.05	-0.03	0.04	0.05	0.00	0.05	0.06
Physical characteristics								
Age	26.55	27.67	-1.12	0.88	32.17	31.50	0.67	1.32
Height (meters)	1.58	1.59	-0.01	0.01	1.70	1.68	0.03	0.02
Skin tone	2.30	2.36	-0.06	0.22				
Very beautiful	0.08	0.20	*-0.12*	*0.07*				
Beautiful	0.44	0.53	-0.09	0.13				
Education and Income								
Less than high school	0.02	0.06	-0.03	0.04	0.01	0.00	0.01	0.03
High school	0.09	0.06	0.04	0.07	0.10	0.00	0.10	0.08
Post-secondary	0.00	0.00	0.00	0.01	0.06	0.00	0.06	0.06
College	0.53	0.50	0.03	0.12	0.42	0.46	-0.04	0.14
Master's	0.28	0.33	-0.05	0.11	0.18	0.23	-0.05	0.11
PhD	0.06	0.06	0.00	0.06	0.22	0.31	-0.09	0.12
Other degree	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.03
Humanities/Arts	0.57	0.75	-0.18	0.13	0.04	0.09	-0.05	0.07
Commerce	0.13	0.06	0.06	0.08	0.41	0.27	0.14	0.15
Science	0.30	0.19	0.11	0.12	0.55	0.64	-0.09	0.16
Other field	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Log wage	5.56	5.41	0.15	0.14	5.61	5.61	0.00	0.21
Log income	8.68	9.16	-0.48	0.60	9.45	9.22	0.23	0.39
Location								
Calcutta	0.82	0.60	0.22	0.18	0.78	0.40	*0.38*	*0.19*
West Bengali	0.39	0.40	-0.01	0.13	0.38	0.56	-0.17	0.17
Demands mentioned								
Only within caste	0.10	0.05	0.05	0.07	0.09	0.07	0.02	0.08
Caste no bar	0.32	0.42	-0.10	0.11	0.24	0.29	-0.05	0.12
No dowry demanded	0.01	0.05	-0.04	0.03	0.10	0.14	-0.04	0.08
Ads which omit...								
Caste	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.02
Age	0.01	0.05	-0.04	0.03	0.03	0.14	**-0.11**	**0.05**
Height	0.03	0.11	-0.07	0.04	0.11	0.14	-0.04	0.09
Education	0.08	0.05	0.03	0.06	0.19	0.07	0.12	0.11
Field	0.25	0.16	0.10	0.10	0.30	0.21	0.09	0.13
Residence	0.84	0.74	0.11	0.09	0.51	0.64	-0.13	0.14
Family origin	0.23	0.21	0.02	0.10	0.28	0.36	-0.08	0.12
Wage	0.85	0.63	**0.22**	**0.09**	0.57	0.50	0.07	0.14
Income	0.98	0.89	**0.08**	**0.04**	0.73	0.79	-0.05	0.12
Skin tone	0.21	0.26	-0.06	0.10				
Beauty	0.27	0.21	0.06	0.10				

Differences in italics are significant at 10%, those in bold, at 5%.

Table C.2: **Characteristics of ads who agreed and refused second round interview**

Variable	Ads placed by females				Ads placed by males			
	Means		**Difference**		**Means**		**Difference**	
	Agreed	Refused	Mean	Sd. Error	Agreed	Refused	Mean	Sd. Error
Number of responses	25.643	18.844	*6.80*	*3.51*	85.551	71.217	14.33	17.17
Caste								
Brahmin	0.25	0.25	0.00	0.08	0.23	0.36	-0.13	0.09
Baidya	0.04	0.06	-0.02	0.04	0.06	0.08	-0.02	0.05
Kshatriya	0.03	0.00	0.03	0.03	0.03	0.00	0.03	0.03
Kayastha	0.39	0.31	0.08	0.09	0.28	0.28	0.00	0.10
Baisya and others	0.18	0.16	0.03	0.07	0.21	0.12	0.09	0.09
Sagdope and others	0.07	0.16	-0.09	0.05	0.13	0.04	0.09	0.07
Other castes	0.02	0.03	-0.01	0.03	0.03	0.00	0.03	0.03
Scheduled castes	0.03	0.03	-0.01	0.03	0.02	0.12	**-0.10**	**0.04**
Physical characteristics								
Age	25.88	26.53	-0.65	0.60	31.92	32.45	-0.53	0.98
Height (meters)	1.58	1.59	-0.01	0.01	1.71	1.70	0.01	0.02
Skin tone	2.30	2.23	0.07	0.16				
Very beautiful	0.10	0.00	*0.10*	*0.06*				
Beautiful	0.42	0.58	-0.15	0.11				
Education and Income								
Less than high school	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.02
High school	0.10	0.03	0.06	0.06	0.10	0.05	0.05	0.07
Post-secondary	0.01	0.00	0.01	0.02	0.04	0.05	-0.01	0.05
College	0.51	0.53	-0.02	0.10	0.42	0.37	0.05	0.12
Master's	0.29	0.37	-0.08	0.09	0.22	0.16	0.07	0.10
PhD	0.07	0.07	0.00	0.05	0.20	0.37	*-0.17*	*0.10*
Other degree	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.02
Humanities/Arts	0.59	0.42	0.17	0.11	0.07	0.06	0.02	0.07
Commerce	0.13	0.27	*-0.14*	*0.08*	0.38	0.28	0.10	0.12
Science	0.28	0.31	-0.03	0.10	0.55	0.67	-0.12	0.13
Other field	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Log wage	5.53	5.73	*-0.21*	*0.12*	5.66	5.57	0.09	0.15
Log income	9.39	8.52	*0.87*	*0.28*	9.52	9.49	0.04	0.33
Location								
Calcutta	0.88	0.60	0.28	0.18	0.78	0.64	0.14	0.14
West Bengali	0.42	0.30	0.11	0.11	0.40	0.26	0.13	0.12
Demands mentioned								
Only within caste	0.09	0.09	0.00	0.06	0.08	0.04	0.04	0.06
Caste no bar	0.34	0.31	0.02	0.09	0.27	0.08	**0.19**	**0.09**
No dowry demanded	0.02	0.00	0.02	0.02	0.10	0.08	0.02	0.06
Ads which omit...								
Caste	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.02
Age	0.01	0.00	0.01	0.01	0.02	0.12	**-0.10**	**0.04**
Height	0.03	0.00	0.03	0.03	0.11	0.20	-0.09	0.07
Education	0.08	0.06	0.01	0.05	0.15	0.24	-0.09	0.08
Field	0.25	0.19	0.06	0.08	0.26	0.28	-0.02	0.10
Residence	0.84	0.84	0.00	0.07	0.51	0.56	-0.05	0.11
Family origin	0.24	0.28	-0.04	0.08	0.31	0.24	0.07	0.10
Wage	0.83	0.88	-0.05	0.07	0.54	0.44	0.10	0.11
Income	0.97	0.97	0.01	0.03	0.74	0.72	0.02	0.10
Skin tone	0.22	0.06	**0.16**	**0.08**				
Beauty	0.27	0.19	0.08	0.08				

Differences in italics are significant at 10%, those in bold, at 5% .

Table C.3: **Caste groupings**

	1. Brahmin	
Brahmin	Kshatriya Brahmin	Rudraja Brahmin*
Kulin Brahmin	Nath Brahmin	Baishnab Brahmin*
Sabitri Brahmin	Rajput Brahmin	Baishnab*
Debnath Brahmin	Gouriya Baishnab*	Nath*
Kanya Kubja Brahmin		
	2. Baidya	
Baidya	Lata Baidya	Kulin Baidya
Rajasree Baidya		
	3. Kshatriya	
Kshatriya	Ugra Kshatriya	Rajput (Solanki) Kshatriya
Poundra Kshatriya	Malla Kshatriya	Jana Kshatriya
Rajput Kshatriya	Barga Kshatriya	
	4. Kayastha	
Kayastha	Rajput Kayastha	Kayastha Karmakar
Kulin Kayastha	Pura Kayastha	Karmakar
Kshatriya Kayastha	Mitra Mustafi	Mitra Barujibi
Kshatriya Karmakar		
	5. Baisya and others	
Baisya	Suri	Teli
Baisya Saha	Suri Saha	Ekadash Teli
Baisya Ray	Rudra Paul	Dadash Teli
Baisya Kapali	Modak	Tili
Baisya Teli	Modak Moyra	Ekadash Tili
Rajasthani Baisya	Banik	Dsadah Tili
Barujibi	Gandha Banik	Marwari
Baisya Barujibi	Kangsha Banik	Malakar
Sutradhar	Khandagrami Subarna Banik	Tambuli
Baisya Sutradhar	Subarna Banik	Rajak
Tantubai	Shankha Banik	Kasari
Baisya Tantubai	Swarnakar	Baisya Tambuli
	6. Sadgope and others	
Sadgope	Yadav	Mahishya
Kulin Sadgope	Yadav Ghosh	Kumbhakar
Kshatriya Sadgope	Goyala	Satchasi
Yadav (Gope)	Gope	
	7. Other (mostly) non-scheduled castes	
Kaibarta	Rajak	Paramanik
Jele	Bauri	Jelia Kaibarta
Napit		
	8. (mostly) Scheduled castes	
Rajbanshi	Namasudra	Karan
Rajbanshi Kshatriya	Sagari	SC
Malo	Sudra	OBC
Mathra	Baisya Rajbanshi	

Table C.4: **Fraction of ad placers omitting given characteristics**

Variable	**Ads placed by females**		**Ads placed by males**	
	Full set (N=14172)	Interviewed (N=506)	Full set (N=8038)	Interviewed (N=277)
Caste	0.02	0.00	0.03	0.01
Age	0.01	0.01	0.02	0.04
Height	0.04	0.04	0.10	0.11
Education	0.10	0.08	0.22	0.18
Field	0.27	0.25	0.39	0.30
Residence	0.86	0.84	0.70	0.52
Family origin	0.29	0.23	0.32	0.29
Wage	0.83	0.84	0.25	0.57
Income	0.98	0.97	0.78	0.74
Skin tone	0.23	0.21		
Beauty	0.25	0.27		

Table C.5: **Fraction of letters and matches omitting given characteristics**

Variables	**Ads placed by females**		**Ads placed by males**	
	Letters (N=5630)	Matches (N=158)	Letters (N=3944)	Matches (N=131)
Caste	0.30	0.01	0.28	0.02
Age	0.04	0.00	0.03	0.00
Height	0.13	0.00	0.08	0.00
Education	0.08	0.00	0.04	0.00
Field	0.20	0.39	0.25	0.42
Residence	0.15	0.00	0.19	0.00
Family origin	0.31	0.03	0.27	0.00
Wage	0.44	0.08	0.86	0.79
Income	0.66	0.31	0.98	0.04
Skin tone			0.14	1.00
Beauty			0.36	1.00

Table C.6: **Rank of the letter**

	Ads placed by females					Ads placed by males				
	Basic	No caste	Main caste	Limited	Logit	Basic	No caste	Main caste	Limited	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Same caste	1.2797***		1.1275**	1.3320***	0.4314***	1.2591***		1.5022***	1.4068***	0.3595***
	(0.2933)		(0.3821)	(0.4171)	(0.0928)	(0.3458)		(0.4292)	(0.4153)	(0.0928)
Same main caste			0.2377					-0.4295		
			(0.3825)					(0.4490)		
Diff. in caste*Higher caste male	-0.0500		-0.0179	-0.0176	-0.0034	-0.4707**		-0.5472**	-0.3732	-0.1421**
	(0.1341)		(0.1437)	(0.1615)	(0.0418)	(0.1699)		(0.1878)	(0.2379)	(0.0461)
Diff. in caste*Lower caste male	0.1070		0.0767	0.0784	0.0281	-0.3310		-0.2548	-0.3634	-0.0976*
	(0.1183)		(0.1280)	(0.1742)	(0.0372)	(0.1705)		(0.1882)	(0.2093)	(0.0458)
Same caste*only within	1.1726		1.1737	1.1669	0.2128	2.1112		2.0985	2.1664	0.7029
	(0.9116)		(0.9117)	(1.0856)	(0.2848)	(1.3256)		(1.3257)	(1.4721)	(0.3674)
Diff. in caste*only within	-0.4459		-0.4471	-0.4554	-0.1670	0.0183		0.0094	-0.1336	0.0874
	(0.3334)		(0.3334)	(0.4955)	(0.1117)	(0.5781)		(0.5782)	(0.6173)	(0.1582)
Same caste*no bar	-0.8681**		-0.8678**	-0.8606*	-0.2911**	-0.8599*		-0.8912*	-0.9364	-0.2521*
	(0.3258)		(0.3258)	(0.3821)	(0.1028)	(0.4315)		(0.4328)	(0.5233)	(0.1156)
Diff. in caste*no bar	-0.1021		-0.1041	-0.0832	-0.0247	0.2092		0.2020	0.1752	0.0734
	(0.1071)		(0.1072)	(0.1240)	(0.0342)	(0.1521)		(0.1523)	(0.2094)	(0.0409)
Diff. in age	0.0345	0.0255	0.0348	0.0217	0.0053	0.5215***	0.5411***	0.5205***	0.4459*	0.1457***
	(0.0405)	(0.0405)	(0.0405)	(0.0856)	(0.0127)	(0.0816)	(0.0820)	(0.0816)	(0.1978)	(0.0218)
Squared diff. in age	-0.0114***	-0.0115***	-0.0114***	-0.0110*	-0.0031***	-0.0284***	-0.0291***	-0.0282***	-0.0263*	-0.0079***
	(0.0023)	(0.0023)	(0.0023)	(0.0054)	(0.0007)	(0.0057)	(0.0057)	(0.0057)	(0.0117)	(0.0015)
Diff. in height	9.5137***	9.8711***	9.4794***	9.8330**	3.5492***	7.2790*	6.8472*	7.2231*	7.6700*	1.9194*
	(2.5694)	(2.5757)	(2.5701)	(3.3196)	(0.8651)	(3.2304)	(3.2517)	(3.2309)	(3.5366)	(0.8796)
Squared diff. in height	-24.5037**	-26.3139**	-24.4011**	-25.3370	-9.5136**	-69.0103***	-68.9625***	-68.8785***	-70.3327***	-18.7289***
	(9.2415)	(9.2562)	(9.2436)	(13.7834)	(3.2019)	(12.3135)	(12.3931)	(12.3145)	(13.6789)	(3.3576)
High school	0.6719	0.9189	0.6811		0.3796	1.7107**	1.7634**	1.7049**		0.4798**
	(0.9403)	(0.9438)	(0.9405)		(0.3366)	(0.6092)	(0.6140)	(0.6092)		(0.1709)
Post-secondary	1.3963	1.7144	1.4059		0.5588	2.5003	2.3729	2.4921		0.6638
	(1.0262)	(1.0290)	(1.0264)		(0.3629)	(1.4645)	(1.4709)	(1.4645)		(0.3922)
Bachelor's	1.4920	1.7376	1.4965		0.6384	2.7817**	2.9152**	2.7961**		0.7474**
	(1.0213)	(1.0243)	(1.0214)		(0.3635)	(0.8894)	(0.8959)	(0.8896)		(0.2434)
Master's	2.3654*	2.6088*	2.3650*		0.9383*	3.9425***	4.0203***	3.9590***		1.0457***
	(1.0533)	(1.0564)	(1.0534)		(0.3739)	(0.9236)	(0.9303)	(0.9237)		(0.2527)
PhD	2.6963*	2.9129**	2.6967*		1.0487**	4.2363***	4.2562***	4.2333***		1.2354***

Continued on next page

	Ads placed by females					Ads placed by males				
	Basic	No caste	Main caste	Limited	Logit	Basic	No caste	Main caste	Limited	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	(1.0810)	(1.0842)	(1.0811)		(0.3828)	(1.0650)	(1.0720)	(1.0650)		(0.2918)
Same education	0.5329*	0.5361*	0.5340*		0.1369*	0.2423	0.1380	0.2433		0.0577
	(0.2091)	(0.2100)	(0.2092)		(0.0662)	(0.2995)	(0.3013)	(0.2995)		(0.0803)
Male more educated	0.8218*	0.8550*	0.8256*		0.2317*	0.3416	0.2331	0.3442		0.0886
	(0.3315)	(0.3327)	(0.3316)		(0.1065)	(0.4169)	(0.4194)	(0.4169)		(0.1120)
Non-rankable degree	1.8538	2.1751*	1.8618		0.7512*	2.6315**	2.6192**	2.6275**		0.7227**
	(0.9855)	(0.9886)	(0.9857)		(0.3497)	(0.8065)	(0.8122)	(0.8065)		(0.2225)
Science	1.0444***	0.9810***	1.0454***		0.3522***	0.7039***	0.6512***	0.7092***		0.2050***
	(0.1882)	(0.1887)	(0.1882)		(0.0600)	(0.1928)	(0.1931)	(0.1929)		(0.0516)
Commerce	0.3640	0.3573	0.3646		0.1096	1.1107***	1.1203***	1.1076***		0.3257***
	(0.1948)	(0.1956)	(0.1948)		(0.0622)	(0.2600)	(0.2612)	(0.2600)		(0.0698)
Other field	0.1361	0.1378	0.1388		0.0921	1.1653	1.2332	1.1686		0.3351
	(0.4631)	(0.4654)	(0.4632)		(0.1476)	(0.7950)	(0.7994)	(0.7950)		(0.2213)
Calcutta	0.4690***	0.4953***	0.4703***	0.4923**	0.1738***	0.6515***	0.6240**	0.6501***	0.6288**	0.1741***
	(0.1204)	(0.1206)	(0.1205)	(0.1744)	(0.0383)	(0.1891)	(0.1897)	(0.1891)	(0.2223)	(0.0509)
Same location	0.4846	0.4160	0.4831	0.4059	0.1181	-0.1912	-0.2096	-0.1944	-0.2074	-0.0551
	(0.3086)	(0.3097)	(0.3086)	(0.3802)	(0.0959)	(0.2876)	(0.2893)	(0.2877)	(0.3468)	(0.0777)
Same family origin	0.2665	0.3861*	0.2770	0.2767	0.0712	0.7190***	0.8573***	0.7150***	0.8038**	0.1903**
	(0.1710)	(0.1710)	(0.1710)	(0.1963)	(0.0538)	(0.2156)	(0.2163)	(0.2156)	(0.2646)	(0.0580)
Log income	0.8761***	0.8254***	0.8782***		0.2906***					
	(0.1310)	(0.1308)	(0.1310)		(0.0431)					
Log wage	0.9205***	0.9451***	0.9221***		0.2988***					
	(0.1258)	(0.1262)	(0.1259)		(0.0397)					
Skin tone						-0.4585***	-0.4657***	-0.4581***	-0.4994***	-0.1292***
						(0.1005)	(0.1012)	(0.1005)	(0.1531)	(0.0271)
Beautiful						0.2045	0.2127	0.2095	0.1766	0.0404
						(0.1885)	(0.1893)	(0.1885)	(0.2272)	(0.0505)
Very beautiful						0.5376	0.5587	0.5363	0.4247	0.1614*
						(0.2934)	(0.2951)	(0.2934)	(0.3056)	(0.0787)
Predicted income				3.2509*					0.8807	
				(1.3526)					(2.2178)	
N	5094	5094	5094	5094	5094	3520	3520	3520	3520	3520

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Ads placed by females (males) received letters by males (females): the first five columns refer to decisions made by females regarding prospective grooms, the last five to decisions made by males regarding prospective brides. Standard errors in parentheses.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.7: **Probability of writing to a particular ad**

	Ads placed by females				Ads placed by males			
	Ad placer selection		Respondent selection		Ad placer selection		Respondent selection	
	LP	Logit	LP	Logit	LP	Logit	LP	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same caste	0.0206***	3.4296***	0.1080***	2.1627***	0.0319***	2.3853***	0.1956***	2.2002***
	(0.0013)	(0.3504)	(0.0022)	(0.0672)	(0.0014)	(0.2043)	(0.0049)	(0.0895)
Diff. in caste*Higher caste male	-0.0013	-1.7058	0.0001	0.0609*	-0.0004	0.2302	0.0236***	0.5106***
	(0.0014)	(1.1849)	(0.0009)	(0.0308)	(0.0013)	(0.3532)	(0.0016)	(0.0353)
Diff. in caste*Lower caste male	-0.0011	-2.0820	-0.0092***	-0.3236***	-0.0020	-0.7402*	0.0014	-0.0809*
	(0.0014)	(1.1721)	(0.0007)	(0.0254)	(0.0012)	(0.3519)	(0.0018)	(0.0380)
Same caste*only within	0.0029	13.0267			-0.0059	14.5443		
	(0.0038)	(770.0985)			(0.0033)	(984.4139)		
Diff. in caste*only within	0.0004	-0.0170			0.0011	0.2650		
	(0.0008)	(368.9421)			(0.0007)	(324.9982)		
Same caste*no bar	-0.0046**	-1.4258***			-0.0010	-0.4298		
	(0.0015)	(0.3972)			(0.0016)	(0.2442)		
Diff. in caste*no bar	-0.0003	-0.1701			0.0007	0.3169**		
	(0.0003)	(0.1420)			(0.0004)	(0.1003)		
Diff. in age	0.0003***	0.2974***	0.0042***	0.4822***	0.0005***	0.4746***	0.0085***	0.6196***
	(0.0001)	(0.0562)	(0.0002)	(0.0158)	(0.0002)	(0.0546)	(0.0005)	(0.0228)
Squared diff. in age	-0.0000***	-0.0234***	-0.0005***	-0.0395***	-0.0000***	-0.0398***	-0.0005***	-0.0484***
	(0.0000)	(0.0043)	(0.0000)	(0.0011)	(0.0000)	(0.0044)	(0.0000)	(0.0017)
Diff. in height	0.0435**	17.6596**	0.3241***	13.3879***	0.0452***	9.7321***	0.3539***	6.0564***
	(0.0167)	(5.9477)	(0.0256)	(1.0314)	(0.0099)	(2.0036)	(0.0413)	(0.8609)
Squared diff. in height	-0.1922***	-75.6526***	-1.2001***	-50.3339***	-0.2013***	-43.4930***	-1.9223***	-32.4783***
	(0.0528)	(20.1851)	(0.0747)	(3.3084)	(0.0414)	(8.3431)	(0.1723)	(3.8381)
High school	0.0013	0.7340	0.0176***	0.4294***	-0.0001	13.1424	-0.0135	-0.1717
	(0.0022)	(0.8006)	(0.0040)	(0.1206)	(0.0029)	(702.6814)	(0.0098)	(0.2239)
Post-secondary	-0.0010	0.2473	-0.0159*	-0.7547**	0.0020	14.0290	0.0117	-0.1526
	(0.0035)	(1.0634)	(0.0065)	(0.2810)	(0.0033)	(702.6813)	(0.0118)	(0.2490)
Bachelor's	-0.0006	0.1855	-0.0115***	-0.2506*	-0.0017	13.2529	-0.0360***	-0.6465**
	(0.0021)	(0.7795)	(0.0035)	(0.1125)	(0.0029)	(702.6813)	(0.0095)	(0.2180)
Master's	0.0024	0.8934	-0.0101*	-0.1507	0.0034	13.9488	-0.0378***	-0.7335**
	(0.0023)	(0.8084)	(0.0039)	(0.1256)	(0.0033)	(702.6813)	(0.0109)	(0.2379)
PhD	-0.0005	0.3537	-0.0151***	-0.1832	0.0048	14.0380	-0.0229*	-0.5667*
	(0.0027)	(0.8864)	(0.0045)	(0.1425)	(0.0035)	(702.6813)	(0.0111)	(0.2423)
Same education	0.0022	0.5264	0.0191***	0.5524***	0.0032*	0.7805**	0.0448***	0.8407***
	(0.0012)	(0.2759)	(0.0019)	(0.0575)	(0.0013)	(0.2434)	(0.0047)	(0.0864)
Male more educated	0.0016	0.4578	0.0014	0.0406	0.0021	0.5918	0.0324***	0.7051***
	(0.0016)	(0.4240)	(0.0030)	(0.0915)	(0.0020)	(0.3213)	(0.0062)	(0.1133)
Non-rankable degree	-0.0031	-13.2632	-0.0242*	-0.5629	-0.0018	13.2663	-0.0534	-0.5984
	(0.0131)	(4420.5696)	(0.0098)	(0.4140)	(0.0049)	(702.6816)	(0.0281)	(0.4275)
Science	0.0004	0.0622	-0.0013	0.0553	0.0022	0.2396	-0.0084	-0.0976
	(0.0008)	(0.1794)	(0.0013)	(0.0395)	(0.0012)	(0.1661)	(0.0055)	(0.0939)
Commerce	0.0009	0.2188	0.0013	0.0450	-0.0015	-0.3376	-0.0186***	-0.2452**
	(0.0012)	(0.2561)	(0.0018)	(0.0539)	(0.0013)	(0.1743)	(0.0055)	(0.0945)
Other field	0.0013	0.0839	-0.0053	-0.0701	0.0085**	1.0443**	-0.0602***	-0.5009
	(0.0035)	(0.7779)	(0.0066)	(0.1701)	(0.0032)	(0.3378)	(0.0178)	(0.2599)
Calcutta	0.0097***	1.7482***	-0.0043	-0.1346	0.0097***	1.1826***	0.0062	0.0029
	(0.0017)	(0.4223)	(0.0038)	(0.1150)	(0.0012)	(0.1721)	(0.0049)	(0.0871)
Same location	-0.0007	0.0442	0.0051	0.2150*	-0.0051	-0.4259	0.0088	0.1428
	(0.0026)	(0.5239)	(0.0029)	(0.0889)	(0.0032)	(0.4468)	(0.0046)	(0.0822)
Same family origin	0.0053***	1.3955***	0.0194***	0.4990***	0.0058***	0.8628***	0.0259***	0.3742***
	(0.0008)	(0.2287)	(0.0012)	(0.0364)	(0.0009)	(0.1545)	(0.0027)	(0.0463)
Log income					0.0024**	0.2556*	0.0044	-0.0708
					(0.0009)	(0.1187)	(0.0037)	(0.0683)
Log wage					0.0041***	0.8576***	0.0010	0.0260
					(0.0005)	(0.1070)	(0.0020)	(0.0352)
Skin tone	-0.0012**	-0.3719**	-0.0033***	-0.0927***				
	(0.0004)	(0.1179)	(0.0007)	(0.0219)				
Beautiful	-0.0011	-0.2338	0.0016	0.0264				
	(0.0007)	(0.1671)	(0.0012)	(0.0369)				
Very beautiful	0.0008	0.0304	0.0047	0.0523				
	(0.0015)	(0.3025)	(0.0024)	(0.0683)				
N	49025	49025	147546	144543	70337	69617	53043	52407

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the respondent/ad placer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the respondent/ad placer and a dummy for both individuals not providing caste, age, height, education, location and family origin. Ads placed by females (males) received letters by males (females): the first four columns refer to decisions made by males regarding which ad placed by females they should write to, the last four to decisions made by females regarding which ads placed by males they should contact. Standard errors in parentheses. * significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.8: **Number of responses received to an ad**

	Ads placed by females		Ads placed by males	
	OLS	Poisson	OLS	Poisson
	(1)	(2)	(3)	(4)
Baidya	0.0199	1.4363	-0.4018***	-32.5365
	(0.0554)	(4.5688)	(0.0387)	(22.6938)
Kshatriya	-0.3880***	-6.4094	-0.4774***	-32.4609
	(0.1017)	(7.0018)	(0.0746)	(38.5897)
Kayastha	0.1941***	4.8539*	0.1565***	14.8425
	(0.0242)	(2.2215)	(0.0176)	(12.0916)
Baisya	-0.2298***	-4.2818	-0.0679**	-6.3319
	(0.0313)	(2.5611)	(0.0214)	(13.7648)
Sagdope	-0.0900*	-2.0499	-0.0344	-3.5924
	(0.0360)	(3.2275)	(0.0253)	(15.8213)
Other non-scheduled castes	-0.5491***	-8.1897	-0.6427***	-28.3260
	(0.1107)	(7.2236)	(0.0673)	(30.0856)
Scheduled castes	-0.0659	-1.2732	-0.5098***	-39.0446
	(0.0670)	(5.5995)	(0.0421)	(23.3959)
Age	-0.0401***	-0.8096**	0.0119***	0.8895
	(0.0031)	(0.2490)	(0.0016)	(1.0717)
Height	1.5551***	35.4319	-0.4142***	-17.6774
	(0.2196)	(19.5507)	(0.1239)	(79.5235)
High school	-0.1107	-1.8582	0.8501***	19.0770
	(0.0761)	(6.5589)	(0.1762)	(55.5553)
Post-secondary	-0.4580	-10.6578	1.6886***	82.9122
	(0.2403)	(20.2488)	(0.1781)	(61.3144)
Bachelor's	-0.0769	-1.2923	1.5513***	67.2765
	(0.0774)	(6.7409)	(0.1756)	(56.9136)
Master's	-0.1423	-2.8572	1.8182***	89.1902
	(0.0808)	(7.0390)	(0.1768)	(58.7970)
PhD/Professional degrees	-0.2741**	-5.4127	1.7035***	77.3746
	(0.0926)	(7.8143)	(0.1767)	(58.3160)
Non-rankable degree	-1.0200***	-14.9420	1.2666***	40.0588
	(0.1777)	(10.7632)	(0.1896)	(69.6573)
Science	0.0463	1.2457	0.2546***	22.4205
	(0.0253)	(2.2666)	(0.0421)	(26.3598)
Commerce	-0.0520	-1.1006	-0.0265	-1.1862
	(0.0346)	(3.0170)	(0.0433)	(26.8366)
Other field	-0.6742*	-5.9297		
	(0.2846)	(14.3313)		
Calcutta	0.4087***	8.6102	0.1608***	20.7122
	(0.0684)	(5.3780)	(0.0164)	(13.4021)
From West Bengal	0.1941***	4.6963*	0.4275***	29.7894
	(0.0228)	(2.0787)	(0.0271)	(15.4041)
Log income			-0.2129***	-16.0723
			(0.0180)	(11.4682)
Log wage			0.0190	3.6086
			(0.0200)	(13.2790)
Skin tone	-0.2570***	-5.1665***		
	(0.0166)	(1.2562)		
Very beautiful	0.2804***	9.0867*		
	(0.0369)	(3.8408)		
Beautiful	0.0147	0.3033		
	(0.0243)	(2.1623)		
N	5788	5788	4075	4075

Standard errors in parentheses. All regressions include dummies indicating non-response for each characteristics. *significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.9: **Responses for letters, top four castes only**

	Ads placed by females			**Ads placed by males**		
	Considered-OLS	Considered-Logit	Rank	Considered-OLS	Considered-Logit	Rank
	(1)	(2)	(3)	(4)	(5)	(6)
Same caste	0.1636***	0.8372***	1.6650***	0.1047*	0.6521**	0.9490*
	(0.0408)	(0.2017)	(0.3041)	(0.0503)	(0.2180)	(0.4200)
Diff. in caste	0.0203	-0.0389	0.2100	-0.0307	0.1188	-0.6039**
	(0.0157)	(0.0862)	(0.1274)	(0.0204)	(0.0989)	(0.1996)
Same caste*only within	0.2760		4.0097*	0.2206		2.5592
	(0.2504)		(1.6520)	(0.1946)		(1.5047)
Diff. in caste*only within	0.1630		1.5846**	0.0173		-0.2654
	(0.0907)		(0.6090)	(0.0827)		(0.6165)
Same caste*no bar	-0.1214		-1.4500**	-0.0283		-0.4768
	(0.0774)		(0.4943)	(0.0868)		(0.7489)
Diff. in caste*no bar	-0.0013		-0.0133	-0.0526		-0.2027
	(0.0301)		(0.1612)	(0.0347)		(0.2678)
Diff. in age	0.0086	0.1785*	0.0384	0.0424**	0.2239**	0.5249***
	(0.0115)	(0.0824)	(0.0551)	(0.0138)	(0.0783)	(0.0941)
Squared diff. in age	-0.0021**	-0.0237***	-0.0124***	-0.0016	-0.0075	-0.0296***
	(0.0008)	(0.0061)	(0.0034)	(0.0010)	(0.0054)	(0.0064)
Diff. in height	1.7176***	11.5875***	12.8167***	0.4528	9.9158*	6.4163
	(0.4304)	(2.7654)	(2.9819)	(0.5064)	(4.2931)	(3.8687)
Squared diff. in height	-4.7533**	-32.3551***	-36.7084***	-5.5546**	-57.2542***	-69.2712***
	(1.5071)	(9.5394)	(10.5597)	(1.8509)	(16.0106)	(14.5440)
High school	0.0893	-0.3359	0.3344	0.1458	0.6317	2.3437**
	(0.2058)	(1.0614)	(1.0421)	(0.1319)	(0.8511)	(0.7957)
Post-secondary	0.1455	-0.0292	0.9657	1.0020		2.8634
	(0.2204)	(1.1724)	(1.1656)	(0.7954)		(1.7153)
Bachelor's	0.0994	-0.1983	0.9457	0.1373	0.3398	2.8282*
	(0.2228)	(1.1747)	(1.1653)	(0.1754)	(1.0892)	(1.1618)
Master's	0.2457	0.6397	1.7441	0.2074	0.7712	3.9660***
	(0.2286)	(1.2091)	(1.2018)	(0.1799)	(1.1094)	(1.1982)
PhD	0.3103	0.9926	1.9778	0.3754*	2.0243	5.6290***
	(0.2335)	(1.2364)	(1.2347)	(0.1875)	(1.1387)	(1.3764)
Same education	0.0698	0.3108	0.5517*	0.0544	0.2778	0.1380
	(0.0400)	(0.2295)	(0.2502)	(0.0516)	(0.2602)	(0.3726)
Male more educated	0.0683	0.3453	1.1132**	-0.0048	-0.1850	0.2927
	(0.0642)	(0.3564)	(0.3964)	(0.0727)	(0.3859)	(0.5242)
Non-rankable degree	0.2176	0.5038	1.6034	0.3889*	1.8667	3.6022***
	(0.2114)	(1.0908)	(1.0982)	(0.1595)	(0.9668)	(1.0440)
Science	0.1027**	0.6910***	1.1189***	0.0266	0.2026	0.4503
	(0.0339)	(0.1962)	(0.2215)	(0.0320)	(0.1624)	(0.2406)
Commerce	0.0690	0.4884*	0.2930	0.0442	0.2986	0.8302*
	(0.0356)	(0.2064)	(0.2310)	(0.0411)	(0.2131)	(0.3260)
Other field	-0.0211	0.2345	0.1823	0.0806	-0.0493	0.4942
	(0.0953)	(0.5211)	(0.5432)	(0.1210)	(0.7079)	(1.0121)
Calcutta	0.0363	0.2345	0.4769***	0.0472	0.2776	0.6114**
	(0.0224)	(0.1239)	(0.1432)	(0.0318)	(0.1689)	(0.2353)
Same location	0.1162*	0.7043*	0.9203*	-0.0082	-0.0137	-0.1505
	(0.0576)	(0.3370)	(0.3757)	(0.0489)	(0.2607)	(0.3615)
Same family origin	0.0121	0.1294	0.1625	0.0969**	0.6508***	0.9472***

Continued on next page

	Ads placed by females			Ads placed by males		
	Considered-OLS	Considered-Logit	Rank	Considered-OLS	Considered-Logit	Rank
	(1)	(2)	(3)	(4)	(5)	(6)
	(0.0311)	(0.1733)	(0.2085)	(0.0344)	(0.1945)	(0.2728)
Log income	0.1254***	0.2514*	1.0116***			
	(0.0222)	(0.1185)	(0.1564)			
Log wage	0.1176***	0.4247**	0.9331***			
	(0.0235)	(0.1306)	(0.1528)			
Skin tone				-0.0343*	-0.2055*	-0.5198***
				(0.0171)	(0.0927)	(0.1261)
Beautiful				0.0214	0.1621	0.0731
				(0.0313)	(0.1644)	(0.2377)
Very beautiful				0.0472	0.4497	0.5465
				(0.0527)	(0.2594)	(0.3878)
N	2295	2045	2191	3944	1474	3570

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Standard errors in parentheses. Ads placed by females (males) received letters by males (females): the first three columns refer to decisions made by females regarding prospective grooms, the last three to decisions made by males regarding prospective brides.

* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.10: **Dowries and probability of being considered**

	Full Regression		Parsimonious	
	Main effects in sample that does not mention dowries	Interaction of characteristics with no request for dowry	Main effects in sample that does not mention dowries	Interaction of characteristics with no request for dowry
	(1)	(2)	(3)	(4)
Same caste	0.0836**	0.1363	0.0887***	0.1971
	(0.0264)	(0.1080)	(0.0265)	(0.1070)
Diff. in caste*Higher caste male	0.0128	0.0089	0.0144	-0.0170
	(0.0143)	(0.0463)	(0.0144)	(0.0454)
Diff. in caste*Lower caste male	-0.0258*	0.0801	-0.0243	0.1018*
	(0.0124)	(0.0458)	(0.0124)	(0.0450)
Diff. in age	-0.0025	0.0031	-0.0040	0.0110
	(0.0049)	(0.0190)	(0.0049)	(0.0188)
Squared diff. in age	-0.0008**	-0.0001	-0.0008**	-0.0006
	(0.0003)	(0.0014)	(0.0003)	(0.0014)
Diff. in height	1.3842***	-1.9984	1.4127***	-2.1377*
	(0.2817)	(1.0405)	(0.2822)	(1.0249)
Squared diff. in height	-3.9449***	6.9149	-3.9571***	8.1506*
	(0.9871)	(3.6745)	(0.9880)	(3.5935)
High school	0.0776	-0.1167		
	(0.1100)	(0.1386)		
Post-secondary	0.1334	-0.2867		
	(0.1191)	(0.2939)		
Bachelor's	0.1239	-0.3886		
	(0.1187)	(0.2535)		
Master's	0.2513*	-0.4281		
	(0.1225)	(0.2641)		
PhD	0.2923*	-0.6111*		
	(0.1254)	(0.2697)		
Same education	0.0421	-0.3778		
	(0.0242)	(0.0638)		
Male more educated	0.0515	0.0639		
	(0.0383)	0.0882		
Non-rankable degree	0.2018			
	(0.1149)			
Science	0.0961***	0.0377		
	(0.0222)	(0.0809)		
Commerce	0.0467*	0.0654		
	(0.0232)	(0.0827)		
Other field	0.0232	0.0253		
	(0.0526)	(0.3418)		
Calcutta	0.0886***	0.1042*	0.0821***	-0.0916
	(0.0158)	(0.0482)	(0.0143)	(0.0520)
Same location	0.0792***	-0.0945	0.0442	0.0179
	(0.0143)	(0.0533)	(0.0358)	(0.0953)
Same family origin	0.0500	0.0535	0.0440*	-0.0142*
	(0.0358)	(0.0977)	(0.0199)	(0.0570)
Log income	0.0422*	-0.1274*		
	(0.0198)	(0.0583)		
Log wage	0.1084***	-0.0160		
	(0.0149)	(0.0565)		
Predicted income			0.3490***	0.0018
			(0.0198)	(0.0747)
No dowry	-0.3008		0.1042	
	(0.5804)		(0.7096)	
F-test: Same coefficients		1.24		1.34
N	5056		5056	

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Columns (1) and (2) represent the coefficients of a single regression. Columns (3) and (4) also represent a single regression. The main effects of each characteristics in the sample that does not mention dowries is presented in columns (1) and (3). The coefficients in columns (2) and (4) correspond to the coefficient of the interaction term between the letter stating that it has no dowry demand and each characteristic. Ads placed by females received letters by males: this table refers to decisions made by females regarding prospective grooms. Standard errors in parentheses.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.11: **Difference in individuals' characteristics by marital status**

	Simulated		Observed		
	2.5 ptile	97.5 ptile	Mean	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)	(5)
	Panel A: Women, without search frictions				
Age	*1.0111*	*2.7490*	0.8976	0.3009	1.5290
Height	*-0.0240*	*-0.0044*	-0.0034	-0.0114	0.0044
Caste	**-0.1459**	**1.6301**	0.0760	-0.2741	0.4090
Education level	-1.0636	-0.5316	-0.1557	-0.3510	0.0492
Arts and Social Science	*0.0921*	*0.3320*	0.0162	-0.0902	0.1186
Commerce	*-0.1795*	*-0.0650*	-0.0414	-0.1142	0.0330
Science	*-0.2495*	*-0.0146*	0.0274	-0.0674	0.1162
Other field	**-0.0158**	**0.0306**	-0.0022	-0.0214	0.0132
From West Bengal	**-0.1552**	**0.0305**	-0.0076	-0.1019	0.0879
Kolkota	*-0.4872*	*-0.1024*	-0.0287	-0.2095	0.1357
Skin rank	0.4543	0.8081	0.0209	-0.1364	0.1726
Very beautiful	**-0.0837**	**0.0167**	-0.0118	-0.0681	0.0417
Beautiful	**-0.2614**	**0.0103**	-0.0158	-0.1172	0.0867
Income	**-10223**	**7805**	-6274	-10950	-1310
Log wage	**-0.1012**	**0.0740**	0.0028	-0.1314	0.1356
"Quality"	-0.1106	-0.0811	-0.0051	-0.0190	0.0091
	Panel B: Women, with search frictions				
Age	**0.6167**	**2.2867**	0.8976	0.3009	1.5290
Height	*-0.0221*	*-0.0074*	-0.0034	-0.0114	0.0044
Caste	**-0.0406**	**1.6823**	0.0760	-0.2741	0.4090
Education level	-0.9793	-0.4801	-0.1557	-0.3510	0.0492
Arts and Social Science	*0.1004*	*0.3413*	0.0162	-0.0902	0.1186
Commerce	*-0.2081*	*-0.0737*	-0.0414	-0.1142	0.0330
Science	*-0.2395*	*-0.0357*	0.0274	-0.0674	0.1162
Other field	**-0.0248**	**0.0294**	-0.0022	-0.0214	0.0132
From West Bengal	**-0.1567**	**0.0486**	-0.0076	-0.1019	0.0879
Kolkota	*-0.3968*	*-0.1018*	-0.0287	-0.2095	0.1357
Skin rank	0.4204	0.7431	0.0209	-0.1364	0.1726
Very beautiful	**-0.1021**	**0.0084**	-0.0118	-0.0681	0.0417
Beautiful	**-0.2493**	**0.0441**	-0.0158	-0.1172	0.0867
Income	*-1347*	*6925*	-6274	-10950	-1310
Log wage	**-0.1473**	**0.0676**	0.0028	-0.1314	0.1356
"Quality"	-0.1042	-0.0743	-0.0051	-0.0190	0.0091
	Panel C: Men, with search frictions				
Age	**-1.3211**	**0.5338**	0.4442	-0.7127	1.5818
Height	**-0.0245**	**0.0213**	-0.0037	-0.0194	0.0129
Caste	**-2.0046**	**0.1326**	-0.1284	-0.6251	0.3746
Education level	*-1.2875*	*-0.2950*	-0.1665	-0.5052	0.2118
Arts and Social Science	**-0.0755**	**0.1353**	-0.0692	-0.1162	-0.0264
Commerce	**-0.1400**	**0.4701**	0.1212	-0.0224	0.2598
Science	**-0.5747**	**0.0294**	-0.0520	-0.1977	0.0987
Other field	**-0.0136**	**0.0708**	0.0000	0.0000	0.0000
Family origin	**-0.4124**	**0.2000**	0.0173	-0.1250	0.1613
Calcutta	**-0.4754**	**0.2105**	0.0358	-0.1176	0.1738
Income	*-9135*	*-1158*	-12683	-41640	1333
Log wage	*-0.8809*	*-0.1959*	-0.1174	-0.3229	0.0741
"Quality"	-0.1388	-0.0442	-0.0193	-0.0428	0.0057

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval. Entries in italic have overlapping confidence intervals with the observed distribution.

Table C.12: **Couples' characteristics, variances of the algorithm**

	Women propose			**Balanced sex ratio**		
	Mean	2.5 ptile	97.5 ptile	Mean	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)	(5)	(6)
Age diff.	**6.06**	**5.63**	**6.61**	*5.15*	*4.86*	*5.40*
Age corr.	**0.89**	**0.82**	**0.94**	**0.84**	**0.75**	**0.91**
Height diff.	**0.11**	**0.11**	**0.12**	**0.12**	**0.11**	**0.13**
Height corr.	0.84	0.77	0.90	0.81	0.75	0.86
Same caste	0.90	0.75	0.99	0.90	0.81	0.95
Caste diff.	**0.36**	**-0.02**	**1.13**	**0.04**	**-0.05**	**0.17**
Caste corr.	**0.78**	**0.30**	**1.00**	**0.87**	**0.58**	**0.99**
Same education	**0.54**	**0.21**	**0.83**	**0.59**	**0.30**	**0.79**
Education diff.	-0.21	-0.53	0.09	**0.10**	**-0.17**	**0.32**
Education corr.	**0.44**	**0.17**	**0.68**	**0.48**	**0.27**	**0.67**
Same family origin	1.00	0.99	1.00	0.99	0.97	1.00
Family origin diff.	**-0.00**	**-0.01**	**0.01**	**0.00**	**-0.01**	**0.02**
Family origin corr.	0.99	0.97	1.00	0.99	0.93	1.00
Same residence	**0.72**	**0.00**	**1.00**	**0.76**	**0.2**	**1.00**
Location corr.	**0.36**	**-0.35**	**1.00**	**0.28**	**-0.29**	**1.00**
Log wage diff.	-0.14	-0.35	0.06	-0.44	-0.67	-0.23
Log wage corr.	**0.08**	**-0.18**	**0.36**	**0.16**	**-0.14**	**0.41**
Income diff.	**-54700**	**-692003**	**14000**	**-6802**	**-46000**	**23000**
Income corr.	**0.11**	**-1.00**	**1.00**	**-0.20**	**-1.00**	**1.00**
Quality diff.	0.17	0.16	0.18	0.18	0.17	0.19
Quality corr.	**0.23**	**0.07**	**0.41**	**0.36**	**0.14**	**0.52**

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval. Entries in italic have overlapping confidence intervals with the observed distribution.

Table C.13: **Couples' characteristics, variances of the algorithm**

	Heterogeneous coefficients			**With residuals**		
	Mean	2.5 ptile	97.5 ptile	Mean	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)	(5)	(6)
Age diff.	4.61	4.10	5.17	**5.84**	**5.37**	**6.28**
Age corr.	0.14	-0.08	0.34	**0.48**	**0.27**	**0.65**
Height diff.	**0.12**	**0.11**	**0.13**	**0.11**	**0.11**	**0.12**
Height corr.	0.00	-0.10	0.12	**0.35**	**0.21**	**0.49**
Same caste	0.42	0.36	0.48	0.49	0.34	0.63
Caste diff.	*-0.33*	*-0.54*	*-0.10*	**0.28**	**-0.14**	**0.63**
Caste corr.	0.27	0.09	0.43	*0.49*	*0.15*	*0.75*
Same education	*0.32*	*0.22*	*0.39*	**0.37**	**0.27**	**0.46**
Education diff.	*0.09*	*-0.11*	*0.28*	-0.25	-0.44	-0.05
Education corr.	0.01	-0.13	0.15	*0.08*	*-0.06*	*0.24*
Same family origin	**0.68**	**0.59**	**0.78**	**0.77**	**0.67**	**0.86**
Family origin diff.	**0.04**	**-0.05**	**0.13**	**0.01**	**-0.06**	**0.09**
Family origin corr.	**0.36**	**0.17**	**0.56**	**0.52**	**0.33**	**0.70**
Same residence	**0.50**	**0.29**	**0.69**	**0.50**	**0.28**	**0.72**
Location corr.	**0.04**	**-0.32**	**0.39**	**0.01**	**-0.35**	**0.58**
Log wage diff.	-0.51	-0.78	-0.28	-0.42	-0.67	-0.14
Log wage corr.	**0.01**	**-0.27**	**0.36**	**0.01**	**-0.30**	**0.39**
Income diff.	**12467**	**-6000**	**75600**	**17610**	**-18833**	**166166**
Income corr.	**0.06**	**-1.00**	**1.00**	**0.07**	**-1.00**	**1.00**
Quality diff.	0.17	0.15	0.18	*0.14*	*0.13*	*0.16*
Quality corr.	**0.01**	**-0.10**	**0.13**	**0.07**	**-0.04**	**0.17**

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval. Entries in italic have overlapping confidence intervals with the observed distribution.

Figure C.1: **Correlations between coefficients of the considered and rank regressions, ads placed by females**

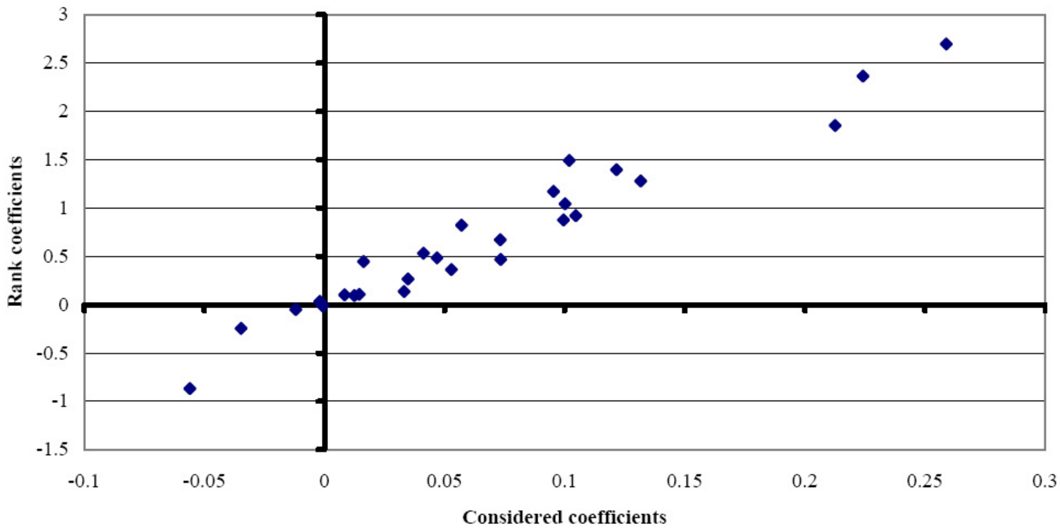


Figure C.2: **Correlations between coefficients of the considered and rank regressions, ads placed by males**