Essays on Matching, Marriage and Human Capital Accumulation

by

Jeanne Lafortune

B.A., McGill University (2002)
M.A., University of Toronto (2003)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

Author ....
Department of Economics
15 May 2008

Certified by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Certified by ....
David Autor
Associate Professor, Labor Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Elisha Gray II Professor, Economic History and Applied Economics
Chairman, Department Committee on Graduate Students

Essays on Matching, Marriage and Human Capital Accumulation

by

Jeanne Lafortune

Submitted to the Department of Economics
on 15 May 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis explores the link between human capital accumulation and the functioning of marriage markets.

The first chapter studies the effect of marriage market conditions on pre-marital investment. After showing how a change in the sex ratio can alter incentives for investments, I test this prediction using exogenous variation in the marriage market sex ratio, brought about by immigration, exploiting the preference of second generation Americans for endogamous matches. I find that a worsening of marriage market conditions spurs higher pre-marital investments, measured by years of education, literacy and occupational choice. Overall, the results suggest that there are substantial returns to education in the marriage market, and that both men and women take these returns into account when making education decisions.

The second chapter studies the role played by caste and other attributes in arranged marriages among middle class Indians. Using interview data from a sample of parents who placed matrimonial ads in a popular Bengali newspaper, we estimate preferences for each attribute. We then compute a set of stable matches and find it quite similar to the actual matches observed in the data, suggesting a relatively frictionless marriage market. There is a very strong preference for in-caste marriage but, because this preference is horizontal rather than vertical and because the groups are fairly balanced, in equilibrium, the cost of insisting on marrying within one's caste is small which allows castes to remain a persistent feature of this marriage market.

Finally, the third chapter estimates the effect of marriage delay on fertility by exploiting state laws that restricted age of marriage in the first half of the 20th century. These state laws strongly predict the age at first marriage for both males and females. Moreover, they also influence the age of one's spouse. Using these laws to instrument the age at first marriage for both spouses, I find that delaying marriage of a female by one year reduces fertility by 0.35. The effect of male's timing on fertility outcomes is smaller. Ignoring the effect that these laws have on one spouse's age would lead to an overestimate of the effect.

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Thesis Supervisor: David Autor
Title: Associate Professor, Labor Economics

Acknowledgments

My years at MIT have been influenced by many individuals without whom this process would have been much more difficult.

First, I would like to express my gratitude to my two advisors, Esther Duflo and David Autor. My progress in this doctoral program was mirrored by the change in our relationship: from being a student and research assistant, I became advisee and co-author. Their enthusiasm and passion as teachers were contagious. As researchers, they taught me the importance of thoroughly considered and carefully conducted empirical research. Their mentoring, even while on leave, helped me tremendously in progressing in my research. I was particularly touched by their willingness to explore topics that were not their main areas of expertise to help me pursue this research agenda.

I would also like to emphasize the important role played by other faculty members during my stay at MIT. Josh Angrist and Abhijit Banerjee both significantly, although very differently, contributed to this thesis by their constant support and advice. Josh's philosophy on empirical work has shaped my approach to research. Without Abhijit's careful advice, the theoretical framework of the first chapter of this thesis would have been much less complete. I'd also like to thank faculty members who provided me with various advices during lunches and teas: Tavneet Suri, Ben Olken, Alan Manning, Steve Pischke, Michael Piore, Amy Finkelstein, Whitney Newey and Peter Temin. My thanks also go to Aloysius Siow who made me discover family economics.

My classmates were also instrumental in the completion of this degree, not only for their knowledgeable assistance but also for their friendship and moral support. Without the encouragement of Tal Gross and Patrick Warren, in particular, I would have been unable to pursue research with the same enthusiasm, in particular during my third year. I would also like to thank personally Konrad Menzel for his excellent and extremely helpful econometric advice. I would also like to recognize how Tom Wilkening's help was key for me to acquire the computational skills required for the second chapter of this thesis. For fun times and great conversations, I would also like to thank: Carmen Taveras, Chris Smith, Neil Bhutta, Amanda Kowalski, Tom Chang, Rick Hornbeck, Trang Nguyen, Maisy Wong, Jim Berry, Nicolás Arregui, Suman Basu, Mauro Alessandro and Arthur Campbell. I would also like to thank members of TCC for giving me the sense of a community away from home: my thanks go particularly to Fr. Clancy, Fr. Paul, Fadi Kanaan and Elizabeth Basha. Finally, for Evelia Manac'h and Anne-Sophie Dumesnil who have continued being my friends despite the long distance separating us, I am very grateful.

I would also like to thank the staff at the MIT Department of Economics, at the MIT Dewey Library and the Harvard Schlesinger Library for their help and assistance over this period. I also acknowledge financial support from the Social Science and Humanities Research Council of Canada and from MIT.

Finally and more importantly, I'd like to thank my family for helping me during this process. For my parents, who gave me the strong values and faith that anchored me during this journey. For my brother and sister, and their families, who reminded me of the beauty and joy that life brings. And, more especially, for my husband Jose, without whom this whole adventure would have been much more difficult. I appreciated his generous and always extremely useful advice, his love, patience and encouragement... and his willingness to engage (and lose) in "household bargaining" (aka, rock, paper, scissors).

Contents

List of Tables

List of Figures

Chapter 1

Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market

1.1 Introduction

Several studies have shown that marriage market conditions (such as divorce laws and sex ratios) affect post-marital behavior (Chiappori et al. 2002 and Angrist 2002). These results are generally interpreted as a rejection of the so called "unitary" model of the household where households, once formed, behave like a single individual. Because marriage market conditions change the outside option of each spouse, they alter bargaining weights and lead to modifications in the way household surplus is shared. However, there is little empirical work on the impact of these factors on pre-marital investments. This is surprising, since if individuals are forward-looking, these conditions should be anticipated and potentially modify pre-marital decisions. For example, if one foresees having a lower share of the post-marital output, one could increase one's pre-marital investment in order to compensate for this loss. Moreover, several studies have demonstrated that educational investments appear to respond to perceived returns in practice (see Foster and Rosenzweig 1996 and Nguyen 2007 for returns to education in the labor force and Foster and Rosenzweig 2001 for returns to education in marriage markets). Changes in marriage market conditions may thus impact upon agents' behavior before the union is formed. This paper investigates theoretically and empirically how changes in sex ratios (here defined as the ratio of males to females) modify incentives for pre-marital investments.

Theoretically, a change in the sex ratio can be expected to affect pre-marital investments through two channels: its effect on the probability of matching (which would hold even in a unitary model) and its effect on anticipated bargaining power. To better frame these two channels, I present a simple model. The timing of the model is as follows: first, pre-marital investments are undertaken by each individual, random matching then pairs individuals into couples and finally the output is shared according to rules that may depend on pre-marital investments and external conditions.

Sex ratios affect whether and with whom one can match. The model shows that for any relative risk aversion parameter larger than one, an increase in the sex ratio will lead men to increase and women decrease their pre-marital investments because of the matching effect. If the sex ratio is higher, a man has a higher probability of remaining single. The income effect of having no partner dominates the effect that a partner has on one's returns for high enough risk aversion. Thus, when one's marriage prospects get poorer, one's investment incentives are increased.

Secondly, sex ratios may also alter incentives for pre-marital investments because they modify the balance of power within a household. The model assumes that the division of the marital output occurs such that the bargaining leads to an ex-post Pareto optimal allocation, in the same spirit as in the "collective model" of the household. The bargaining weights may depend on an external determinant of bargaining power, as suggested by Browning and Chiappori (1998) and Chiappori et al. (2002). However, this paper also allows individuals' pre-marital investments to influence the way the output is shared. I restrict the way these investments influence one's bargaining share by assuming that only the ratio of one's own investment to that of one's spouse influences the sharing factor.

The standard framework linking bargaining power and investments draws upon the work of Grossman and Hart (1980), in which agents with linear utility functions are engaged in a contractual arrangement. In that framework, an increase in one's bargaining power always leads one to invest more since the additional bargaining power translates into a larger share of the returns on that investment. Since the utility is linear, there is no income effect stemming from obtaining a bigger proportion of the surplus. However, while risk neutrality may be a good approximation in the context of contracting between firms, it may not be appropriate in the context of spouses engaging in marital bargaining where risk aversion is likely to be present.

When the utility function is concave, a rise in the sex ratio decreases the incentives for male investment through a lower bargaining power (and hence return) as emphasized by Grossman and Hart (1980). This corresponds to a substitution effect. However, because the lower bargaining power also translates into smaller incomes, this increases the incentives for investment due to decreasing marginal utility. In the context of the model, this income effect is found to dominate the substitution effect if the elasticity of intertemporal substitution is less than one. Finally, the change in the sex ratio also modifies the incentives for one's spouse and in order for that response not to undo the direct effect of the bargaining power, it suffices to assume that the investments are gross substitutes in consumption.

Most estimates of either the relative risk aversion parameter or the elasticity of intertemporal substitution (which in this case are the inverse of each other) in the literature suggest that the restrictions mentioned above will hold (see for example Hall 1988, Beaudry and van Wincoop 1996 and Vissing-Jorgensen and Attanasio 2003) .

Note that if pre-marital investments modify post-marital outcomes, one would observe that the sex ratio affects post-marital outcomes, even outside a bargaining model. Furthermore, even within a bargaining framework, the estimated effect of marriage market conditions on post-marital outcomes may not properly measure the full impact of changes in bargaining power because part of this shift in bargaining is anticipated and counteracted by a change in pre-marital investments.

The model implies that for realistic values of the elasticity of intertemporal substitution, a rise in the sex ratio leads men to increase their pre-marital investments and (by an analogous argument) women to decrease them. This paper explores whether there is evidence of this pattern in the context of ethnic marriage markets in the United States around the turn of the twentieth century. Did second generation Americans modify their human capital acquisition decision when faced with a plausibly exogenous shift in the sex ratio of their state-level marriage market?

To answer this question, I exploit the fact that a large fraction of the children of immigrants, here referred to as second-generation Americans, tend to marry within their own ethnicity. Therefore, waves of newly arrived immigrants impact on their ethniticy's marriage markets (as in Angrist 2002). While Angrist looks at national ethnic markets and instruments using flows at entry, this study focuses on state-level, within ethnic group variation, which allows one to control for many potential confounders of the effect of a change in sex ratios. The variation in sex ratios of immigrants at this level is large and influences significantly the marriage market conditions of second-generation Americans. Since immigrants may select their location based on labor and marriage market conditions which also affect the second generation, this shock may be endogenous. To control for endogeneity, this paper constructs an instrument based on the fact that immigrant flows by country within a larger ethnic group are persistent. Each country within a group has located, in the past, to various destinations. Furthermore, over the course of the early twentieth century, the sex ratio of new immigrants has varied substantially and differently across countries. Consequently, one can construct an instrument that allocates shifts in the flows of immigrants to different states using past shares, akin to the strategy used by Card (2001). This variation proves to be highly predictive of both the flow of newly arrived immigrants and their gender composition.

Using this strategy, this paper finds that shifts in sex ratios influence pre-marital investment decisions of men, whether defined in terms of years of education, literacy or occupational choices. In states and ethnic groups where the number of males per female increases in their state of birth due to gender-biased immigration, young adult males acquire more formal education, are more literate, and pursue higher ranked occupations. A change from a balanced sex ratio among immigrants to one where men are twice as numerous as women leads men to increase their educational investment by 0.5 years and women to decrease it by (an insignificant) 0.05 years.

These results are robust to various changes in the start and end dates of the period observed, in the states selected and to variations in the instrument.

As in previous studies (for example, Angrist 2002, Chiappori et al. 2002, Amuedo-Dorantes and Grossbard 2007 and Oreffice and Bercea 2006), this paper also finds that post-matching labor supply decisions are affected by a change in the sex ratio, although the estimated impact is weaker and less significant than previously estimated, possibly due to the difference in empirical strategy. In particular, women's labor force attachment is reduced. A doubling in the sex ratio of newly arrived immigrants from a balanced level leads to a fall of about 4 percent in female labor force participation, of 1.4 hours worked per week, and of 1.3 weeks worked per year. The labor supply response for men appears to be generally positive, but smaller in magnitude and insignificant. The indirect effect of the sex ratio on labor supply through educational attainment, however, appears significant, particularly for males, which suggests that using marriage market conditions as proxies for ex-post bargaining power may lead to inaccurate conclusions regarding the importance of bargaining power.

Finally, this paper combines the model developed and the empirical estimates obtained to calibrate the returns to education in the marriage market. The fact that marriage market conditions influence educational decisions suggests that there are some returns to education in the marriage market, whether stemming from a joint production function or through the effect of pre-marital investments on bargaining weights. Defining returns to schooling on the marriage market as any returns that would not be captured if one was single, I find that these make up around 40 to 60 percent of total returns. These returns are thus important in magnitude and may help to explain why, in this context and many others, women are as educated as men although they have very low rates of labor force participation. It may reflect both the importance of education as an input in household tasks such as child rearing as well as a method to strengthen one's position within the household.

This paper is organized as follows. Section 1.2 summarizes the existing literature. Section 1.3 then introduces the model and derives comparative statics, while Section 1.4 presents the data and explains the empirical strategy. The subsequent section presents the results of the regressions and section 1.6 then uses both the estimates and the theoretical model to separate the returns to education stemming from the labor market vis-à-vis those from the marriage market. The last section concludes and suggests avenues for future research.

1.2 Literature review

This paper is related to the growing theoretical literature linking education and marriage markets in order to address changes in the educational attainment gap between genders (Chiappori et al. 2007 and Pena 2006). In contrast to this literature, the model in this paper assumes symmetry in the production function in order to focus more closely on the effect of the sex ratio. The theoretical work most related to this paper is that of Iyigun and Walsh (2007) who construct a model where the sex ratio can generate gender gaps in educational attainment. They assume a competitive marriage market where consumption levels are independent of spousal investments and this implies that investments are Pareto efficient. This means that their model cannot generate monotone comparative statics for investments with respect to the sex ratio. By contrast, under the specification used in this paper, bargaining may lead to inefficient investment levels and parameters can be selected to ensure monotone comparative statics.

Empirically, this paper relates to a wave of new studies that have explored effects of changes in sex ratios on non-labor outcomes, mostly marital and fertility decisions (Porter 2007, Brainerd 2006, Kvasnicka and Bethmann 2007). They all use large shifts in fertility or mortality rates which altered sex ratios and find that when the sex ratio increases women tend to marry more and to be less likely to have out-of-wedlock births. Porter (2007) also finds that higher sex ratios lead women to marry "better" mates in terms of health, age and height. Angrist (2002) studies the effect of a national shock to the ethnic sex ratio brought about by immigration and uses as an instrument for the gender composition of immigrants the sex ratio at entry. If immigrants leave their home country for reasons that are exogenous to the local marriage and labor market, this instrument identifies the causal effect of changes in sex ratios on post-marital behavior. Using this strategy, he finds that both men and women are more likely to get married and that women's labor supply is reduced while overall incomes are increased when the sex ratio rises. However, no study has yet explored pre-marital investments as a potential outcome.

Finally, this study also contributes to existing work exploring the returns to education in the marriage market. Foster and Rosenzweig (2001) show in a general equilibrium framework that agricultural productivity growth raises the demand for educated wives and confirm it empirically. Behrman et al. (1999) suggest that much of this response is due to the capacity of better educated mothers to teach their children. An older branch of this literature has looked at earnings correlations with own and spousal education. Some studies found positive correlation between own earnings and spousal education (Benham 1974 for the United States, Tiefenthaler 1997 for Brazil, Neuman and Ziderman 1992 for Israel) suggesting that in particular in the case of entrepreneurs, one's earnings tend to rise with spousal education. Also, marriage market conditions seem to influence human capital acquisition (Boulier and Rosenzweig 1984 for example). However, no study has yet quantified the return to education in the marriage market.

1.3 Model

1.3.1 General model set-up

Let us assume a setting where each man (m) and woman (f) is endowed with an initial wealth w. Individuals have an additive utility function over two periods[1]:

$$u(c_1, c_2) = u(c_1) + E\left(u(c_2)\right).$$

For simplicity, assume that the utility function has constant elasticity of intertemporal substitution/constant relative risk aversion given by the parameter σ:

$$u\left(c_k\right) = \frac{c_k^{1-\sigma}}{1-\sigma}, \qquad \sigma > 0, \qquad k = 1, 2.$$

In the first period, an individual can invest in a productive asset i at a cost of 1. Her consumption in the first period is thus given by:

$$c_1 = w - i.$$

[1]It is assumed for simplicity that the discount factor is 1; none of the results derived below depend on this assumption.

In the second period, individuals pair and can share resources. The joint output is given by the function h which is assumed to be increasing in both investment levels, twice-continously differentiable and symmetric:

$$\frac{\partial h(i,i')}{\partial i} = \frac{\partial h(i',i)}{\partial i} > 0.$$

In addition, the production function is always positive when one individual has positive investments and offers positive returns even at very low levels of investment:

$$\begin{aligned}
h\left(i^j,0\right) &> 0, \qquad j = m,f \\
\lim_{i^j \to 0}\left(\frac{\partial h\left(i^j,i^{j'}\right)}{\partial i^j}\right) &> 0, \qquad j = m,f \\
\frac{\partial h\left(i^j,0\right)}{\partial i^j} &> 0, \qquad j = m,f.
\end{aligned}$$

Notice that these restrictions exclude the Cobb-Douglas production function, for example, but are perfectly compatible with a constant elasticity of substitution function.

Once paired, individuals, through post-matching bargaining, arrive at a Pareto optimal sharing (as in Browning and Chiappori 1998, Chiappori et al. 2002). This implies that the post-matching division of the total output will be given by the maximization of the following program:

$$Max \quad \mu\left(i^m,i^f,z\right)u\left(c_2^m\right) + \left(1-\mu\left(i^m,i^f,z\right)\right)u\left(c_2^f\right) \quad s.t. \quad c_2^m + c_2^f = h\left(i^m,i^f\right).$$

Pareto weights are allowed to depend on 3 elements: male and female pre-marital investments and potentially the sex ratio denoted by z.

One can show that the optimal allocations are given by:

$$\begin{aligned}
c_2^m &= c_2^m\left(i^m,i^f,z\right) = \frac{\mu^{\frac{1}{\sigma}}}{(1-\mu)^{\frac{1}{\sigma}} + \mu^{\frac{1}{\sigma}}} h\left(i^m,i^f\right) \\
c_2^f &= c_2^f\left(i^m,i^f,z\right) = \frac{(1-\mu)^{\frac{1}{\sigma}}}{(1-\mu)^{\frac{1}{\sigma}} + \mu^{\frac{1}{\sigma}}} h\left(i^m,i^f\right)
\end{aligned}$$

which imply that one's consumption is a share of the total output and that these shares are determined monotonically by the Pareto weight μ. Further assume that the following conditions for these weights:

$$\begin{aligned} \mu\left(0, i^f, z\right) &= 1 - \mu\left(i^m, 0, z\right) = 0 \\ \frac{\partial \mu}{\partial i^m} &> 0 > \frac{\partial \mu}{\partial i^f}. \end{aligned} \tag{1.1}$$

The first condition indicates that when one is paired with a partner who does not invest (that is when one is single), one captures the full output of the pair. The third imposes that one's investment increases one's share of the output.

Finally, weights will always be such that one always consumes at least what they can obtain from being single. If it was not the case, individuals would elect not to marry and thus this would not be an equilibrium. Formally,

$$c_2^k \geq i^k, \qquad k = f, m.$$

The main assumption of this model is that the consumption exhibits constant returns to scale in the investment levels, that is

$$c_2^j\left(\lambda i^m, \lambda i^f, z\right) = \lambda c_2^j\left(i^m, i^f, z\right), \qquad j = m, f.$$

This will imply that the household production function also has constant returns to scale. More intuitively, this also implies that the Pareto weights (μ) will be homogenous of degree zero in male and female investments: only the ratio of investments influences the share one receives. Thus, in a couple where the male has twice the amount of investment as his wife, this man captures the same share of the total output, no matter what are the absolute levels of investments of both parties. Finally, imposing that consumption must rise at a decreasing rate with own investment

$$\frac{\partial^2 c_2^j}{\partial i^{j2}} < 0, \qquad j = m, f$$

implies, because of the assumption of constant returns to scale, that

$$\frac{\partial^2 c_2^j}{\partial i^m \partial i^f} > 0, \qquad j = m, f.$$

In the first period, spouses decide their optimal investment level based on the following maximization problem

$$Max \quad u\left(w - i^k\right) + E\left(u\left(c_2^k\left(i^m, i^f, z\right)\right)\right), \qquad k = m, f$$

taking the sex ratio (z) and the future spouse's investment as given.[2]

Once the investments have been made, individuals match randomly. This can be rationalized by the existence of search frictions that prevent individuals from finding their perfect partner. This excludes the possibility of using investments to capture a better spouse. Because of this, the probability that males and females will stay single is independent of the investment level and given by

$$\begin{aligned} p^m(z) &= \begin{cases} \frac{z-1}{z} & if\ z > 1 \\ 0 & if\ z \leq 1 \end{cases} \\ p^f(z) &= \begin{cases} 1-z & if\ z < 1 \\ 0 & if\ z \geq 1 \end{cases} . \end{aligned}$$

One can also see this $p(z)$ as the fraction of period 2 one will spend being single if there are frictions once matched and one can possibly lose one's partner. The first order conditions to this maximization are given by

$$\left(w - i^k\right)^{-\sigma} = p^k(z)\left(h\left(i^k, 0\right)\right)^{-\sigma} + \left(1 - p^k(z)\right)\left(c_2^k\right)^{-\sigma}\frac{\partial c_2^k}{\partial i^k}, \qquad k = m, f \tag{1.2}$$

Notice that because the matching is random, one cannot invest in order to capture a higher skilled wife which is why there is no term in the first order conditions that relates own investment to

[2]From our assumptions that $c_2^k \geq i^k$ and that $\frac{\partial^2 c_2^k}{\partial i^m \partial i^f} > 0$, we know that the return to investment will always be at least 1 and so that savings will not occur.

that of one's spouse.

Investments would be ex-ante Pareto optimal in this case if one was to be the full residual claimant of the returns since the total output available to the household would be maximized in that case. However, in this model, the investment will never be Pareto efficient since $\frac{\partial c_2^j}{\partial i^j} = \frac{\partial h}{\partial i^j}$ cannot hold simultaneously for both spouses, as in Acemoglu (1996). There are two factors that distort the decision away from the optimal one. First, one only receives a share of the total output and thus does not capture the full return to one's investment because part of the benefits of this investment will be captured by one's spouse. This would lead one's investment to be below the Pareto optimal level. On the other hand, because investments can be employed to obtain a larger share of the output, overinvestment may also occur. The investment levels selected are thus not ex-ante Pareto optimal unlike in Peters and Siow (2002) or in Iyigun and Walsh (2007). The bargaining process here does not eliminate the "public good" nature of the pre-marital investment as suggested by Bergstrom et al. (1986).

Lemma 1 *There exists a unique pure strategy Nash Equilibrium to this game.*

Proof. See Appendix 1.A. ■

The existence of a Nash Equilibrium depends on the assumption that the consumption function exhibits constant returns to scale (which bounds the degree of complementarity between inputs in the consumption function) and that single individuals receive a positive return (which prevents the existence of a "no-investment" equilibrium).

1.3.2 Comparative statics

A change in the sex ratio modifies the incentives for pre-marital investment through three distinct channels: changes in the probability of marriage, in the relative bargaining weight as well as in the anticipated investment level of one's spouse.

Effect of a change in the probability of matching

First assume that spousal Pareto weights are independent of the sex ratio. The only effect that the sex ratio has is then through the probability of one matching. Formally,

$$\left.\frac{\partial i^k}{\partial z}\right|_{i^{k'}} = \frac{\frac{\partial p^k}{\partial z}\left(\left(h\left(i^k,0\right)\right)^{-\sigma}-\left(c_2^k\right)^{-\sigma}\frac{\partial c_2^k}{\partial i^k}\right)}{\sigma\left(w-i^k\right)^{-\sigma-1}+p^k\left(\sigma\left(h\left(i^k,0\right)\right)^{-\sigma-1}\right)+\left(1-p^k\right)\left(c_2^k\right)^{-\sigma-1}\left(\sigma\frac{\partial c_2^k}{\partial i^k}^2-c_2^k\frac{\partial^2 c_2^k}{\partial i^{k2}}\right)} \quad k=f,m. \tag{1.3}$$

Importantly, this will only affect the investment level of an individual who is on the short side of the market. That is

$$\begin{aligned}\left.\frac{\partial i^m}{\partial z}\right|_{i^f} &= 0 \quad if \quad z<1 \\ \left.\frac{\partial i^f}{\partial z}\right|_{i^m} &= 0 \quad if \quad z>1.\end{aligned}$$

Proposition 1 *The investment level of the individuals on the short side of the market will increase as the number of potential spouses available to them decreases if $\sigma > \tilde{\sigma}$, where $\tilde{\sigma} < 1$.*

Proof. See Appendix 1.A. ■

This result can be explained intuitively. When an individual is single, she has a lower income which would entice her to invest more. On the other hand, the return to her investment is lower because she does not have a spouse to increase the value of this investment. If the agent is sufficiently risk averse, her desire to insure her consumption in case she remains single dominates the substitution effect.

Effect of a change in bargaining power

Now, adding in the effect that the sex ratio has on investments through bargaining power, assume that

$$\frac{\partial \mu}{\partial z} < 0.$$

Thus, z is a factor that decreases the consumption of men and increases that of females given investment levels. Furthermore, assume that the sex ratio is limited in the way it can decrease the return to male investment by:

$$\frac{\partial^2 c_2^j}{\partial i^m \partial z}\frac{\partial c_2^j}{\partial i^f} > \frac{\partial c_2^j}{\partial i^m}\frac{\partial^2 c_2^j}{\partial i^f \partial z}, \qquad j = m, f. \tag{1.4}$$

This implies that the effect of the sex ratio on the return one receives from the spouse's investment must be significant enough in magnitude compared to that on own investment.[3] Thus, it cannot be that the sex ratio penalizes greatly the return that can be obtained from own investment but not from that of the spouse.

In this case, the effect of z on pre-marital investment through the channel of bargaining, conditional on spousal investment, is given by

$$\left.\frac{\partial i^k}{\partial z}\right|_{i^{k'}} = \frac{(1-p^k)(c_2^k)^{-\sigma-1}\left(-\sigma\frac{\partial c_2^k}{\partial i^k}\frac{\partial c_2^k}{\partial z}+c_2^k\frac{\partial^2 c_2^k}{\partial i^k \partial z}\right)}{\sigma(w-i^k)^{-\sigma-1}+p^k\left(\sigma(i^k)^{-\sigma-1}\right)+(1-p^k)(c_2^k)^{-\sigma-1}\left(\sigma\frac{\partial c_2^k}{\partial i^k}^2-c_2^k\frac{\partial^2 c_2^k}{\partial i^{k2}}\right)}, \qquad k = f, m. \tag{1.5}$$

Proposition 2 *Conditional on spousal investment, an increase in bargaining power will lead an individual to decrease their investment level as long as $\sigma > \bar{\sigma}$, where $\bar{\sigma} < 1$*

Proof. From (1.5),

$$\left.\frac{\partial i^k}{\partial z}\right|_{i^{k'}} \propto \left(-\sigma\frac{\partial c_2^k}{\partial i^k}\frac{\partial c_2^k}{\partial z} + c_2^k\frac{\partial^2 c_2^k}{\partial i^k \partial z}\right)$$

Appendix 1.A demonstrates that when (1.4) holds, a sufficient condition for $\left.\frac{\partial i^m}{\partial z}\right|_{i^f} > 0$ and vice-versa for females is that $\sigma > 1$. ■

A rise in the sex ratio as a shift in bargaining power towards females influences the investment decision through two channels. Males have lower consumption for any value of investment which increases their incentives for investment through this income effect. On the other hand, this increase in z also reduces the return to investment and through this channel, leads to a lower investment level. For the income effect to dominate and thus for males to increase their investment when the sex ratio rises, σ must be sufficiently large. This is akin to the effect of a productivity shock on investment decisions in a macroeconomic model.

[3]This can be shown to be equivalent to the conditions $\frac{\partial^2 \alpha}{\partial i^m \partial z} < \min\left\{\frac{-1}{1-\alpha}\frac{\partial \alpha}{\partial i^m}\frac{\partial \alpha}{\partial z}, \frac{1}{\alpha}\frac{\partial \alpha}{\partial i^m}\frac{\partial \alpha}{\partial z}\right\}$ where α is defined as $c_2^m = \alpha(i^m, i^f, z) * h(i^m, i^f)$. Thus, the effect that the sex ratio has on one's ability to modify their weight in the decision process through education cannot be too large in magnitude.

Spousal response

Finally, it must also be that the spousal response will not undo the effect of the bargaining power as presented above. A sufficient, although not necessary, condition for this to occur is that investments be strategic substitutes. That is when one is faced with a spouse who has invested more, the income effect brought about by this is larger than the incentives embedded in the complementarity of investments and this leads one to reduce one's investment. Formally, investments will be strategic substitutes if

$$\begin{aligned}\frac{\partial i^k}{\partial i^{k'}} &= \frac{\left(1-p^k\right)\left(c_2^k\right)^{-\sigma-1}\left(-\sigma\frac{\partial c_2^k}{\partial i^k}\frac{\partial c_2^k}{\partial i^{k'}}+c_2^k\frac{\partial^2 c_2^k}{\partial i^k \partial i^{k'}}\right)}{\sigma\left(w-i^k\right)^{-\sigma-1}+p^k\left(\sigma\left(i^k\right)^{-\sigma-1}\right)+\left(1-p^k\right)\left(c_2^k\right)^{-\sigma-1}\left(\sigma\frac{\partial c_2^k}{\partial i^k}^2-c_2^k\frac{\partial^2 c_2^k}{\partial i^{k2}}\right)}<0, \qquad k=f,m\\ \Leftrightarrow\ & \sigma\frac{\partial c_2^k}{\partial i^k}\frac{\partial c_2^k}{\partial i^{k'}}>c_2^k\frac{\partial^2 c_2^k}{\partial i^k \partial i^{k'}}.\end{aligned}$$

This implies that investments cannot be so highly complementary that the substitution effect dominates the income effect. This translates into a fairly intuitive condition

$$\sigma>\frac{1}{\rho_c\left(i^m,i^f\right)}$$

where $\rho_c\left(i^m,i^f\right)$ is the elasticity of substitution of investments within the consumption function. If investments are gross substitutes, $\sigma>1$ is then a sufficient condition. This rules out the possibility that individuals are the full residual claimant of their investment or receive a return above the return they produce on the couple's output (as in Wells and Maher 1998). In both these cases, investments are strategic complements.

Thus, for $\sigma>1$ and when investments are gross substitutes, an increase in the sex ratio will lead to a decrease in female investment and an increase in male pre-marital investments. Appendix 1.B presents a special case of the model presented above: one where individuals Nash bargain over the surplus.

1.3.3 Perfect competition

One could remove the assumption of random matching and turn to a model where there is assortative matching. However, in that case, because there are no search frictions, one's consumption will be determined by market forces rather than bargaining. If even one man with a given investment is single, all the other men with the same investment level as his will earn a single man's payoff. If that was not the case, a single man could offer to any woman to pair with them and offer him only ε more than his current pay-off and every woman would accept.

This also means that the individual is the full residual claimant on his marginal contribution since:

$$c_2^j = h\left(i^m, i^f\right) - c_2^{j'}$$

and $c_2^{j'}$ is a price outside the control of the agent himself. I will further assume that when $z = 1$, the output is shared equally between spouses (any share $\in [0, 1]$ would be an equilibrium). Assume again that the output function h has constant returns to scale and has positive marginal return when own investments are 0 and that one receives $c_2^k = i^k$ if single.

Proposition 3 *Under perfect competition, when the sex ratio increases, men increase their investment and women decrease their investment as long as* $\sigma > 1$.

Proof. See Appendix 1.A. ∎

Thus, the result obtained above also holds in a context where there is assortative matching and perfect competition.

1.3.4 A different outside option

The previous sections assume that a higher sex ratio will lead males to be more likely to be single. However, in the empirical context that follows, it may be more relevant to think of individuals as being pushed to another marriage market (that of natives). Assume that the second period utility of a member of an ethnic group is given by consumption minus a fixed penalty γ if he marries someone from another ethnic group. An increase in the sex ratio leads men to be more likely to marry native women. This gives them less utility which creates an incentive for higher investment levels. Furthermore, in this particular application, the investment levels in this native pool are higher and this encourages further investment due to complementarity. In this case the size of the preferred marriage pool would be also important since the impact of the sex ratio within your own marriage market may depend on the likelihood of marrying within one's group. Thus, even in this case, tighter marriage market conditions will lead to higher investments, whether or not the sex ratio influences post-matching bargaining.

1.3.5 Ex-post outcomes and the sex ratio

This analysis also highlights that if the sex ratio affects pre-marital investments, the effect of the sex ratio on post-marital consumption levels will not represent the effect of bargaining power. It is a mixture of the bargaining power effect, the effect of one's investment level on post-marital outcome and the effect of one's spouse's. Formally,

$$\frac{dc_2^k}{dz} = \frac{\partial c_2^k}{\partial z} + \frac{\partial c_2^k}{\partial i^k}\frac{\partial i^k}{\partial z} + \frac{\partial c_2^k}{\partial i^{k\prime}}\frac{\partial i^{k\prime}}{\partial z}.$$

1.4 Data and empirical strategy

Having established a framework where changes in the sex ratio modify individuals' pre-marital investments, I now investigate empirically the link between sex ratios and pre-marital investments.

As in Angrist (2002), this paper uses data from second-generation Americans born around the turn of the century (from 1885 to 1915). This identification strategy is based on the observation that second-generation Americans tend to marry within their ethnic group (40 percent of second generation males and 45 percent of females among a slightly older cohort marry within their own ethnicity). Thus, for this population, the relevant marriage market includes new waves of immigration. Because marriage markets are fairly local, I focus on state-level within ethnic group marriage markets. From 1900 to 1930, the sex ratio of newly arrived immigrants varied greatly transforming the balance of power within each state's ethnic marriage markets. These waves occurred at the moment when the sample of second generation individuals was making educational and marriage decisions. Because location choices of immigrants may be endogenous, this paper instruments for both the flow and the sex ratio of new immigrants using the fact that immigrants locate near existing networks (as in Card 2001 and justified by Munshi 2003), which leads past immigrant stocks in a particular state to predict current immigrant flows.

1.4.1 Basic specification

The basic regression of this study relates pre-marital outcomes of second generation Americans to two characteristics of their marriage market: its sex ratio and its total size. The second variable provides an estimate of the effect of market thickness which may influence decisions as explained in Section 1.3.4. In addition, it captures any effect that overall own-ethnic immigration has on local conditions, either through the marriage or the labor market.

In order to control for potential confounding factors that affect sampled individuals, the regressions include fixed effects for cohort, state and ethnic group as well as for cohort*state, cohort*ethnic group and state*ethnic group. They also include dummies for age, year of birth, year of the Census and for nativity of parents. The estimation equation is given by

$$y_{kjst} = \alpha * sexratio_{jst} + \beta * flow_{jst} + \psi * X_{kjst} + \sigma_{js} + \phi_{st} + \gamma_{jt} + \varepsilon_{kjst} \qquad (1.6)$$

where the left hand-side variable is an outcome for an individual k, of ethnic group j, born in state s, of cohort t. Conceptually, this regression contrasts the change in outcomes over time among individuals from a given state of two different ethnic groups.

The marriage market size and sex ratio may be endogenous. Immigrants potentially select their state of residence based on labor and marriage market conditions. Female immigrants may choose to immigrate to a state where women's bargaining power is larger. Males may elect locations where there are good work opportunities. Since these factors influence choices made by second generation Americans, it introduces a bias in the estimation of Equation (1.6).

To alleviate this problem, I construct an instrument in a similar spirit as that of Card (2001). Since individuals from the same country of origin tend to form networks, they also tend to migrate to similar locations (Munshi 2003). Past location choices are thus a good predictor of future immigration decisions. As long as past waves of immigrants did not select the state of migration based on future marriage market conditions for their children, using these shares provide an exogenous source of variation. Various countries of birth are included within each ethnic group and each had previously selected different locations. Since over the period, the sex ratio of immigrants within an ethnic group varied by country of birth, the combination of this variation and differences in past location shares provides the source of variation that the instrument will exploit. In short, this instrumental variable strategy assumes that individuals tend to locate where their fellow countrymen live but marry within the entire ethnic group. If all countries of birth within one ethnic group selected the same locations, there would be no geographical variation to exploit.

More precisely, two instruments were constructed as follows. All male and female immigrants were allocated separately to a given state for each period and country of origin based on the 1900 concentration of that country in that particular state. If 10 percent of all Norwegians were located in Minnesota in 1900, 10 percent of all men and women immigrants from Norway arriving after 1900 are assigned to Minnesota. This generates a predicted flow of males and females by country of birth. Summing for all countries within an ethnic group, one obtains a measure of the predicted flow of immigrant of each gender for each state, ethnic group, and immigration period cell. The instrument for the flow of immigrants of a given ethnic group is then obtained by adding the predicted flow of males and that of females. The sex ratio instrument is built by dividing the predicted flow of males by that of females. Equations (1.7) and (1.8) define formally the instruments:

$$pred\_sexratio_{ist} = \frac{\sum_{j \in i} \left(\frac{imm_{js1900}}{imm_j} \right) * males_{jt}}{\sum_{j \in i} \left(\frac{imm_{js1900}}{imm_j} \right) * females_{jt}} \tag{1.7}$$

$$pred\_flow_{ist} = \sum_{j \in i} \left(\frac{imm_{js1900}}{imm_j} \right) * (males_{jt} + females_{jt}) \,. \tag{1.8}$$

The strategy can best be illustrated by an example using the Scandinavian ethnic group in two key states: Illinois and Wisconsin. In 1900, Illinois had 10.2 percent of the Danes, only 1.3 percent of the Finnish, 8.9 percent of the Norwegians but 17.3 percent of all Swedes. Wisconsin, on the other hand, had a similar fraction of the Danes (10.5 percent), slightly more Finnish (3.5), a much larger share of the Norwegians (18.2) and only 4.6 percent of the Swedes. Figure 1-1 presents the evolution of the sex ratio among all four countries over the period studied and the predicted sex ratio of this ethnic group in both states. Because Illinois had a high concentration of Swedes in 1900, the evolution of its predicted sex ratio is highly influenced by the changes in the sex ratio of Sweden immigrants. On the other hand, Wisconsin follows much more closely that of Norwegians. Figure 1-2 shows that the same argument holds for flows.

This identification strategy relies on one key assumption: that immigrants before 1900 did not select these locations because they anticipated the changes in marriage and labor market conditions for that particular ethnicity after 1900. This assumption will not be violated if immigrants select locations that were more attractive for their ethnic group before 1900 but remained similarly attractive over the next 30 years. It will also not be violated if immigrants anticipated shocks for their ethnicity that were short-lived so that by 1905, no remnants of these shocks were found. Finally, it would also not violate the exclusion restriction if pre-1900 immigrants selected states in anticipation of better conditions for all ethnic groups but not particularly for their particular ethnic group because regressions control for state-time fixed effects.

In addition, it must also be the case that, once controlling for the total number of immigrants, no other characteristics of the immigrants change at the same time as the sex ratio by location. This could be violated if when more men than women enter the United States, these men tend to be of lower/better quality. Little information on immigrants' quality is available to test whether this is violated, except for immigrants' literacy as measured by the Census. No correlation was found between that measure for either gender and the actual or the instrumented sex ratio of immigrants.

1.4.2 Data

Outcome measures

All outcomes, obtained from IPUMS files between 1900 and 1970, are presented with a detailed description in Appendix Table 1.C.1. First, marital outcomes are collected: marital status, measures of marital stability (divorce rates and number of marriages) and country of birth of one's spouse. Unfortunately, ethnicity of spouse's parents is not available in either 1940 or 1950 so it is difficult to classify spouses as second generation Americans of a particular group and thus measure this broader definition of endogamy. Because pre-marital investments may be modified because marriage is delayed when the marriage market is tight, leaving more time to acquire education, age at first marriage is also measured. To alleviate the problem of sample selection (age is only measured if one is already married), this variable is restricted to individuals older than 35, for whom most first unions have already been entered into.

Various measures of human capital investment as proxies for pre-marital investments are considered: literacy, years of schooling and occupational choices. Literacy should be acquired before marriage and could affect post-matching output (see Behrman et al. 1999 for an example in India). A more continuous measure of human capital investments is the highest grade a person has attained. While it provides a more detailed categorization of the level of skills acquired, schooling may also be obtained partially after marriage although there is little evidence of this in my sample. First, this sample has an average schooling level below high school completion (9.5 years for females and for males) and the average age at first marriage is 23 for females and 27 for males. Also, while 22 percent of the individuals aged 15-25 attend school, only 1 percent of the married males and 3 percent of the married females report being in school. Finally, two occupational indices measuring the "quality" of the current occupation are available. These variables could reflect pre-marital investments because the quality of an occupation is correlated with on-the-job training and past human capital accumulation although it could also reflect some labor supply decisions. To alleviate potential problems linked to these measures capturing post-marital investments, they are measured only for those aged 15-25 except in the case of education where education was only available at later ages.

To measure post-matching outcomes, this paper uses labor supply of all individuals aged 25 and above. I do not restrict the analysis to married individuals because this would potentially introduce selection bias. For all individuals, a variable indicating labor force participation and employment is available. In addition, measures of weeks worked last year and hours worked last week can be obtained.

Table 1.1 gives the main summary statistics for each outcome. The rate of non-marriage after age 35, around 10 percent, is much above that of natives (about 5 percent).[4] Divorce rates are low but widowhood is not uncommon for women. The age at first marriage is around 23 for women and 27 for men. About 8 percent of second generation women and 3 percent of males are currently married to first generation immigrants from their ethnic group (which accounts for most marriages between second generation and immigrants). This is somewhat low but the denominator includes all singles and widows, to avoid selection bias. This is also lower than the total endogamous marriages which include all marriages within second-generation individuals as well. Literacy is very high among second generation Americans (close to 99 percent of them are literate) but varies considerably across ethnicities, with non-European groups having much lower levels. Men and women are both achieving about the same level of schooling (9.5 years) and if anything, women are more educated than men. This is a fact that holds for natives as well. Labor supply attachment by woman is quite low. Slightly more than 30 percent of women were in the labor force compared to 79 percent for men. While 61 percent of men worked full time, only 18 percent of women did.

Marriage market measures

A key decision in implementing the above framework is the appropriate empirical definition that should be used for a "marriage market". In this setting, a marriage market is assumed to be a given ethnic group within a state in a particular cohort. This definition of marriage market is quite restrictive but as long as what happens to one's market is more relevant than what occurs in another group, this approximation will capture relevant variation.

The marriage market is first defined within an age cohort. The second generation sample born between 1885 and 1915 is divided into 5 year intervals. I maintain the assumption that people marry within their age cohort.[5]

Marriage markets are fairly local with more than 65 percent of sampled individuals married to someone who was born in the same state as them.[6] State is the lowest geographical unit for

[4]Previous studies have noticed that second generation immigrants have the lowest rate of marriage (Groves and Ogburn 1928, Haines 1996 and Landale and Tolnay 1993).

[5]Around 50 percent of married individuals younger than 40 are matched to someone of the same age group.

[6]This is almost as large as the proportion of individuals still living in their state of birth. One finds very small proportion of "out-of-state" marriages for individuals who are still living in their state of birth.

which place of birth is available in the IPUMS files. Furthermore, mobility is limited: more than 70 percent of individuals in this sample still live in their state of birth and this figure increases to 85 percent among those less than 20.

Marriage markets must also include a definition of ethnic groups. From 1900 to 1970, the IPUMS files include information on parents' country of origin. Using this variable, each second generation individual is associated with a particular ethnicity based on father's ethnicity.[7] Using all countries of birth, the sample was divided into 9 ethnic groups, summarized in Appendix Table 1.C.2. This division was inspired by that used by Angrist (2002) and based on Pagnini and Morgan (1990), with required modifications.[8] Using marriage patterns of previous immigration waves, these groupings were found to correspond closely to the patterns observed in the data. In almost all cases, the percent of individuals marrying someone within their ethnic group but not from their own country of birth was much higher than the prevalence of those countries in the sample. The regressions below were performed with slight differences in the allocation of ethnicities to ethnic group with very similar results.

Two sets of marriage market conditions were constructed using the IPUMS files for 1910, 1920 and 1930: one for immigrants only, the other incorporating second generation individuals as well to which we will refer to as "foreign stock". The former were classified by their country of birth, year of immigration (grouped into 5-year periods) and state of current residence. The latter were classified by their state of birth, their father's place of birth and grouped with immigrants such that these immigrants arrived while the second generation individuals are in their late teens (age 15-19), an age at which schooling and marriage decisions are made.[9] Only immigrants arriving between the ages of 10 and 25 are included since they are more likely to

[7]While Angrist (2002) uses mother's ethnicity, I employ father's ethnicity because in 1960 and 1970, only father's ethnicity is reported when the father is foreign born. This is of little importance, however, because 95 percent of foreign born parents share a common country of birth.

[8]East European Jews are grouped by nationality because it is difficult to identify them after 1930. Also, two countries of birth per ethnic group were required since the instrument relies on differences in 1900 location choices within ethnic groups across countries of birth. Immigrants from Ireland were joined with other British Isles. Italians were grouped with other Catholic Southern European countries: Spain and Portugal. Finally, Mexicans were included with other immigrants from the Caribbean, Central and South America.

[9]To avoid double-counting for the flow indicator, only the 1910 Census is used to compute the flow of immigrants between 1900 and 1909, the 1920 Census for immigrants arriving between 1910 and 1919 and so forth. However, since the sex ratio may suffer more from measurement error because it is a ratio, all three waves of the Census were employed to construct that measure.

be part of the marriage pool of the cohort of second generation Americans.[10] For each ethnic group-state-immigration period, the above methodology produces a measure for the number of immigrants and their gender. Measures of total flow of immigrants and total flow of foreign stock are then built by summing all individuals in each state-ethnic group-period cell. Sex ratios were defined as the number of males per female in each cell.[11]

Instrument construction

Equations (1.7) and (1.8) employ as national flow measures the sum of all state flows as defined previously. Location shares were obtained from the 1900 Census tables (United States Census Office 1901).[12] Ideally, the 1890 shares would have been used but a few key countries were only tabulated starting in 1900 and for countries which were similarly identified in both periods, shares are almost identical.

Immigrants were concentrated in some key states with 72 percent of all immigrants locating in 10 states in 1900–see Appendix Table 1.C.3. The location of immigrants varied importantly by ethnic group, the traces of which can still be found in the ethnic composition of today's population. The relative concentration of ethnicities also varied from the most concentrated (Hispanics, with 94 percent living in 10 top states) to the least concentrated (British ancestry, at 75 percent).

More importantly for the instrument, variation in location choice across countries of birth and within ethnicity arises. For example, among those of British ancestry, English Canadians located mostly in Massachusetts and in Michigan while Australians elected California, the Welsh primarily settled in Pennsylvania and the Irish, in New York and Massachusetts. Even between Poles and Russians, where the same three states are preferred locations (New York, Pennsylvania and Illinois), the Poles were distributed equally across the three states while the Russians were much more concentrated in New York.

Table 1.2 summarizes the distribution of the endogenous variables and of the instruments.

[10] Also, a variant using different age distribution by gender such that females are between 10 and 22 but men between 13 and 25 was used to better match spousal age differences and produced very similar results.

[11] If the cell is empty, the sex ratio is set to 1. If there are only men, the sex ratio is equal to 1.5 times the number of males. Neither adjustment is crucial; similar results were obtained with various modifications.

[12] Because these shares are computed using the full population of immigrants and not just a small public-use sample, they are robust to the "small cell bias" as argued by Aydemir and Borjas (2006).

As can be seen, the major immigrant groups over this period were the Russians, Poles and Romanians, followed by Southern Europeans and Germans. Including second generation individuals, those of Germans and British descent are far more numerous, reflecting the importance of past immigration waves. The sex ratios of Others (mostly Asians) and Italians were among the highest at almost two men per women while the Francophone and those of British ascendance had close to a balanced sex ratio. The sex ratios of the total foreign stock are more balanced but the same differences across ethnic groups emerge.

1.5 Results

1.5.1 First stage

The instruments are very highly predictive of their respective endogenous variable. An increase of one in the predicted flow leads to an increase of about 0.87 in the actual number of immigrants arriving over that period and of about 0.57 for the total foreign stock. Similarly, an increase of one in the predicted sex ratio measure is linked to an increase of about 0.84 among immigrants and about 0.43 among total foreign stock. These are all significant at 0.1 percent level. These results are presented in Table 1.3.

The robustness of the first stage is tested through various specifications presented in Table 1.4.[13] To verify that the share of immigrants not only predicts the behavior of immigrants shortly after 1900, column (1) ignores the first two periods of immigration and finds that this omission does not change the robustness of the first stage. Column (2) restricts itself to the first four periods of immigration and again finds similar results. Although the relationship between the instrument and the actual marriage market measure is stronger for some ethnic group than for others, removing any ethnic group does not alter the significance of the first stage. The first stage is also robust to the transformation of the instruments and the endogenous variables in logarithmic form. Finally, removing the key immigration-receiving state over this period (New York) does not change the results. All empty cells where the sex ratio was imputed were dropped without significantly altering the coefficient on the sex ratio. To ensure that this result

[13]It only presents the results for the marriage market measures based on immigrants but very similar results were obtained for the ones based on total foreign stock.

does not stem from correlated measurement error in the flow of immigrants, another source of data was used to construct the flow of immigrants for the instrument. This measure included all immigrants (irrespective of their year of birth and their age at arrival) by country of birth, arrival period and gender but only for Foreign Born Whites. As can be seen in column (6), the instrument conserves predictive power, although the precision goes down due to the imperfect measure of flows it constitutes.

1.5.2 Outcomes

Marital outcomes

I first explore the causal effect of marriage market conditions on marital outcomes. The first panel of Table 1.5 presents regressions where marriage market variables are defined to include only immigrants while the bottom section relates to conditions measured for the total foreign stock. Surprisingly, both men and women are more likely to have ever been married when the sex ratio rises (although this is similar to Angrist 2002). The probability of a man ever having been married rises by 3.3 percent when the sex ratio of immigrants doubles from a balanced level. The next two columns shed some additional light on this result. Men are also more likely to be divorced and more likely to have been married more than once (the opposite being true for women). The rise in the probability of having been married more than once is almost comparable in size with the effect of the sex ratio on the rate of marriage. Thus, the effect of the increased sex ratio among this population appears to be more linked to marital stability than to the formation of relationships.

As presented in the last section of the model, marriage rates may also be unaffected if the outside option is not to remain single but rather to elect a less desirable marriage market. Column (4) of Table 1.5 indicates that men are significantly less likely to marry an immigrant of their own ethnicity when the sex ratio is higher. A change from a balanced sex ratio to one where there are twice as many male as female immigrants decreases the probability of a man to be married to a female immigrant of his own ethnicity by 1.7 percent. The effect is smaller and imprecisely estimated for females. Marriage market size strongly increases the probability of marrying an immigrant as predicted from the model. Also, age at first marriage does not appear to be modified suggesting that any effect of the sex ratio on pre-marital investments will not mechanically stem from a delay in marriage timing. Marriage market sizes appear to hurry the timing of marriage of women but delay that of men. When one's preferred marriage market expands, it may be easier to select optimally the timing of one's marriage. If females prefer marrying earlier than men, this could explain these results.

Pre-marital investments

The main outcomes of interest relate pre-marital investments to marriage market conditions. Table 1.6 first presents correlations obtained from an OLS regression. The effects observed here are in the predicted direction and significant in some cases for females but the coefficients are extremely small. The OLS results should be biased towards zero if immigrants elect locations where they have more bargaining power and more mating possibilities. Then, locations with a larger number of male immigrants are also those in which second generation men have less incentive to invest in human capital. There is weak evidence that this is the case, as immigrant men elect states where there are more second generation females and fewer second generation males of their own ethnic group. Also, the correlation between the sex ratio and the probability that a female marries an immigrant of her ethnic group is fairly strong while the IV result presented above is small and insignificant, which may be indicative that immigrants select locations where the marriage prospects are good.

Once one purges endogeneity using instruments, the coefficients are larger in magnitude for both genders (except for literacy). The results now indicate that a marriage market favoring women leads men to be more literate, to have more years of completed education and select more highly paid occupations. All these are significant at least at 10% significance level. The coefficients for females are negative except for literacy but are neither very large nor significant.

This suggests that a shift from a sex ratio among immigrants of 1 to 2 (more than two standard deviations in the sex ratio) leads to a 1.7 percent point increase in the probability that a male is literate and, on average, to about half a year more of education. Furthermore, young men were selecting much more highly ranked occupations when faced with higher sex ratios. Women's responses are smaller in magnitude, except for occupational choices. Marriage market size appears to lead both genders to select much more highly paying occupations, in particular for women. This is surprising but may reflect the fact that immigrants fill low-paid occupations and push second generation Americans to higher-paying ones. Overall, omitting the flow measure usually renders the effect of the sex ratio more significant.

Table 1.7 tests in various ways the robustness of these results.[14] The first column removes the first immigration period and finds very similar results, indicating that the result is not driven by the early years of the period in question. Dropping the last period does not modify the point estimate for education by much but does increases the standard errors while it greatly increases the size of the effect of the sex ratio on the occupational ranking variable. Removing the major immigrant-receiving state over this period (New York), if anything, strengthens the relationship. Adding dummies for each country of origin (rather than ethnic group dummies as in the base specification) does not weaken the pattern observed. Restricting attention to older or younger respondents leaves the results unchanged. Similarly, ignoring a particular Census year does not affect the results.[15] Although not presented here, variants of the instrument were explored with similar results. For example, although gender-specific shares by country of birth were not available from the Census tables, overall immigrant sex ratios by state were obtained. If one allocates immigrants based on the interaction of a state's attractiveness for a particular gender and its attractiveness for a particular ethnicity, the results are very similar to the ones presented above.[16]

Labor supply

The model presented above argues that the change in pre-marital investments due to altered sex ratios stems from a desire to offset partially the expected effect of the sex ratio on post-marital outcomes. Having found a significant effect of marriage market conditions on human capital decisions, this paper now turns to proxies of post-marital outcomes. The OLS regressions presented in Table 1.8 suggest that higher sex ratios among immigrants are correlated with higher

[14] It focuses on males and on two specific measures, years of education and Duncan Index of occupational choices, although the results are similar for females and other outcome measures.

[15] These last two variants are only presented for the educational variable because the occupational score was restricted to individuals between the ages of 15 and 25.

[16] An interesting data set including the intended state of residence of immigrants at the port of entry was also collected. Unfortunately, the first stage using this data proved to be too weak to be of use for this paper.

labor force participation of both men and women. This could be either an overestimate or an underestimate of the real causal effect. It would be an overestimate if male immigrants select to locate in states where the labor market is booming. On the other hand, if men tend to locate in areas where they have more bargaining power, they would select locations where males are working less and the OLS would be a lower bound on the magnitude of the causal estimate.

The right-hand side panel of Table 1.8 presents the results of the instrumental variable regressions. The causal effect of the sex ratio appears to lead women to reduce their labor inputs while men increase theirs. These results indicate that a doubling of the sex ratio (from 1 to 2) lead women to be 4 percent less likely to be in the labor force and 3 percent less likely to be employed. This is smaller than the 9 percent found by Angrist (2002) which included females aged 16-33, an age at which labor supply is much more variable and potentially influenced by pre-marital decisions. A rise in the sex ratio of immigrants from a balanced level to one where immigrant men are twice as numerous reduces hours worked per week and the number of weeks per year by about 1.3. These results are significant only at 10 percent significance. For males, a change from a balanced sex ratio to one where men are twice as numerous leads to no effect for either employment or labor force participation and raises hours worked per week and weeks per year by about 0.5, although these are very imprecisely estimated and insignificant. The OLS results are usually lower, although not significantly so, than the IV for males and higher for females as expected if immigrants locate based on the bargaining conditions of the marriage market.[17]

1.5.3 Labor supply, pre-marital investment and mate selection

The previous section found that the sex ratio had modest albeit imprecisely measured effects on labor supply. One could conclude that this implies little evidence of ex-post bargaining. However, one must also take into account that the effect of the sex ratio measured by the above regression includes not only the ex-post bargaining effect but also any effect that the sex ratio may have had on post-marital outcomes through its effect on education. Economic theory does not predict whether education increases or decreases labor supply. The income effect decreases

[17]Selecting only married females would show a much clearer pattern where females greatly reduce their labor supply. However, it is unclear whether this would stem from selection or ex-post bargaining.

labor supply. On the other hand, the substitution effect increases the number of hours spent working.

To isolate the effect of education on labor supply in this population, I use compulsory schooling laws as tabulated by Lleras-Muney (2002) as instrument for education in a sample of individuals born between 1900 and 1924, a slightly younger cohort than the one studied above. Labor supply and education are measured in the 1940-1970 IPUMS files. Two sets of results are obtained: one for the full sample and another for second generation Americans. The results presented in Table 1.9 use as instruments a set of dummy variables for each minimum number of years of schooling required by the state.[18] The first stage suggests that each additional year of compulsory schooling leads men to increase their level of schooling by about 0.05 years and women to do so by about 0.8 years.[19] The IV estimates suggest that education decreases labor supply, whether measured in terms of labor force participation rates or hours worked. The estimates are fairly large suggesting that one more year of education reduces hours worked per week by about 0.5 hours for females and 1.5 hours for males, but are only significant for males. Among females, the results are stronger when the sample is restricted to second generation individuals, increasing the magnitude and the significance of the effect to about 1.5 hours. The first stage is much weaker in this subsample for males and the effect for hours worked falls to about 0.5. Results for labor force participation are much smaller and weaker, in particular for males.

The next set of regressions attempts to measure the overall effect of both spouses' education. It is restricted to married individuals for that reason. The instrument is based on the compulsory schooling that affected each spouse in his or her state of birth. Two caveats must be mentioned. First, the first stages are much weaker in this context than before, simply because there are a few spouses who were subject to different compulsory schooling laws (since individuals tend to marry within their state and within a relatively close age cohort). The compulsory schooling laws affecting females tend to be a better predictor of the education of both spouses. Second, even if both educational levels are instrumented, this regression does not control for the potential endogeneity of the match. Nevertheless, these results are presented as a robustness check on the previous estimates. They suggest that for both genders, one's own education decreases labor

[18]Similar results were obtained by using a continuous measure of the minimum number of years of schooling.
[19]This is very similar to the first stage presented jointly for both genders by Lleras-Muney (2002).

supply while that of one's spouse tends to attenuate this effect.

Combining these estimates with the ones from the above section, a doubling of the sex ratio, through the educational channel itself, decreases the number of hours worked by males by about 0.5-0.75 hours per week.[20] The effect for females is in the same direction albeit much smaller. This suggests that the effect of the sex ratio on labor supply obtained in the previous section is underestimating the true effect of the sex ratio on post-matching outcomes, as predicted by the model presented above. The effect of the sex ratio on post-matching outcomes, once purged of the effect it has through changes in pre-marital investments and matching patterns, then provides an estimate of the effect of an external shifter in bargaining power on post-marital labor supply.

Furthermore, in a case where only matching influenced the choice of pre-marital investment, the effect of the sex ratio on labor supply, for example, is entirely driven by its effect on one's own and spouse's education because the sex ratio does not alter ex-post decisions. These results are thus not in accordance with a hypothesis where only matching is at play.[21]

This exercise is meant as an illustration of the importance of considering the link between pre-marital behavior and post-marital decisions. It suggests that using changes in sex ratios as proxies for ex-post bargaining power without taking into effect the potential link that marriage market conditions have on pre-marital behavior and matching patterns may lead to misleading inference. It would have been best to include education in the above labor supply regressions and use another source of exogenous variation to instrument for it. Unfortunately, the sample of second generation Americans employed in this study was too small to use a measure of compulsory schooling as an additional instrument for the educational attainment of an individual in the labor supply regression.

[20]A similar range of values would be given for males if using the effect of both own and spousal education on labor supply decisions.

[21]Other evidence that matching is not solely driving the results was obtained. First, while the above matching model suggests that the effect of the sex ratio is largest when one is on the short side of the market, no evidence of this was found. Second, the effect of the sex ratio appears to be larger in larger communities, which is inconsistent with a matching model since a similar change in the sex ratio implies many fewer potential mates in a small than in a large community.

1.6 Returns to education in the marriage market

The results presented above suggest that marriage market conditions influence pre-marital investments. Assuming this is due to a reaction to changes in the incentives imbedded within the marriage market, these results can be used to infer returns to education in the marriage market.

1.6.1 General framework

Let us define the returns to education in the marriage market as any additional benefit that is given by one's human capital investment that would not be observed if one were single. First, there could be additional benefits captured once married simply because the educational investments of each spouse are complementary in the household production function (from utility derived from conversations, from the role of parental education in child-rearing or even because of learning spillovers). Because those benefits are shared between spouses, the "public good" aspect of this return may lead individuals to underinvest compared to the optimal level. Secondly, marriage market returns could arise if one's bargaining weight depends on one's educational level. Thus, if single, one's education simply affects the output produced but if married, it affects both the output and the share of it one can capture. In this setting, the spouses simply play a zero-sum game where education does not have any additional productive element but serves as a negotiation tool. This would lead individuals to overinvest.[22]

Disentangling the various sources of incentives for human capital investment is not easy.[23] Nevertheless, as an illustration, this section attempts to derive some estimates of the importance of marriage market-related returns to education by combining the empirical estimates found above with the model developed in Section 1.3. I fit the parameters of the model to be the most consistent with the observed educational choices of males, females and their spouse and the measured effect of the sex ratio on education. The estimated parameters are then used to

[22] Finally, marriage market may also stimulate investments through competition between individuals of the same gender if the matching is not random. It can be shown, however, that in this case, a rise in the sex ratio would lead to a fall in males' investment because as the sex ratio rises, the value of the benefit of more education (i.e. a spouse) falls because fewer females are available. Nevertheless, in this case, the gender who is on the short side of the market may over invest in education simply to compete with one another.

[23] Studying the effects of a policy that increases education of only one gender on the other gender's investment decision would be a key input in this analysis.

compute the fraction of returns to education in the marriage market. Formally, let me define total returns to education as

$$\frac{\partial \log c_2^k\left(i^m, i^f\right)}{\partial i^k} = \frac{1}{c_2^k\left(i^m, i^f\right)} \frac{\partial c_2^k}{\partial i^k}.$$

To separate marriage and labor market returns, I take the marriage market returns to correspond to returns not captured by a single individual, that is

$$\frac{1}{c_2^k\left(i^m, i^f\right)} \left(\frac{\partial c_2^k\left(i^m, i^f\right)}{\partial i^k} - \frac{\partial c_2^k\left(i^k, 0\right)}{\partial i^k}\right).$$

Notice that in the models presented below, $c_2^k\left(i^k, 0\right) = i^k$ and thus the labor market returns will be given by $1/c_2^k$. I further parameterize the model by assuming that the household production is given by

$$h\left(i^m, i^f\right) = \left(i^{m\alpha} + i^{f\alpha}\right)^{\frac{1}{\alpha}},$$

a constant elasticity of substitution function (CES) with an elasticity given by $\frac{1}{1-\alpha}$.

1.6.2 Spouse selection model

Let me first consider a model where the sex ratio affects the matching patterns but not the way the household surplus is shared. Formally, male and female consumption are given by

$$\begin{aligned} c_2^m &= i^m + \lambda\left(\left(i^{m\alpha} + i^{f\alpha}\right)^{\frac{1}{\alpha}} - i^m - i^f\right) \\ c_2^f &= i^f + (1-\lambda)\left(\left(i^{m\alpha} + i^{f\alpha}\right)^{\frac{1}{\alpha}} - i^m - i^f\right) \end{aligned} \tag{1.9}$$

In addition, if one matches outside one's preferred marriage market, one receives a penalty of γ in utility terms. Notice that in this case, the first order condition is given by

$$\left(w^m - i^m\right)^{-\sigma} = p(z)\left(c_2^{mN} - \gamma\right)^{-\sigma} \frac{\partial c_2^{mN}}{\partial i^m} + (1 - p(z))\left(c_2^{mO}\right)^{-\sigma} \frac{\partial c_2^{mO}}{\partial i^m} \tag{1.10}$$

where $p(z)$ is the probability of marrying a native, c_2^{mN} corresponds to the consumption when married to a native female, c_2^{mO} to the consumption level when married to a female of one's own ethnicity and γ the utility cost of marrying out of one's ethnic group. In this case, one can also find the effect of a change in z which is given by

$$\frac{\partial i^m}{\partial z} = \frac{\frac{\partial p}{\partial z}\left(\left(c_2^{mN}-\gamma\right)^{-\sigma}\frac{\partial c_2^{mN}}{\partial i^m}-c_2^{mO-\sigma}\frac{\partial c_2^{mO}}{\partial i^m}\right)+(1-p)\left(c_2^{mO}\right)^{-\sigma-1}\left(-\sigma\frac{\partial c_2^{mO}}{\partial i^m}\frac{\partial c_2^{mO}}{\partial i^f}+c_2^{mO}\frac{\partial^2 c_2^{mO}}{\partial i^m \partial i^f}\right)\frac{\partial i^{fO}}{\partial z}}{-SOC^m} \tag{1.11}$$

where the denominator simply corresponds to the second order condition. Using Equations (1.10) and (1.11), the mean value of i^m, i^{fN}, i^{fO} from the data and the computed estimate of $\frac{\partial i^m}{\partial z}$ and $\frac{\partial i^{fO}}{\partial z}$ from above, I find the set of parameters α and σ that offer the best fit.[24] An assumption must also be made about the average initial wealth of each individual (w). Since educational investment ranges from 0 to 18 with an average slightly above 9, the results below will use variations in the average wealth ranging from 22 to 30 since one should never want to invest more than half of one's wealth based on the model presented above. Three values of bargaining weights (λ) are also evaluated. Finally, two more parameters are required in this case. I must calibrate the cost of marrying outside one's ethnic group and impose that this cost be 9, slightly less than the consumption level one would receive if there were no complementarity between investment levels. Similar results were obtained with other values. Finally, an estimate of $\frac{\partial p}{\partial z}$ was computed to correspond to the estimated effect of the sex ratio on spousal education level since the effect of the sex ratio on the probability of marrying a member of one's own ethnic group (immigrant and second generation Americans) could not be estimated.

The top panel of Table 1.10 presents these results. Both for males and females, estimates of the parameter α are very comparable and vary between 0.26 and 0.49. This is somewhat surprising because females modified their education by a much smaller fraction in response to a change in the sex ratio in the estimates presented above. The reason for this result is that despite the fact that they have barely modified their behavior, they are now facing men who have changed their behavior substantially ($\frac{\partial i^{mO}}{\partial z}$ is large and positive). In response to this females would want to decrease their investment decision by a large fraction. To match the small decrease observed in the data, men and women must be fairly high complements in the

[24]Formally, the parameters selected minimize the sum of the squared errors in the two equations. Imposing that the first order condition holds with equality and then finding the set of parameters which offers the best fit for the comparative static equation offers very similar estimates of α although lower estimates of σ.

production function. Men's effect operates mostly through a change in probability of marrying a native and this leads to similar estimates.

These parameter estimates then imply fairly substantial returns to education when married (between 2 and 5 percent). These estimates suggest that about 40-60 percent of all returns are obtained because of the role education plays within the household production function.

1.6.3 Bargaining power model

For purpose of comparison, let me now assume that the sex ratio only affects one's bargaining power within the household and that the sex ratio has no influence on marital patterns. Consumption levels are now given by

$$\begin{aligned} c_2^m &= i^m + \beta(z)\left(\left(i^{m\alpha} + i^{f\alpha}\right)^{\frac{1}{\alpha}} - i^m - i^f\right) \\ c_2^f &= i^f + (1-\beta(z))\left(\left(i^{m\alpha} + i^{f\alpha}\right)^{\frac{1}{\alpha}} - i^m - i^f\right). \end{aligned} \tag{1.12}$$

Further assume that the sharing factor is $\beta(z) = \exp((\ln \lambda) * z)$, $\quad \lambda < 1$. This is a suitable parameterization since it implies that $\beta(0) = 1$ and $\beta(\infty) = 0$.

Using this framework, the first order conditions for each gender is given by

$$(w - i^m)^{-\sigma} = (c_2^m)^{-\sigma}\left(\frac{\partial c_2^m}{\partial i^m}\right) \tag{1.13}$$

and the effect of the sex ratio on investments given by

$$\frac{\partial i^m}{\partial z} = \frac{(c_2^m)^{-\sigma-1}\left[\left(-\sigma\frac{\partial c_2^m}{\partial i^m}\frac{\partial c_2^m}{\partial z} + c_2^m\frac{\partial^2 c_2^m}{\partial i^m \partial z}\right) + \left(-\sigma\frac{\partial c_2^m}{\partial i^m}\frac{\partial c_2^m}{\partial i^f} + c_2^m\frac{\partial^2 c_2^m}{\partial i^m \partial i^f}\right)\frac{\partial i^f}{\partial z}\right]}{SOC^m}. \tag{1.14}$$

Equations (1.13) and (1.14) are then used as above to find the parameter values of σ and α most consistent with the empirical results obtained.

Although not shown here, the sex ratio increases the investment levels of the spouses of both second generation males and females. Because of this, the direct effect of the sex ratio on male investment in this setting is partially counterbalanced by the fact that they are now paired with higher investment females. Females, on the other hand, experience both the direct and the spousal effect in the same direction. Because of this, the algorithm must estimate men and women's investment to be fairly complementary similar to the case where only matching was at play.

Estimates of the key parameters are shown at the bottom of Table 1.10. The first two panels include the estimates obtained from individual calibrations for both men and women. Those results suggest that the parameter of the CES function is fairly close to 0.4 which implies fairly high rates of return to education in the marriage market. Although obtained from different sets of estimates, the results for men and women are surprisingly consistent with each other. Panel C thus solves simultaneously for the four sets of restrictions and finds similar results, except for the case where men have more bargaining power where the returns to education in the marriage market are estimated to be much larger. On average, results from Table 1.10 imply that marriage market returns to education are of the order of about 2-5 percent and correspond to about 50 percent of total returns with slightly larger shares for women than for men. Thus, this model, while making very different assumptions, generates very similar results to the ones presented above where no bargaining power was assumed.

Furthermore, one can, in this case, compare the level of investment observed (about 9.5 years of schooling for both men and women) to the one that would maximize the sum of male and female utility if education decisions could be made jointly. The optimal estimated investment levels, in this case, are well above the observed level. This implies that individuals invest less than would be optimal because of the "public good" nature of the household production. Also, the use of education as a bargaining tool which would lead to over-investment is dominated by this effect. However, this is driven by the assumption of Nash Bargaining, which imposes that returns to investments are lower than optimal.

Despite the fact that these estimates stem from calibrations and would benefit greatly from being refined using more empirically-based techniques, they nevertheless point to substantial returns to education related to the marriage market. Furthermore, they emphasize that household production may produce human capital externalities, a channel that is yet to be explored.

1.7 Conclusions

Overall, the results of this paper suggest that substantial returns to education derived from the marriage market, whether it be through household production or bargaining power effects. Furthermore, men and women appear to understand, and respond to, the incentives for human capital accumulation embedded in marriage market conditions.

I first derive a framework to explore the relationship between marriage market conditions and pre-marital investments. The model shows that the usual effect of bargaining power on investment within the incomplete contracts framework does not hold for fairly reasonable values of the elasticity of intertemporal substitution. In addition, because sex ratios may affect both bargaining power and matching patterns, a change in the sex ratio would have a similar effect on pre-marital investments under both a unitary framework and a bargaining model. Finally, the model highlights that using marriage market conditions as proxies for ex-post bargaining may not be appropriate as post-marital outcomes will be influenced by marriage market conditions even without any effect through a change in bargaining power.

Empirical support for these conclusions was found in the data. Using shocks to one's marriage market coming from immigration, this paper shows that an increase in the sex ratio increases pre-marital investments for males and lowers them for females, although only significantly so for males. This is confirmed by a variety of outcome measures and is robust to changes in specifications. It appears to stem, at least partially, from changes in bargaining power. In addition, the magnitude of these shifts combined with estimates of the effect of education on labor supply suggest that the interpretation of the effect of marriage market conditions on post-marital outcomes is difficult, in particular for the case of males. Finally, these empirical estimates merged with the structure of the model suggest that returns to education in the marriage market are substantial.

These results provide interesting insights into the determinants of educational decisions. The importance of incentives linked to returns received once married may partially explain why the educational gap by gender is not always correlated with either difference in labor force attachment or differences in wages between men and women. Furthermore, while conventional wisdom maintains that women's educational attainment will increase as their bargaining power in developing countries, the results of this paper indicate that this may not be the case.

The conclusions of this paper also suggest that our understanding of the household would be enhanced by a more careful analysis of how marriage market conditions may affect both the process of household formation and pre-marital decisions as well as post-marital outcomes. For example, while divorce laws have been previously envisaged as strong determinants of ex-post bargaining power within the household, little is known about how these may modify matching patterns and other decisions undertaken before the union is formed. More research is warranted. Furthermore, the fact that marriage market conditions may affect post-marital outcomes through modifications in pre-marital conditions even when no post-matching bargaining occurs cautions the use of such measures as tests of the unitary framework.

Finally, these findings may also shed light on the persistence of skewed sex ratios. Willis (1999), for example, suggests that out-of-wedlock births may be more likely when the sex ratio is lower and when men have fewer economic opportunities. This has been used to explain the high rates of single motherhood among inner city African-Americans where the bias in the marriage market sex ratio in favor of males is due to high male incarceration rates and fast population growth (Guttentag and Secord 1983). This paper suggests that this low sex ratio would lead males to invest less in their human capital and thus offer them even worse economic outcomes. This would reinforce the existing gap between male and female economic opportunities and thus generate even worse marriage pools for African-American females. Similarly, markedly high sex ratios in the context of Asian countries would be predicted to induce lower educational attainment by females. Because of this, parents may be less likely to rely on their girls for future economic support and this could reinforce a pre-existing cultural bias for boys. These are fruitful topics left for further research.

1.8 References

Acemoglu, D. (1996). A microfoundation for social increasing returns in human capital accumulation. *The Quarterly Journal of Economics 111*(3), 779–804.

Amuedo-Dorantes, C. and S. Grossbard (2007). Cohort-level sex ratio effects on women's labor force participation. *Review of Economics of the Household 5*(3), 249–278.

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? Evidence from America's second generation. *The Quarterly Journal of Economics 117*(3), 997–1038.

Aydemir, A. and G. Borjas (2006). Attenuation bias in measuring the wage impact of immigration. Sabanci University working paper series.

Beaudry, P. and E. van Wincoop (1996). The intertemporal elasticity of substitution: An exploration using a US panel of state data. *Economica 63*(251), 495–512.

Behrman, J. R., A. D. Foster, M. R. Rosenzweig, and P. Vashishtha (1999). Women's schooling, home teaching, and economic growth. *The Journal of Political Economy 107*(4), 682–714.

Benham, L. (1974). Benefits of women's education within marriage. *The Journal of Political Economy 82*(2), S57–S71.

Bergstrom, T., L. Blume, and H. Varian (1986). On the private provision of public goods. *Journal of Public Economics 29*(1), 25–49.

Boulier, B. L. and M. R. Rosenzweig (1984). Schooling, search, and spouse selection: Testing economic theories of marriage and household behavior. *The Journal of Political Economy 92*(4), 712–732.

Brainerd, E. (2006). Uncounted costs of World War II: The effect of changing sex ratios on marriage and fertility of Russian women. unpublished mimeo, available at`http://urban.hunter.cuny.edu/RePEc/seminar/old/rfwomen.pdf`.

Browning, M. and P.-A. Chiappori (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica 66*(6), 1241–1278.

Card, D. (2001). Immigrant inflows, native outflows, and the local market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Chiappori, P.-A., B. Fortin, and G. Lacroix (2002). Marriage market, divorce legislation, and household labor supply. *The Journal of Political Economy 110*(1), 37–72.

Chiappori, P.-A., M. F. Iyigun, and Y. Weiss (2007). Public goods, transferable utility and divorce laws. unpublished mimeo, available at `ftp://repec.iza.org/RePEc/Discussionpaper/dp2646.pdf`.

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

Foster, A. D. and M. R. Rosenzweig (1996). Technical change and human-capital returns and investments: Evidence from the green revolution. *American Economic Review 86*(4), 931–53.

Grossman, S. J. and O. D. Hart (1980). Takeover bids, the free-rider problem, and the theory of the corporation. *The Bell Journal of Economics 11*(1), 42–64.

Groves, E. R. and W. F. Ogburn (1928). *American Marriage and Family Relationships.* Henry Holt & Co.

Guttentag, M. and P. F. Secord (1983). *Too Many Women? The Sex Ratio Question.* Newbury Park, California: Sage Publications.

Haines, M. R. (1996). Long term marriage patterns in the United States from colonial times to the present. available at `http://www.nber.org/papers/h0080.pdf`.

Hall, R. E. (1988). Intertemporal substitution in consumption. *The Journal of Political Economy 96*(2), 339–357.

Iyigun, M. and R. P. Walsh (2007). Building the family nest: Premarital investments, marriage markets, and spousal allocations. *Review of Economic Studies 74*(2), 507–535.

Kvasnicka, M. and D. Bethmann (2007). World war II, missing men, and out-of-wedlock childbearing. SFB 649 Discussion Paper 2007-053.

Landale, N. S. and S. E. Tolnay (1993). Generation, ethnicity, and marriage: Historical patterns in the Northern United States. *Demography 30*(1), 103–126.

Lleras-Muney, A. (2002). Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. *Journal of Law & Economics 45*(2), 401–35.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the U.S. labor market. *The Quarterly Journal of Economics 118*(2), 549–599.

Neuman, S. and A. Ziderman (1992). Benefits of women's education within marriage: Results for Israel in a dual labor market context. *Economic Development and Cultural Change 40*(2), 413–424.

Nguyen, T. (2007). Information, role models and perceived returns to education: Experimental evidence from Madagascar. available at `http://econ-www.mit.edu/grad/trang/research`.

Oreffice, S. and B. Bercea (2006). Quality of available mates, education and intra-household bargaining power. Fondazione Eni Enrico Mattei Working Paper, available at `http://www.feem.it/NR/rdonlyres/23FB2DA0-9D56-4623-919E-EF744D3B8D34/2157/13308.pdf`.

Pagnini, D. L. and S. P. Morgan (1990). Intermarriage and social distance among U.S. immigrants at the turn of the century. *The American Journal of Sociology 96*(2), 405–432.

Pena, X. (2006). Assortative matching and the education gap. University of Georgetown mimeo, available at http://www12.georgetown.edu/students/xp/matching.pdf.

Peters, M. and A. Siow (2002). Competing premarital investments. *The Journal of Political Economy 110*(3), 592–608.

Porter, M. (2007). Imbalance in China's marriage market and its effect on household bargaining.

Tiefenthaler, J. (1997). The productivity gains of marriage: Effects of spousal education on own productivity across market sectors in Brazil. *Economic Development and Cultural Change 45*(3), 633–650.

United States Census Office (1901). *Census Reports, Twelfth Census of the United States 1900*, Volume 1. United States Census Office.

Vissing-Jorgensen, A. and O. P. Attanasio (2003). Stock-market participation, intertemporal substitution, and risk-aversion. *American Economic Review 93*(2), 383–391.

Wells, R. and M. Maher (1998). Time and surplus allocation within marriage. Massachusetts Institute of Technology Working Paper.

Willis, R. J. (1999). A theory of out-of-wedlock childbearing. *The Journal of Political Economy 107*(6), S33–S64.

1.9 Tables and figures

Table 1.1: **Summary statistics-Outcomes**

	Males			Females		
	N. Obs.	Mean	Sd	N. Obs.	Mean	Sd
General characteristics						
Age	203954	40.31	16.66	219800	41.26	17.22
Mother foreign	203954	0.79	0.41	219800	0.78	0.41
Father foreign	203954	0.88	0.32	219800	0.88	0.32
Pre-marital investments						
Literate	59946	0.99	0.10	62272	0.99	0.09
Duncan Index	59946	19.99	19.64	62272	15.82	22.48
Wage Index	59946	16.71	12.37	62272	9.14	11.15
Years of education	109252	9.44	3.34	121352	9.45	3.08
Labor supply						
Employed	127283	0.81	0.39	157528	0.30	0.46
In the labor force	144008	0.86	0.34	139844	0.29	0.45
Hours last week	105985	34.61	20.95	119725	10.73	18.06
Weeks last year	107553	39.53	18.61	119752	14.31	21.41
Marital status						
Never married (after 35)	113709	0.12	0.32	123310	0.11	0.32
Divorced	203954	0.02	0.14	219800	0.02	0.14
Widowed	203954	0.03	0.16	219800	0.11	0.31
Married more than once	87070	0.09	0.29	104453	0.09	0.29
Age at first marriage (older than 35)	48712	26.92	6.33	59017	23.33	5.80
Married to an immigrant of own ethnic group	203679	0.03	0.18	219564	0.08	0.26

All summary statistics are weighted by Census sample-line weights

Table 1.2: **Summary statistics-Marriage market conditions and instrument**

Ethnic group	**Immigrants**		**Foreign stock**		**Instrumented**	
	Flow (000s)	Sex ratio	Flow (000s)	Sex ratio	Flow (000s)	Sex ratio
British Ancestry	0.138	0.941	0.605	0.918	0.124	0.851
	[0.136]	[0.546]	[0.483]	[0.158]	[0.108]	[0.067]
Francophone	0.029	1.012	0.112	0.948	0.027	0.882
	[0.033]	[0.734]	[0.102]	[0.424]	[0.030]	[0.174]
South Europeans	0.252	1.536	0.451	1.347	0.256	1.427
	[0.294]	[1.118]	[0.396]	[1.010]	[0.317]	[0.365]
Hispanics	0.129	1.306	0.195	1.174	0.162	1.375
	[0.113]	[0.783]	[0.163]	[0.594]	[0.180]	[0.185]
Scandinavian	0.059	1.724	0.274	1.142	0.069	1.330
	[0.056]	[1.425]	[0.239]	[0.347]	[0.075]	[1.127]
Germanic	0.183	1.404	0.704	1.038	0.177	1.262
	[0.220]	[0.697]	[0.555]	[0.260]	[0.225]	[0.104]
Russians and others	0.407	1.068	0.707	1.065	0.399	1.044
	[0.495]	[0.429]	[0.635]	[0.314]	[0.520]	[0.271]
Other Europe	0.095	1.410	0.175	1.329	0.093	1.252
	[0.101]	[1.337]	[0.131]	[1.215]	[0.112]	[0.643]
Other countries	0.043	2.609	0.058	2.191	0.036	2.257
	[0.039]	[2.249]	[0.055]	[1.800]	[0.040]	[1.323]

Standard deviations in brackets. All summary statistics are weighted by the size of the foreign stock in each cell.

Table 1.3: **First stage**

	Sex ratio of immigrants (1)	Sex ratio of foreign stock (2)	Flow of immigrants (3)	Flow of foreign stock (4)
Predicted sex ratio of immigrants	0.835***	0.434***	-0.011	-0.007
	(0.179)	(0.084)	(0.007)	(0.023)
Predicted flow of immigrants	0.165	0.051	0.868***	0.573***
	(0.199)	(0.069)	(0.041)	(0.059)
N. Obs	2343	2343	2343	2343
R-squared	0.379	0.362	0.986	0.967

Standards errors clustered at the state level in parentheses. All regressions include state, ethnic groups, immigration period fixed effects and all double interactions. All regressions are weighted by the size of the total foreign stock in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 1.4: **First stage-Robustness checks for immigrant measures**

	1910-1929 (1)	1900-1919 (2)	In logs (3)	Without NY (4)	No missing (5)	Census table (6)
Panel A: Sex ratio						
Predicted sex ratio	0.916***	0.798***	0.693***	0.867***	0.867***	0.571**
	(0.246)	(0.197)	(0.126)	(0.187)	(0.159)	(0.201)
Predicted flow	0.130	0.098	-0.046	0.619	0.170	-0.100
	(0.212)	(0.226)	(0.078)	(0.403)	(0.241)	(0.142)
R-squared	0.273	0.442	0.529	0.360	0.366	0.341
Panel A: Flow						
Predicted sex ratio	-0.010	-0.013	0.013	-0.013*	-0.012	-0.009
	(0.006)	(0.012)	(0.194)	(0.006)	(0.009)	(0.023)
Predicted flow	0.907***	0.846***	0.466***	0.801***	0.861***	0.648***
	(0.034)	(0.065)	(0.104)	(0.138)	(0.045)	(0.035)
R-squared	0.989	0.986	0.966	0.951	0.984	0.967
N. Obs	1556	1606	1748	2289	1748	1909

Standards errors clustered at the state level in parentheses. All regressions include state, ethnic groups, immigration period fixed effects and all double interactions. All regressions are weighted by the size of the total foreign stock in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 1.5: **Marriage market outcomes**

	Ever married	Divorced	Ever married twice	Married to own ethnic immigrant	Age at first marriage
	(1)	(2)	(3)	(4)	(5)
			Immigrants		
Panel A: Males					
Sex ratio of immigrants	0.033*	0.005	0.026*	-0.017	-0.027
	(0.014)	(0.003)	(0.011)	(0.009)	(0.371)
Flow of immigrants	0.002	-0.003	0.004	0.020*	0.881**
	(0.013)	(0.003)	(0.012)	(0.009)	(0.296)
N. Obs	203954	203954	87070	203679	48712
Panel B: Females					
Sex ratio of immigrants	0.001	-0.002	-0.004	-0.002	-0.009
	(0.009)	(0.005)	(0.010)	(0.015)	(0.301)
Flow of immigrants	0.008	-0.001	-0.010	0.026*	-0.463*
	(0.011)	(0.004)	(0.016)	(0.011)	(0.210)
N. Obs	219800	219800	104453	219564	60276
			Foreign Stock		
Panel A: Males					
Sex ratio of foreign stock	0.081*	0.011	0.145	-0.050*	-0.036
	(0.034)	(0.010)	(0.094)	(0.024)	(1.091)
Flow of foreign stock	0.020	-0.003	0.011	0.035	1.700**
	(0.034)	(0.006)	(0.028)	(0.018)	(0.590)
N. Obs	203954	203954	87070	203679	48712
Panel B: Females					
Sex ratio of foreign stock	0.003	-0.007	-0.012	-0.006	-0.063
	(0.029)	(0.015)	(0.032)	(0.051)	(1.023)
Flow of foreign stock	0.016	-0.003	-0.021	0.048*	-0.817*
	(0.021)	(0.007)	(0.031)	(0.022)	(0.369)
N. Obs	219800	219800	104453	219564	60276

Standards errors clustered at the state level in parentheses. All regressions include state, ethnic groups, immigration period fixed effects and all double interactions. Also includes age, year of birth and year of Census fixed effects and dummies for parents' nativity. All regressions are weighted by the Census sample-line weight.

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

Table 1.6: **Pre-marital investments**

	Literacy (1)	Duncan Index (2)	Wage Index (3)	Highest Grade (4)	Literacy (5)	Duncan Index (6)	Wage Index (7)	Highest Grade (8)
				Immigrants				
Panel A: Males		**OLS**				**IV**		
Sex ratio of immigrants	0.000	-0.015	0.085	0.026	0.017*	2.351*	0.788	0.443*
	(0.000)	(0.130)	(0.069)	(0.017)	(0.008)	(0.971)	(0.398)	(0.173)
Flow of immigrants	0.003	2.761*	1.382**	0.049	0.001	1.887	0.960	-0.039
	(0.005)	(1.224)	(0.445)	(0.145)	(0.007)	(1.684)	(0.614)	(0.208)
N. Obs	59946	59946	59946	109252	59946	59946	59946	109252
Panel B: Females								
Sex ratio of immigrants	-0.001*	-0.116	-0.106*	-0.001	0.008	-0.596	-0.696	-0.052
	(0.000)	(0.082)	(0.049)	(0.010)	(0.010)	(0.697)	(0.490)	(0.085)
Flow of immigrants	-0.003	3.700**	1.821*	-0.037	-0.002	3.562*	2.077**	-0.101
	(0.005)	(1.529)	(0.749)	(0.157)	(0.007)	(1.401)	(0.758)	(0.207)
N. Obs	62272	62272	62272	121352	62272	62272	62272	121352
				Foreign Stock				
Panel A: Males		**OLS**				**IV**		
Sex ratio of foreign stock	0.000	0.387	0.226	-0.037	0.039	5.542	1.862	1.789*
	(0.001)	(0.200)	(0.118)	(0.032)	(0.021)	(2.968)	(1.147)	(0.874)
Flow of foreign stock	-0.009	0.785	0.682	-0.284*	0.004	3.793	1.895	-0.022
	(0.006)	(1.121)	(0.662)	(0.133)	(0.011)	(2.490)	(0.980)	(0.578)
N. Obs	59946	59946	59946	109252	59946	59946	59946	109252
Panel B: Females								
Sex ratio of foreign stock	-0.001	0.412	0.0446	-0.007	0.051	-3.484	-4.333	-0.164
	(0.003)	(0.556)	(0.250)	(0.027)	(0.062)	(4.454)	(3.345)	(0.282)
Flow of foreign stock	0.001	-0.677	0.508	-0.291	-0.002	6.682*	3.817*	-0.212
	(0.004)	(1.027)	(0.597)	(0.168)	(0.010)	(3.145)	(1.701)	(0.398)
N. Obs	62272	62272	62272	121352	62272	62272	62272	121352

Standards errors clustered at the state level in parentheses. All regressions include state, ethnic groups, immigration period fixed effects and all double interactions. Also includes age, year of birth and year of Census fixed effects and dummies for parents' nativity. All regressions are weighted by the Census sample-line weight.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 1.7: **Pre-marital investments-Robustness checks for males**

	All	Excluding oldest cohort	Excluding youngest cohort	Without NY	With ethnicity dummies	Younger than 65	Excluding 1940
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Highest grade attained			
Sex ratio of immigrants	0.443*	0.409*	0.538	0.470**	0.510*	0.431*	0.465**
	(0.173)	(0.170)	(0.525)	(0.165)	(0.224)	(0.177)	(0.164)
Flow of immigrants	-0.039	0.036	-0.336	0.584	-0.028	-0.130	0.182
	(0.208)	(0.220)	(0.307)	(0.422)	(0.186)	(0.208)	(0.221)
N. Obs	109252	101195	79446	89533	109252	86807	92703
				Duncan Index			
Sex ratio of immigrants	2.351*	2.390*	5.576*	1.809	2.768*		
	(0.971)	(0.935)	(2.647)	(0.949)	(1.114)		
Flow of immigrants	1.887	1.725	-1.254	-2.111	1.755		
	(1.684)	(1.661)	(2.588)	(2.861)	(1.715)		
N. Obs	59946	49068	53005	49818	59946		

Standards errors clustered at the state level in parentheses. All regressions include state, ethnic groups, immigration period fixed effects and all double interactions. Also includes age, year of birth and year of Census fixed effects and dummies for parents' nativity. All regressions are weighted by the Census sample-line weight.

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

Table 1.8: **Labor supply**

	In LF (1)	Employed (2)	Hours (3)	Weeks (4)	In LF (5)	Employed (6)	Hours (7)	Weeks (8)
				Immigrants				
Panel A: Males		**OLS**				**IV**		
Sex ratio of immigrants	0.002	0.002	0.145*	0.145*	-0.003	-0.004	0.534	0.523
	(0.001)	(0.001)	(0.065)	(0.065)	-0.007	(0.009)	(0.690)	(0.359)
Flow of immigrants	0.012	0.009	1.504*	1.504*	0.004	0.003	1.412	0.494
	(0.014)	(0.012)	(0.616)	(0.616)	(0.012)	(0.012)	(0.735)	(0.773)
N. Obs	144008	127283	105985	107553	144008	127283	105985	107553
Panel B: Females								
Sex ratio of immigrants	0.001	0.001	0.038	0.004**	-0.038	-0.028	-1.387	-1.308
	(0.002)	(0.002)	(0.062)	(0.001)	(0.019)	(0.018)	(0.827)	(0.771)
Flow of immigrants	-0.003	-0.007	0.579	0.001	-0.003	-0.005	0.520	0.367
	(0.010)	(0.011)	(0.710)	(0.023)	(0.015)	(0.015)	(0.846)	(1.142)
N. Obs	157528	139844	119725	119752	157528	139844	119725	119752
				Foreign Stock				
Panel A: Males		**OLS**				**IV**		
Sex ratio of foreign stock	0.002	0.001	-0.158	-0.000	-0.009	-0.015	2.504	2.242
	(0.002)	(0.003)	(0.225)	(0.004)	(0.021)	(0.027)	(3.755)	(1.943)
Flow of foreign stock	-0.014	-0.012	-0.433	-0.020	0.007	0.005	2.993	1.053
	(0.011)	(0.011)	(0.822)	(0.024)	(0.022)	(0.021)	(1.758)	(1.713)
N. Obs	144008	127283	105985	107553	144008	127283	105985	107553
Panel B: Females								
Sex ratio of foreign stock	-0.001	-0.001	-0.272	-0.003	-0.109*	-0.084	-4.398	-4.113
	(0.004)	(0.004)	(0.181)	(0.006)	(0.051)	(0.046)	(2.323)	(2.149)
Flow of foreign stock	0.008	0.008	1.052	0.000	-0.018	-0.018	0.775	0.481
	(0.013)	(0.014)	(0.679)	(0.015)	(0.019)	(0.020)	(1.243)	(1.806)
N. Obs	157528	139844	119725	119752	157528	139844	119725	119752

Standards errors clustered at the state level in parentheses. All regressions include state, ethnic groups, immigration period fixed effects and all double interactions. Also includes age, year of birth and year of Census fixed effects and dummies for parents' nativity. All regressions are weighted by the Census sample-line weight.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 1.9: **Effect of education on labor supply**

	In LF (1)	Hours (2)	In LF (3)	Hours (4)	In LF (5)	Hours (6)	In LF (7)	Hours (8)
					Full sample			
Panel A: Males	**OLS**		**IV**		**OLS**		**IV**	
Own education	0.007***	0.688***	-0.014	-1.439**	0.006***	0.531***	-0.026	-3.665**
	(0.000)	(0.032)	(0.013)	(0.521)	(0.000)	(0.023)	(0.029)	(1.352)
Spouse's education					0.003***	0.370***	0.024	2.833**
					(0.000)	(0.019)	(0.018)	(1.007)
F-test (instruments): own			11.79***				10.20***	
F-test (instruments): spouse							28.29***	
N. Obs	682362	658532	682362	658532	510355	491396	510355	491396
Panel B: Females								
Own education	0.016***	0.669***	-0.018	-0.511	0.024***	0.966***	-0.030	-1.926
	(0.001)	(0.038)	(0.015)	(0.659)	(0.001)	(0.029)	(0.044)	(1.849)
Spouse's education					-0.010***	-0.345***	0.034	2.546
					(0.001)	(0.025)	(0.051)	(2.159)
F-test (instruments): own			9.40***				16.48***	
F-test (instruments): spouse							7.54***	
N. Obs	655981	647750	655981	647750	499269	492756	499269	492756
					Only second generation Americans			
Panel A: Males	**OLS**		**IV**		**OLS**		**IV**	
Own education	0.003***	0.517***	0.008	-0.575	0.004***	0.473***	-0.005	-3.236
	(0.000)	(0.044)	(0.027)	(1.349)	(0.000)	(0.042)	(0.034)	(2.181)
Spouse's education					0.001***	0.311***	0.017	3.087
					(0.000)	(0.026)	(0.027)	(1.874)
F-test (instruments): own			5.58***				11.31***	
F-test (instruments): spouse							20.44***	
N. Obs	104819	102015	104819	102015	76478	74340	76478	74340
Panel B: Females								
Own education	0.010***	0.380***	-0.045**	-1.695**	0.019***	0.695***	-0.087	-2.664
	(0.002)	(0.058)	(0.014)	(0.535)	(0.002)	(0.057)	(0.044)	(1.553)
Spouse's education					-0.009***	-0.323***	0.091	2.243
					(0.001)	(0.040)	(0.060)	(2.080)
F-test (instruments): own			14.03***				6.05***	
F-test (instruments): spouse							4.75***	
N. Obs	99519	98201	99519	98201	74828	73789	74828	73789

Standards errors clustered at the state level in parentheses. All regressions include state, year of birth and Census year fixed effets as well as age and age squared. All regressions are weighted by the Census sample-line weight.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 1.10: **Calibration results**

	$\lambda = 1/3$			$\lambda = 1/2$			$\lambda = 2/3$		
	w=22	w=26	w=30	w=22	w=26	w=30	w=22	w=26	w=30
					Spouse selection model				
Panel A: Males									
α	0.36	0.32	0.29	0.42	0.37	0.34	0.46	0.41	0.37
σ	4.88	4.96	4.75	4.93	4.83	4.86	4.82	4.98	4.59
Returns in marriage	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.05
Panel B: Females									
α	0.34	0.29	0.26	0.40	0.34	0.30	0.45	0.38	0.33
σ	5.00	4.98	4.77	4.95	4.93	4.99	5.00	5.00	4.99
Returns in marriage	0.04	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.05
					Bargaining model				
Panel A: Males									
α	0.40	0.30	0.27	0.51	0.39	0.33	0.63	0.47	0.39
σ	8.44	7.81	8.32	8.03	8.41	8.03	7.06	8.31	8.40
Returns in marriage	0.04	0.04	0.05	0.02	0.03	0.04	0.02	0.02	0.03
Panel B: Females									
α	0.45	0.36	0.31	0.42	0.34	0.29	0.39	0.31	0.27
σ	8.48	8.08	8.17	8.33	8.29	7.47	8.50	8.17	8.14
Returns in marriage	0.03	0.04	0.04	0.03	0.04	0.04	0.03	0.04	0.04
Panel C: Joint									
α	0.30	0.23	0.20	0.51	0.39	0.33	0.18	0.13	0.11
σ	4.46	3.83	3.67	8.05	8.50	8.50	2.56	2.22	2.00
Male returns in marriage	0.03	0.04	0.05	0.02	0.03	0.04	0.05	0.05	0.05
Female returns in marriage	0.04	0.05	0.05	0.02	0.03	0.04	0.05	0.05	0.05

Figure 1-1: **Sex ratios of Scandinavians by country and state**

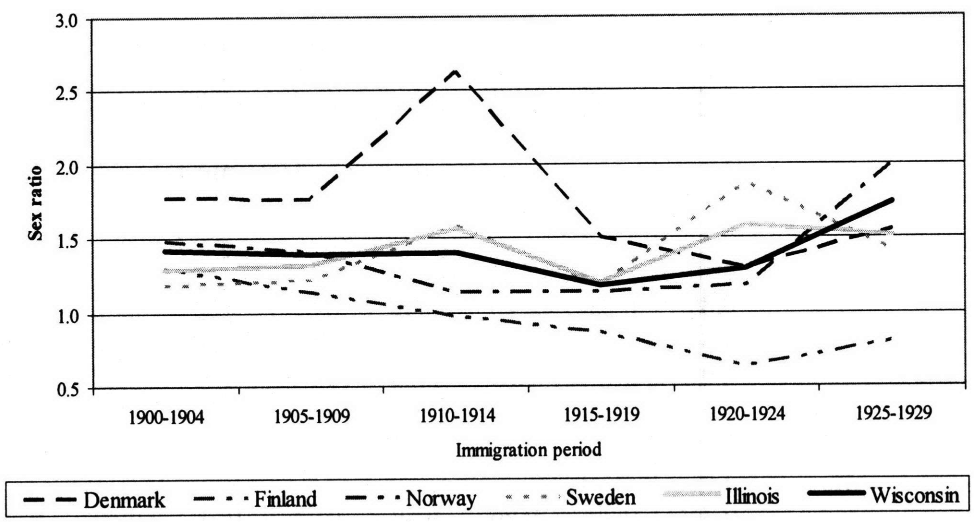
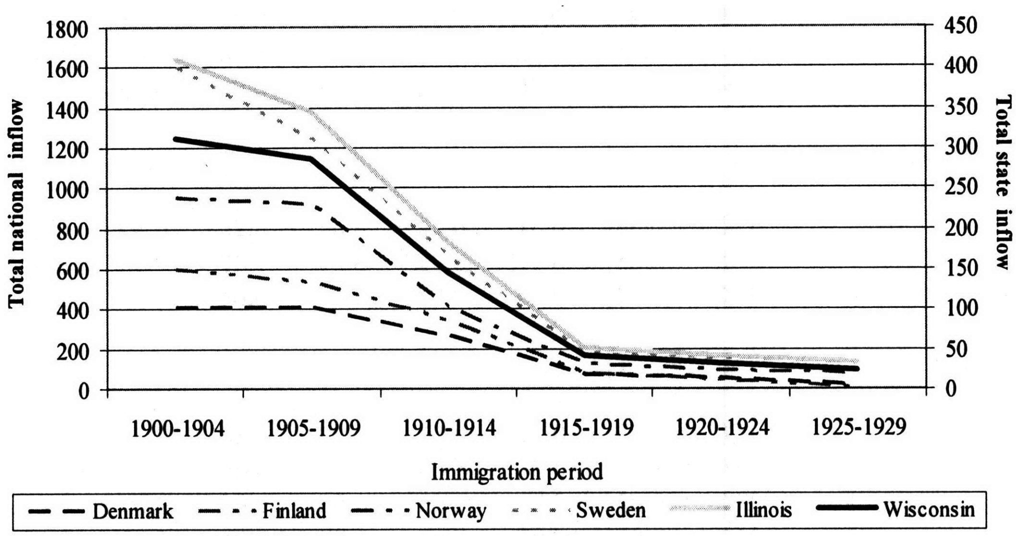


Figure 1-2: **Flows of Scandinavians by country and state**



1.A Omitted proofs of results

Proof of Lemma 1. There will be no strategic behavior between men or between women since the matching is random: thus men do not compete against other men to capture the best females. Assume without loss of generality that the sex ratio is above 1. Thus, $p^m(z) > 0$ and $p^f(z) = 0$. Let $p = p^m(z)$. The first order conditions (1.2) define best response functions $i^f(i^m)$ and $i^m(i^f)$ for both husband and wife.

Define a Nash Equilibrium as

$$i^{f*} = i^f\left(i^m\left(i^{f*}\right)\right)$$

The function

$$\bar{i}^f - i^f\left(i^m\left(\bar{i}^f\right)\right)$$

is strictly increasing iff

$$1 > \frac{\partial i^f}{\partial i^m}\frac{\partial i^m}{\partial i^f}$$

which implies

$$1 > \left(\frac{(1-p)\left(-\sigma\frac{\partial c_2^m}{\partial i^m}\frac{\partial c_2^m}{\partial i^f}+c_2^m\frac{\partial^2 c_2^m}{\partial i^m\partial i^f}\right)}{\sigma\left(\left(\frac{w-i^m}{c_2^m}\right)^{-\sigma-1}+p\left(\frac{i^m}{c_2^m}\right)^{-\sigma-1}\right)+(1-p)\left(\sigma\frac{\partial c_2^m}{\partial i^m}^2-c_2^m\frac{\partial^2 c_2^m}{\partial i^{m2}}\right)}\right)\left(\frac{\left(-\sigma\frac{\partial c_2^f}{\partial i^m}\frac{\partial c_2^f}{\partial i^f}+c_2^f\frac{\partial^2 c_2^f}{\partial i^m\partial i^f}\right)}{\sigma\left(\frac{w-i^f}{c_2^f}\right)^{-\sigma-1}+\left(\sigma\frac{\partial c_2^f}{\partial i^f}^2-c_2^f\frac{\partial^2 c_2^f}{\partial i^{f2}}\right)}\right)$$

$$1 > \frac{(1-p)\left(\sigma^2\frac{\partial c_2^m}{\partial i^m}\frac{\partial c_2^f}{\partial i^f}\frac{\partial c_2^m}{\partial i^f}\frac{\partial c_2^f}{\partial i^m}-\sigma\left(c_2^f\frac{\partial c_2^m}{\partial i^m}\frac{\partial c_2^m}{\partial i^f}\frac{\partial^2 c_2^f}{\partial i^m\partial i^f}+c_2^m\frac{\partial c_2^f}{\partial i^m}\frac{\partial c_2^f}{\partial i^f}\frac{\partial^2 c_2^m}{\partial i^m\partial i^f}\right)+c_2^m c_2^f\frac{\partial^2 c_2^m}{\partial i^m\partial i^f}\frac{\partial^2 c_2^f}{\partial i^m\partial i^f}\right)}{A+(1-p)\left(-\sigma\left(c_2^f\frac{\partial c_2^m}{\partial i^m}^2\frac{\partial^2 c_2^f}{\partial i^{f2}}+c_2^m\frac{\partial c_2^f}{\partial i^f}^2\frac{\partial^2 c_2^m}{\partial i^{m2}}\right)+\sigma^2\frac{\partial c_2^m}{\partial i^m}^2\frac{\partial c_2^f}{\partial i^f}^2+c_2^m c_2^f\frac{\partial^2 c_2^m}{\partial i^{m2}}\frac{\partial^2 c_2^f}{\partial i^{f2}}\right)} \qquad where \quad A > 0$$

This will hold because male and female consumption exhibits constant returns to scale which imply

$$\frac{\partial^2 c_2^m}{\partial i^{m2}}\frac{\partial^2 c_2^f}{\partial i^{f2}} = \left(-\frac{i^f}{i^m}\frac{\partial^2 c_2^m}{\partial i^m\partial i^f}\right)\left(-\frac{i^m}{i^f}\frac{\partial^2 c_2^f}{\partial i^m\partial i^f}\right) = \frac{\partial^2 c_2^m}{\partial i^m\partial i^f}\frac{\partial^2 c_2^f}{\partial i^m\partial i^f}$$

and also

$$c_2^f\frac{\partial c_2^m}{\partial i^m}^2\frac{\partial^2 c_2^f}{\partial i^{f2}} + c_2^m\frac{\partial c_2^f}{\partial i^f}^2\frac{\partial^2 c_2^m}{\partial i^{m2}} = -\left(\frac{i^m}{i^f}c_2^f\frac{\partial c_2^m}{\partial i^m}^2\frac{\partial^2 c_2^f}{\partial i^m\partial i^f} + \frac{i^f}{i^m}c_2^m\frac{\partial c_2^f}{\partial i^f}^2\frac{\partial^2 c_2^m}{\partial i^f\partial i^m}\right)$$

$$= -\left(c_2^f \frac{\partial c_2^m}{\partial i^m} \frac{\partial^2 c_2^f}{\partial i^m \partial i^f} \left(\frac{c_2^m}{i^f} - \frac{\partial c_2^m}{\partial i^f} \right) + c_2^m \frac{\partial c_2^f}{\partial i^f} \frac{\partial^2 c_2^m}{\partial i^f \partial i^m} \left(\frac{c_2^f}{i^m} - \frac{\partial c_2^f}{\partial i^m} \right) \right)$$

$$< c_2^f \frac{\partial c_2^m}{\partial i^m} \frac{\partial c_2^m}{\partial i^f} \frac{\partial^2 c_2^f}{\partial i^m \partial i^f} + c_2^m \frac{\partial c_2^f}{\partial i^m} \frac{\partial c_2^f}{\partial i^f} \frac{\partial^2 c_2^m}{\partial i^m \partial i^f}$$

and by our assumption that

$$\frac{\partial \mu}{\partial i^m} > 0 > \frac{\partial \mu}{\partial i^f}$$

this ensures that

$$\frac{\partial c_2^m}{\partial i^m} \frac{\partial c_2^f}{\partial i^f} > \frac{\partial c_2^m}{\partial i^f} \frac{\partial c_2^f}{\partial i^m}$$

There will be a unique Nash Equilibrium if in addition

$$0 < i^f \left(i^m (0) \right)$$

$$w > i^f \left(i^m (w) \right)$$

which holds since by the concavity of the utility function and the fact that the return to investment is strictly positive, one always wants to invest a strictly positive amount and consume a strictly positive amount in the first period. ■

Proof of Proposition 1. From Equation (1.3) it follows that

$$\frac{\partial i^k}{\partial z} \propto \frac{\partial p(z)}{\partial z} \qquad if \quad i^k < c_2^k \left(\frac{\partial c_2^k}{\partial i^k} \right)^{\frac{-1}{\sigma}}$$

which can be rewritten as

$$\left(\frac{\partial c_2^k}{\partial i^k} \right)^{\frac{1}{\sigma}} < \frac{c_2^k}{i^k}$$

$$\frac{1}{\sigma} \ln \left(\frac{\partial c_2^k}{\partial i^k} \right) < \ln \left(\frac{c_2^k}{i^k} \right)$$

$$\frac{\ln \left(\frac{\partial c_2^k}{\partial i^k} \right)}{\ln \left(\frac{c_2^k}{i^k} \right)} = \tilde{\sigma} < \sigma$$

Using the fact that c_2^k has constant returns to scale

$$\begin{aligned}
\tilde{\sigma} &= \frac{\ln\left(\frac{\partial c_2^k}{\partial i^k}\right)}{\ln\left(\frac{i^k \frac{\partial c_2^k}{\partial i^k} + i^{k'} \frac{\partial c_2^k}{\partial i^{k'}}}{i^k}\right)} \\
&= \frac{\ln\left(\frac{\partial c_2^k}{\partial i^k}\right)}{\ln\left(\frac{\partial c_2^k}{\partial i^k} + \frac{i^{k'}}{i^k}\frac{\partial c_2^k}{\partial i^{k'}}\right)} \\
&< 1
\end{aligned}$$

As long as $\frac{\partial c_2^k}{\partial i^{k'}} > 0$ since $\frac{\partial c_2^k}{\partial i^k} > 1$. ■

Proof of Proposition 2. To show that $\sigma > 1$ is a sufficient condition for

$$\sigma > \bar{\sigma} = \frac{c_2^m \frac{\partial^2 c_2^m}{\partial i^m \partial z}}{\frac{\partial c_2^m}{\partial i^m}\frac{\partial c_2^m}{\partial z}} \tag{1.A.1}$$

Let me use the fact that the consumption functions exhibit constant returns to scale and thus this is equal to

$$\bar{\sigma} = \frac{\left(i^m \frac{\partial c_2^m}{\partial i^m} + i^f \frac{\partial c_2^m}{\partial i^f}\right)\frac{\partial^2 c_2^m}{\partial i^m \partial z}}{\frac{\partial c_2^m}{\partial i^m}\left(i^m \frac{\partial^2 c_2^m}{\partial i^m \partial z} + i^f \frac{\partial^2 c_2^m}{\partial i^f \partial z}\right)}$$

$$\begin{aligned}
\bar{\sigma} &= 1 + \frac{i^f\left(\frac{\partial c_2^m}{\partial i^f}\frac{\partial^2 c_2^m}{\partial i^m \partial z} - \frac{\partial c_2^m}{\partial i^m}\frac{\partial^2 c_2^m}{\partial i^f \partial z}\right)}{\frac{\partial c_2^m}{\partial i^m}\left(i^m \frac{\partial^2 c_2^m}{\partial i^m \partial z} + i^f \frac{\partial^2 c_2^m}{\partial i^f \partial z}\right)} \\
\bar{\sigma} &< 1 \; if \quad \frac{\partial c_2^m}{\partial i^f}\frac{\partial^2 c_2^m}{\partial i^m \partial z} > \frac{\partial c_2^m}{\partial i^m}\frac{\partial^2 c_2^m}{\partial i^f \partial z}
\end{aligned}$$

A similar derivation would lead us to conclude that $\sigma > 1$ is a sufficient condition for $\sigma \frac{\partial c_2^f}{\partial i^f}\frac{\partial c_2^f}{\partial z} < c_2^f \frac{\partial^2 c_2^f}{\partial i^f \partial z}$. ■

Proof of Proposition 3. When an individual is single, he will invest

$$i = \frac{w}{2}$$

When the sex ratio is 1 and each individual is matched with someone identical to them and the surplus is shared equally. The Nash Equilibrium exists and is unique and, given the fact that the production function is symmetric, is equal to

$$-\left(w-i\right)^{-\sigma}+\left(\frac{1}{2}h\left(i,i\right)\right)^{-\sigma}\frac{\partial h\left(i,i\right)}{\partial i}=0 \tag{1.A.2}$$

$$\left(w-i\right)^{-\sigma}=\left(i\frac{\partial h\left(i,i\right)}{\partial i}\right)^{-\sigma}\frac{\partial h\left(i,i\right)}{\partial i}$$

$$i=\frac{w}{1+\left(\frac{\partial h(1,1)}{\partial i}\right)^{\frac{\sigma-1}{\sigma}}}$$

This will be less than $\frac{w}{2}$ if $\sigma>1$ since

$$1+\left(\frac{\partial h\left(1,1\right)}{\partial i}\right)^{\frac{\sigma-1}{\sigma}}>2$$

$$\left(\frac{\partial h\left(1,1\right)}{\partial i}\right)^{\frac{\sigma-1}{\sigma}}>1$$

If the sex ratio is in one's favor ($z>1$ for women, $z<1$ for men), one is match with a partner who invests like a single individual (and thus more than when $z=1$) and can be offered the single individual pay-off. Their first-order condition is given by

$$-\left(w-i\right)^{-\sigma}+\left(h\left(i,i'\right)-i'\right)^{-\sigma}\frac{\partial h\left(i,i'\right)}{\partial i}=0 \tag{1.A.3}$$

Because the second order condition is negative, the solution to (1.A.3) will be a lower investment level than that to (1.A.2) if

$$-\left(w-i\right)^{-\sigma}+\left(\frac{1}{2}h\left(i,i\right)\right)^{-\sigma}\frac{\partial h\left(i,i\right)}{\partial i}>-\left(w-i\right)^{-\sigma}+\left(h\left(i,i'\right)-i'\right)^{-\sigma}\frac{\partial h\left(i,i'\right)}{\partial i}$$

using the fact that $h\left(i^m,i^f\right)$ has constant returns to scale this implies

$$\left(i\frac{\partial h\left(i,i\right)}{\partial i}\right)^{-\sigma}\frac{\partial h\left(i,i\right)}{\partial i}>\left(i\frac{\partial h\left(i,i'\right)}{\partial i}+i'\left(\frac{\partial h\left(i,i'\right)}{\partial i'}-1\right)\right)^{-\sigma}\frac{\partial h\left(i,i'\right)}{\partial i}$$

and because $\frac{\partial h(i,i')}{\partial i'} > 1$ this will hold if

$$i^{-\sigma}\left(\frac{\partial h(i,i)}{\partial i}\right)^{1-\sigma} > i^{-\sigma}\left(\frac{\partial h(i,i')}{\partial i}\right)^{1-\sigma}$$

which will hold if $\sigma > 1$ since $i' > i$ and thus $\frac{\partial h(i,i')}{\partial i} > \frac{\partial h(i,i)}{\partial i}$. ■

1.B Nash bargaining

A special and traditional case of the model presented above is one where the sharing decision is made through Nash bargaining and thus where consumption levels will be given by

$$c_2^m = i^m + \beta(z)\left(h\left(i^m, i^f\right) - i^m - i^f\right)$$
$$c_2^f = i^f + (1-\beta(z))\left(h\left(i^m, i^f\right) - i^m - i^f\right)$$

We will assume that the sharing parameter is influenced by bargaining power. It is easy to show that in this case, returns to investments when single are equal to 1 and thus positive. Define the surplus to be shared as

$$g(i^m, i^f) = h\left(i^m, i^f\right) - i^m - i^f$$

Assume $h(\cdot)$ displays constant returns to scale and thus $g(\cdot)$ also does.

Proposition 4 *There exists a unique pure strategies Nash Equilibrium in this setting.*

Proof. Since g exhibits constant returns to scale in investments, so will $c_2^m$and c_2^f. It is easy to show that conditions (1.1) are satisfied since

$$c_2^m(i^m, 0, z) = i^m \Longrightarrow \mu(i^m, 0, z) = 1$$
$$c_2^f\left(0, i^f, z\right) = i^f \Longrightarrow \mu\left(0, i^f, z\right) = 0$$

and defining

$$c_2^m\left(i^m, i^f, z\right) = \alpha\left(i^m, i^f, z\right) h\left(i^m, i^f\right)$$

where

$$\alpha\left(i^m, i^f, z\right) = \beta(z) + \frac{i^m(1-\beta(z)) - \beta(z)\, i^f}{h(i^m, i^f)}$$

$$\frac{\partial \mu}{\partial i^m} = \frac{\partial \mu}{\partial \alpha}\left(\frac{(1-\beta(z))\left(h\left(i^m, i^f\right) - i^m \frac{\partial h}{\partial i^m}\right) + \beta(z)\, i^f \frac{\partial h}{\partial i^m}}{h(i^m, i^f)^2}\right) > 0$$

$$\frac{\partial \mu}{\partial i^f} = \frac{\partial \mu}{\partial \alpha}\left(\frac{-\beta(z)\left(h\left(i^m, i^f\right) - i^f \frac{\partial h}{\partial i^f}\right) - i^m(1-\beta(z))\frac{\partial h}{\partial i^f}}{h(i^m, i^f)^2}\right) < 0$$

■

A rise in the sex ratio will lead to an increase in men's investments if (1.A.1) holds which in this case implies

$$\sigma > 1 > \min\left\{\frac{\left(i^m + \beta g\right)\frac{\partial g}{\partial i^m}}{\left(1+\beta\frac{\partial g}{\partial i^m}\right) g}, \frac{\left(i^f + (1-\beta) g\right)\frac{\partial g}{\partial i^f}}{\left(1+(1-\beta)\frac{\partial g}{\partial i^f}\right) g}\right\}$$

And this is satisfied since $g\left(i^m, i^f\right)$ exhibits constant returns to scale. Investments will be strategic substitutes if

$$\sigma > \min\left\{\frac{\left(i^m + \beta g\right)\frac{\partial^2 g}{\partial i^m \partial i^f}}{\left(1+\beta\frac{\partial g}{\partial i^m}\right)\frac{\partial g}{\partial i^f}}, \frac{\left(i^f + (1-\beta) g\right)\frac{\partial^2 g}{\partial i^m \partial i^f}}{\left(1+(1-\beta)\frac{\partial g}{\partial i^f}\right)\frac{\partial g}{\partial i^m}}\right\}$$

for which a sufficient condition is

$$\sigma > \frac{1}{\rho_g}$$

where ρ_g is the elasticity of substitution between the two inputs in the surplus function $g\left(i^m, i^f\right)$.

1.C Appendix tables

Table 1.C.1: **Data description**

Variables	Census years	Age sampled	Details
Marital outcomes			
Ever married	1900-70	15+	
Currently married to same ethnic immigrant	1900-70	15+	
Currently divorced	1900-70	15+	
Number of marriages	1910 1940-60	15+	
Age at first marriage	1930-40 1960-70	35+	
Pre-marital investments			
Literacy	1900-30	15-25	Literacy in any language
Highest grade achieved	1940-70	25+	Only available from 1940, when the youngest cohort is 25
Duncan index	1900-30	15-25	Based on a measure of prestige linked to wage and education
Wage index	1900-30	15-25	Based on 1950 wages
Post-marital labor supply			
In the labor force	1910-70	25+	
Employed	1910-10 1930-70	25+	
Hours worked per week	1940-70	25+	Transformed from intervals to a continuous variable by selecting the mid-point of the interval
Weeks worked per year	1940-70	25+	

Table 1.C.2: **Ethnic group composition**

Ethnic group	**Countries of Birth**
British Ancestry	Australia, English Canada, English, Ireland, Scotland and Wales
Francophone	Belgium, French Canada and France
South Europeans	Italy, Spain and Portugal
Hispanics	Mexico, Cuba, Other West Indies, Central America and South America
Scandinavian	Denmark, Finland, Norway and Sweden
Germanic	Austria, Germany, Luxembourg, Netherlands and Switzerland
Russians and others	Russia, Poland and Romania
Other Europe	Bohemia (Czechoslovakia), Greece, Hungary and Other Europe
Other Countries	Africa, Atlantic Islands, China, India, Japan, Other Asia, Pacific Islands, Turkey and Other countries

Table 1.C.3: **Spatial distribution of immigrants by ethnic group, 1900**

Ethnic group	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	TOP 10
British ancestry	NY	MA	PA	IL	MI	NJ	OH	CA	CT	MN	75.4
	(19.4)	(14.5)	(11.2)	(6.9)	(6.6)	(4.5)	(3.9)	(3.4)	(3.0)	(2.1)	
French	MA	MI	NY	NH	IL	RI	ME	WI	CT	NJ	75.6
	(21.9)	(10.7)	(9.2)	(7.0)	(6.8)	(5.1)	(4.9)	(3.6)	(3.4)	(2.8)	
South Europeans	NY	PA	MA	NJ	CA	IL	CT	LA	RI	OH	86.3
	(34.8)	(12.7)	(8.0)	(8.0)	(6.8)	(4.5)	(3.7)	(3.4)	(2.2)	(2.2)	
Hispanics	TX	AZ	FL	CA	NY	NM	PA	MA	LA	NJ	93.9
	(51.9)	(10.3)	(8.6)	(7.4)	(6.5)	(4.9)	(1.2)	(1.2)	(0.9)	(0.9)	
Germanic	NY	IL	PA	WI	OH	NJ	MI	IA	MN	MO	77.1
	(18.7)	(11.8)	(9.4)	(8.4)	(7.4)	(4.6)	(4.4)	(4.2)	(4.2)	(3.9)	
Scandinavians	MN	IL	WI	IA	NY	MI	ND	MA	NE	SD	76.3
	(21.9)	(12.9)	(9.4)	(6.4)	(6.0)	(5.3)	(3.8)	(3.8)	(3.6)	(3.1)	
Russians and	NY	PA	IL	MA	WI	NJ	MI	OH	CT	MN	83.6
others	(29.1)	(15.8)	(12.0)	(6.0)	(4.5)	(4.2)	(4.0)	(3.1)	(2.7)	(2.1)	
Other Europeans	NY	PA	IL	OH	NE	NJ	WI	MN	IA	TX	85.9
	(20.1)	(16.1)	(14.4)	(9.7)	(5.1)	(5.1)	(4.7)	(4.2)	(3.5)	(3.1)	
Other countries	HI	CA	MA	NY	OR	WA	MT	PA	AK	IL	86.4
	(34.6)	(24.5)	(5.6)	(5.5)	(5.3)	(0.42)	(1.9)	(1.9)	(1.5)	(1.4)	

For each ethnic group, the first row represents the states with the highest concentration and the second, the actual concentration in each state. The last column measures the share of all immigrants from that ethnic group located in the ten most popular states for that ethnic group.

Chapter 2

Marry for what? Caste and Mate Selection in Modern India

This chapter is joint with Abhijit Banerjee, Esther Duflo and Maitreesh Ghatak.

2.1 Introduction

Marriage is a crucially important economic decision. In developing countries, where many women do not work, marriage is arguably the single most important determinant of her and her offspring's economic welfare. In India, the setting for this study, several studies have shown that marriage is indeed taken as a very serious economic decision, managed by parents more often than by the prospective spouses.[1] Rosenzweig and Stark (1989) show that parents marry daughters in villages where income co-vary less. Foster and Rosenzweig (2001) show that demand for healthy women in the marriage market influence investments in girls.

Yet, despite the economic importance of this decision, "status"-like attributes, such as castes, continue to play a seemingly crucial role in determining marriage outcomes in India. In a sample of married couples we interviewed in Kolkata in 2005-2006, 70 percent were from the same caste. In a recent opinion poll carried by CNN-IBN (the Indian subsidiary of CNN) in a representative sample 15141 individuals across India, 74 percent of respondents declared to be opposed to

[1]For example The CNN-IBN opinion poll mentioned below found that more than 72% of Indian parents think that parents should have the last say in marriage decisions. 69% oppose dating.

inter-caste marriage. The institution is so prevalent that matrimonial ads in Indian newspapers are classified under caste headings, making it immediately obvious where a prospective brides and groom can find someone from their own caste.

Cole et al. (1992) analyze marriage as a matching institution which gives men the ability to enjoy a non-marketed non-storable endowment which women possess in return for sharing his income with the woman. They show that an "aristocratic equilibrium" can exist, in which both men and women marry based on "status" (a rank which is initially exogenously assigned) rather than on income (on the man's side) and the endowment (on the woman's side). This rank is inherited from father to son as long as a man of a given rank in status marries a woman who is of the same rank. The equilibrium is sustained by the fear that the offsprings of mixed rank couples will lose their status.

The aristocratic equilibrium in this model has a clear similarity to the caste system, where offsprings of an inter-caste couple are supposed to lose their caste.[2] Cole et al. (1992) suggest that this equilibrium may be characterized by low productivity, because the incentive to work hard in order to marry a "high quality" woman is suppressed.

Such an equilibrium will however not exist when the distribution of wealth is such that a low-status/high-wealth person finds it sufficiently profitable to deviate from the social norm and marry a woman with high endowment (the woman may agree in order to consume more) at the cost of their offspring's future status. Economic growth and the diversification of earnings opportunity has significantly lowered the correlation between caste and income in India. In other settings, such as occupational choice, the traditional role of castes is eroding, and there is a distinct tension between the social pressure to continue to act according to caste rules and the incentives provided by the modern world (Munshi and Rosenzweig 2006). Will the same forces also progressively lead to a decline in the role of caste in marriage decisions, as the constraints it imposes become too costly to be sustained in equilibrium? Or to reverse the question, is it the case that the "aristocratic" (caste hierarchic) equilibrium is still in force and constitutes a significant drag on the process of growth?

This paper sheds light on these questions. We analyze an unusual data set on the arranged

[2] The formal rule may be that the children of an inter-caste couple inherit the caste of the father, but in practice, they tend to be discriminated against.

marriage market we collected in Kolkata, the capital of the state in West Bengal, India. We interviewed a sample of 783 people who placed matrimonial ads in the major Bengali newspaper, *Anandabazar Patrika*, which, with its circulation of 1.2 million is the largest circulated single edition daily newspaper in India.[3] All ad-placers are parents who are placing an ad on behalf of their sons or daughter. The sample is representative of the educated urban upper-middle class: 85 percent of both the prospective grooms and brides have a college degree, and average income of 9800 rupees per month compared to 1935 rupees per month for the country at current prices during the year 2004-05. Fathers who report an occupation have on average a log occupational wage of 5.8 compared to the median NSS for formal sector workers of 4.5 in 2004.[4] Only 7 percent of parents are from different castes although about 30 percent of their siblings married someone from another caste.

At the first interview, we collected information on the prospective groom or bride, as well as information on the responses they received to their ad, their subjective ranking of those responses, and with which ones they were planning on following up. We also asked them which ad in the newspaper they were planning to respond to themselves. At a second interview, a year later, we asked them whether they were married or engaged, and the characteristics of their (prospective or actual) spouses if they were married.

The responses received to their ad, the ads they were planning to respond to, and the ranking they gave to the letter they received, provide three independent ways to assess the relative importance given to different attributes (caste, education, beauty, proxies for wealth, etc...). For example, using either a linear probability or a fixed effect logit model, we estimate how the probability that an ad placer decides to give further consideration to a response he received depends on a series of attribute of the ad placer, the response, and the interaction of the two. An advantage of this data set is that the entire information set available to the ad-placer is also available to us (at the time we initially interviewed them, they had just received the letter, and they had not yet met the prospective groom or bride or their parents). A disadvantage is that we do not observe dowries. Dowries are illegal and frowned upon in this group (middle-class

[3]We estimate that its circulation represent about one sixth of the literate bengali speaking population of greater Kolkata.

[4]Central Statistical Organization, 2006.

urban Bengalis), which made it impossible to collect data on them. However, precisely because they are not very frequent in this group, dowries are probably not a very important part of the story.[5] More importantly, even if dowries do play a role as equilibrium prices, our analysis will still be valid. This is because, at the time the respondents decide how to respond to a particular letter or to an ad, they do not yet know what the dowry would be (dowry demands are never mentioned in ads or in the letters the respondent receive –except in the case of 7 percent to 10 percent of men who mention at the outset, in the ad or in the letter, that they will not accept a dowry) only the expected dowry they would have to pay to marry someone with these characteristics.[6] We argue below that this might allow us to recover their true preferences over the observed attributes even if expected dowry (or some other unobserved attribute) is correlated with what they observed.

These alternative ways to estimate the reduced-form preferences for castes versus other attributes lead to very similar qualitative conclusions.[7] Both women and men prefer educated partners. Men prefer women who describe themselves as beautiful or very beautiful, and whose skin tone is lighter. Women prefer men who earn more, or are in higher paying occupations. A striking result is that the preferences for marrying within one's castes appear to be particularly strong: for example, we find in one specification that parents of a prospective bride would be willing to trade off the difference between no education and a master degree to avoid marrying outside their caste. For men seeking brides, it is twice the effect of the difference between a self-described "very beautiful" woman and a self-described "decent looking" one. On the other hand, perhaps surprisingly, we find less clear preference for marrying "above" one's caste, in particular for women (men do seem to have some preference for marrying up).

These results suggest that castes continue to play an extremely important role in structuring

[5] We have so far failed to locate a comprehensive study on dowry in this population. However, we note that while Kolkata has 12% of the population of the largest metropolitan cities in India, it has only 1.9% 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 by her in-laws following negotiation failure about the dowry. To the extent that the prevalence of dowry death is indicative of the prevalence of dowry, it suggests that they are less prevalent in Kolkata than elsewhere.

[6] In this sense, we are in a similar situation as Hitsch et al. (2006) or Fisman et al. (2006), Fisman et al. (2008) who examine dating in the US: when considering whether to date an attractive woman or not, their subjects probably factor in how expensive the meal they will have to pay will be.

[7] We borrow the term "Reduced-form preference" from Cole et al. (1992). It signals the fact that the preference for caste may not be a "deep" preference parameter, but a feature of the equilibrium, where caste serves as a focus point to allocate non-marketed goods.

people's preferences for marriage partners in contemporary India, even among this educated, relatively affluent, group. But does this necessarily mean that caste has a large effect on marital matching? Do people end up marrying someone very different (in terms of attributes other than caste) from those who they would have married absent this regard for caste? In other words do we actually see the distortion in choices that drives the results in Cole et al. (1992) ?

A simple model, developed in section 2, helps clarify what is at issue here. We show that in the case where preferences for caste are primarily "horizontal", in the sense that people care more about marrying someone from the same caste than about marrying "up", preference for in-caste marriage does not change the equilibrium matching patterns as long as castes are "balanced" in the sense (made more precise below) that the distribution of male attributes and female attributes within each caste bear the same relation to the distribution of those attributes in the overall population.[8] This will be true even if the "price" of caste (how much people are willing to give up in terms of partner quality to marry within caste) is very high. The reason is that with horizontal preferences people prefer to marry in caste and by the balanced population assumption anyone they could realistically expect to marry outside their caste, has a corresponding person within their own caste.

By contrast if caste is primarily vertical, then preference for in-caste marriage or marrying up in caste affects the entire pattern of who matches with whom. This is will also be the case if the population is highly unbalanced, because then even though people want to marry within caste, there may not be any suitable candidates available for them to do so.

Since we can estimate preferences we can actually ask whether the situation on the ground is closer to the horizontal preference-balanced population world where preference for caste matching does not "matter" very much in equilibrium, or the vertical preference/unbalanced population world where it does. To do this we use a Gale-Shapley (Gale and Shapley 1962) algorithm to compute the set of stable matches implied by the preferences we estimate (Hitsch et al. 2006 perform the same exercise for the on-line dating market in the US).

Note that the Gale-Shapley algorithm gives us the set of stable matches implied by these preferences under the assumption that utility is not transferable, and therefore that an individual

[8]In other words it is not, for example, the case that all the women from one caste are at the 90th percentile of the population distribution in terms of the relevant attributes while all the men in that caste are at the 30th.

cannot compensate her partner for being a worst match by paying her a higher price. If in reality the families could compensate a prospective partner for a "bad" match along the characteristics we observe with a monetary transfers (i.e. a dowry adjustment), we would observe that the Gale-Shapley set of stable matches do not look at all like the actual matches. In fact, it is encouraging that the set of stable matches approximates fairly well the set of actual marriages we observe in the data, with some exceptions, which we discuss in the paper.

To investigate the role of caste in equilibrium, we perform several exercises with the Gale-Shapley algorithm. First, we compute the set of stable matches ignoring the caste preferences. The percentage of intra-caste marriage drops dramatically (showing that caste is not just a proxy for other characteristics households also care about), but the matches otherwise look very similar to what they were allowing people to match within caste. Second, in the set of stable matches, we regress each characteristic to a dummy for whether the match is "within caste". This gives us an indication of the "equilibrium price" people actually pay to marry within their caste. For none of the characteristics we look at do we see a significant coefficient: this indicate that, in equilibrium, there is no cost to marry within one caste, even though household's willingness to pay to avoid not marrying with the caste is very high. Moreover we observe that these patterns are also observed in the data on actual marriages, though this (unlike what we observe in the data generated by our algorithm) can be driven by unobservables. Finally we demonstrate that this method for estimating the "price" has some power by showing (in the data generated by our algorithm) that men have to pay in terms of other attribute (e.g. beauty) to marry a more educated wife.

Thus, while individuals seem willing to pay large amounts in terms of education, beauty, etc.. to marry within their caste, they do not have to do so in equilibrium. This implies that caste, operating through marriage, is not a significant constraint on marriage as an institution to match people with other characteristics. Moreover this explains why the role of caste in marriage has not been weakened by economic forces - essentially there is no trade-off between economic wellbeing and caste. This implies that the "aristocratic" equilibrium could be quite persistent in this context.

And yet, 30 percent of people in our sample *do not* marry within their caste. They apparently do not gain much by marrying out of caste, so why do they do it? In part, this comes from heterogeneity in caste preferences, with some people preferring to marry outside. But there is something else. A substantial fraction the marriages that are not within caste are "love marriage" (40 percent of the children of our respondent eventually marry through another channel than the ads and 20 percent enter into a "love marriage", meaning that they find their spouses themselves). So the institution that capitalism is not able to destroy may be endangered by love.

The remainder of the paper proceeds as follows. Section 2.2 first sketches a model where caste and other attributes interact on the marriage market. Section 3.4 presents the data while Section 2.4 elaborates on the methodology and the results of preference estimation. Section 2.5 highlights the results of the stable matches and Section 2.6 uses these results to derive conclusions regarding the equilibrium. Finally, Section 3.7 concludes.

2.2 Model

In this section we develop a simple model of marriage. Our goal is to identify some useful properties of the choice problem faced by decision-makers in the marriage market as well as the equilibrium matching pattern, in a world where individuals care about the caste of their partner, as well as some standard characteristics (e.g., education, beauty). These will motivate our empirical analysis and help us interpret some of the results.

A key modeling decision is whether to assume that we are in a non-transferable utility (NTU) environment (as in studies of the US matching market studied for example by Hitsch et al. (2006), Fisman et al. (2006) and Fisman et al. (2008)or the TU environment more traditional in the literature (e.g., Becker 1973, Lam 1988 or more recently, Anderson 2003).[9]

The standard view, mentioned above, is that dowry is not particular important in the pop-

[9]In contrast, to explain the phenomenon of dowry inflation, Anderson (2003) constructs a model where women have a strong preference for marrying in an upper caste (and low caste women are not sensitive to income among high caste men). Dowry inflation follows then from an increase in the heterogeneity of income among men. This assumption does not appear consistent with what we find in this data set. One possibility is that the preference we estimate already discount for the expected dowry payment the family of the brides anticipate they will have to pay if they marry up. Sufficient anticipated dowry payment would make the brides indifferent between higher and lower caste men.

ulation we study–middle-class Bengalis–which inclines towards the NTU approach.[10] This is consistent with the fact that no one in our data asks for a dowry or offers one, but since dowry is both illegal and socially frowned upon, it is hardly surprising. Indeed to the extent that dowry exists in this population it is unlikely to be divulged, and therefore the prevalent view (that dowry is not very important) may be biased. To not entirely foreclose the possibility of transfers, we take the following approach: Our estimation of preferences is based on recording the observable characteristics of those who get chosen (to get a call back or a letter) out of a set of "applicants". We first observe that as long as there enough people who prefer not to demand transfers (a not insignificant part of our sample actually spend money (in the form of ad space) to explicitly mention that they do *not* want a dowry), it makes sense to first choose everyone who you would have chosen *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 another person is small which, given the large numbers people contact, seems plausible. Proposition 1 below makes this argument explicit for the case where there is one unobservable variable (need not be the dowry demand/offer) which is potentially correlated with the observables.

Assuming that the conditions of Proposition 1 hold, what we observe in the data is people's true ordering between those whom they consider and those whom they reject. Based on this ranking we infer people's preferences over a range of attributes. Given these preferences we then construct the standard "equilibrium" of a NTU matching game, namely the Gale-Shapley stable match which we compare with the actual matches we observe. On the whole the model performs well, giving some credence to the NTU assumption. We therefore only model the NTU world, though the possibility of some transfers is implicitly allowed in the formulation of proposition 1.

2.2.1 Set up

Men and women are differentiated by "caste". Men and women are differentiated by "caste". The caste of an individual is $i \in \{1, 2\}$. They are ranked in descending order: $i = 1$ is the highest

[10] Of course the TU environment can be relevant even in the absence of dowries or brideprice, so long as there is some other "currency" which can be used to make ex ante transfers (e.g., household chores, location decision).

caste, followed by $i = 2$. We allow some members of both castes being caste-neutral i.e., they do not put any weight on the caste of their potential partner.

Men and women are assumed to be differentiated according to a "vertical" characteristic that affects their attractiveness to a potential partner. The characteristic of men will be denoted by $x \in [0, B]$ and the characteristic of women will be denoted by $y \in [0, B]$ were $B > 0$. We can think of these as education levels of men and women, or, income and beauty. Other things constant, everyone prefers a higher attribute partner. Following the tradition of Becker, we are also going to allow these characteristics to be complementary in the payoff of men and women.

The payoffs of men and women are both governed by the quality of the match. We assume that this has two (multiplicatively) separable elements, one governed by the vertical characteristics, $f(x, y)$, and the other by caste, $A(i, j)$. We assume that the function $f(x, y)$, is twice continuously differentiable, increasing and concave with respect to both arguments, and a positive cross-partial derivative (i.e., it is supermodular). A standard example would be the Cobb-Douglas: $f(x, y) = x^a y^{1-a}$ where $0 < a < 1$

The function $A(i, j)$ captures the quality of a match for a individual of caste i (man or woman) who is matched with a partner of type j. This is defined as follows:

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

where $\alpha \geq 0$. It is readily verified that so 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 matching function is flexible, and allows there being 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 those within the same caste. Otherwise, the higher the caste of the partner (lower is j) the higher is 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. In the marriage literature, a high β will be viewed as the case of hypergamy and a high γ will be viewed as the case of endogamy.

Therefore we have:

$$\begin{aligned} A(1,1) &= 1+\alpha\beta \\ A(2,2) &= 1 \\ A(1,2) &= 1-\alpha\gamma \\ A(2,1) &= 1+\alpha\beta-\alpha\gamma. \end{aligned}$$

Notice that $A(1,1) > A(2,2)$ and $A(2,1) > A(1,2)$ when $\beta > 0$: otherwise caste preferences are purely horizontal with the same "penalty" $\alpha\gamma$ for any inter-caste marriage. Similarly, if $\gamma = 0$ then one high caste partner in a match raises the payoff from the caste component to $1+\alpha\beta$. We assume $\alpha\gamma < 1$.

We also assume that some members of the population, drawing from both caste-groups, have caste-neutral preferences. That is, for these individuals, $\alpha = 0$. 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 $j = 1,2$. For those who are caste-conscious, they value a caste-neutral individual of caste i $(i = 1,2)$ in the same way as they would a caste-conscious individual of caste i $(i = 1,2)$.

Given these two elements that govern the quality of a match, we assume that the payoff of a man of caste i whose quality is x and who is matched with a woman of caste j whose quality is y is given by:

$$u^M(i,j,x,y) = A(i,j)f(x,y)$$

and correspondingly, the utility of a woman of caste j whose quality is y and who is matched with a man of caste i whose quality is x is given by:

$$u^W(i,j,x,y) = A(j,i)f(x,y).$$

Several observations are in order.

First, we assume that the non-caste component (or, lets say the standard component) of the quality of a match, $f(x,y)$ is the *same* for a man and a woman. This is clearly most relevant to settings where this aspect of a match is a pure public good (e.g, children, joint activities), or in a transferable utility world, where a match generates output that can be perfectly divided.[11]

Second, the caste component and the standard component interact with each other: in particular, a "good" caste-specific match will have higher marginal product of the standard attributes. This formulation allows the two components (caste and non-caste) to be additive as well as multiplicative: e.g., $f(x,y) = 1 + x^a y^{1-a}$. The purely separable case (i.e., $u^G(i,j,x,y) = A(i,j) + f(x,y)$, $G = M, W$) turns out not to be very interesting, as we discuss later.

Third, the caste matching function is symmetric for men and women. That is, a man of caste 1 marrying a woman of caste 2, gets the same payoff that a woman of caste 1 would get from marrying a man of caste 2.

2.2.2 Adding unobserved characteristics

The model focuses on the case where, other than caste, people differ on a single characteristic. It is straightforward to extend it to a vector of characteristics for each gender. However, as noted above, there may be other (payoff relevant) characteristics (such as demand for dowry) that are not observed by the parties at this stage. For example, suppose men are also differentiated by the characteristic z which is not observed at the first stage (i.e., in terms of our data, in the ad or in the response letter), which is 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.

For a simple illustration, 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:

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

[11] In a NTU world, if men and women get very different payoffs from the standard component of a match, it is hard to provide much in the way of characterization. In any case, our results go through if men and women put different weights on the standard component of a match but these weights are not very different.

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 ($f(x', y)\{pH + (1-p)L\} - c$ 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 5 *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 reflect 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.

2.2.3 The price of caste

In the data we observe the trade-offs people make between caste and other observables in selecting the set of people they are prepared to explore further. Here we want to develop a simple notion of the "price" of caste that corresponds to this trade-off, i.e., the extent of partner quality one is willing to give up to marry within caste. Consider a man of type x who belongs to caste 1. Suppose the best match he has is a woman of quality y from his own caste. Then he is indifferent between marrying a woman of quality y within his own caste and a woman of caste 2 if the attribute of this woman is higher by the margin ε given by:

$$(1+\alpha\beta)f(x,y) = (1-\alpha\gamma)f(x,y+\varepsilon). \tag{2.1}$$

We can solve $\varepsilon(x,y,\beta,\gamma)$ from this equation. This can be interpreted as the "supply" price of caste: this is the price at which a high caste person (here, a man) will agree to marry a low caste person.

For $\alpha = 0$, the supply price of caste is zero. Lets consider the case where $\alpha > 0$. Clearly, $\varepsilon(x,y,\beta,\gamma)$ is increasing in β and γ. It is also increasing in y : if a person already has an attractive match within his own caste (by concavity of $f(x,y)$ with respect to y) the quality differential has to be large for this person to want to marry inter-caste. We need more structure on the function $f(x,y)$ to characterize the effect of x on $\varepsilon(x,y,\beta,\gamma)$. Totally differentiating (2.1), we obtain:

$$\frac{\partial\varepsilon}{\partial x} = \frac{f(x,y+\varepsilon)}{f(x,y)}\left[\frac{\frac{\partial f(x,y+\varepsilon)}{\partial x}}{f(x,y+\varepsilon)} - \frac{\frac{\partial f(x,y)}{\partial x}}{f(x,y)}\right].$$

The sign of this expression depends on whether $\frac{\frac{\partial f(x,y)}{\partial x}}{f(x,y)}$ is increasing in y or not, i.e.,

$$\frac{f(x,y)f_{xy}}{f_x f_y} - 1 \gtreqless 0.$$

If $f(x,y) = [x^\rho + y^\rho]^{\frac{1}{\rho}}$ with $\rho \leq 1$ (i.e., a member of the CES family) then ε is non-decreasing in x so long as $\rho \leq 1$ (i.e., x and y are not very substitutable).[12] If the function $f(x,y)$ is

[12]Recall that $\rho = 1$ implies x and y are perfect substitutes, $\rho = 0$ is the case of the Cobb-Douglass, and $\rho \to -\infty$ implies x and y are perfect complements (Leontief).

multiplicatively separable, then it directly follows that ε is independent of x.

Now let us consider a woman of type y' who belongs to caste 2. Suppose the best match she can find in her own caste group is x'. Then she is indifferent between marrying a man of quality x' within her own caste and a man of caste 1 if the attribute of this man is not lower than the margin δ :

$$(1+\alpha\beta-\alpha\gamma)f(x'-\delta,y')=f(x',y').$$

We can solve $\delta(x',y',\beta,\gamma)$ from this equation. This can be interpreted as the "demand" price of caste: this is the price a person of low caste is willing to pay to marry a higher caste person. As before, for $\alpha=0$, the demand price of caste is 0.

Clearly, for $\alpha>0$, the demand price is decreasing in β and increasing in γ. It is also increasing in x' : if a woman has an attractive match within her own caste (by concavity of $f(x,y)$ with respect to x) she can bear the loss of a drop of quality better. As before, the effect of y' on δ is ambiguous and depends on the substitution possibilities between x and y.

Observing a high supply price is consistent with both strongly vertical and strongly horizontal preferences. By contrast a high demand price suggests that preferences are vertical.

Once we have the concepts of demand price and supply price, the following implication is straightforward:

Observation 1 *A inter-caste marriage takes place if and only if* $\varepsilon\leq\delta$.

That is, the quality gain a man (woman) needs to marry down cannot exceed the quality loss a woman (man) is willing to tolerate for marrying up.

If we take $f(x,y)=x^a y^{1-a}$ then we can explicitly solve for ε and δ :

$$\varepsilon=\left[\left(\frac{1+\alpha\beta}{1-\alpha\gamma}\right)^{\frac{1}{1-a}}-1\right]y$$

and

$$\delta=\left[1-\left(\frac{1}{1+\alpha\beta-\alpha\gamma}\right)^{\frac{1}{a}}\right]x'.$$

The following implications are straightforward:

Observation 2 *If $\beta = 0$ (a purely horizontal world), $\delta \leq 0 \leq \varepsilon$, whereas if $\gamma = 0$ (a purely vertical world), $\delta \geq 0, \varepsilon \geq 0$ for all $\beta > 0$.*

Observation 3 *The supply price of caste is increasing in β and γ, whereas the demand price of caste is increasing in β and decreasing in γ.*

Together, observations 2 and 3 suggest that inter-caste marriages are more likely in a world where caste is more vertical. We turn to this in more detail in the now.

2.2.4 Matching in a balanced population

Other than preferences, the distribution of the population in terms caste and quality would clearly affect the equilibrium matching pattern and the associated equilibrium price of caste. We begin our analysis by focusing only on the role of preferences. For this we assume that the distribution of x and y within each caste is balanced. For example, in the two-type case, let $x \in \{L, H\}$ and $y \in \{L, H\}$ with $H > L$. Let m_k^i is the number of men of type k ($k = L, H$) in caste i and w_k^i is the number of women of type k ($k = L, H$) in caste i. Then a balanced population assumption implies that $m_k^i = w_k^i$ for all $k = L, H$ and for all $i \in \{1, 2\}$. If x and y are continuous then let $F_m^i(x)$ denote the distribution function of x for men in caste i and correspondingly, let $G_w^i(x)$ denote the distribution function of y for women in caste i. The balanced population assumption is $F_m^i(v) = G_w^i(v)$ for all $v \in [0, B]$ and for all $i \in \{1, 2\}$.

This formulation looks more artificial than it needs to be: rather than thinking of x and y as the physical values of education and beauty we could see them as the percentile levels in the population distribution of education and beauty, which would make it more natural for them to have the same range. Even with this clarification, it remains that this is a strong assumption. We will come back briefly to what would happen if it fails.

Let the distribution of x and y within each caste be balanced. This is best illustrated by the two-type case: let $x \in \{L, H\}$ and $y \in \{L, H\}$ with $H > L$. Then a balanced population assumption implies that any man whose type is z ($z = L, H$) in caste i can find a woman whose caste is i and whose type is z. We begin with the following simple observation:

Observation 4 *With balanced population within each caste group, if marriage is restricted to* *within caste, the equilibrium displays assortative matching.*

Since the thought experiment is to restrict attention to within caste matches only, the result follows immediately from the assumption of $f(x,y)$ being increasing in both arguments. If a L type man is matched with a H type woman (or vice versa) somewhere else a H type man must be matched with a L type woman, and this assignment cannot be stable as a H type woman and a H type man can form a pair that will make them both better off.[13]

Next, let us consider the case of preferences that are additively separable in the caste and the non-caste components. We have the following observation:

Observation 5 *With balanced population within each caste group, if preferences are additively separable, i.e.,* $u^G(i,j,x,y) = A(i,j) + f(x,y)$, $G = M, W$, *then it is not possible to get inter-caste marriages unless both parties are caste-neutral, i.e.,* $\alpha = 0$.

Proof. To see this, consider a L-type person in caste 1 who might want to marry a H-type person in caste 2. This will be the case i if $A(1,1) + f(L,L) \leq A(1,2) + f(L,H)$. To persuade the H-type person in caste 2, who by assumption of balanced population has a default match of a H-type, the following condition must hold: $A(2,2) + f(H,H) \leq A(2,1) + f(L,H)$. A *necessary* condition for these two inequalities to be satisfied is $A(1,1) + A(2,2) + f(L,L) + f(H,H) \leq A(1,2) + A(2,1) + 2f(L,H)$ but that is impossible given that both $A(i,j)$ and $f(x,y)$ satisfy complementarity. ■

From now on we assume that preferences are multiplicative in the caste and non-caste component. Let us consider the possibility of inter-caste marriage. We show that when the horizontal component of caste preferences is as important as the vertical component, we will observe assortative matching in equilibrium, which is also what we would observe if caste were entirely irrelevant:

Proposition 6 *With balanced population within each caste group, if the horizontal component in preferences,* γ, *is at least as important as the vertical component* β, *i.e.,* $\gamma \geq \beta$:

[13]This is under the assumption of NTU. With TU, as is well known from Becker (1973), to get assortative matching x and y would need to be complements.

(i) inter-caste marriages can never take place that involve at least one caste-conscious individual ($\alpha > 0$);

(ii) those with caste-neutral preferences are indifferent between marrying within caste or outside;

(iii) the equilibrium displays assortative matching and so the equilibrium price of caste is zero.

Proof. (i) Given balanced population within each caste group, for inter-caste marriages to take place a caste-conscious individual of caste 2 must be keen to marry someone from caste 1 and be willing to sacrifice some amount of partner quality for this. This cannot occur when $\gamma \geq \beta$ as that implies $A(2,1) \leq 1 = A(2,2)$, i.e., the demand price of caste is non-positive.

(ii) This follows directly from the balanced population assumption and the fact that $\alpha = 0$.

(iii) Given (i) and (ii) there is no strict incentive marry outside caste (caste-neutral individuals may be indifferent) and given the balanced population assumption within each caste group, assortative matching results. This immediately implies that the equilibrium price of caste is zero: we would not observe an individual sacrificing partner quality in order to marry outside caste. ■

For those who are caste-conscious, with horizontal preferences, there is no strict preference for marrying outside caste. Within the caste neutral group, given balanced population, people will be indifferent between marrying some of their own caste vs. someone from another caste (for the same partner quality). Given this some of the marriages may be inter-caste.[14]

We now turn to the case where inter-caste marriages may emerge in equilibrium even with balanced populations. From the above results we know that for this to happen, it must be the case where β is relatively large compared to γ (i.e., caste is primarily vertical, not horizontal).

Let us begin with an allocation that involves assortative matching within each specific caste group. A strict Pareto-improvement will result if caste-conscious caste 2 individuals are matched with caste-neutral caste 1 individuals of the opposite sex who are of the same quality level. Assuming that the caste-neutral population is small relative to the caste-conscious population (in particular, the size of the group of caste-neutral individuals of caste 1 is small relative to the

[14]Since individuals are indifferent, other idiosyncratic factors can play a tie-breaking role and lead to inter-caste marriages.

size of the group of caste-conscious individuals of caste 2 for each quality level) there will still be some caste-conscious caste 2 individuals left who were not able to match with a caste-neutral individual of caste 1 who is of the same quality level. Their next best option would be to match with a caste-neutral caste 1 individual of lower quality. A caste-conscious H-type person in caste 2 (say, a woman) would prefer marrying a caste-neutral L-type man of caste 1, if

$$f(H,H) \leq A(2,1)f(L,H).$$

The latter will be persuaded as

$$f(L,L) < f(L,H).$$

Therefore, we have the following Proposition:

Proposition 7 *With balanced population within each caste group, the size of the group of caste-neutral individuals being small relative to the population of caste-conscious individuals, and the vertical component in preferences, β, being at least as important as the horizontal component β, i.e., $\gamma \geq \beta$ inter-caste marriages involving at least one caste-conscious individual ($\alpha > 0$) will always take place. If preferences are sufficiently vertical ($1 + \alpha(\beta - \alpha\gamma) \geq \frac{f(H,H)}{f(L,H)}$) then the equilibrium price of caste will be positive.*

The above Proposition is in stark contrast with the previous Proposition: as long as there are low caste individuals who value marrying up in caste, and as long as there are some caste-neutral individuals of the upper caste, there are gains from trade. If the preference for marrying up in caste is strong enough, then some caste-conscious individuals of caste 2 will be willing to marry a caste-neutral individual from caste 1 even if that involves sacrificing partner quality.

However, this result gives an incomplete characterization of the matching outcome. For example, there could be remaining caste-conscious caste 2 individuals, whose choice would be to stick to someone of the same quality within the caste, or try to match with a caste-conscious caste 1 individual of lower quality as given the balanced population assumption he/she is not going to persuade a caste-conscious H-type person in caste 1 to match with him/her. We now turn to a more complete characterization of this case ($\beta > \gamma$).

Consider a caste-conscious H-type person in caste 2 (say, a woman). As before, she would prefer marrying a caste-conscious L-type man of caste 1 if

$$f(H,H) \leq A(2,1) f(L,H).$$

However, a caste-conscious L-type man in caste 1 will be persuaded if

$$A(1,1) f(L,L) \leq A(1,2) f(L,H).$$

As $\alpha\gamma < 1$ by assumption, $A(2,1) = 1 + \alpha\beta - \alpha\gamma \leq \frac{1+\alpha\beta}{1-\alpha\gamma} = \frac{A(1,1)}{A(1,2)}$ with the strict inequality holding for $\gamma > 0$. A *necessary* condition for these two inequalities to be satisfied is[15]

$$(1+\alpha\beta) f(L,L) + f(H,H) \leq (2 + \alpha\beta - 2\alpha\gamma) f(L,H).$$

Clearly, β has to be high enough relative to γ for this to be satisfied: for example, for $\beta \leq 2\gamma$ the condition is not satisfied as $f_{xy}(x,y) \geq 0$.

The two conditions can be combined as

$$\frac{f(H,H)}{f(L,H)} \leq 1 + \alpha\beta - \alpha\gamma < \frac{1+\alpha\beta}{1-\alpha\gamma} \leq \frac{f(L,H)}{f(L,L)}.$$

A necessary condition for this to be satisfied is

$$\frac{f(H,H)}{f(L,H)} < \frac{f(L,H)}{f(L,L)}.$$

Clearly, for symmetric production functions (i.e., $f(x,y) = f(y,x)$), these are equal and so inter-caste marriages cannot take place.

To obtain a more precise characterization, let us work with $f(x,y) = x^a y^{1-a}$. To simplify notation, let us also set $\gamma = 0$ (in which case $1+\alpha\beta-\alpha\gamma = \frac{1+\alpha\beta}{1-\alpha\gamma} = 1+\alpha\beta$). Then the condition simplifies to

$$\theta^a \leq 1 + \alpha\beta \leq \theta^{1-a}.$$

[15] In the TU case, this condition is both necessary and sufficient.

where $\theta \equiv \frac{H}{L}$. We assume $a < \frac{1}{2}$, otherwise this can never hold. Let us define the following two thresholds for β:

$$\begin{aligned}\overline{\beta}_1 &\equiv \frac{\theta^a - 1}{\alpha} \\ \overline{\beta}_2 &\equiv \frac{\theta^{1-a} - 1}{\alpha}.\end{aligned}$$

Now we are ready to state:

Proposition 8 *With balanced population within each caste group, purely vertical preferences ($\gamma = 0$), Cobb-Douglass preferences over quality $f(x,y) = x^a y^{1-a}$ with $a \in \left[0, \frac{1}{2}\right]$, and the size of the caste-neutral group being small:*

(i) inter-caste marriages involving a caste-conscious individual of caste 2 and a caste-neutral individual of caste 1 who is of lower quality will take place if $\beta \geq \overline{\beta}_1$;

(ii) inter-caste marriages involving a caste-conscious individual of caste 2 and a caste-conscious individual of caste 1 who is of lower quality will take place if $\beta \in [\overline{\beta}_1, \overline{\beta}_2]$ where $0 < \overline{\beta}_1 < \overline{\beta}_2$;

(iii) the equilibrium price of caste will be positive and will decrease the greater the share of caste-neutral individuals;

(iv) Observed inter-caste marriages will take place between low quality men (women) of the high caste and high quality women (men) of the lower caste. High quality men and women in the upper caste and low quality men and women in the lower caste will tend to marry within caste.

Proof. The proof of parts (i) and (ii) follow directly from the discussion preceding the Proposition.

(iii) Since there will be non-assortative matching under the conditions stipulated in (i) and (ii), the equilibrium price of caste will be positive: some high quality individuals of caste 2 will marry low quality individuals of caste 1. Since we have two quality levels, the effect of the size of the caste neutral population on the equilibrium price of caste is discrete: as it goes up above a certain threshold, all caste 2 individuals who want to marry up in caste will find a caste-neutral caste 1 individual of the same quality and so the price of caste will be zero. Otherwise it will be positive.

(iv) Clearly high type men and women of caste 1 who are caste-conscious marry each other: there are no gains from deviation in terms of caste or quality. Now a low quality high caste woman (man) has the choice of marrying a low quality high caste man, a high quality low caste man (woman), or a low quality low caste man (woman). The last option is clearly dominated by the first. Under the parameter assumptions, the second option dominates the first option. Analogously, for a high quality low caste man (woman), the choice is between marrying a high quality low caste woman (man), a low quality high caste woman (man), or a low quality low caste woman (man). Once again, second option dominates. This leaves low caste men and women of low quality marrying each other. ■

The intuition is as follows. Unless caste preferences are vertical up to some minimum level, there is no reason for a high quality woman of low caste to give up a high quality mate in her own caste and settle for a low quality mate from the upper caste. However, if caste preferences are vertical beyond a certain threshold then inter-caste marriages will no longer take place. Now the price at which a low quality man from the high caste will be willing to marry a high quality woman from the low caste ("demand price") will be higher than what a high quality woman from the low caste is willing offer since she values a fall in quality more (her own quality being high).[16]

Observe that if a is small (men's role in the marital payoff function is minimal) then $\overline{\beta}_1 < 0$, while $\overline{\beta}_2 > 0$. Therefore, inter-caste marriages will take place if β is not too high in this case.

Also, if men and women both play equally important roles in the marital payoff function then inter-caste marriages will not take place. The value of caste must be high enough to offset the loss from having a lower quality husband for a high quality bride, but the loss in terms of marrying a lower caste woman should not be high enough to outweigh the gain from having a high quality bride for a low quality high caste man. If both genders play equally important roles, this double coincidence will not take place.

Proposition 4 has the following implication:

[16]In a TU world, caste preferences being sufficiently vertical will lead to inter-caste marriages. With free side transfers, it is *as* if that caste preferences and quality preferences are separable. In a NTU world, this minimum threshold will be higher than the TU case, since no side transfers are possible and the only method of compensation is providing a sufficiently high quality differential to the low quality mate from the high caste to induce him to marry her.

Observation 6 *The equilibrium price of caste for a high (low) caste individual is less (greater) than the average supply (demand) price of caste in that caste group.*

This follows from part (iv) of the Proposition. Recall that the demand and supply prices (ε and δ) are increasing in the quality of the existing match within caste. With balanced population, only lower quality men (women) will marry someone from the lower caste in equilibrium when the relevant conditions on parameters apply. This means the equilibrium price of caste will be lower than the ex ante (or notional) average supply price at which a caste 1 individual would be willing to marry inter-caste. The same argument applies for a low caste individual in reverse. Since only a higher quality man or woman will marry inter-caste, for caste 2 individuals, the equilibrium price of caste will be less than the ex ante or average demand price of caste for caste 2 individuals.

2.2.5 Matching in an unbalanced population

The simple vertical-horizontal dichotomy of the previous section is only possible because we assumed a balanced population. With balanced population, naturally preferences are the only determinant of the equilibrium allocation. In the absence of a balanced population, other than preference parameters, the distribution of the population will affect the equilibrium outcomes. In this section we explore the implications of this possibility.

With a balanced population, preferences need to be sufficiently vertical for inter-caste marriages to take place (ignoring caste-neutral individuals who are, by definition, indifferent between marrying inter-caste or not other things being equal). When the assumption of a balanced population is relaxed, inter-caste marriages can take place for all types of preferences, including purely horizontal ($\beta = 0$). With balanced population one always has the option of marrying someone of the corresponding quality level within the same caste. As a result, inter-caste marriages take place when a low caste person values marrying up in caste sufficiently to agree to marry someone of lower quality from the upper caste. With unbalanced population, one is not guaranteed to find someone of the corresponding quality level within the same caste and this raises the likelihood of inter-caste marriages. Therefore, we will not observe assortative matching even if we restrict marriage to within caste only. This creates an additional reason for inter-caste marriages to take place. Obviously, it needs some complementarities in the quality-specific sex ratios. For example, if very beautiful low caste women cannot find a suitably qualified low caste men, there must be qualified men in the upper caste who do not find sufficiently beautiful women from within their own caste.

To see this point most starkly, consider the case where preferences are purely horizontal (i.e., $\gamma > \beta = 0$) so that in a balanced population matches will be assortative, and no inter-caste marriages will take place. Also, for simplicity, let us assume that everyone is caste-conscious ($\alpha > 0$).

As before, suppose there are two quality levels, L and H for both castes. Consider first individuals in caste 1. H-type individuals who are lucky enough to find H-type individuals from within the same caste are clearly not going to be interested in inter-caste marriage. Suppose some of them cannot find a partner of corresponding quality within caste 1. In that case their option is to marry a L-type individual from within the same caste or a H-type individual of the opposite sex from caste 2 (L-type individuals from caste 2 are dominated by L-type individuals from caste 1). The latter is more attractive if:

$$(1-\alpha\gamma)f(H,H) \geq f(H,L).$$

For $f(x,y) = x^a y^{1-a}$ this condition simplifies to

$$\gamma \leq \overline{\gamma}$$

where $\overline{\gamma} \equiv \frac{\theta^a - 1}{\alpha\theta^a}$. Recall that we assumed $\alpha\gamma < 1$. As $\theta > 1$, $0 < \overline{\gamma} < \frac{1}{\alpha}$. With purely horizontal preferences, the demand and supply prices for caste 2 individuals are the same. Therefore this is the same condition for a H-type person from caste 2 of the opposite sex to agree to marry this individual. Assuming the payoff from being single to be zero, for a L-type individual in caste i who cannot find a L-type individual of the opposite sex within the same caste (and, by transitivity, a H-type person of the opposite sex within the same caste) will be willing to marry L type individual of the opposite sex from caste $j \neq i$. The latter will agree if he/she too cannot find a L-type match from their own caste group. The payoff of both parties will be

$(1 - \alpha\gamma)f(L, L) > 0$ (as we assume $\alpha\gamma < 1$).

Recall that a balanced population assumption implies that $m_k^i = w_k^i$ for all $k = L, H$ and for all $i \in \{1, 2\}$. If $m_k^i > w_k^i$ and $w_k^i > m_k^i$ for some k ($k = L, H$) and $i \neq j$ then we define the sex ratio for quality level k to be complementary across the two caste groups. Now we are ready to state:

Proposition 9 *With unbalanced population, and complementary inter-caste sex ratios for at least some quality level k, inter-caste marriages will take place even with purely horizontal preferences ($\gamma > 0 = \beta$) so long as $\gamma \leq \overline{\gamma}$. Inter-caste marriages, if they take place, will be assortative and the equilibrium price of caste will be zero.*

Proof. This follows from the fact that given the assumption $\gamma \leq \overline{\gamma}$, a H-type man in caste i prefers to marry a H-type woman in caste j rather than marrying a L type woman in caste i, and vice versa. Also, as $\overline{\gamma} < \frac{1}{\alpha}$, a L-type man in caste i and a L-type woman in caste j prefer marrying each other rather than staying single. Given this assortative matching directly follows, and so the equilibrium price of caste will be zero. ∎

Therefore in the unbalanced population case, so long as sex ratios are complementary across caste groups for at least some quality level there will be inter-caste marriages even with purely horizontal preferences. If $\beta > 0$, that will reinforce this tendency. If sex ratios are not complementary for any quality level then not a lot can be said in general. Among other factors, the outcome would depend on the aggregate sex ratio.

The above analysis assumed only two quality levels. The basic intuition goes through with more quality levels. For example, if there is an intermediate quality level M such that $H > M > L$ then we will have a richer set of possibilities. Still, with complementary sex ratios, inter-caste marriages will tend to be assortative: a man of type H from caste 1 will marry someone who is type M from caste 2 only when he cannot find either a H-type or a M-type woman from his own caste, which is not very likely.

If these inter-caste marriages take place, which are more likely? By our previous analysis, the price of caste will be the highest for a H-type since he/she is matched with, at worst, a M type. Clearly, if they still find it worthwhile to do this, so will M types matched with L types and L types who are single.

This suggests two reasons why ex ante price of caste will be lower than equilibrium price of caste.

First, for any given type (say H) if he does marry inter-caste he will be marrying a H-type given that M was his best match within caste. Therefore, compared with someone of the same type who was luckier and found a H type within his own caste, his stated price of caste will be lower (this follows from the fact that $\varepsilon(x, y, \beta, \gamma)$ is increasing in y for the same x).

Second, of all types, the relatively lower types are likely to marry inter-caste. The price of caste of a x type who is matched with at worst, a $x - \triangle$ type within his/her own caste, to marry inter-caste and find someone of type x is increasing in x.This is another reason why the average stated caste prices will be lower than the observed prices of caste.

What kind of type distributions are consistent with the scenario above? It is fair to assume that beauty is distributed identically across castes but education or income may not be. Suppose both caste groups have population size normalized to 1 and in both groups there are $1/3$ H type women, $1/3$ M type women and a $1/3$ L type women. However, if there are lots of qualified men in one caste (say, more than $1/2$) and lots of unqualified men in the other caste (again, more than$1/2$) then we will have a scenario that is similar to the one we described.

Finally, what will happen in a hypothetical world where caste preferences just disappeared (the $A(i, j)$ function becomes equal to 1 for all i, j) compared to a world where they exist? With unbalanced population within caste, but balanced population for all castes taken together, all marriages will be assortative. So if the actual type distribution has many quality levels (and not just three) and the gaps between these quality levels are small, then very few inter-caste marriages will take place, unless someone who is a high type is matched with someone who is considerably lower than him/her within his/her own caste (and finds someone with a parallel situation from the other caste). Now we can see that taking away caste will lead to full assortative matching, and so with respect to the intital population lots of inter-caste marriages will take place.

2.2.6 Discussion

There are two broad implications from the above analysis that are important for interpreting our empirical results.

First, with horizontal preferences ($\beta < \gamma$), everyone demands compensation to marry outside caste and as a result, demand price always exceeds supply price for all groups, and so there are no-intercaste marriages. Moreover, in this case, if everyone became caste neutral (i.e., $\alpha = 0$ so that for all i and j, $A(i, j) = 1$) the same pattern of matching will be observed (given the balanced population assumption).

Compare this with a world where preferences are significantly vertical (i.e., the premise of Propositions 3 and 5 holds). Now inter-caste marriages will take place. In this case, if everyone becomes caste-neutral, there will be significant changes in the pattern of matching as now there will be assortative matching in terms of x and y for the whole population.

Second, in the horizontal world, if we observe intercaste marriages it is because there are some caste-neutral people. The equilibrium price of caste therefore be zero. If preferences are sufficiently vertical to observe intercaste marriages outside the caste-neutral group, the equilibrium price of caste will be positive - people will be willing to "pay" in terms of partner quality to marry up in terms of caste.

Third, when the population is not balanced, then one can get inter-caste marriages even with purely horizontal preferences. A sufficient condition for this complementary inter-caste sex ratios for at least some quality level. In this case, inter-caste marriages will tend to be assortative and the equilibrium price of caste will tend to be low.

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 horizontal or vertical. Then, using these estimates, we will demonstrate the equilibrium consequences that these caste preferences generate for marital pairing.

2.3 Setting and data

This section summarizes the way the data was collected and how the variables used throughout the empirical exercise were constructed.

2.3.1 The search process

The starting point for data collection was 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. 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 phone or by mail), and elect whether or not to follow up with a particular response. Note that while both men and women place ads, "groom wanted" ads constitute almost 75 percent of all ads placed, and "bride wanted" ads received four times as many responses. When both parties are interested, the set of parents meet, then the children meet. The process takes time: in our sample, within a year of placing an ad, 44 percent of the interview sample were married or engaged (in 29 percent of the case however, they had placed a single ad). 65 percent of those married are married through an ad, the rest having met through relatives or, in 20 percent of the cases, on their own (which is referred to as "love" marriage).

2.3.2 Sample and data collection

The first step was to code the information in all the ads published in the Sunday edition over this time period (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 then selected a random sample from these ads, after excluding ads placed under the heading "Christian" or "Muslims" in the newspaper. Importantly, we also restricted the sample to the ads which did not mention a phone number, and requested all responses to be sent at the newspaper PO Box or to a personal mailing address.[17] This restriction was necessary to make sure that the letters received in response to an ad reflect all the relevant information the ad placer has on the respondent. About 43 percent of all ads included a phone number (sometimes in addition to a PO Box, sometimes as the only way to contact the ad placer). Comparing the characteristics of ads with and without phone numbers, we find little differences between those who include a phone number and those who do not, except in terms of geographical location: more ad placers with phone numbers were from Kolkata.

From this set, we sampled 784 ads and conducted detailed interviews with the ad placers (usually the parent, uncle or older brother of the prospective groom or bride). 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 ad 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. Detailed information was collected on the prospective groom or bride, his family and the search process for a marriage partner.[18] In particular, ad placers were asked whether they also replied to other ads and, when they did, to identify the ad 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 male and 23 for female 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 sample was re-interviewed, and we collected information regarding their current marital status and their partner's choice. Only 33 ads out

[17] Only a small fraction of ads included only a personal mailing address (35 out of 783 ads in our random sample, 1796 out of 22,210 in the ad placer sample).

[18] The questionnaire is available on line at http://web.mit.edu/~jlafor/Public/Questionnaire/.

of the entire sample could not be contacted. Appendix Table 2.A.1 compares the characteristics of these ad placers compared to those who could be found. There is little evidence of differences between the two groups. At most, ad placers from Kolkota and women who had not mentioned their occupation and incomes were more likely to be found in the second round. At the time of the second round interview, 346 out of the prospective brides or grooms in the original sample were married or engaged. Out of these, 289 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. In a very small number of cases, the ad placer was able to provide either the ad placed by the match or the letter the match sent by mail. This sample, however, was too small for us to use in the analysis. Table 2.A.2 compares the characteristics of the ad placers who agreed to an interview to those who did not. Once more, there appears to be little systematic differences between the two groups.

2.3.3 Variable construction

Ads and letters provide very rich information, which was coded in the following way.

First, we coded caste information. In the newspaper, most ads are placed under a specific section for each caste. The text of the ad then typically does not mention the caste of the ad placer. If an ad was placed under a heading that clearly identified one caste and did not mention its caste, this ad placer is assumed to be of this particular caste. If caste was explicitly mentioned in the ad, we used what was mentioned in the ad. The information on castes is readily available, directly or indirectly, in the overwhelming majority of ads (98 percent). In the letter, caste is explicitly mentioned in about 70 percent of the cases.

There are numerous castes and sub-castes in India. Ad placers or letters can be more or less specific in identifying themselves. There is a hierarchy between broad castes groups, but within each broad group, there is much dispute on the proper ranking. Castes were thus grouped into eight ordered 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 2.A.3 presents the classification. We use this coding to construct an indication of the distance between the caste of respondent and that of the ad placers. The summary statistics are presented in Table 2.1. The majority of the ad placers are Kayashta (more than 30 percent) and Brahmin (more than 25 percent) while Baisya, Sagdope and other similar castes include each more than 10 percent of the ad placers. The other groups are much smaller in sizes.

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 they mentioned in their ad. If they only mentioned a broad group, they are assumed to be of any of the specific subcastes. For example, a self-identified Kulin Brahmin is considered to be from a different caste as a self-identified Nath Brahmin (though the hierarchical distance between them is set to zero), but is considered to be of the same caste as someone who simply identified themselves as a Brahmin. In practice, the distinctions between sub-castes matters most for the lower castes, where the broad groups join differentiated subgroups, and where people typically identify themselves with a specific narrow group.

Another relevant information is the stated preferences regarding castes. 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 20-30 percent explicitly specify their willingness to unions outside their own caste by the use of phrases such as "caste no bar". The remaining 40-50 percent do not make any mention of preferences regarding castes.

Second, we coded information provided on education levels. Educational attainment was classified into 7 categories: less than high school, high school completion, non-university post-secondary, bachelor's, master's, PhD or professional degree and non-classifiable degree.[19] In addition, we also coded, when available, the field in which the degree was obtained. We sorted these into 4 groups: Humanities and Social Sciences (B.A, B.Ed, M.A, etc), Commerce (B.Comm, MBA), Science (B.Sc., B.Eng, M.Sc., etc) and other fields (Law, religion, etc).

Third, we coded the available information on earnings 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 their occupation, we

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

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

Fourth, we coded information on the origin of the family (East or West Bengal) and the current location of the prospective bride or groom (Kolkata, Mumbai, Other West Bengal, or other –mainly abroad).

Fifth, a very large fraction of prospective bride's ads specify physical characteristics of the woman, using fairly uniform language and the same broad characteristics. Skin color was coded into four categories (from "extremely fair" to "dark"). General beauty was divided into three categories ("very beautiful", "beautiful" and "decent").

Finally, ads occasionally mention a multitude of other characteristics, such as "gotras" (a group within which one is not supposed to inter-marry), astrological signs, blood type, family characteristics, family members mentioned, personality traits, female skills, previous marital history and number of children, specific demands, etc... These were coded as well. However, each of these is rarely mentioned including or excluding them does not affect our results.

2.3.4 Summary statistics

Table 2.1 presents summary statistics for both our interview sample and the full set of ads.

Our sample 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. 90 percent of both men and women have at least a bachelor's degree. Women rarely mention their occupation. When they do, their occupational score (5.51) is similar to that of men and 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. While few women provide their income, a few include a description of their occupation and although their occupational score is lower than that of men, the difference (among those who reveal it) is quite small.

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. Skin tone is mentioned in 86 percent of the cases, beauty, in over 70 percent of the ads. There appears to be little boasting about physical appearance, however: more ads describe the bride as being "decent" than either "beautiful" or "very beautiful".

Since our sampling strategy excluded all the ads that did not mention a phone number, it is important to compare their characteristics with the overall sample of ads, to assess the impact of this selection rule on the make up of our sample. Generally, the interview sample looks very similar to the overall sample of ad placer. There are three significant differences. First, perhaps not surprisingly, an individual who is interviewed is more likely to live in Kolkata. This is probably because ad placers mention a phone number when they cannot collect the letters so easily themselves. Second, men are much less likely to report their occupation (57 percent of them do not report it in the interview sample, while 27 percent do not in the general sample), though their occupational score is similar when they do report it. Finally, and perhaps most importantly, they are much more likely to mention in their ad that they will only marry within their castes (33 percent versus 10 percent for men; 43 percent versus 9 percent for women). It is therefore important to keep in mind that our sample is more likely to be a more traditional sample than the sample of people who place ads in newspaper.

Table 2.2 presents similar statistics for two different samples: the sample of people who wrote a letter in response to an ad ("the letter writers") and the sample of actual spouses. Note that the information on the spouse was collected from interviews with the ad placer (few families could show us the original ad or letter of the spouse). In terms of their characteristics, both of these samples look very similar to the sample of ad placers. In the few dimensions where the ad placer and the interview sample differ, the letter looks more similar to the interview sample, except for the Kolkata location (50 percent to 55 percent of the letter writers mention that the prospective spouse lives in Kolkata; 15 percent to 20 percent do not mention anything in the letter). A few prospective grooms (7 percent) explicitly mention that they will not demand a dowry. None mentions that they want a dowry.

This table also shows comparisons between the ad placer and the letter they have received, as well as with their eventual spouse. In this table, as well as in the remainder of the paper, all differences are presented in terms of the difference between the characteristic of the man and the characteristics of the woman. Since the sampling was stratified with unequal weights, each letter is weighted by the inverse of its probability of selection.

We begin by describing how the respondents compare to the ad placers. It is relatively common to write to someone from a different caste. Two thirds of the letters which mention castes are from someone from the same caste as the ad placer. 79 percent of the ad placers have received at least one letter written by someone from another caste among those we sampled. On average, men tend to write to castes above theirs (the difference in caste between men and women is negative, indicating that the man is from a higher caste); when they write outside of castes, women write equally up and down. In 37 percent to 44 percent of the cases, the letter writer has the same education as the ad placer. When they don't have the same education as the men they write to, women tend to have less education than them. Men seem equally likely to write women who are more or less educated than them. Not surprisingly, men write to somewhat younger and shorter women then themselves, and vice versa. These differences reflect the average difference in the population.

Turning to the actual matches, we observe somewhat different patterns: First, while there are still a number of matches that are not within castes, the fraction of within caste marriage is higher than that of letters that are coming from within the castes: 72 percent of the prospective grooms and 68 percent of the prospective brides who are married after a year have done so within their own narrow caste. This fraction increases to 76 percent and 72 percent respectively if we use the broad classification. Second, men who marry outside of caste tend to marry a lower caste bride, and women who marry outside of caste tend to marry a higher caste groom. Females 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. 72 percent to 75 percent of the brides and grooms are from the same family origin (West or East Bengal).

2.4 Estimating preferences

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

2.4.1 Basic empirical strategy

The first goal of this paper 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:

$$U(X_j, X_i) = \alpha X_j + \beta f\left(X_i, X_j\right) + \mu_i + \varepsilon_{ij} \tag{2.2}$$

We use various strategies to attempt to estimate the parameters of equation (2.2).

First, the ad placers provided us with their ranking of each ad. If we assume that the ranking are truthful, a higher ranking prospective spouse j than for prospective spouse j' must indicate that i prefers j to j'. A first possible strategy is to estimate an equation similar to (2.2) in the sample of letters, using the rank provided by the ad placer as the dependent variable. We run this estimation with ordered probit, and with OLS.

There is a danger that these ranks do not reflect the respondent's true preferences, since they are just a response to an interviewer. We have however in our data several indications of individuals' revealed preference for a spouse versus another. First, we know whether an ad placer is following up with a particular letter or not. We thus have information that he preferred this letter to the letters he did not consider. Second, for ad placers who have also replied to ads, we know which ad they decided to reply to (and we of course also know the universe of ads they could have replied to). Third, we know that a letter writer decided to reply to an ad. Finally, we also know how many replied an ad received.

Hitsch et al. (2006) show that under the assumption that if an individual i contacted j rather than j', which was also available to him, it implies that i prefers j to j', the parameters of equation (2.2) can be estimated using a fixed effect conditional logit estimation (where the dependent variable is 1 if individual i contacted individual j, and 0 otherwise) if ε_{ij} has the standard logistic distribution.

The regressions we estimate thus take takes the following form:

$$y_{ij} = \alpha X_j + \beta f(X_i, X_j) + v_i + \epsilon_{ij}, \tag{2.3}$$

where y_{ij} is a dummy equal to 1 if ad placer i replied to letter j, for example. In the empirical exercise, 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 the difference between men and women), and its square. We estimate equation (2.3) using a conditional logit with fixed effect for each person i, and OLS with fixed effects.[20]

The assumption that choices reflect preferences is of course not innocuous: in particular, it rules out strategic behavior, for example the fact that an ad placer does not respond to an ad because they think that person is "too good for them".

We have three variables to perform this exercise: an ad-placer i writes back to a letter writer j (in the set of letters he receives); and an ad placer i writes to ad j (in the set of available ads); a letter writer writes to an ad j (in the set of available ads). The last data source is a data set similar to what we would have obtained if we had run a randomized experiment by placing fake profiles on a web site, and varying the attributes one by one (similar to Bertrand and Mullainathan 2004).[21] However, this last data set suffers from measurement error, because we did not sample all the letters received by each ad placer. The two other sources do not suffer from this problem. The data on ad-placer's responses to the letter has two advantages over the data on which ad placer replied to each ad. 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

[20] 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.

[21] In this case, there would not be huge advantage to running an experiment, however, since we do observe the same information as a letter writers.

expressed interest in the ad placer. We will thus present the results from the ad placer response to the letter in the main text, and the results using the responses of ad placers to other ads and using the letter writers responses to the ad are presented in appendix. The results are very consistent, but we will underline the main differences below.

Finally, we also have data on the number of letters an ad placer receives: this can be used to estimate a count model (which we estimate with a Poisson model and with OLS), but it is not possible to introduce heterogeneity in preferences in this estimation.

There are three major possible objections to the interpretation of these results in terms of relative preference particular attributes. First, as we mentioned, behavior could be strategic, in which case the choices of whom to respond to may not reflect preferences. In a market where time is important, people could avoid wasting their time by writing to someone who will reject them, or could write to different people with the view of constructing an optimal portfolio of prospects (with some high value but unlikely prospects for example, and enough good matches to ensure at least one acceptable match). Second, ad placers could interpret responses to their ad as signaling some unobserved quality of the match. For example, if a suitor with very good observed characteristics is writing to a woman with poor observed characteristics, this woman could infer that there is something wrong with the person who is writing to them. This creates a correlation between the error term and the attributes in equation (2.2), even though we have the same information set as the household. Third, even assuming that the choice reflects actual preferences, this preference may take into account expected dowry. If this is the case, the trade-off between different attributes may not be representative of actual preferences. Below, we review these three objections in more details, and present evidence that, in our view, strengthens the argument that this strategy probably uncovers actual preferences.

2.4.2 Results: Ad placers's response to letters and letter ranking

Table 2.3 presents the results fixed effects and conditional logit regressions, where the binary decision of whether or not an ad placer i respond to a letter j is regressed on a set of characteristics of the letter, and its interactions with the ad placer's.

Columns 1 to 5 present the specifications for the groom wanted ads (these are ads placed on behalf of a woman, and letters are sent on behalf of men), and columns 6 to 10 present the specifications for the bride wanted ad (placed on behalf of a man). Recall that in both cases, differences are presented in terms of the difference between the characteristics of the man and the characteristics of the female. A positive difference in education for example, means that the prospective groom is more educated than the prospective bride.[22] Most categorical variables are dummied out. The excluded categories are "less than high school" for education, outside of Kolkata for residence, and "decent" for beauty. All variables are set to zero if the letter did not mention the characteristic, and we include a dummy variable to indicate whether each variable was missing. All models were estimated with and without including a series of additional covariates (such as indication on the culture of the family, its wealth level, astrological sign etc...). To save space we focus on the more parsimonious specification in the tables; the results are extremely similar when these additional controls are included.

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. Men prefer women who describe themselves as beautiful or very beautiful, and seem to have a strong preference for lighter-skin brides. As Hitsch et al. (2006), we find that looks matters. For example, the OLS estimate suggests that the probability to be called back would be higher for a very light-skinned woman without an education than for a dark skin woman with a college degree. Both men and women prefer a spouse who lives in Kolkata (recall that most of our families are from Kolkata as well), and whose family comes from the same part of 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 estimate indicate that a woman is 13 percent more likely to call back a prospective groom if he is from the same caste, controlling for all other attributes. A man is 17 percent more likely to call back a woman from his caste. These are large difference, 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

[22]Also a positive difference between the man's and woman's caste indicate that the man is of a *higher* caste.

bride wanted ad 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 degree. Men are willing to sacrifice three shade of skin tones to marry someone within their caste (Column 6). These ratios are very similar from the logit coefficients.

Particularly important given our theoretical framework, this preference for homogeneity in caste is stronger than the preferences for marrying "up". Conditional on marrying out of their caste, women prefer men who are as close to their caste as possible: among men who are of a higher caste, they prefer the smallest difference possible, among those of a lower caste, they prefer the highest possible caste. Men prefer the highest caste women possible if they can't find a match within their caste, particularly if they are of a lower caste than the prospective bride. The magnitudes of the coefficient on the difference in caste, however, are much smaller than those for being of the same caste.

One possibility is that several of the variables in these regressions are 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 column (4) and (9) a more parsimonious specification. We first constructed a predicted income, by regressing in the entire data set of letter log income (when reported) on all the education variables, and the occupational score (including dummies when they are not reported). We then construct for each ad placer and letter writer a predicted income, and include this variable instead of all the education, income, and wage variables. Predicted income has a strong and significant impact on the probability of call back, but this regression does reveal that caste plays an important role relative to income. Even for males, one's predicted income would have to be at least 1.5 times larger to compensate being from a different caste.

To display graphically the trade-off between the different attributes. Figures 2-1 and 2-2 show indifference curves, drawn using the conditional logit estimates. They display the age difference, height difference, education, and income a prospective spouse need to have to keep the ad placer indifferent when his or her caste changes, expressed in standard deviations. In both cases, the cost of keeping caste is very marked. To remain indifferent between two prospective brides, one of the same caste and one from a caste one below, the second one must have 3 standard deviation more education, must be 5 standard deviation more closer in age or earn 6 standard deviation more income. The differences are slightly less marked for female preferences but still very marked for same caste. For both genders, there seems to be less of penalty attached to marrying individuals of a higher caste than of a lower one, in addition to the penalty of marrying outside one's caste. This is somewhat related to the findings of Fisman et al. (2008) who find strong same-race preferences among female speed daters that is unrelated to physical attractiveness. Similarly, Hitsch et al. (2006) also find same-race preferences, particularly for women.

Table 2.4 presents similar regressions, using the ranking of the ad provided by the ad placers as the dependent variable. [23] The results from these regressions are virtually homothetic to the ones presented in the previous table, as evidenced by Figures 2-3 and 2-4, which show a regression of the coefficients in Table 2.3 on those in Table 2.4. Appendix Tables 2.A.4 and 2.A.5 present similar regressions, using the other choice variable at our disposal (letter writer response to ad; ad placer response to other ads; number of letters received by an ad). In all these specifications, the importance of caste in the choice is at least as important as in this table. For example, in Appendix Table 2.A.4 being of the same caste increases the probability that an ad chooses to reply to another ad by 2-3 per cent. In the same appendix table, being of the same caste increases the chance that a letter writer replies to an ad placers by about 20 per cent. Turning to the effects of the other variables, there are interesting differences between these specifications and the ones presented in the main text, which we discuss in more details below.

2.4.3 Do these coefficient really reflect preferences?

We argue that these estimates provide us with information on the relative preferences for different attributes. There are two main objections to this interpretation. First, ad placer's choice to respond to letter j rather than to letter j' may not indicate that she would derive more utility from being matched to letter j than to letter j', and instead reflect their assessment

[23]The sample size is a bit smaller, due to the fact that some ad placers refused to provide ranking and the interviewers did not rank the letters in the same way in the early interviews.

that they may be wasting their time writing to j', because j' will not write back. We argue below that there seems to be little evidence of strategic behavior in our sample. Second, and related, while we observe every characteristic observed by the ad placer, we need to take into account the inference that the ad placer is making when observing that he is getting a letter from a specific person. It could be the case, for example, that if someone from a high caste decides to contact an ad-placer from a low caste, it signals something very negative about this person. Using our data on the eventual matches of this people, we will look for evidence that people who write to people outside their own caste are in any way different from those who do not.

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 replying to a letter that appears to be "too good" because they think there is little chance of that relationship progressing. 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. An immediate reaction is thus less likely to occur.

Nevertheless, the issue is further investigated here. We first compute an absolute measure of "quality" of the letter.

To do so, we regress 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_j a dummy indicating whether letter j is considered by ad placer i, we run:

$$P_{ij} = X_j\beta + \epsilon_{ij}$$

without any fixed effect for the ad placer.

We form two versions of this indicator: with and without including the caste of the letter writer. The results presented here use those without caste but similar results were obtained with the caste variables included. The quality indicator is then $Q_j = X_j\hat{\beta}$. We also predict the quality of the ad-placer, using the same coefficients $Q_i = X_i\hat{\beta}$.

Figures 2-5 and 2-6 plot the probability of considering a letter based on the quality of the ad placer and that of the letter. If the responses displayed strategic behavior, we would expect that low quality ad placers would be less likely to consider high quality letters. In fact, Figures 2-5 and 2-6 show little difference in the relative probability of considering letters of different quality by the quantile of quality of the ad placer, although higher quality ad placers appear to consider on average a smaller fraction of letters of all quality levels. If anything, lower quality ad placers seem to respond to a higher fraction of higher quality respondents. Combining this with the letters received by each ad placer's quality, this implies that the eventual number of letters considered are about evenly shared among the lowest level of ad placer quality and then become more and more skewed towards higher quality respondents for higher quality ad placers. Further evidence is provided by Table 2.5 where similar regressions as the ones presented above are presented but this time restricting the sample to letters where the quality of the ad placer and the quality of the letter writers are relatively close. Overall, the behavior of the ad placer seems to be fairly similar when looking at the overall sample compared to this lower relative quality one, either in terms of considering letters or ranking them. The preference of prospective grooms for brides of a similar caste falls slightly but that of women for men increases by a small fraction. The female preference for science graduates is also lowered. Finally, the preference for income rises while that for wages falls. Overall, however, the differences are small and not indicative of any strategic behavior on the part of the ad placers.

Interestingly, the decision to respond to an ad (displayed in the appendix tables) seems to reflect more strategic behavior than the choice of whether to respond to a letter an ad placer received. For example, in the decision of whether an ad placer replies to another ad, and in the decision of whether a letter writer replies to another ad (Appendix Table 2.A.4), education loses its previous importance and appears to potentially decrease one's attractiveness. Similarly, a commerce degree now seems to decrease the likelihood of being selected. This seems to be evidence of strategic behavior at the stage of responding to an ad. Moreover, the fact that the coefficient of the "same caste" dummy is also higher in this sample may reflect in part caste-based search.

Likewise, when we estimate the number of letters an ad placers received (Appendix Table

2.A.5), many results are similar to the ones we find for ad-placers' choices (beauty, skin tone, education for men and being from a large caste, all increase the number of responses), but other variables which were previously important become insignificant or change sign (female education, male income). Finally, 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 of the between quality of an add and overall number of response is an inverse-U. 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 at the point of first contact. 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 each ad placer considers that each of the 40 letters they receive over the course of their search is a potential prospect, and that they do not behave strategically whom to respond to (they respond to about 30 percent of the letters they receive).[24]

What does caste signal?

One of our main empirical result 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 many variables, perhaps unobserved by the ad placer and us, but reflecting a prospective spouse background and culture. People would then match within their castes to marry people like them. However, the strong preference for caste does not seem to be affected by controlling for a host of variables including cultural variables (ability to sing, etc...) (result omitted to save space, but available from the authors) and it remains very strong within the four highest caste, who are culturally and economically quite homogenous (Table 2.6). It therefore does not appear that caste is just a proxy for cultural similarity. Furthermore, Columns (3) and (8) of the Tables 2.3 and 2.4 also include a dummy

[24]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.

variable for being from the same big main caste. The results suggest that it is the small caste which matters for preference. If caste was a proxy for cultural identity, large caste groupings should be stronger than smaller groups.

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 unobservably bad 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 people from other caste are observationally different from those who do not, or whether people who receive letter from people from other castes are observationally different from those who do not. In Columns 1 and 3 of Panel A in Table 2.7, we show the average quality index Q for ad placers who have indicated to us that they have written to at least one letter from another caste (or one letter to a caste below them, or one letter to a caste above them) versus those who have written to only 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 observable differences between people who write out of castes 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 of course still leaves open the possibility that they 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 a person: 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 (or a caste below, or a caste above) to that of people who have only written within their caste. In an alternative specification, we also regress the quality 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 2.7) suggest that the ultimate marriage outcome of those who write out of castes (or below, or above caste) are no different that those of those who do not (panel A). Likewise, those to whom people from other caste write marry with people of the same observable quality (panel B). This is strong indication that writing out of caste does not sends 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 reflect a true preference for eventually marrying within the same caste. This preference seems to be related to caste itself, rather than characteristics castes would 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.

2.4.4 Do these preferences reflect dowry?

We have so far ignored dowries, for the reasons discussed in some detail in Section 2. None of those arguments are however entirely water-tight. The argument in Proposition 5, for example, depends on the assumption that exploring all the potentially attractive options is cheap enough.

One way to check the validity of this argument is to test one of its testable implications: those who either say that they do not want dowry or say that they will not offer dowry should get the same responses as everyone else. To verify this conjecture in the data we re-estimate the preferences in the sample of letters that explicitly mentions not wanting a dowry, and comparing the overall results. We do this in Table 2.8 where we interact not wanting a dowry with each characteristic. The full specification is presented in column (1) and (2), and the parsimonious specification is presented in columns (3) and (4).[25] The even columns correspond to the interaction terms and the odd columns to main effect. The results are noisier for the interactions than for the main effects given the sample size, but overall, we cannot reject 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, while it is not significant), while the role of predicted income does not change. This give us if anything an even larger marginal rate of substitution between caste and income, which is the opposite of what would have been predicted in our model if families needed

[25] 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.

to compensate a rich groom with a higher dowry (but would not need to do so for caste when tastes are similar).

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 more or less letters, their attributes as mentioned in the letter are valued similarly and the quality of their responses and their eventual match is not significantly different than others, except for female ad placers who receive slightly worse applicants when they do not offer a dowry (results not reported to save space, but available form the authors).

2.5 Stable matching estimates

Following Hitsch et al. (2006), in this section, we compute the set of stable matches implied by the preferences we just estimated. 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 rather prefer being with them than with their current spouse.

2.5.1 Empirical strategy

The pool of men and women attempting to match within this market is defined as the entire set of ads within the dates of the survey, from October 2002 and March 2003. Although this is a simplification, it appears to be a good approximation of the actual market: most people both place and reply to ad (75 percent of our sample had replied to at least one ad). Furthermore, most people (40 percent) only post an ad once, so that there is no repetition.

We now need 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 our the parameters in equation (2.2) 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. We use both the estimates coming from the ranking and the decision to consider or not a letter [26]

$$
\begin{aligned}
U_{ij}^{m} &= \hat{\alpha}_m X_j + \hat{\beta}_m f\left(X_i, X_j\right) \\
U_{ij}^{f} &= \hat{\alpha}_f X_i + \hat{\beta}_f f\left(X_i, X_j\right)
\end{aligned}
\tag{2.4}
$$

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

$$
\begin{aligned}
R_{ij}^{m} \quad &= \quad n \quad if \quad \left\{ \begin{array}{c} U_{ij'}^{m} > U_{ij}^{m} > U_{\widetilde{ij}}^{m} \\ and \quad R_{ij'}^{m} = n-1 \quad and \quad R_{\widetilde{ij}}^{m} = n+1 \end{array} \right\} \\
R_{ij}^{f} \quad &= \quad n \quad if \quad \left\{ \begin{array}{c} U_{i'j}^{m} > U_{ij}^{m} > U_{\widetilde{ij}}^{m} \\ and \quad R_{i'j}^{m} = n-1 \quad and \quad R_{\widetilde{ij}}^{m} = n+1 \end{array} \right\}
\end{aligned}
$$

Applying this methodology for all males and females in the sample, this generates a full set of ordinal preferences for each ad placer with respect to all ad placers of the opposite gender.

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 men have been matched. Since they are the long side of the market, some women will remain single.

In this setting, ties will occur. This is due to the fact that some people are, based on the characteristics chosen in the main regression, identical one to another. These ties are broken randomly. However, this is not of great importance in this context (unlike what has been discussed in other settings, see Erdil and Ergin 2008). Since ties are generated by individuals

[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.

who have exactly the same preferences, randomizing who is selected does not create any problem: if individuals A and B are identical and have the same preferences, it is irrelevant for our purpose whether person C is matched with A or with B.

In order to obtain confidence intervals for the results of the matching algorithm, preference estimation from the previous section were bootstrapped. Then, using each of the 1000 iterations of the bootstrap, the algorithm was separately computed. This resulted in 1000 stable matchings 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.

We introduce search frictions in the following way. First, we constrain males to contact individuals close to their unconstrained optimal choice (within 1000 ranks). 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.

Two other important variations were introduced, to explore the role that caste preference play in equilibrium. In the first case, caste matching was imposed on all individuals. Any suitor who approaches a female of a different caste is immediately rejected. This provides a benchmark equilibrium in the case of perfect caste matching. Symmetrically, "caste-blindness" is also considered by removing any caste-related coefficients from the preference parameters when computing equation (2.4). This allows us to simulate what the equilibrium would look like if caste was simply ignored.

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 sample. This was done simply because our overall sample is small and this insure that whatever difference observed between the algorithm and the observed data does not stem from any difference between the samples. Results are extremely similar if we compare the algorithm results for the entire set of ads to the sampled outcomes.

2.5.2 Results

This section presents the stable matches estimated with the algorithm as described above. It suggests that the observed outcomes are fairly similar to what is predicted by a Gale-Shapley algorithm despite the simplifications it imposes.

Who stays single?

Table 2.9 display 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. Columns 1 and 2 show the values of the difference at the 2.5 percent and 97.5 percent of the distribution in the bootstrap simulation when we use the preferences parameters estimated with from the "considered" data (Table 2.3). Columns 3 and 4 repeat the same exercise with the preferences estimated from the "rank" data (Table 2.4). In all cases, we use the linear model although similar results were obtained with the non-linear specification. Column 5 present the mean differences in the actual sample with the confidence interval around that mean shown in Columns 6 and 7.

In most cases, the differences between married and singles observed in the stable matching have the same signs as the actual differences. Older, shorter, darker skinned, less beautiful and less educated women are more likely to be single in both the stable matches and the actual data. Commerce graduates are also less likely to be single. Being from West Bengal, being beautiful or very beautiful, and occupational wage and income reported in the ad does not affect the probability to be married or single. For 7 out of the 16 variables, the actual difference between single and married in our data lies within the confidence interval of the stables matches. In 5 more cases, the confidence intervals overlap.

There are two variables for which the stable matching algorithm gets the sign wrong. The most important one is the role of caste.[27] While we predict that the singles would be of a lower caste than those who are married, it is not true in the real data, where the singles are, if anything, of slightly higher castes.

In most cases where the point estimate of the difference in the actual data does not lie

[27]The other one being whether a woman has a science degree.

within the bounds of the stable matches estimate, the stable matches overestimate the differences between the variable. This probably reflects the fact that other factors than these attributes eventually determines whether or not people decide to marry: this will thus dampen the role of the variable in the case of actual matches.

As a first pass to investigate this possibility, panel B introduces search frictions. The resulting characteristics of married and single female is actually quite similar in both scenario (possibly because the search frictions do not do much). There are now 6 cases where the point estimates in the data are within the bound of the stables matches, and 6 where the confidence interval overlap.

Panel C repeats the exercise for males. Since men are the short side of the market, without any search friction, all men are married. The algorithm results are thus only presented in the case of search frictions. The signs are now congruent for all the variables, and the observed means differences between single and married fits within the 95 percent predicted by the stable matching algorithm in eight out of thirteen characteristics although the algorithm does not produce very tight predictions. The main characteristics have the expected signs on the change to be married however: males who are more educated, have a science degree, and report higher income or wages, are less likely to remain single, both in reality and as the results of the matching algorithm.

Who marries whom?

We now compare the characteristics of the couples in the stable matches and in our actual sample. Table 2.10 displays the main results. Columns 1 and 2 present the lower and upper bound for the stable matches, using the "considered" response to estimate the preferences, columns 3 and 4 repeat the exercise for the estimates based on ranking. Columns 5 to 7 present the data for comparing ad placers and the people they consider. Columns 8 to 10 present the data on the 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, and the confidence intervals overlap in 15 cases, even though for many variables, the bounds on the stable matches are quite tight.

Not surprisingly, a dominant feature is the tendency to marry within one's caste. The stable matching based on the considered data predicts that 77 percent to 87 percent of the couples will have the same caste, while the estimates based on ranking predicts that 67 percent to 84 percent of the couples will have the same caste. In practice, almost 70 percent of the couple are from the same caste.

Turning to the other pattern, the prediction regarding age are roughly similar in the simulations and in the data. Husbands are almost 6 years older than their wives on average. Height differences are slightly underestimated and the correlations are a little bit too high. Both the data and the simulation 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, and probably comes from the fact that for women, we only have education in the regression, while for men, we have education, income, and wage. As we discussed, these three variables may be colinear, which may lead to underestimating the importance of education in the groom wanted regression.

Comparing our indices of quality, we find that males have higher quality than females although this measure is slightly overestimated compared to the observed data. 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 improves slightly the fit of the algorithm result. Although the results are not altered greatly, they are modified in a way that usually increases their resemblance to the observed data. The education and wage differences become more positive with search frictions than they were without them. Height differences are now including the observed data in the case where considered probabilities are used as preference parameters. Family origin matching is still overestimated when compared to the observed matches.

We also computed the equilibrium under two variants, presented in Table 2.A.6. 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. Furthermore, while not shown, the characteristics of who remains single and who finds a match are almost identical when women proposed. This is encouraging, since finding very different results when men and women propose would have suggested a multiplicity of equilibria in our marriage market. Finally, we also imposed a balanced sex ratio by randomly selecting a subset of females equal to the number of male ads in the sample. While this creates some differences in the algorithm, the results are still fairly similar to the ones presented in the main tables.

2.6 The role of caste preferences in equilibrium

In Section 2.4, we saw that there was a strong preference for marrying within one's caste. Men were willing to sacrifice up to 4 categories of education and women more than 300 percent of a man's income in order to remain within one's caste. In section 4, we saw that indeed, about 70 percent of the marriages take place within caste. While individual appear to be ready to pay a high price to marry within their caste, do they end up paying it in equilibrium? More generally, does the preference for marrying within caste affects other dimension of matching?

In Section 2.2, the theoretical model emphasized that the equilibrium role of caste crucially depends on whether preferences for caste are horizontal or vertical. Section 2.4 has then argued that the estimate we obtain from the estimation of preferences suggest that the desire for endogamy is much larger than that for hypergamy in this context, that is that the preference for caste is horizontal.

The theoretical model discussed above also suggests that one important element is whether the distribution of male and female "quality" is balanced across castes. In this context, 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 used the "quality" measure defined above (without any caste parameter) and compared the overall distribution of quality by caste for males and females among the interview sample. We find that the distributions are fairly similar for all major caste groups (Brahmin, Kayastha, Baisya and Sagdope).but are less similar for caste groups with fewer observations. These results hold whether one compares the distribution in quality among the interview sample or among the letter writers.

Finally, the model we elaborated earlier also suggests that the equilibrium price will be low when there is a group who does not have caste preferences. We find that in our data, between 25 and 30 per cent of individuals are willing to marry outside their caste. This roughly corresponds to the number of matches observed that are not within one's caste, although not all individuals who say they would be willing to marry outside their caste eventually do so (and vice versa).

Given these pieces of evidence, what do the algorithm results tell us about the actual role of caste in the matching equilibrium? Table 2.11 takes one cut at this issue. The first columns of panel A of Table 2.11 reproduce columns 1 and 2 of the first panel of Table 2.10. The second panel constrains all marriages to take place within one's caste. Panel C entirely 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 affects very much how matches look like along the other dimensions. As expected, the correlations in age, height, education increase as the preferences for castes diminishes (they are the highest when matches are restricted to be within caste, and the lowest when preferences for caste is "shut down"), but the gradient is fairly low, and very few of the other variables are affected.

Moreover, the proportion of within-caste marriage falls by a large fraction when preferences are caste-blind. This suggests that castes do not proxy for other attributes. There are many potential matches for each person, both within and outside her caste.

Columns (3) to (10) present the algorithm results by key caste groups. These results suggest that the conclusions drawn above are fairly similar across caste groups, despite the fact that the sub-castes within the Baisya and the Sagdope groups are relatively smaller than those within the Brahmin. However, imposing caste-blindness appears to affect more importantly smaller castes than Brahmins or Kayashtas. In particular, the first two groups still marry within caste in 20-40 percent of the cases. Baisya and Sagdope, on the other hand, almost rarely marry within their castes once caste preferences are omitted. Some correlations among the Sagdope, in particular age and education correlations, appear to fall once one imposes within-caste matching.

Overall it seems that once the algorithm removes caste information from the preferences, the individuals marry almost identical individuals but from another caste. This would suggest that the equilibrium price of caste ought to be low. To further study this pattern, we look at the actual matching patterns of our sample. We found 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, that is that individuals who have less of a preference for caste would select to marry outside their caste. Since their "cost" of caste matching is lower, this is what we would measure in equilibrium. Thus, we turn to the results of the algorithm to attempt to alleviate this concern since in this context, there is no unobservable taste determinants. To do so, a regression was run for each iteration of the algorithm. This regression controlled for all of the ad placer's characteristics and compared various measures of quality of the match for the pairs that were within caste to those that were not. Table 2.12 presents the mean and the standard deviation of the coefficients on whether or not the couple was within the same caste. These results suggest very small, insignificant and often in the wrong direction prices of caste matching. For example, individuals who marry within their own caste are also more likely to marry more educated individuals.

As a comparison, the equilibrium price of education is computed as well in a similar fashion. The left hand panel of Table 2.12 suggests that as opposed to caste, individuals are forced to make a trade-off between for example beauty and educational level of a female. A man who marries a female who has more education also marries one who is older, less beautiful and darker skinned. Little correlation is found between a prospective groom's education and other qualities. One should note, however, that the tradeoff in equilibrium in this case is still smaller than the one observed from preferences.

We thus find that the equilibrium price of caste is very small and that altering the way caste is perceived by individuals does not transform the overall matching equilibrium importantly. This is consistent with our theoretical model and the estimated preferences we obtained.

2.7 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.

A number of conclusions follow from this: 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 when caste becomes less important inequality might increase along other dimensions as we see more assortative matching. Given that the matching is already close to being assortative this is probably not an important concern.

2.8 References

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. available at `http://www.iza.org/conference_files/worldb2007/bargain_o1569.pdf`.

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

Bertrand, M. and S. Mullainathan (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review 94*(4), 991–1013.

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.

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

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 Vol. 75*(1), 117–132.

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

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 available at `http://ssrn.com/paper=719121`.

Hitsch, G. J., A. Hortacsu, and D. Ariely (2006). What Makes You Click? Mate Preferences and Matching Outcomes in Online Dating. MIT Sloan Research Paper No. 4603-06, available at `http://ssrn.com/paper=895442`.

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.

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.

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.

2.9 Tables and figures

Table 2.1: **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
Other degree	0.00	0.04	0.01	0.10	0.01	0.08	0.01	0.10
Humanities/Arts	0.66	0.47	0.58	0.49	0.12	0.33	0.05	0.21
Commerce	0.11	0.31	0.12	0.33	0.37	0.48	0.40	0.49
Science	0.28	0.45	0.30	0.46	0.55	0.50	0.55	0.50
Other field	0.01	0.11	0.01	0.07	0.02	0.15	0.00	0.00
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
Location								
Calcutta	0.51	0.50	0.80	0.40	0.50	0.50	0.76	0.43
Other residence								
West Bengali	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	0.43	0.50	0.10	0.30	0.33	0.47
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
Ads which omit...								
Caste	0.02	0.13	0.00	0.04	0.03	0.16	0.01	0.08
Age	0.01	0.10	0.01	0.12	0.02	0.13	0.04	0.20
Height	0.04	0.19	0.04	0.19	0.10	0.30	0.11	0.31
Education	0.10	0.30	0.08	0.27	0.22	0.42	0.18	0.39
Field	0.27	0.44	0.25	0.43	0.39	0.49	0.30	0.46
Residence	0.86	0.35	0.84	0.37	0.70	0.46	0.52	0.50
Family origin	0.29	0.45	0.23	0.42	0.32	0.47	0.29	0.45
Wage	0.83	0.38	0.84	0.37	0.25	0.43	0.57	0.50
Income	0.98	0.13	0.97	0.16	0.78	0.41	0.74	0.44
Skin tone	0.23	0.42	0.21	0.41				
Beauty	0.25	0.43	0.27	0.44				

Statistics are computed only among individuals reporting a given characteristics

Table 2.2: **Summary statistics-Respondents**

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
Other degree	0.03	0.18	0.00	0.00	0.04	0.19	0.00	0.00
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
Humanities/Arts	0.13	0.33	0.52	0.50	0.63	0.48	0.79	0.41
Commerce	0.34	0.47			0.11	0.31		
Science	0.51	0.50	0.48	0.50	0.25	0.43	0.21	0.41
Other field	0.02	0.14	0.00	0.00	0.01	0.12	0.00	0.00
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
Location								
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
West Bengali	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				
Letters which omit								
Caste	0.30	0.46	0.01	0.11	0.28	0.45	0.02	0.12
Age	0.04	0.20	0.00	0.00	0.03	0.17	0.00	0.00
Height	0.13	0.33	0.00	0.00	0.08	0.27	0.00	0.00
Education	0.08	0.27	0.00	0.00	0.04	0.19	0.00	0.00
Field	0.20	0.40	0.39	0.49	0.25	0.43	0.22	0.42
Residence	0.15	0.36	0.00	0.00	0.19	0.40	0.00	0.00
Family origin	0.31	0.46	0.03	0.18	0.27	0.44	0.00	0.00
Wage	0.44	0.50	0.08	0.28	0.86	0.35	0.79	0.41
Income	0.66	0.47	0.31	0.46	0.98	0.14	0.04	0.19
Skin tone					0.14	0.35	1.00	0.00
Beauty					0.36	0.48	1.00	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 characteristics. 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 2.3: **Probability of considering a 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	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)

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)
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)
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***	
				(0.0193)					(0.0228)	
N	5628	5628	5628	5628	5628	3944	3944	3944	3944	3944

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 2.4: **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**	1.3319***	0.4314***	1.2591***		1.5022***	1.4072***	0.3595***
	(0.2933)		(0.3821)	(0.2942)	(0.0928)	(0.3458)		(0.4292)	(0.3492)	(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**	0.3725*	-0.1421**
	(0.1341)		(0.1437)	(0.1345)	(0.0418)	(0.1699)		(0.1878)	(0.1710)	(0.0461)
Diff. in caste*Lower caste male	0.1070		0.0767	-0.0784	0.0281	-0.3310		-0.2548	0.3626*	-0.0976*
	(0.1183)		(0.1280)	(0.1188)	(0.0372)	(0.1705)		(0.1882)	(0.1724)	(0.0458)
Same caste*only within	1.1726		1.1737	1.1670	0.2128	2.1112		2.0985	2.1633	0.7029
	(0.9116)		(0.9117)	(0.9163)	(0.2848)	(1.3256)		(1.3257)	(1.3420)	(0.3674)
Diff. in caste*only within	0.4459		0.4471	0.4552	0.1670	-0.0183		-0.0094	0.1361	-0.0874
	(0.3334)		(0.3334)	(0.3350)	(0.1117)	(0.5781)		(0.5782)	(0.5843)	(0.1582)
Same caste*no bar	-0.8681**		-0.8678**	-0.8602**	-0.2911**	-0.8599*		-0.8912*	-0.9396*	-0.2521*
	(0.3258)		(0.3258)	(0.3267)	(0.1028)	(0.4315)		(0.4328)	(0.4362)	(0.1156)
Diff. in caste*no bar	0.1021		0.1041	0.0831	0.0247	-0.2092		-0.2020	-0.1763	-0.0734
	(0.1071)		(0.1072)	(0.1074)	(0.0342)	(0.1521)		(0.1523)	(0.1538)	(0.0409)
Diff. in age	0.0345	0.0255	0.0348	0.0214	0.0053	0.5215***	0.5411***	0.5205***	0.4463***	0.1457***
	(0.0405)	(0.0405)	(0.0405)	(0.0406)	(0.0127)	(0.0816)	(0.0820)	(0.0816)	(0.0817)	(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)	(0.0023)	(0.0007)	(0.0057)	(0.0057)	(0.0057)	(0.0057)	(0.0015)
Diff. in height	9.5137***	9.8711***	9.4794***	9.8311***	3.5492***	7.2790*	6.8472*	7.2231*	7.6700*	1.9194*
	(2.5694)	(2.5757)	(2.5701)	(2.5784)	(0.8651)	(3.2304)	(3.2517)	(3.2309)	(3.2590)	(0.8796)
Squared diff. in height	-24.5037**	-26.3139**	-24.4011**	-25.3582**	-9.5136**	-69.0103***	-68.9625***	-68.8785***	-70.3860***	-18.7289***
	(9.2415)	(9.2562)	(9.2436)	(9.2646)	(3.2019)	(12.3135)	(12.3931)	(12.3145)	(12.4198)	(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***
	(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)

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)
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***	0.4926***	0.1738***	0.6515***	0.6240**	0.6501***	0.6294***	0.1741***
	(0.1204)	(0.1206)	(0.1205)	(0.1206)	(0.0383)	(0.1891)	(0.1897)	(0.1891)	(0.1906)	(0.0509)
Same location	0.4846	0.4160	0.4831	0.4077	0.1181	-0.1912	-0.2096	-0.1944	-0.2105	-0.0551
	(0.3086)	(0.3097)	(0.3086)	(0.3094)	(0.0959)	(0.2876)	(0.2893)	(0.2877)	(0.2906)	(0.0777)
Same family origin	0.2665	0.3861*	0.2656	0.2767	0.0712	0.7190***	0.8573***	0.7150***	0.8015***	0.1903**
	(0.1710)	(0.1710)	(0.1710)	(0.1718)	(0.0538)	(0.2156)	(0.2163)	(0.2156)	(0.2177)	(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***	-0.4995***	-0.1292***
						(0.1005)	(0.1012)	(0.1005)	(0.1014)	(0.0271)
Beautiful						0.2045	0.2127	0.2095	0.1762	0.0404
						(0.1885)	(0.1893)	(0.1885)	(0.1907)	(0.0505)
Very beautiful						0.5376	0.5587	0.5363	0.4229	0.1614*
						(0.2934)	(0.2951)	(0.2934)	(0.2965)	(0.0787)
Predicted income				3.2430***					0.9296***	
				(0.1715)					(0.2302)	
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 2.5: **Responses to "not too good" letters**

	Ads placed by females				Ads placed by males			
	Considered		Rank		Considered		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same caste	0.1073*	0.1134**	1.0817*	1.2763***	0.0884	0.1498***	1.2144*	1.4484***
	(0.0451)	(0.0364)	(0.4438)	(0.3404)	(0.0489)	(0.0418)	(0.5085)	(0.4270)
Diff. in caste* Higher caste male	0.0464*	0.0253	0.2376	0.0389	0.0570*	0.0186	0.7497**	0.4847*
	(0.0197)	(0.0166)	(0.1888)	(0.1524)	(0.0243)	(0.0203)	(0.2536)	(0.2100)
Diff. in caste* Lower caste male	0.0027	-0.0058	-0.0291	-0.1165	0.0373	0.0431*	0.6135*	0.5529**
	(0.0175)	(0.0146)	(0.1714)	(0.1356)	(0.0233)	(0.0200)	(0.2464)	(0.2060)
Same caste*only within	-0.0906	-0.0344	-1.0448	-0.6513	0.1245	0.1138	0.4840	0.6478
	(0.1408)	(0.1273)	(1.2780)	(1.1149)	(0.1851)	(0.1679)	(1.8022)	(1.6123)
Diff. in caste*only within	0.0036	0.0062	0.3854	0.5496	0.0096	0.0088	0.5102	0.6311
	(0.0492)	(0.0473)	(0.4439)	(0.4123)	(0.0797)	(0.0751)	(0.7765)	(0.7210)
Same caste*no bar	-0.0733	-0.0527	-0.9739*	-1.0054**	0.0027	-0.0206	-0.4229	-0.9570
	(0.0508)	(0.0415)	(0.4908)	(0.3853)	(0.0574)	(0.0499)	(0.6295)	(0.5286)
Diff. in caste*no bar	0.0031	0.0069	0.1017	0.1457	-0.0265	-0.0066	-0.5458*	-0.3208
	(0.0163)	(0.0135)	(0.1559)	(0.1243)	(0.0206)	(0.0175)	(0.2236)	(0.1847)
Diff. in age	0.0058	0.0053	0.0372	0.0696	0.0435***	0.0436***	0.5121***	0.4841***
	(0.0060)	(0.0051)	(0.0560)	(0.0459)	(0.0120)	(0.0105)	(0.1297)	(0.1103)
Squared diff. in age	-0.0008*	-0.0009**	-0.0097***	-0.0117***	-0.0023*	-0.0021**	-0.0270*	-0.0245**
	(0.0003)	(0.0003)	(0.0028)	(0.0025)	(0.0009)	(0.0008)	(0.0105)	(0.0085)
Diff. in height	0.9198*	0.7934*	9.2645*	6.8037*	0.7503	0.9038*	6.2082	7.4802*
	(0.4189)	(0.3334)	(4.1113)	(3.2594)	(0.4284)	(0.3641)	(4.3149)	(3.5929)
Squared diff. in height	-3.2350	-2.0427	-25.9230	-13.3929	-6.1195***	-6.0644***	-66.2058***	-65.7108***
	(1.7081)	(1.2791)	(16.7790)	(12.7629)	(1.4949)	(1.3248)	(15.1818)	(13.3146)
High school	-0.0930	-0.0507	-0.0679	0.3281	0.1697	0.1437	2.9543*	2.0051**
	(0.2237)	(0.1441)	(2.0167)	(1.2549)	(0.1245)	(0.0766)	(1.2073)	(0.7601)
Post-secondary	0.0173	0.0473	1.0474	1.2573	0.3295	0.2195	4.5315*	2.4932
	(0.2323)	(0.1522)	(2.1097)	(1.3380)	(0.2200)	(0.1573)	(2.2618)	(1.6627)
Bachelor's	-0.0341	0.0017	1.3182	1.2914	0.1965	0.1959	4.4956**	2.9271**
	(0.2323)	(0.1523)	(2.1078)	(1.3368)	(0.1488)	(0.1041)	(1.4671)	(1.0621)
Master's	0.0745	0.1415	2.1164	2.3877	0.3004*	0.2742*	5.8510***	4.1727***
	(0.2374)	(0.1559)	(2.1598)	(1.3715)	(0.1530)	(0.1080)	(1.5109)	(1.1016)
PhD	0.1705	0.1858	3.2869	2.9018*	0.3640	0.3425**	6.2600**	5.9120***
	(0.2413)	(0.1597)	(2.1997)	(1.4062)	(0.1920)	(0.1321)	(1.9928)	(1.4177)
Same education	0.0579	0.0432	0.3489	0.5761*	-0.0065	0.0194	0.1562	0.3351
	(0.0342)	(0.0273)	(0.3252)	(0.2501)	(0.0496)	(0.0373)	(0.5013)	(0.3735)
Male more educated	0.0488	0.0224	0.2172	0.5776	0.0116	0.0001	0.4938	0.5975
	(0.0564)	(0.0448)	(0.5369)	(0.4083)	(0.0611)	(0.0491)	(0.6235)	(0.5000)
Non-rankable degree	0.0831	0.0986	1.3728	1.6644	0.2916*	0.2564*	3.5910*	2.9083**
	(0.2284)	(0.1482)	(2.0635)	(1.2959)	(0.1482)	(0.0999)	(1.4593)	(0.9985)
Science	0.0574*	0.0727**	0.9701***	0.9189***	0.0444	0.0406	0.5336*	0.7062**
	(0.0281)	(0.0234)	(0.2711)	(0.2158)	(0.0236)	(0.0209)	(0.2476)	(0.2152)
Commerce	0.0558*	0.0535*	0.4692	0.3747	0.0074	0.0618	0.5900	1.2313**
	(0.0279)	(0.0238)	(0.2654)	(0.2190)	(0.0466)	(0.0356)	(0.5229)	(0.3771)
Other field	0.0839	0.0639	0.1661	0.4733	-0.2849	-0.0266	0.6582	1.8935
	(0.0881)	(0.0684)	(0.8389)	(0.6334)	(0.2053)	(0.1164)	(2.3068)	(1.2467)
Calcutta	0.0441*	0.0601***	0.5010*	0.5145***	0.0626*	0.0605**	0.9589**	0.6954**
	(0.0205)	(0.0160)	(0.1957)	(0.1468)	(0.0287)	(0.0232)	(0.3092)	(0.2414)
Same location	0.0715	0.0400	0.2603	0.3765	-0.0179	-0.0207	-0.0462	-0.1084
	(0.0468)	(0.0387)	(0.4501)	(0.3577)	(0.0389)	(0.0331)	(0.4131)	(0.3410)
Same family origin	0.0336	0.0349	0.4720	0.1820	0.0913**	0.0691**	0.5997	0.6442*
	(0.0265)	(0.0218)	(0.2558)	(0.2019)	(0.0309)	(0.0249)	(0.3307)	(0.2602)
Log income	0.1641***	0.1494***	1.3992***	1.2974***				
	(0.0281)	(0.0222)	(0.2655)	(0.2022)				
Log wage	0.0951***	0.0860***	0.8867***	0.8047***				
	(0.0212)	(0.0168)	(0.2037)	(0.1536)				

Continued on next page

	Ads placed by females				Ads placed by males			
	Considered		Rank		Considered		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Skin tone					-0.0529***	-0.0421***	-0.4603**	-0.5388***
					(0.0143)	(0.0118)	(0.1494)	(0.1209)
Beautiful					0.0151	0.0170	0.4348	0.1823
					(0.0262)	(0.0219)	(0.2757)	(0.2241)
Very beautiful					0.0915	0.0855*	0.4869	0.6153
					(0.0505)	(0.0419)	(0.5124)	(0.4259)
Diff. in quality less than ptile	50th	75th	50th	75th	50th	75th	50th	75th
N	2767	4141	2488	3753	2048	2909	1762	2553

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 four columns refer to decisions made by females regarding prospective grooms, the last four to decisions made by males regarding prospective brides.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 2.6: **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)

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)
Same family origin	0.0121	0.1294	0.1625	0.0969**	0.6508***	0.9472***
	(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 2.7: **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 2.8: **Dowries and probability of being considered**

	Full Regression		Parsimonious	
	Main effects in sample that does not mention dowries	Interaction of characteristics with not requesting a dowry	Main effects in sample that does not mention dowries	Interaction of characteristics with not requesting a 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)

Continued on next page

	Full Regression		Parsimonious	
	Main effects in sample that does not mention dowries	Interaction of characteristics with not requesting a dowry	Main effects in sample that does not mention dowries	Interaction of characteristics with not requesting a dowry
	(1)	(2)	(3)	(4)
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 2.9: **Difference in individuals' characteristics by marital status**

	Considered		Rank		Observed		
	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile	Mean	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Women, without search frictions							
Age	**0.8759**	**2.6992**	**0.7551**	**2.4377**	0.9215	0.2566	1.5865
Height	*-0.0246*	*-0.0063*	*-0.0279*	*-0.0087*	-0.0035	-0.0119	0.0049
Caste	*0.1842*	*1.0929*	*0.3150*	*1.3770*	-0.0772	-0.4235	0.2691
Education level	-1.0987	-0.6624	-1.1754	-0.8123	-0.1486	-0.3630	0.0658
Arts and Social Science	0.1242	0.3326	0.1567	0.3597	0.0148	-0.0899	0.1195
Commerce	*-0.1693*	*-0.0849*	-0.1783	-0.1108	-0.0416	-0.1118	0.0285
Science	*-0.2599*	*-0.0151*	*-0.2626*	*-0.0398*	0.0292	-0.0677	0.1260
Other field	**-0.0146**	**0.0318**	**-0.0167**	**0.0131**	-0.0023	-0.0180	0.0133
From West Bengal	**-0.1472**	**0.0299**	**-0.1596**	**0.0178**	0.0090	-0.1115	0.0935
Kolkota	*-0.5348*	*-0.1621*	*-0.4795*	*-0.1288*	-0.0290	-0.2126	0.1546
Skin rank	0.4877	0.8295	0.4159	0.8036	0.0214	-0.1407	0.1835
Very beautiful	**-0.0858**	**0.0059**	**-0.0895**	**0.0154**	-0.0141	-0.0707	0.0425
Beautiful	**-0.2190**	**0.0428**	**-0.2097**	**0.0477**	-0.0188	-0.1248	0.0873
Income	**-11265**	**3915**	*-1121*	*3915*	-6267	-11449	-1084
Log wage	**-0.0770**	**0.0860**	**-0.0768**	**0.0966**	0.0065	-0.1470	0.1599
"Quality"	-0.1134	-0.0838	-0.1048	-0.0644	-0.0050	-0.0088	0.0187
Panel B: Women, with search frictions							
Age	**0.4462**	**2.1565**	**0.2880**	**1.7310**	0.9215	0.2566	1.5865
Height	*-0.0240*	*-0.0079*	*-0.0264*	*-0.0118*	-0.0035	-0.0119	0.0049
Caste	*0.1853*	*0.9895*	0.3430	1.3190	-0.0772	-0.4235	0.2691
Education level	-1.0220	-0.6292	-1.1027	-0.7500	-0.1486	-0.3630	0.0658
Arts and Social Science	0.1341	0.3701	0.1684	0.3923	0.0148	-0.0899	0.1195
Commerce	*-0.2080*	*-0.0937*	-0.2237	-0.1119	-0.0416	-0.1118	0.0285
Science	*-0.2660*	*-0.0049*	*-0.2657*	*-0.0269*	0.0292	-0.0677	0.1260
Other field	**-0.0190**	**0.0294**	**-0.0223**	**0.0125**	-0.0023	-0.0180	0.0133
From West Bengal	**-0.1417**	**0.0363**	**-0.1565**	**0.0102**	0.0090	-0.1115	0.0935
Kolkota	*-0.4092*	*-0.1001*	*-0.3302*	*-0.0840*	-0.0290	-0.2126	0.1546
Skin rank	0.4921	0.7767	0.4204	0.7433	0.0214	-0.1407	0.1835
Very beautiful	**-0.1042**	**0.0016**	**-0.0931**	**0.0176**	-0.0141	-0.0707	0.0425
Beautiful	**-0.2086**	**0.0773**	**-0.2020**	**0.0575**	-0.0188	-0.1248	0.0873
Income	*-1347*	*3853*	*-1347*	*3853*	-6267	-11449	-1084
Log wage	**-0.1301**	**0.0820**	**-0.1418**	**0.0861**	0.0065	-0.1470	0.1599
"Quality"	-0.1081	-0.0809	-0.0999	-0.0620	-0.0050	-0.0088	0.0187
Panel C: Men, with search frictions							
Age	**-1.0919**	**0.5233**	*-1.2496*	*0.3194*	0.4175	-0.6997	1.5346
Height	**-0.0179**	**0.0125**	**-0.0179**	**0.0161**	-0.0040	-0.0206	0.0126
Caste	**-0.1533**	**2.0519**	**-0.2714**	**1.6719**	0.1195	-0.3815	0.6205
Education level	*-1.2680*	*-0.5757*	-1.4264	-0.7888	-0.2399	-0.6066	0.1268
Arts and Social Science	**-0.0738**	**0.0811**	**-0.0736**	**0.0714**	-0.0696	-0.1308	-0.0084
Commerce	**0.1040**	**0.4386**	*0.1287*	*0.4776*	0.1201	-0.0281	0.2683
Science	-0.5674	-0.2112	-0.5976	-0.2303	-0.0505	-0.2014	0.1004
Other field	**-0.0149**	**0.0224**	**-0.0156**	**0.0334**	0.0000	0.0000	0.0000
Family origin	**-0.2584**	**0.1309**	**-0.2580**	**0.1846**	0.0197	-0.1223	0.1617
Calcutta	**-0.5658**	**0.2069**	**-0.2901**	**0.2087**	0.0363	-0.1122	0.1847
Income	*-8887*	*-2954*	*-9171*	*-2845*	-13560	-42033	14912
Log wage	-0.9925	-0.4129	-1.0500	-0.5386	-0.1141	-0.3196	0.0915
"Quality"	-0.1306	-0.0583	-0.1255	-0.0502	-0.0193	-0.0427	0.0041

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 2.10: **Couples characteristics, simulated and observed**

	Considered		Rank		Observed-considered			Observed-matched		
	2.5 ptile	97.5 ptile	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)	(10)
					Panel A: Without search frictions					
Age diff.	**5.3394**	**6.2323**	**5.3812**	**6.2363**	5.9032	5.8191	5.9873	5.6993	5.3476	6.0510
Age corr.	**0.7990**	**0.9242**	*0.8540*	*0.9419*	0.8331	0.8144	0.8507	0.6521	0.5700	0.7341
Height diff.	**0.1043**	**0.1235**	**0.1032**	**0.1221**	0.1201	0.1178	0.1223	0.1237	0.1146	0.1328
Height corr.	0.8108	0.9085	0.8187	0.9023	0.3825	0.3473	0.4188	0.3880	0.2875	0.4886
Same caste	0.8682	0.9732	*0.7646*	*0.9389*	0.7506	0.7333	0.7679	0.6937	0.6396	0.7478
Caste diff.	**0.0444**	**0.4856**	0.1626	0.6931	0.0916	0.0504	0.1328	-0.0071	-0.1584	0.1443
Caste corr.	**0.6536**	**0.9600**	**0.4668**	**0.8318**	0.8450	0.8202	0.8682	0.7599	0.6873	0.8324
Same education	**0.2529**	**0.7882**	**0.2527**	**0.7495**	0.4487	0.4299	0.4675	0.4380	0.3778	0.4982
Education diff.	-0.5093	0.0084	-0.4060	0.0164	0.3385	0.3120	0.3823	0.2902	0.1393	0.4410
Education corr.	**0.2368**	**0.6001**	**0.1597**	**0.5543**	0.4202	0.3778	0.4620	0.3564	0.2383	0.4746
Same family origin	0.9898	1.0000	0.9773	1.0000	0.7839	0.7655	0.8024	0.7644	0.7060	0.8229
Family origin diff.	**-0.0047**	**0.0092**	**-0.0058**	**0.0153**	0.0054	-0.0154	0.0263	0.0433	-0.0208	0.1073
Family origin corr.	0.9769	1.0000	0.9502	1.0000	0.5407	0.4959	0.5814	0.5147	0.3932	0.6361
Same residence	**0.0000**	**1.0000**	**0.0000**	**1.0000**	0.4687	0.4346	0.5028	0.4831	0.3834	0.5829
Location corr.	**-1.0000**	**0.4891**	**-0.4985**	**0.4961**	0.0441	-0.0393	0.1195	-0.0566	-0.2246	0.2142
Log wage diff.	-0.4990	-0.0826	-0.4941	-0.0804	0.1375	0.0811	0.1939	0.2462	0.1349	0.3575
Log wage corr.	**-0.1670**	**0.4222**	**-0.1542**	**0.4106**	0.0687	-0.0720	0.2017	0.1855	-0.1284	0.4993
Income diff.	**-11375**	**10300**	**-6000**	**18800**	9277	-3842	22397	28374	-16	56764
Income corr.	**-0.6231**	**1.0000**	**-1.0000**	**1.0000**	0.5760	0.4923	0.8139	0.4474	0.0837	0.8110
Quality diff.	*0.1299*	*0.1554*	0.1377	0.1638	0.1026	0.0983	0.1069	0.1202	0.1069	0.1336
Quality corr.	**0.0941**	**0.4640**	**0.1143**	**0.4730**	0.0386	-0.2434	0.3383	0.1950	0.0714	0.3187
					Panel B: With search frictions					
Age diff.	**5.2017**	**6.2993**	**5.3119**	**6.3414**	5.9032	5.8191	5.9873	5.6993	5.3476	6.0510
Age corr.	**0.7700**	**0.9167**	*0.8369*	*0.9379*	0.8331	0.8144	0.8507	0.6521	0.5700	0.7341
Height diff.	**0.1036**	**0.1241**	**0.1014**	**0.1220**	0.1201	0.1178	0.1223	0.1237	0.1146	0.1328
Height corr.	0.7833	0.8920	0.7846	0.8904	0.3825	0.3473	0.4188	0.3880	0.2875	0.4886
Same caste	0.8869	0.9874	*0.7513*	*0.9464*	0.7506	0.7333	0.7679	0.6937	0.6396	0.7478
Caste diff.	**0.0040**	**0.4286**	*0.1013*	*0.6970*	0.0916	0.0504	0.1328	-0.0071	-0.1584	0.1443
Caste corr.	**0.6889**	**0.9915**	**0.5025**	**0.8790**	0.8450	0.8202	0.8682	0.7599	0.6873	0.8324
Same education	**0.2325**	**0.7870**	**0.2029**	**0.7515**	0.4487	0.4299	0.4675	0.4380	0.3778	0.4982
Education diff.	-0.4397	0.1527	-0.2729	0.1772	0.3385	0.3120	0.3823	0.2902	0.1393	0.4410
Education corr.	**0.2223**	**0.6350**	**0.1207**	**0.6053**	0.4202	0.3778	0.4620	0.3564	0.2383	0.4746
Same family origin	0.9799	1.0000	0.9715	1.0000	0.7839	0.7655	0.8024	0.7644	0.7060	0.8229
Family origin diff.	**-0.0061**	**0.0149**	**-0.0109**	**0.0189**	0.0054	-0.0154	0.0263	0.0433	-0.0208	0.1073
Family origin corr.	0.9524	1.0000	0.9346	1.0000	0.5407	0.4959	0.5814	0.5147	0.3932	0.6361
Same residence	**0.0000**	**1.0000**	**0.0000**	**1.0000**	0.4687	0.4346	0.5028	0.4831	0.3834	0.5829
Location corr.	**-0.7262**	**1.0000**	**-0.5000**	**0.5080**	0.0441	-0.0393	0.1195	-0.0566	-0.2246	0.2142
Log wage diff.	-0.3845	0.0484	-0.3982	0.0424	0.1375	0.0811	0.1939	0.2462	0.1349	0.3575
Log wage corr.	**-0.1770**	**0.4803**	**-0.2289**	**0.4747**	0.0687	-0.0720	0.2017	0.1855	-0.1284	0.4993
Income diff.	**-6000**	**188000**	**-6750**	**238001**	9277	-3842	22397	28374	-16	56764
Income corr.	**-1.0000**	**1.0000**	**-1.0000**	**1.0000**	0.5760	0.4923	0.8139	0.4474	0.0837	0.8110
Quality diff.	*0.1310*	*0.1653*	0.1405	0.1783	0.1026	0.0983	0.1069	0.1202	0.1069	0.1336
Quality corr.	**0.0543**	**0.4191**	**0.0688**	**0.4390**	0.0386	-0.2434	0.3383	0.1950	0.0714	0.3187

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 2.11: **Couples characteristics and the impact of caste, by caste**

	All castes		Brahmin		Kayastha		Baisya		Sagdope	
	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Without restrictions										
Age diff.	5.3394	6.2323	5.4830	6.3200	5.3668	6.1957	5.5092	6.2090	5.4749	6.1827
Age corr.	0.7990	0.9242	0.8677	0.9515	0.8697	0.9512	0.7453	0.8808	0.8018	0.9160
Height diff.	0.1043	0.1235	0.1086	0.1276	0.1035	0.1227	0.1057	0.1196	0.1065	0.1208
Height corr.	0.8108	0.9085	0.8590	0.9303	0.8466	0.9214	0.7170	0.8425	0.7740	0.8790
Same caste	0.8682	0.9732	0.7340	0.9899	0.9661	0.9991	0.9229	0.9946	0.7696	0.9790
Same education	0.2529	0.7882	0.2187	0.8429	0.2055	0.8016	0.3053	0.7483	0.2652	0.7877
Education diff.	-0.5093	0.0084	-0.5910	0.0262	-0.6129	-0.1270	-0.5431	-0.1430	-0.4906	-0.0257
Education corr.	0.2368	0.6001	0.3086	0.6688	0.2840	0.6453	0.2693	0.5692	0.2372	0.5628
Log wage diff.	-0.4990	-0.0826	-0.3596	-0.1905	-0.3894	-0.2215	-0.5133	-0.2609	-0.3747	-0.1432
Log wage corr.	-0.1670	0.4222	0.0651	0.2787	0.0120	0.2131	-0.0285	0.2019	-0.0442	0.2387
Quality diff.	0.1299	0.1554	0.1286	0.1512	0.1375	0.1513	0.1266	0.1488	0.1203	0.1452
Quality corr.	0.0941	0.4640	0.1419	0.4386	0.1034	0.3954	0.1456	0.3845	0.1365	0.3860
Panel B: With forced caste matching										
Age diff.	5.3814	6.2504	5.3744	6.5029	5.2848	6.2702	5.2521	6.4215	4.9047	6.2835
Age corr.	0.7856	0.9130	0.8176	0.9438	0.8413	0.9483	0.6697	0.8998	0.7200	0.9207
Height diff.	0.1050	0.1237	0.1050	0.1278	0.1033	0.1247	0.1012	0.1254	0.1039	0.1294
Height corr.	0.7998	0.8978	0.8624	0.9426	0.8350	0.9399	0.6714	0.8734	0.6927	0.9031
Same caste	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Same education	0.2612	0.7835	0.2034	0.8487	0.2127	0.8216	0.2959	0.7273	0.2143	0.8148
Education diff.	-0.4933	-0.0132	-0.6792	0.0508	-0.6028	0.0202	-0.5000	0.0556	-0.3333	0.4037
Education corr.	0.2538	0.6059	0.2106	0.7548	0.1849	0.6601	0.1375	0.5903	-0.1395	0.7290
Log wage diff.	-0.5338	-0.0920	-0.6701	0.0481	-0.7318	0.4171	-0.8300	-0.1611	-0.7702	0.3437
Log wage corr.	-0.1424	0.4106	-0.4029	0.4733	-0.8488	0.8865	-0.1616	0.9073	-0.9447	0.9537
Quality diff.	0.1297	0.1562	0.1218	0.1702	0.1118	0.1514	0.1286	0.1719	0.1040	0.1671
Quality corr.	0.0980	0.4547	0.0327	0.5188	0.0353	0.4921	0.0893	0.4734	-0.0952	0.5946
Panel C: Caste-blinded										
Age diff.	5.3867	6.2850	5.2343	6.2655	5.4810	6.4838	5.2844	6.3530	5.2500	6.3714
Age corr.	0.8818	0.9611	0.8382	0.9536	0.8706	0.9624	0.8910	0.9714	0.8947	0.9741
Height diff.	0.1039	0.1234	0.1031	0.1245	0.1037	0.1235	0.1004	0.1225	0.1026	0.1280
Height corr.	0.8937	0.9529	0.8887	0.9605	0.8849	0.9573	0.8900	0.9630	0.8797	0.9658
Same caste	0.1552	0.2357	0.1829	0.3690	0.2165	0.3904	0.0000	0.0862	0.0000	0.1622
Same education	0.2019	0.8503	0.2047	0.8731	0.2043	0.8507	0.2222	0.8969	0.1430	0.8846
Education diff.	-0.5890	0.0268	-0.6240	0.0842	-0.6621	0.0299	-0.5911	0.1031	-0.5963	0.3513
Education corr.	0.2913	0.6902	0.2479	0.7807	0.2161	0.7153	0.2584	0.7994	-0.0391	0.7909
Log wage diff.	-0.4723	-0.0717	-0.6604	0.0217	-0.6750	0.3825	-0.7236	-0.0225	-0.6789	0.4324
Log wage corr.	-0.1366	0.4105	-0.3681	0.5017	-0.6788	0.8421	-0.2646	0.7928	-0.8874	0.8542
Quality diff.	0.1284	0.1562	0.1315	0.1780	0.1091	0.1529	0.1304	0.1775	0.0834	0.1501
Quality corr.	0.0888	0.5048	0.0301	0.5254	0.0588	0.5425	0.0929	0.5813	-0.0936	0.6616

Table 2.12: **Distribution of costs of...**

	Keeping caste		Education	
	Male	Female	Male	Female
Education	-0.0757	0.0373		
	(0.3816)	(0.3033)		
Height difference	-0.0001	0.0083	0.1488	2.7930
	(0.0090)	(0.0106)	(2.6600)	(2.2407)
Age difference	0.2053	-0.1221	-0.0667	**-0.1878**
	(0.6059)	(0.5748)	(0.0571)	**(0.0364)**
Income	-2628.65	36.7885	-0.0025	
	(35954.27)	(629.96)	(0.0080)	
Wage	-0.1232	0.0836	0.2847	
	(0.2368)	(0.4030)	(0.1802)	
Very beautiful		-0.0134		**-0.3645**
		(0.1166)		**(0.1175)**
Beautiful		0.0671		-0.1266
		(0.2069)		(0.0940)
Skin tone		-0.0684		0.1472
		(0.3362)		(0.0885)

Standard deviation of the distribution in parameters in parentheses. Bold entries mark significance at 5% or more.

Figure 2-1: **Indifference curve of males**

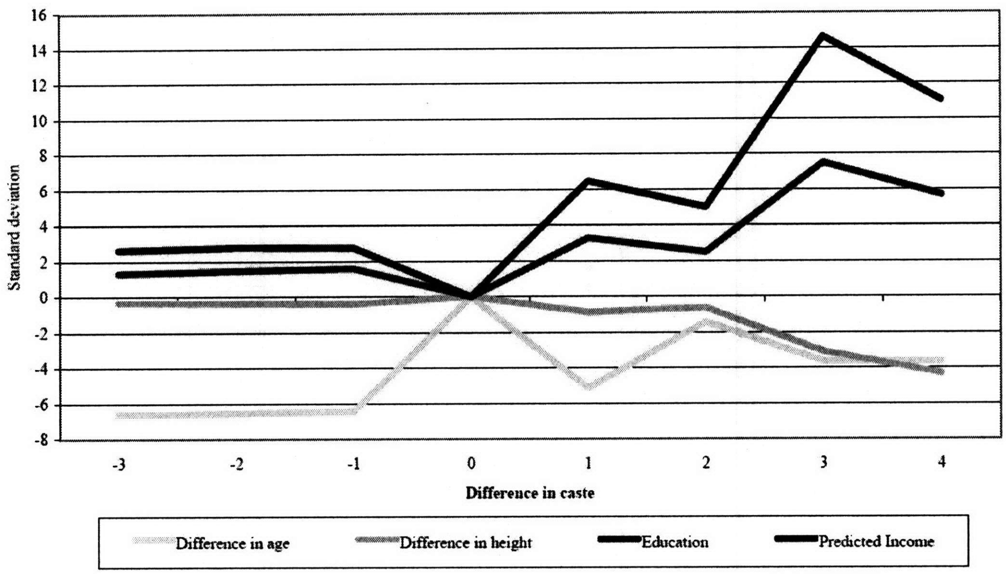


Figure 2-2: **Indifference curve of females**

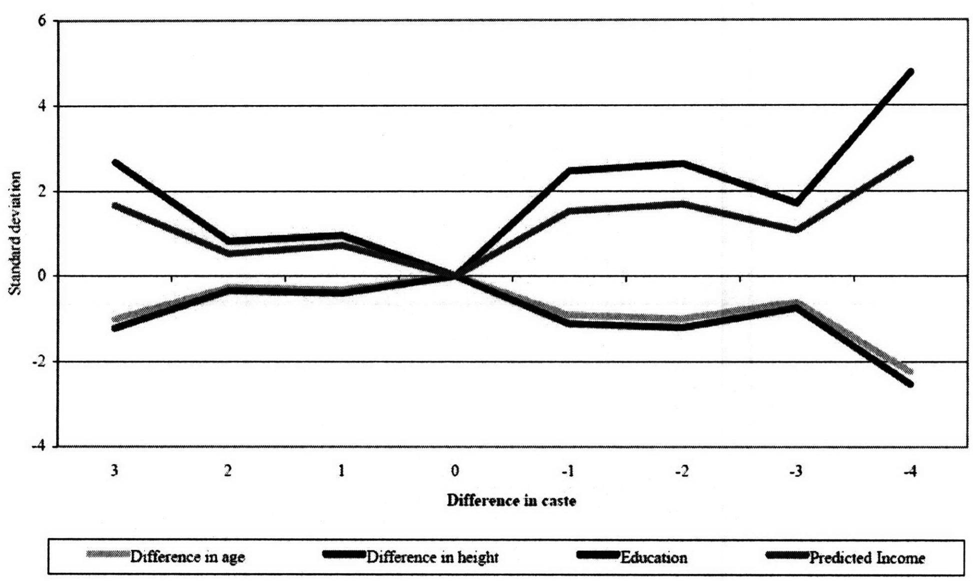


Figure 2-3: **Correlations between coefficients of the considered and rank regressions, ads placed by females**

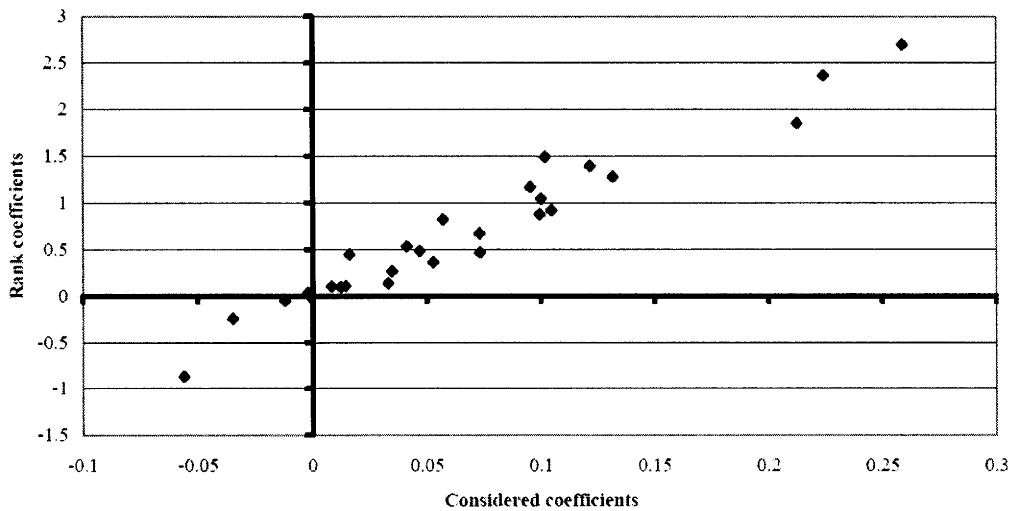


Figure 2-4: **Correlations between coefficients of the considered and rank regressions, ads placed by males**

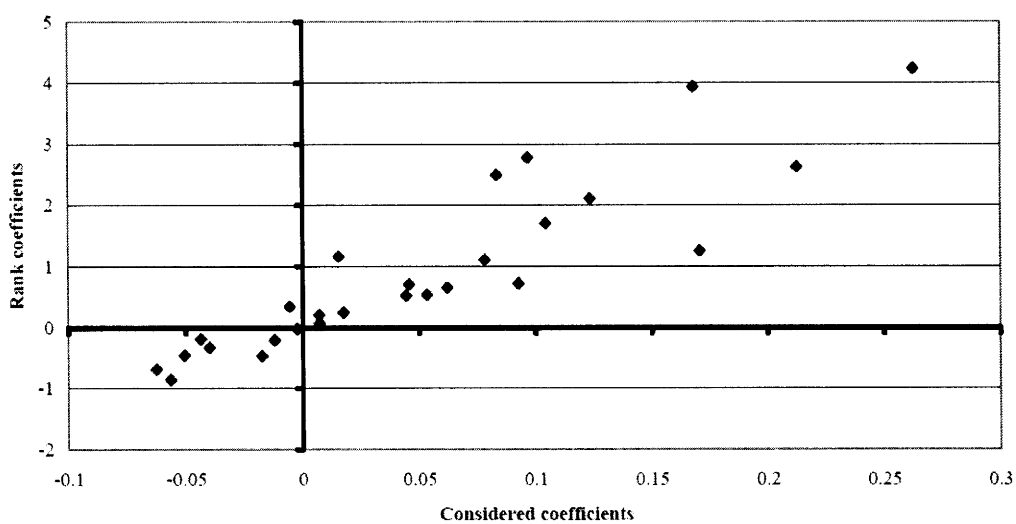


Figure 2-5: **Proportion of considered letters by quality of the letter and ad placer, ads placed by females**

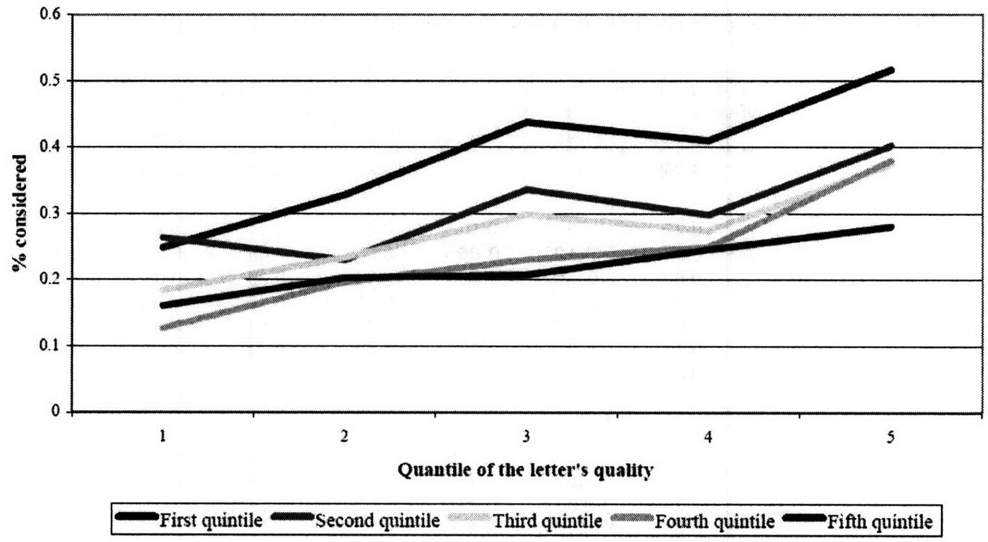
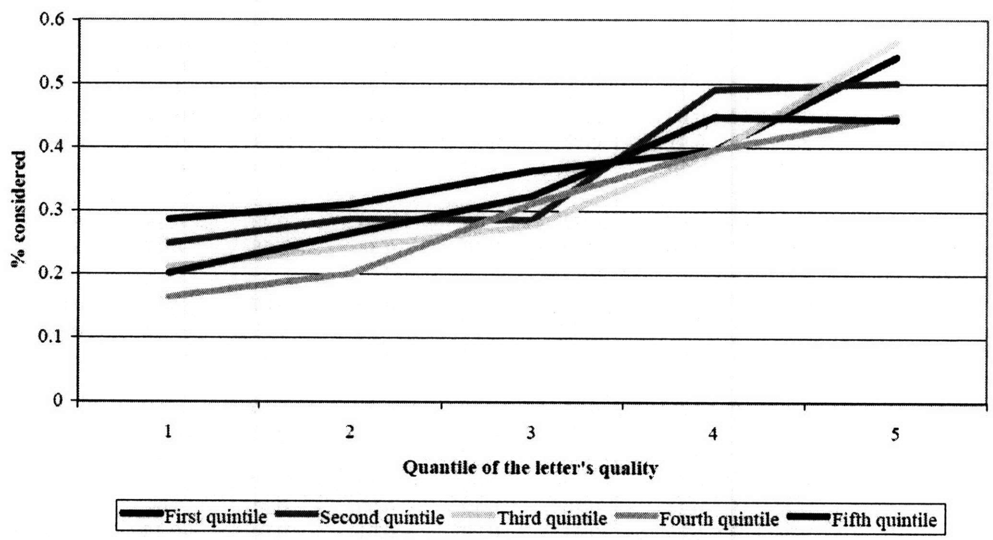


Figure 2-6: **Proportion of considered letters by quality of the letter and ad placer, ads placed by males**



2.A Appendix tables

Table 2.A.1: **Characteristics of ads by attrition status in second round 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 2.A.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 2.A.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 2.A.4: **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 2.A.5: **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 2.A.6: **Couples characteristics, variances of the algorithm**

	Women propose		**Balanced sex ratio**	
	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)
Age difference	**5.4765**	**6.4272**	4.5947	5.3435
Age correlations	**0.8079**	**0.9376**	**0.7370**	**0.8997**
Height difference	**0.1049**	**0.1222**	**0.1128**	**0.1297**
Height correlations	0.7752	0.8955	0.7536	0.8742
Same caste	0.8439	0.9556	0.8598	0.9631
Caste difference	*0.1111*	*0.6316*	**-0.0743**	**0.1620**
Caste correlation	**0.5680**	**0.9296**	**0.5714**	**0.9756**
Same education level	**0.2090**	**0.8019**	**0.3248**	**0.7812**
Education difference	-0.5250	-0.0098	**-0.0656**	**0.4133**
Education correlations	**0.2591**	**0.6586**	**0.3659**	**0.7289**
Same family origin	0.9893	1.0000	0.9579	1.0000
Family origin difference	**-0.0067**	**0.0064**	**-0.0064**	**0.0347**
Family origin correlations	0.9766	1.0000	0.9079	1.0000
Same residence	**0.0000**	**1.0000**	**0.0000**	**1.0000**
Location correlations	**-0.7986**	**1.0000**	**-0.8419**	**1.0000**
Log wage difference	*-0.3380*	*0.0815*	-0.4980	-0.0539
Log wage correlations	**-0.2233**	**0.3461**	**-0.1700**	**0.3497**
Income difference	**-491999.30**	**40416.89**	**-0.02**	**14500.29**
Income correlations	**-1.0000**	**1.0000**	**-1.0000**	**1.0000**
Quality difference	0.1566	0.1758	0.1662	0.1887
Quality correlation	**0.0785**	**0.4057**	*0.2705*	*0.5355*

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.

Chapter 3

Why Wait? Male and Female Marriage Timing and Fertility

3.1 Introduction

In 1965, the United Nations Office of the High Commissioner for Human Rights recommended that no member state allow marriages below the age of fifteen. This was due to the perception that youthful marriages lead to negative consequences. For example, teenagers who get married have lower socio-economic status than those who marry at a later date. An important part of the policy debate surrounding this issue, both in the United States and in developing countries, is the link between these early marriages and fertility. It is often assumed that young women who marry will have more children than those who delay marriage to a later date. However, this may not necessarily be the case: early marriages could simply lead to having children at an earlier age. If early marriages have less stability, they could eventually lead to lower fertility through divorce. Are we then to assume that early marriages increase the number of children born?

This question is not easy to answer because of the obvious endogeneity of the marriage timing decision. Marrying earlier or later may very well be influenced by unobservable factors which may in turn determine one's fertility decisions. This paper attempts to solve this endogeneity problem by using changes in United States laws restricting young adults' capacity to marry.

These varied by state over the course of the twentieth century. The changes in these restrictions by state over time offer a source of variation that is plausibly unrelated to individual fertility decisions. Furthermore, these laws varied by gender as well: males and females generally faced different minimum ages and rules pertaining to females were often modified at different moments than those for males. This setting thus offers an opportunity to differentiate the impact of early marriages on males from that on females.

This paper first presents a simple model which highlights why legislations that affect only one gender may also affect a jointly decided outcome, such as fertility, through its effect on the other spouse. It is a simple model where there's a trade-off in the amount of information one learns about a potential partner by delaying marriage and the loss of marital output this delay entails. When a legislation imposes a delay on one side of the market, that gender delays marriage which affects the eventual outcome. However, it may also very well alter the age at which one's spouse enters the union. It could even modify the matching patterns through the fact that this imposed delay increases the amount of information about the quality of potential spouses. This model thus frames the empirical strategy that follows.

Secondly, this study presents a comprehensive set of state legislations on ages of consent to marriage. Information on state rules from 1890 to 1980 were collected without interruption for almost all states. This is a much larger time period than the laws documented by Blank et al. (2007) who restrict their attention to the period from 1950 to 1980. It is also more inclusive than the set presented by Dahl (2005) which restricted itself to only one type of such laws in a restricted number of states and only for females.

These legal limitations are complex but can be classified into three distinct categories. First, most states set an age at which individuals can marry without the authorization of their parents. I will refer to these as "no-consent" ages. Second, most states also set a minimum age under which a teenager cannot get married, even with the consent of the parents. These ages will be called "consent" ages. Finally, many states offer exceptions to these rules if authorized by a judge. Only a few other states have a third set of laws restricting the power of judges to grant marriage licenses to individuals below a particular age. This study collected detailed information on the first two sets of laws only.

These rules suggest that there was wide variation in the limits imposed by each state. Some states continued to use for a long period into the twentieth century the common-law ages of 12 and 14 in place since colonial times. Others increased these ages before the beginning of the twentieth century. Interestingly, there are both increases and decreases in the ages of consent to marriage. While Blank et al. (2007) emphasize that from 1950 to 1980, the no-consent ages converged to a similar level among all states, there are other periods where divergence occurred. More importantly for the empirical strategy used, these laws varied by gender both in level and in the timing of the legislative changes.

Using this data set of laws, I measure their influence on ages at first marriage in a difference-in-difference framework. These laws are highly correlated with the age at which individuals enter into their first union as measured retrospectively by the United States Census. This study uses this data source rather than the Vital Statistics because the latter has been shown to be potentially biased by avoidance behavior (see Blank et al. 2007). An increase in one year in the consent age limit leads to a fairly small increase in the age at which individuals marry (about 0.02 to 0.04) and is particularly weaker for males than for females. No-consent age limits appear to have a stronger influence, particularly for males. Increasing that age limit leads to an increase in the marriage age of about 0.05. The effects are not perfectly linear, however. Increasing the minimum age from 14 to 18 has larger effects than raising it from 18 to 21, for example. The first stage is mostly robust to the introduction of state-specific linear trends and to the addition of other state-specific controls. It does not appear to depend on the inclusion of a particular state. It is stronger over the period of tighter legislations and lower average ages. It also appears more binding for Whites than for Blacks.

What is particularly interesting is that these laws also appear to have altered the age at which one's spouse enters the union. The effects are, of course, much weaker than for one's own laws. Increasing the minimum legal age for girls leads to an increase in the age at first marriage of boys of about 0.04 for consent laws but none for non-consent. Legal restrictions on male ages translate into delayed marriages of girls of about 0.02 years for an increase of one year in the no-consent age. These results are fairly robust to specification changes and to the inclusions of additional controls but fairly dependent on the time period used.

In addition to affecting marital ages, these laws also appear to have had a significant effect on fertility decisions. Increasing the minimum age for marriage decreases the number of children born to females by about 0.05 children for a change in the female consent laws and of about 0.05 for male non-consent laws. The effect size is slightly lower when looking at the children currently living in the household. Raising the minimum age with parental consent for females leads to observing about 0.01 fewer chidren, about 0.02 fewer children for an increase in the male no-consent laws. The effects are larger for males but less precisely estimated. The effect of these laws on fertility timing are not very precisely estimated, but suggest that the increased fertility is driven by the fact that earlier marriages continue to have children for a longer period. The reduced form effect of the laws is robust to the inclusion of state-specific trends and of compulsory schooling laws. It also does not depend on the periods observed. Also, it is only significant for individuals for whom the law is likely to bind. A change in the laws does not appear to affect the behavior of women with some college education and those who marry past age 25. It also only affects the behavior of White women: over this period, their age at first marriage was in general lower than that of minority women. It could also be that non-marital fertility was not frowned upon in the African-American community as much as it was among Whites.

These laws are then used as instruments for both female and male ages at first marriage. For each individual, two sets of laws are assumed to affect age at first marriage. The laws that limited one's own behavior are combined with those that would have been likely to affect one's spouse, given that there is, on average, a three year age difference between spouses over this period. With this strategy, one finds that only the age at which a female enters a relationship significantly influences fertility decisions. A delay of one year in the female age at first marriage translates into 0.35 fewer children ever born and a similar fewer number of children in the household. The age at which a male marries has a smaller and insignificant effect. This appears to be driven by the fact that women who marry earlier continue to have children for a longer period. Men who marry later, on the other hand, delay both the birth of the first and the last child. Finally, delaying marriage also leads to increased labor force participation and higher wage income, particularly for females. These results are robust to a variety of specification changes.

Interestingly, if one were to assume that these laws only affect the behavior of the gender limited by the legal change and use these laws as an instrument for own age at first marriage, the magnitude of the estimated effect would be much larger. This would suggest that delaying one's first union by one year translates into a decreased fertility of about 0.7 children, a much larger estimate than what was previously found in the literature and also significantly larger than the OLS estimate of 0.1.

Few works have so far documented a causal relationship between marital timing and fertility. The only available evidence is provided by Field (2004) as well as Field and Ambrus (2006). Those papers use variation in the age at menarche in Bangladesh as an instrument for age at first marriage of women. They find that a delay in the age of marriage for females decreases fertility, increases their schooling and literacy level as well as the quality of their marital life. Their estimate of the effect of marriage delay on fertility is slightly smaller but similar to the one estimated here (about 0.27).

The type of variation here employed was first explored by Dahl (2005). He uses changes over a shorter period (1935-1969) and only looks at the effect of consent ages for females. His results imply that postponing marriage after adolescence leads females to be 28 percent less likely to be poor as adults. The channel through which this may be occurring, however, is yet to be explored. If delaying marriage leads to decreased fertility and decreased fertility raises one's standard of living, then our results may be complementary.

Theoretically, little work has been devoted to the relationship between marriage and fertility timing. Most existing models that attempt to explain the current trends in delay in marriage and fertility of women model both decisions as simultaneous (see for example Mullin and Wang 2002, Olivettti 2006, Gustafsson 2001 and Conesa 2000). A few models allow for a difference in the age at first marriage and that at first birth (for example Hotz and Miller 1988 and more recently Caucutt et al. 2002) but their focus is on explaining the link between these timing decisions and human capital/labor market outcomes rather than the link between marriage and fertility. Models of marriage timing, on the other hand, usually ignore fertility as a potential outcome (Bergstrom and Bagnoli 1993 and Oppenheimer 1988).

On the other hand, many descriptive and correlation studies have highlighted how delayed marriages translate into better socio-economic outcomes. Studies have documented the recent rise in marriage ages and its potential effect on labor supply, fertility and matching patterns (Goldstein and Kenney 2001, Qian 1998 and Oppenheimer et al. 1997). In the United States, the negative consequences of early marriages on fertility outcomes have been well documented (Kalmuss and Namerow 1994, Martin 2004 and Kiernan 1986). For example, married teen mothers are 40 percent more likely to have a second birth within 24 months of their first birth compared to unmarried teen mothers. Women who marry before the age of 19 are 50 percent more likely to drop out of high school and four times less likely to obtain a college degree (Klepinger et al. 1999). In developing countries, youth marriages are considered particularly problematic for girls. Jensen and Thornton (2003) provide an overview of the various problems that are associated with younger marriages: lower schooling, less reproductive control, higher rates of pregnancy-related mortality and domestic violence.

This study contributes to the existing literature by first exploring the role these laws limiting the marriage capacities of teens may have on fertility. This can be particularly important for developing countries where such policies could be implemented to curtail youthful marriages. The results suggest that modifying the absolute minimum age for girls (with parental consent) is particularly important while it appears the no-consent laws may be more likely to affect the behavior of boys. It also highlights the potential effect of these laws on fertility.

Second, this paper provides some evidence that such an intervention may not only affect the behavior of girls themselves but also who they marry. This is an important conclusion for the study of other policies that may alter the entire marriage market equilibrium through imposing restrictions and changes to only one gender. This could even become important in randomized trials that attempt at estimating the causal effect of delaying marriage age of girls.

The rest of the paper is organized as follows. Section 3.2 first elaborates a model which describes why laws that restrict one gender may also modify the behavior of the other. This motivates the empirical framework presented in Section 3.3. Section 3.4 describes the data employed in this analysis while the following section details the evolution of the laws restricting the marriage of teenagers. The next section presents the results while Section 3.7 concludes and discusses policy implications.

3.2 Model

To better understand the effect of age of consent to marriage legislation, a simple illustrative model is here presented. It highlights how a change in the minimum legal age for marriage may have other effects than simply modifying the age at which individuals marry.

3.2.1 Basic set-up

Assume that we have a set of males and females (of equal measures) in an economy. Each agent has a given quality level q_i which is distributed uniformly for both men and women over the unit space.

Agents live for 2 periods and can marry in either period. Assume for simplicity there is no divorce so that if one marries in the first period, one remains married in the second. Finally, while one knows one's own quality at birth, it is unobservable to others in period 1 but fully observable in period 2.

Individuals who are single earn nothing. One married, individuals receive a pay-off that is a function of their own quality (q_i), and of that of their spouse (q_j). Assume that the pay-off is given by

$$f(q_i, q_j) = q_i q_j$$

which is super-modular such that male and female quality are complements. The discount factor is given by 1.

Proposition 10 *Without government intervention, there exists an infinite number of symmetric equilibria where all individuals of quality less than a given level will marry in the first period and the others will delay until the second.*

Proof. At the beginning of period 2, all remaining single individuals have their quality revealed. Given that the pay-off function is supermodular and that the distribution of quality is identical across males and females, perfect positive assortative matching will be observed. This implies that a male of quality q_i should match with a female of the same quality level. Thus, an individual anticipate that if he postpones marriage until period 2, his pay-off in that period will be given by

$$f(q_i, q_i) = q_i^2.$$

Given that knowledge, who will decide to marry in period 1? The pay-off of delaying marriage for an individual of quality q_i is given by

$$f(q_i, q_i).$$

The pay-off given to an individual who marries in period 1 is

$$2E(f(q_i, q_j))$$

where the expected value is taken with respect to q_j.

Assume an individual believes that all individuals of the other gender with quality less than $\bar{q}$ will marry in the first period. This individual will decide to marry if

$$\begin{aligned} 2q_i E(q_j) &> q_i^2 \\ q_i \bar{q} &> q_i^2 \\ \bar{q} &> q_i \end{aligned}$$

Thus, if one believes that all agents of the other gender with quality less than $\bar{q}$ will marry in the first period, they will marry in the first period as long as their own quality is less than $\bar{q}$. There will thus be an infinite number of equilibria where all individuals of quality less than a given level will marry in the first period and all others will delay until the second. ■

3.2.2 Effect of minimum marriage age legislation

Assume there is an equilibrium where $\bar{q}$ is positive and the government imposes that females cannot marry in the first period.

All women are now forced to delay marriage but males may still prefer to marry in the first period with females who are now older than themselves. Would women in the second period be willing to marry younger men? They must compare the value of marrying with an individual for which they know the quality for sure with one where it will be a gamble. The gamble will only be attractive to women who would marry a very low quality male and are thus very low quality themselves. Now, this low quality pool will not be attractive. Thus, males will delay marriage as well.

In addition to imposing a delay on males' age at marriage, this policy would also affect the type of partners that women who used to marry in the first period before the policy change would marry. In this case, while couples would have formed randomly when marrying in the first period, they would match assortatively once the policy is enacted.

This simple model is meant as an illustration of the various channels through which a law which imposes a legal restriction on the age of marriage for one side of the market may affect the overall matching equilibrium. First, such a law modifies the age at which both males and females marry. It could even alter the type of spouses one marries in a context with imperfect information. Evidence of this is explored in the following sections.

3.3 Empirical strategy

Having established the model framing the analysis, this details the empirical strategy employed to capture the causal effect of marriage delay on fertility outcomes by using the legal environment restricting youthful marriages as an exogenous source of variation.

3.3.1 General framework

The main equation of this paper is one where a fertility outcome (y_{ist}) of an individual i born in state s, at time t, is related to the age at first marriage of the individual (x_{ist}) and that of his or her spouse ($x_{js't'}$). That spouse may have been born in a different state (s') and in a different year (t'). The equation will then be

$$y_{ist} = \alpha * x_{ist} + \beta * x_{js't'} + \psi * X_{ist} + \delta_s + \phi_t + \varepsilon_{ist} \tag{3.1}$$

where ψ_s and δ_t represent state of birth and year of birth fixed effects, X_{ist} represent individual controls which in the main specification will include Census year and race fixed effects and ε_{ist} captures idiosyncratic variation by individual.[1] Standard errors are clustered at the level of the state.

3.3.2 Instrumentation

Obtaining consistent estimates of α and β in equation (3.1) is unlikely through OLS. That is because it is likely that unobserved factors influencing the age at which one marry or the age at which one spouse's marry may also alter fertility decisions, thus leading to a correlation between x_{ist} or $x_{js't'}$and ε_{ist}.

Thus, one must pursue an instrumental variable strategy. The laws limiting one's behavior offer a potential solution to this problem. First, x_{ist} can be instrumented using the minimum legal age at which an individual born in state s in year t may marry (both with and without the consent of his parents). This is a valid instrument in this difference-in-difference framework if trends over time in ages at first marriage would have been parallel, had there not been differences in minimum legal ages.

For each year of birth, state of birth and gender, the minimum age at which an individual could legally marry was computed by comparing the consent and no-consent ages at each year between the ages of 12 and 21. The smallest age at which one was allowed to marry based on this computation was recorded. For example, for an individual who was 19 when the no-consent age was raised from 18 to 21, the minimum no-consent age would be 18. However, if an individual was 19 when the no-consent age was reduced from 21 to 18, then the minimum age at which he could marry would be 19 because he was already passed the new minimum age when that law was modified. These laws are detailed in the next section.

Year of birth was computed as the difference between the Census year and one's age at the time of the survey. Some measurement errors may have been introduced by this computation but this should be minimal. In the same way, the laws were not coded by month but only by year in which they became a law. It is thus possible that an individual would have turned one

[1] This is equivalent to adding a linear function of age. An alternative using dummies for each age was explored with very similar results.

year older after the actual statute became law.

State of birth is, in this analysis, used as a proxy for the state in which one would have been residing in their teenage years. Pre-teen migration would simply introduce noise in the estimation.

Instrumenting for $x_{js't'}$ is slightly more challenging. The laws in place for a person from state s' born in year t' are only valid instruments if one includes fixed effects for s' and t'. Not only would this increase significantly the number of fixed effects, but it would also imply that both the year of birth of the respondent and that of his spouse would be included as controls in the regression. For individuals who married only once and report accurately their age at first marriage, x_{ist} and $x_{js't'}$ are then colinear since i's year of birth plus x_{ist} correspond to the year the first union was celebrated and thus should be equal to j's year of birth plus $x_{js't'}$. Furthermore, the state of birth and age of a spouse may very well be endogenous to the timing decision if, for example, delaying leads to marry spouses with a smaller age difference.

Instead, this paper uses the fact that the average age difference over this time period between a man and a woman is 3 years. Thus, for each individual i, born in state s in year t, one can construct a "potential spouse" $\tilde{j}$ who is born also in state s but in year $\tilde{t} = t + 3$ if i is a male and in year $\tilde{t} = t - 3$ if i is a female.

Thus, the first stage of this instrumentation strategy will thus be

$$x_{ist} = \gamma^1 * NCA_{st} + \rho^1 * CA_{st} + \tau^1 * NCA_{s\tilde{t}} + \theta^1 * CA_{s\tilde{t}} + \sigma^1 * X_{ist} + \eta^1_s + \kappa^1_t + \mu_{ist} \quad (3.2)$$
$$x_{js't'} = \gamma^2 * NCA_{st} + \rho^2 * CA_{st} + \tau^2 * NCA_{s\tilde{t}} + \theta^2 * CA_{s\tilde{t}} + \sigma^2 * X_{ist} + \eta^2_s + \kappa^2_t + \nu_{ist} \quad (3.3)$$

where the η_s, κ_t are fixed effects for state and year of birth, X_{ist} is defined as above and NCA_{kl} and CA_{kl} respectively represent the minimum no-consent and consent ages to marry for an individual born in state k in year l. Again, standard errors will be clustered at the state level.

These regressions should be run on individual outcomes to obtain the appropriate estimate of the variance-covariance matrix. Nevertheless, the sample size was too large for this to be possible in a timely manner. Each estimation was thus performed at the level of a cell which were divided by age, year of Census, year of birth, state of birth, gender and race. The sum of sample-line weights were used to weight each cell of the regression.

3.3.3 Selection issues

Equation (3.1) must be estimated on currently married individuals who both reported their age at first marriage. If legal restrictions affected either the probability of being married or the probability of remaining in a relationship, this strategy will be invalid.

This paper explores this particular issue through a variety of methods. First, both the first stage and the reduced form will be presented for individuals currently married and for the entire sample. Secondly, the probability of being currently married was regressed on the instruments and no effect was found, thus suggesting little potential effect of selection on the results.

3.4 Data

The previous section detailed the empirical strategy used in this paper in order to relate the ages at first marriage of males and females to fertility-related outcomes. The present section explains how each of these variables was constructed.

3.4.1 Age at first marriage

This paper measures the age at which females and males first entered into a marital relationship by using the Integrated Public-Use Microdata Series (IPUMS) from 1960, 1970 and 1980 Census files. These samples are the only ones in which both spouses provided their age at first marriage. The samples are fairly consistent. All ever-married individuals were surveyed. In 1960 and 1970, all individuals above the age of 14 were questioned, in 1980, all those older than 15. Only respondents older than 21 will be included in the analysis so these changes in the sample will be irrelevant for the current analysis.

In all these Census extracts, the question asked for the month and the year of the first marriage and the age is derived from using that date and the date of birth. Only the age is available to IPUMS users.

The information collected through this method is self-reported and may thus be subject to the usual caveats concerning misreporting and measurement errors potentially important to retrospective data. One would expect that this type of error would be more likely among younger marriages since those might have been of shorter duration or simply might have occurred further from the moment of the interview. I find little evidence that youthful marriages are under-reported. If anything, as individuals age, they are more likely to report marrying at younger ages (after controlling for year of birth fixed effects) although the magnitude of the effect is fairly small. Blank et al. (2007) discuss the issue of reliability of self-reported information concerning the age at first marriage and conclude that the source of data employed here appears more accurate than that obtained from the Vital Statistics as the latter may have been altered to falsely comply with the marriage license requirements imposed by each state.

Finally, one should note that this variable has been bottom and top-coded at different values over the various Census extracts. While it is mostly irrelevant for this analysis how marriages occurring after one's ninetieth birthday are coded, censoring of youth marriages may be more of a concern. In 1980, all marriages entered at an age younger than 12 were coded at 12 years old. In 1960 and 1970, it was all unions performed before one's 14th birthday that were coded at 14. It was not found that this difference was very important throughout the analysis but it should be explored further.

Figures 3-1 and 3-2 show the evolution of the median and mean age at first marriage. This figure suggests that contrary to popular beliefs, individuals born in the nineteenth century were not likely to marry at very young ages. Actually, the median and the mean age at first marriage fell dramatically for both genders over this period. If one accounts for the difference in the age at the time of the interview, a fairly similar picture arises. The main difference is that rather than observing a plateau at the end of the period, one would observe a slight increase in both the mean and the median age for both genders.

3.4.2 Outcome measures

Measures of fertility were also obtained from the same Census extracts. First, the number of children born to a woman is provided in all three samples. In 1960, only ever-married females provided a figure, while in 1970 and 1980, it was asked of all females. This will be the main measure of fertility employed in this study. For males and females, however, one can also measure the number of own children living in the same household. This information is available for every single year and individual. To measure fertility timing, it would have been ideal to measure the age at which the first child was born. Unfortunately, that information is not collected by the Census. Nevertheless, this study will use the age of the eldest and youngest own child in the household as proxies for fertility timing. This will be a good approximation as long as many of these children are still living with their parents at the time of the survey. Given that the average age in this sample is well into their 40s, absence of children from the household will potentially be a problem.

Finally, to relate these results to those presented elsewhere, educational attainment (highest grade obtained), employment status and wage income (measured in 1980 constant dollars) will also be employed.

Table 3.1 presents the summary measures of each of these variables for men and women over this period for two distinct samples. One sample includes all men and women; the other, only currently married individuals for which both female and male ages at first marriage were collected. In both cases, the sample is restricted to individuals aged 21 and above to ensure that no individual is still potentially limited by the law. The table highlights that individuals in our sample are mostly White (even more so among currently married) with a large majority of men (0.85) and of women (0.89) having entered at least once into a marriage contract. Men, on average, over this period, married at age 24 and women, at age 21. Among those currently married, the age difference is on average 3 years. Females had on average about 2.4 children by the moment of the interview and about 2.6 children once the sample is restricted to currently married women. The average number of children in the household for both males and females is much smaller, however. Only about one child in the household was found on average: this figure rises to 1.5 among currently married couples. The oldest child in the household (among those for whom a child was found) was about 13.5 years old for couples currently married but about 16 for all males and almost 20 for all females. Similarly, the youngest child in the household was about 10 years old among currently married couples but 13 for all males and almost 17 for all females. The difference by gender is probably due to the fact that more widowed mothers were living with their adult child than widowers. The sample has completed, on average, a few years of high school with the mode of the distribution having obtained a high school degree.

Men and women are similarly educated, in particular once the sample is restricted to currently married individuals. However, men are much more likely to work, in particular among married men, while the opposite is true for females. This translates into large differences between male and female wage income, a difference which is exacerbated among married individuals.

3.5 Legal restrictions on youthful marriages

The instrumentation strategy described above requires that the minimum age at which teenagers and young adults are allowed to marry varied by gender over time by state. When the thirteen colonies became independent in 1776, they inherited the common law regulations of the British Empire on the age at first marriage. Boys and girls were respectively allowed to marry without parental consent at the ages of 14 and 12. Males were given a higher minimum age based on differential ages at puberty. These laws were kept in place in the newly formed United States much longer than they did in Britain. Grossberg (1985) argues low minimum ages continued to prevail because there were no dowries exchanged in the new Continent. Because of this, there was no need to raise the age of marriage without parental consent to protect a father's investment in his daughter from being claimed by an unsanctioned suitor. Shammas (2002) argues that the lack of official registry of marriages (linked to the absence of a state church) made it impossible for states to enforce and monitor restrictions on marriage and thus did not attempt to do so. Whatever the reason might have been, most states continued the tradition of allowing children to marry at very young ages. Many simply never stipulated a legal minimum age before the twentieth century making the common-law ages of 12 and 14 ipso facto the minimum legal ages.

It is important to note that a marriage that did not respect these legal restrictions was not *void* but simply *voidable.* If a couple married while one of the parties was too young to give his or her consent but continued to cohabit after this age of consent was achieved, the union was considered legal and valid. The effect of these laws was thus first to allow adolescents who had married without legal authorization and regretted their decision to be able to annul the union. Secondly, it would also make it more cumbersome for teenagers to obtain a marriage license. In so far as laws were at least partially enforced, they increased the costs related to youthful marriages, as avoidance mechanism (travel, misrepresentation of age) would imply additional efforts.

3.5.1 Data collection

This paper collected information regarding the historical rules determining legal ages at marriage for all 51 states over the period 1882 to 1980, providing a set of legal rules for the longest period yet collected. This is a much larger data set than Blank et al. (2007) who focused on the changes over the 1950-1970 period. This is also more comprehensive than Dahl (2005) who focused only on consent laws, for about 40 states, over the period 1935-1969.

Data collection involved the careful analysis of various secondary sources. Many compendium of marriage laws were published over this period and contained information regarding the laws (including the age of consent to marriage) pertaining to marriage in each state. This constituted the primary source of information over time. This was then supplemented by various sources such as treatises on the status of women, marriage manuals for ministers, etc. Many of them not only reported the current legal age but also the date at which changes took place. If two sources less than five years apart had the same age and quoted the same law, it was assumed that there had been no change in the law over that period. If a change occurred between two data sources and only one legislative change was recorded over the period, it was assumed that the minimum age changed at that point. For any point where uncertainty remained, searches of the legal statutes in Westlaw and LexisNexis were then employed to resolve them. However, as compared to Blank et al. (2007), I could not only resort to legislative search given the long time period studied.

For most states, this generated a set of legal rules for the entire period from 1885 (or their admission to the Union) to 1980. For a few states and time periods, I could not resolve some uncertainties and thus kept those years as missing. All sources employed in this exercises are listed in Appendix 3.A.

These documents were preferred as a source to the *World Almanac* as used by Dahl (2005) for a few reasons. First, that source only started to publish information regarding age of consent to marriage in 1935 and even within that period, many states were not included. Second, several inconsistencies were found between that source and other sources described above and verified through searches in the statutes through Westlaw or LexisNexis. Finally, the *World Almanac* focuses mostly on consent laws and provided little information on no-consent laws.

3.5.2 Legal history

Table 3.2 provides a summary of the rules in place over this period by state. As discussed before, this paper focuses on two different types of laws. The first columns refer to the minimum age at which individuals are allowed to enter into a marriage contract without the authorization of their parents. States differed in the way they set these limits. Sometimes, the law simply prevented individuals younger than the age of majority to marry without parental consent. In that case, changes in state-level age of majority immediately affected no-consent ages. Other statutes specified ages specifically for consenting to marriage. Over time, these laws have not changed dramatically for females over this period. Twenty-one states did not change the no-consent ages over the entire time period. Most states used a minimum age for females of 18 years old but some have also employed 21. A few states had lower ages in the beginning of the twentieth century but those rapidly disappeared. For males, no-consent ages evolved dramatically over time. A majority of states had established a minimum age for males of 21. This was gradually lowered to 18 over the second half of the twentieth century. On the other hand, a few states also had much lower ages in the beginning of the period and increased them into the beginning of the twentieth century. Overall, these laws increased their bindingness over the first-half of the century and then became more and more lax over the subsequent period. While the increase is comparable between males and females, the decrease after 1940 is particularly marked for males. These changes altered the variance in laws faced by teenagers over this period. As shown by the last row of Table 3.2, age of consent to marriage without parental consent experienced a substantial homogenization, as emphasized by Blank et al. (2007), over this period. However, as each state moved towards the new standard of 18, there was a marked increased in variance between 1960 and 1970 (not shown).

The last columns show the consent ages, that is the minimum ages at which a teenager can get married with the consent on his or her parents. Parental "consent" could be demonstrated through various ways, sometimes in person, other times in writing as well. A few states never allowed parents to authorize unions of minors without a court appearance. For those, the minimum age at which a child can marry if he or she secures parental approval *and* court sanction is then recorded. Many other states also allowed marriages below the consent ages if a judge authorized it.[2] However, most of these rules are left to the discretion of a judge without relevance to age and so are not reported here. As shown in the last columns Table 3.2, these laws varied considerably over the period of study, in particular for females. The patterns of evolution of the consent laws are fairly different than those of the no-consent laws. First, it is obvious that many states modified one type of law but kept the other unaltered. Second, as in the case of no-consent laws, the legislation increased the minimum age with parental consent for males over the first half of the period and then decreased it over the remaining 40 years. However, contrary to the evolution of the consent laws, this following decrease was minimal and the average legal age in 1980 remained well above that in 1900. On the other hand, these minimum ages were consistently raised for females over the same period, with a more substantial increase in the beginning of the twentieth century than in the second half of it. The variance in the laws across states also changed over this period. While one observes more and more homogeneity in the laws between 1900 and 1940, after 1940, the variance rises for females. For males, while homogenization continues until 1980, it is much less drastic after 1940 than before.

In addition to an increased homogenization of the laws across states, female and male ages differed in level throughout most of the period. The Supreme Court ruling on Stanton and Stanton (1975) ruled that child support payments should expire at the same age for boys and girls and thus imposed some degree of gender equality in State legislations. Many authors argue that this was an important reason behind the homogenization between female and male no-consent laws which occurred at the end of the period. Arkansas is the only state with a difference in their no-consent laws by 1980. Nevertheless, there was much less homogenization across males and females for the consent ages than for the no-consent rules. In 1980, fifteen states still had different minimum age of consent for marriage (with parental consent) between

[2]One common exception is that many states allow a pregnant woman to be married even if younger than the legal age, sometimes only if she is marrying the father of the child, sometimes irrespective of the identity of the groom.

girls and boys. While this is substantial, it is still much less than in 1900 or even in 1940 where more than forty states had differences between genders.

These trends in age at first marriage presented above are thus in sharp contrast with the trends in the consent and no-consent ages presented here. This suggests that the laws had the most "bite" on the behavior of individuals born between 1900 and 1940 because those were the cohorts for which high ages were still in place and for which lower ages at first marriage were experienced.

A first look at the distribution of ages at first marriage across states over time should provide some evidence of the bindingness of the laws restricting teenagers marriages. Figures 3-3, 3-4, 3-5 and 3-3 explore how binding each set of laws were. They graph the distribution of age at first marriage, by gender, by minimum age imposed. Each graph may include a variety of states and time periods over which the same combination of consent and no-consent laws were in place. Graphs made for the evolution of the distribution of age by state as a function of changes in minimum ages over time produced very similar conclusions.

In Figure 3, states are classified into 3 panels by the minimum age at which one can marry with parental consent. It then shows in each panel, how the distribution of ages change as the minimum age without parental consent evolves. One finds in panel A and B that states which imposed higher no-consent laws had distributions of ages that were shifted to the right. It also highlights that these laws tended to be more binding when placed at or above 18. Panel C suggests that in states where marriages were not allowed until age 18, even with parental consent, there is little difference in cases where the no-consent law was placed at 18 or at 21. Figure 4 plots the same data for females. While it appears that the no-consent laws affected the distribution of ages in this case also, the effect tends to be much more muted than for boys. The distribution of ages when the no-consent minimum age is 21 rarely differs from that when it is 18.

Figures 5 and 6 explore the relationship between the laws setting consent ages and the distribution of marital ages. For males, the effect of consent laws is much less important than that of no-consent minimum ages. In particular, there is little difference in cases where the minimum age is 14 and that when it is 16. However, overall, state-time cells that have larger minimum age appears to have experienced higher ages at first marriage. Figure 6 highlights that these consent laws did bind the behavior of girls as larger minimum ages were accompanied by a shift of the distribution to the right.

All these graphs suggest some degree of bindingness in the laws as the distribution appears to be censored by the legal environment. Nevertheless, these graphs also imply some degree of evasion and lax enforcement as the number of marriages below the imposed minimum is not zero.

3.6 Results

3.6.1 First stage

The first step in this analysis is to demonstrate that the age of consent laws modified the age at first marriage of individuals in the sample. Both Blank et al. (2007) and Dahl (2005) demonstrated that the laws diminished the propensity of individuals to marry at a young age for the period they studied. This paper focuses on the interaction of own and spouse's age and the role of the legal environment in that determination.

The results of the regression equations (3.2) and (3.3) are presented in Table 3.3 and Table 3.4. Table 3.3 presents the results for males and their spouses while Table 3.4 focuses on females and their spouses. In both tables, the first panel introduces the legal rules by separate dummies for each minimum age while Panel B uses a linear measure of the minimum age. The reference category for females is a minimum age of consent of 12 while that for males is a minimum age of 14. The first columns of both tables correspond to the estimation of equation (3.2). The first column corresponds to the estimates of γ^1 and ρ^1 while the second reports estimates of τ^1 and θ^1. Similarly, Columns (5) and (6) report the estimates of equation (3.3), with Column (5) reporting estimates of τ^2 and θ^2 and the regression estimates of γ^2 and ρ^2 can be found in column (6).

The regressions suggest that the laws did in fact influence the age at which individuals married. Males appear to marry at a later age when their own consent age increases and when the consent minimum age of their potential spouse is being raised as well. On the other hand, the own minimum consent age and the no-consent laws likely to affect one's spouse appear to have little influence on the age at which males marry, if not a negative effect. Once one allows for non-linearity in the functional form, spousal rules for no-consent ages appear to be influential. The results for these male's wives are mirroring the previous results. These women marry later when their predicted consent minimum age rises and when the no-consent minimum age of their spouse goes up. Their own no-consent laws also increases their marriage age, even if not by a very large margin.[3] The F-test on the joint significance of the instruments are fairly large and larger in the case of spouse's rules in both cases. They are also larger for males (for which we have the exact laws) than for their wives (where we approximated the laws by the state of birth of the male and the average age difference).

Columns (3) and (4) present, for comparison, the effect of male laws without the addition of the female rules. Consent laws do not appear to change the age at which males marry but are now at least positively correlated. No-consent laws are more strongly causing delay in marriage of males. Similarly, having only spousal consent or no-consent laws–as shown in Columns (7) and (8)– strengthens the relationship. The effect of consent laws is now strongly significant and the effect of the no-consent laws becomes larger and more precisely measured.

Table 3.4 suggests a similar pattern once we look at females. As before, women are strongly influenced by their own consent minimum age and slightly less so by their own no-consent laws. Only the no-consent rules affecting their potential spouse's influence the timing of their marriage. The regression for the age of first marriage for these females' husbands is weaker. Results suggest that the no-consent law we predict would be, on average, likely to influence them, is the most important determinant of their marriage timing. The female consent and no-consent laws both positively contribute to the decision of delaying. Once more, restricting the regressions to only own laws suggest that for females, both consent and no-consent laws are influencing their decision while for their spouses', only the no-consent minimum age are limiting the behavior. Usually, the inclusion of spousal constraints diminish the magnitude of the estimated impact of own laws on the age at first marriage.

In both tables, the effects appear overall to be non linear. First, some smaller cells appear

[3]Since the effect of the laws are smaller on a spouse's age at first marriage than on one's own age, this suggests that these laws are thus also potentially altering age differences between spouses.

to generate strong coefficients such as in the case of a minimum age of 13 or 15 for females. However, grouping these smaller cells with similar ages or dropping observations with those ages does not affect the results significantly. Secondly, it appears that a one year increase in the minimum age does not always translate into the same impact. The male no-consent minimum age, for example, influences marital ages less when increased from 20 to 21 but much more at lower ages. This then explains why, in Panel B, once we turn to linear estimates, the precision of our estimates falls substantially and are very often only significant at 10 percent. These linear estimates suggest that increasing the minimum age without parental consent by one year leads males and females to delay marriage by 0.45 years. The effects of laws with parental consent are slightly smaller, where an increase in one year would lead to a delay in marriage of 0.2 for males and 0.3 for females.

Table 3.5 and 3.6 explore how robust these results are to changes in the specifications and samples. Only the results for the effect of the laws on own ages are reported but similar results were obtained for spousal age regressions. Columns (1) and (2) present the results with state-linear trends. The next columns present the first stage for the full sample (including not currently married individuals). Finally, columns (5) and (6) introduce as additional covariates compulsory schooling laws from Oreopoulos (2007). This reduces somewhat the sample as the number of years and states for which this variable was available is less than those for which I have legislation info regarding marriage rules.

Comparing the first two columns of Table 3.3 to Table 3.5, one finds a fair level of similarities. Adding state-trends strengthens somewhat the effect of no-consent minimum ages for males although the standard error on the linear coefficient increases. However, adding those controls weaken the effect of spousal consent laws. Males consent minimum age appear now much more effective than before while little change can be found in the no-consent minimum ages for spouses. Including all ever married in the sample changes the estimates only slightly. This suggests little possibility of selection bias. Finally, adding schooling laws renders the no-consent minimum age dummies for males and the consent minimum age dummies for females more significant although less linearly related, which leads the linear coefficients to fall and become less significant. However, this is entirely due to the change in sample rather than the introduction of compulsory schooling laws. Compulsory schooling, in itself, increases marriage age by 0.06 years.

Comparing Table 3.6 and the estimates provided in the first two columns of Table 3.4, one finds very similar results for female age at first marriage. Adding state linear trends weakens the first stage more significantly in this case. The effect of female consent ages and spouse no-consent ages are now much less significant than before. On the other hand, little difference is observed by computing the first stage on the full sample although the effect of spousal rules is somewhat smaller than in the currently married sample. Schooling laws alter much less the coefficients in this case but do somewhat increase the standard errors of the linear model.

While not reported, other tests of the first stage were conducted. Past and future laws were included in addition to contemporary ones and the contemporary ones retained their significance. Also, past and future laws were not significant, giving credentials to the hypothesis of parallel trends. The largest states were dropped one at a time to check that none of them are driving the results and this was confirmed. Each of the Census regions were also dropped and results remained broadly consistent although some regions were more important than others for the estimates. However, the first stage is dependent on the period studied. Excluding the oldest cohorts substantially increases the standard errors on the relationship presented above.

3.6.2 Reduced-form effects

Having demonstrated that the legal restrictions on youthful marriages influence the age at which individuals eventually marry, let me now turn to exploring how these laws affect fertility and other economic outcomes. Tables 3.7 and 3.8 present the reduced form effects of the laws on various fertility-related outcomes. Once more, the odd columns show the coefficients on one's own laws while the even ones display the estimates of the effects of spousal laws. Each pair of columns, however, corresponds to only one regression model. In panel A, each law is dummied out while panel B presents the linear coefficients.

The first columns present the effect of minimum age legislation on the number of children in the household. This will be a good measure of fertility for individuals who still have children living with them. For that variable, the dummied out coefficients are rarely significant but the linear model suggests that an increase in the consent minimum age of females and the no-consent minimum age of males both reduce the number of children in the household. This mirrors the findings of the first stage presented above. Effects are relatively small. An increase of one year in the consent minimum age imposed on a female reduces the number of children in the household by about 0.01 while a similar increase in the no-consent minimum age imposed on males decreases the number of children living in the household by about 0.02. Female no-consent ages and male consent ages do not appear to be very influential and are their effects are very imprecisely measured. Looking at the number of children born, the effects are now more significant and much larger as well but for the same variables as the ones for the number of children in the household. The linear model suggests that an increase of one year in the consent age laws of females and one in the no-consent age of males both lead to an decrease in the number of children ever born of about 0.05. The pattern is fairly linear for the consent minimum age of females but much less for the spousal no-consent rules. Overall, non-linearities highlighted above are again visible in this table. For children born, the fertility effects are particularly marked for increasing the minimum consent age from 17 to 18 compared to an increase from 12 to 17.

Effects of the laws on fertility timing (as captured by the ages of children within the household) are much less clear. The dummied coefficients are all very large and very significant suggesting large differences between the ages of children living with their parents in states and time periods where the minimum age is 12 or 14 and those in periods where the ages were larger. However, none of these coefficients are very well aligned, making the linear estimates very noisily estimated. If any conclusions can be drawn, the laws restricting female behavior appear to lead to an increase in the age of both the eldest and the youngest child present in the household. It indicates that these laws did not delay fertility, as measured by this proxy. It also suggests that the laws led to shorter fertility periods with mothers having fewer young children at home when laws restricted the age at which they could marry. This is also visible in the fact that the reduced form effects on fertility are particularly marked when looking at older females compared to younger ones. This suggests that the measured decrease in fertility is not linked to a mechanical relationship. For example, this is not in accordance with a model where a delay in marriage decreases fertility because females have children as soon as they marry and continue to do so until menopause. This is also in contrast to findings in Field (2004) for Bangladesh where most of the decreased fertility was due to a delay in the timing of the first birth.

Table 3.9 performs a series of validity checks for the reduced form using the number of children born. Although not reported, the same pattern was observed for the other outcomes previously presented. The format of this table is the same as before. Adding state linear trends — as in Columns (1) and (2) — decreases the magnitude of the effect of female consent ages and particularly that of male no-consent ages. On the other hand, estimating the regression over the entire sample (including those never married) reduces somewhat the estimates of the effect by very little. However, if one estimates the reduced form on the sample of never married individuals, the laws are found to have no effect on fertility behavior. Including controls for changes in the compulsory schooling laws reduces the size of the coefficients but does not modify their significance, as presented in Columns (5) and (6). Compulsory schooling laws are found to slightly reduce fertility (as shown by Leon 2004) but the effect in this case is insignificant.

As a falsification test, one would like to show that females who were unlikely to be affected by the change in the consent laws also do not appear to alter their fertility decision in response to changes in these laws. While it would have been ideal to use some family characteristics to attempt to proxy the likelihood of being limited by these laws, those were not available in the Census extracts used for this paper. The last four columns of Table 7 focuses on two groups which should not have been influenced by these laws: women who married past the age of 25 and women who completed more than 12 years of schooling (that is attended college). These do not constitute a perfect placebo test because consent and no-consent age legislation could have altered the schooling decisions of females as well as potentially postpone their marriage past the age of 25. The likelihood of these two events, however, is probably small. In particular, there is no evidence that the probability of marrying after age 25 is modified by these laws. Columns (7) and (8) suggest that female rules are not affecting fertility decisions of females who married after age 25. The magnitude of the linear effect is extremely small. The dummies for no-consent laws are very significant but do not follow a logical pattern. However, the effect of the male no-consent age, albeit reduced, is still visible and the linear effect is just short of being significant at the 5 percent level. A very similar pattern arises in Column (9) and (10) when looking at college attendees. The effect of female laws disappears when looking at this particular sample while an increase of one year in the male no-consent age appears to lead to a reduction in fertility of about 0.03.

3.6.3 Instrumental variable results

Combining the reduced form results presented above and the first stage, one can then estimate the causal effect of delaying the age of marriage on fertility outcomes, as presented in Table 3.10. The first three columns include the outcomes for males and the last four, those for females. Because each regression includes a control for one's own year of birth, once age at first marriage is included, the spouse's age at first marriage captures both the effect of having a spouse who delays marriage and having one who is older. Thus, the larger that variable is, the older one's spouse is compared to one's own age and the later this person married.

The top panel presents the results of an OLS model. These estimates suggest that individuals who marry one year later have on average about 0.2 more children in the household, but having a wife who delayed decreases that number while having a husband who married later increases it. Children ever born, however, are negatively correlated with a female's age at first marriage but positively with a husband's age. The estimated correlation is -0.075. Households with husbands who married later have children who are older but the relationship between marriage and fertility timing is more complex for females, since when females marry earlier, their eldest child in the household is younger but their youngest child. older. On the other hand, a male with an older wife has older children. As mentioned before, these differences in the correlations between the measures of fertility suggest that many in our sample are too old to have all their children living with them, making the use of the age of children in the household to measure fertility timing potentially problematic.

The second panel introduces the instrumental variable (IV) results. First, the effect of marriage timing on fertility, as measured by the number of children born, is presented in Column (7). These results suggest that a female delaying by one year her first union will have about -0.35 fewer children. This is significant at 10 percent. Marrying a man who marries one year ;ater will further reduce this by -0.27 children. The effects on the other measures of fertility are more difficult to comprehend. The number of children in the household is reduced by 0.01 for a delay in male and their spouse's marriage. Both of these estimates are very imprecisely estimated, however. For females, it appears that having an older husband is what leads to a decrease in the number of children living in the household. Finally, my estimates of fertility timing are very large but suggest that delaying marriage for a female leads to having older kids in the household while a delay in a male's marriage timing leads to the opposite result. These IV results were obtained by instrumenting both age at first marriage using dummies for each legal restriction. Similar magnitudes were obtained by using the linear instruments but the precisions of those estimates were much weaker.

Table 3.11 explores various robustness checks for the effect of marriage timing on children born. First, one could introduce state linear time trends, as is done in Column (1). The results for female marriage timing is barely altered by this change but the coefficient for males changes sign. Being married to an older man now causes an increase in the number of children born. Column (2) then introduces education as an additional covariate and instruments that covariate using compulsory schooling laws. This reduces the magnitude and the precision of the effect of delaying female marriage. This effect appears to be due, however, more to the change in samples than to the introduction of the new variable. Furthermore, the effect of delaying marriage is larger than that of education (although not statistically significantly so). This may suggest that part of results in Leon (2004) may be explained by the effect of compulsory schooling laws on delayed marriage rather than on education itself. The last two columns highlight that this delay in fertility is visible only among Whites and not among Blacks. The magnitude for Whites is slightly larger at about -0.4 but falls dramatically when only restricting the sample to African-American females. This could be because these laws were not very binding on non-White behavior (the first stage is not significant for Blacks). It could also be due to the fact that issues surrounding legitimacy of children might have been more important for White females over this period than for minorities.

Other checks of the robustness of this strategy were also performed. I attempt to correct for selection by adding a variable for not being married and setting one's age at first marriage and that of one's spouse to zero when that dummy is equal to one. Little evidence of selection was found using this strategy. This is not surprising given that the probability of being married or currently married was not found to be influenced by the instruments.

The first stage and reduced form have highlighted how minimum age laws not only affected the gender that was impacted by them but also the behavior of the other gender who would be likely to marry them. If this is true, one would expect that assuming that the laws only modify outcomes through delaying marriage of the gender affected would lead to overestimating the causal effect of marriage timing (in absolute value). This would be supported by the model presented in Section 3.2. This hypothesis is explored in Table 3.12 which reproduces Table 3.10 but only includes as an endogenous variable one's own age at first marriage and instruments for that variable only with the laws impacting that individual. These results suggest that in most cases, both the OLS and the IV would be larger in magnitude in this strategy than in the one presented above, supporting the hypothesis that the laws are influencing more than only one's own age at first marriage. With that strategy, one would have concluded that delaying marriage for females would lead to a decrease in fertility of -0.7, about twice the size of the current estimate.

Table 3.13 then produces the same comparison for a set of additional economic outcomes. They present the OLS and the IV as in Table 3.10 and then that as in Table 3.12 for three outcomes: highest grade achieved, employment status and wage income. The correlations are interesting in their own right. While delaying marriage is correlated with more human capital for both females and males (and their spouses), the pattern for employment and wage income is more intriguing. Males who delay marriage are more likely to work while females are less likely to do so. The correct IV suggests a few elements. First, there is little evidence that delaying marriage leads to higher human capital accumulation. The coefficients on own age at first marriage is negative but not significant for both males and females. On the other hand, it appears that marrying an individual who is older increases human capital accumulation for both males and females (significant for females at 5 percent but only at 10 percent for males). Females who delay marriage by one year are 4 percent more likely to work and to earn about 473 more dollars per year. No effect on male labor supply is found and the wage income regressions only suggest that for a male to marry a female who is one year older would reduce his wage income by about 1,300 dollars.

If one was to ignore the potential effect of these laws on the other gender, Panel C suggests that one would then overestimate the effect of marriage delay by a substantial fraction. One would conclude that delaying marriage for males leads to about 0.8 more years of education and about 2,400 fewer dollars of wage income. The female effect of marriage timing on labor supply would now be about 6 percent and that on wage income close to 570 dollars.

Overall, the results suggest that marriage delay reduces fertility and could affect other socio-economic outcomes. Furthermore, they highlight the role of these laws as affecting not only the gender that is constrained but also the gender that is anticipating to marrying that particular age group. This is somewhat in contrast to Field and Ambrus (2006) who theoretically argue that their estimate (which uses variation at the individual level) can be used to infer market-wide effects of imposing limits on marriage age.

Finally, the IV estimates remain fairly large and very often well above the ones from the OLS model (although not significantly so). While I've attempted to capture the effect that these laws may have on market equilibrium by looking at both genders simultaneously, there is a possibility that other mechanisms than age are at play. For example, the model suggests that not only ages of both genders would be affected but also potentially the quality of their spouse. The latter effect is not captured by the above strategy and this could potential explain why the estimates are still relatively large.

3.7 Conclusion

This paper has argued that laws constraining the marriages of young adults and teens influence the average age at which individuals and their spouses marry. It has also shown that these laws appear to modify the fertility decisions of females. The IV estimates suggest that if a female delays marriage by one year, this will reduce the number of children born to her by about 0.35 children, an estimate only slightly larger than that of Field (2004). It also appears to increase her labor force attachment and her wage income. These results particularly demonstrate that given the matching process involved in marriage markets, legal restrictions imposed on one gender also modify the decisions of the other and that not accounting for this leads to serious overestimated effects of marriage delay.

These findings are useful in order to better understand the potential impact of imposing minimum ages at marriage in developing countries where youthful marriages are more common. They suggest that such policies may be able to reduce fertility, although their effect on fertility timing is more difficult to predict. Furthermore, they emphasize that such a rule could also influence the overall market equilibrium through changes in matching patterns. This analysis highlights the potential role of that secondary channel.

These results also emphasize the role for a better understanding of the relationship between marriage timing and matching patterns. No theoretical framework currently emphasizes the potential impact of delaying marriage on the type of spouse one may eventually marry. While it is possible one would remain with the same partner but simply delay the official ceremony, it is also possible that imposing a ban on youthful marriages can alter who individuals will marry. It would be particularly interesting to evaluate how assortative matching patterns may evolve because of this and the consequences of such changes in matching patterns. This is left to future research.

3.8 References

Bergstrom, T. C. and M. Bagnoli (1993). Courtship as a Waiting Game. *Journal of Political Economy 101*(1), 185–202.

Blank, R., K. K. Charles, and J. Sallee (2007). Do State Laws Affect the Age of Marriage? A Cautionary Tale About Avoidance Behavior. *NBER Working Paper 13667*.

Caucutt, E. M., N. Guner, and J. Knowles (2002). Why Do Women Wait? Matching, Wage Inequality, and the Incentives for Fertility Delay. *Review of Economic Dynamics 5*, 815–855.

Conesa, J. C. (2000). Educational Attainment and Timing of Fertility Decisions. Unpublished manuscript available at `http://www.ere.ub.es/dtreball/E0278.rdf/at_download/file`.

Dahl, G. (2005). Myopic Matrimony and Drop-Out Decisions: Evidence Using State Laws For Marriage, Schooling, and Work. *NBER Working Paper 11328*.

Field, E. (2004). Consequences of Early Marriage for Women in Bangladesh. Unpublished manuscript, available at `http://ipc.umich.edu/edts/pdfs/FieldEM904.pdf`.

Field, E. and A. Ambrus (2006). Early Marriage and Female Schooling in Bangladesh. Unpublished manuscript, available at `http://www.economics.harvard.edu/faculty/field/files/EarlymarEducation-j.pdf`.

Goldstein, J. and C. T. Kenney (2001). Marriage Delayed or Marriage Forgone? New Cohort Forecasts of First Marriage for U.S. Women. *American Sociological Review 66*, 506–519.

Grossberg, M. (1985). *Governing the Hearth: Law and the Family in Nineteenth Century America.* University of North Carolina Press.

Gustafsson, S. (2001). Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe. *Journal of Population Economics 14*, 225–247.

Hotz, V. J. and R. A. Miller (1988). An Empirical Analysis of Life Cycle Fertility and Female Labor Supply. *Econometrica 56*(1), 91–118.

Jensen, R. and R. Thornton (2003). *Gender, Development and Marriage*, Chapter Early Female Marriage in the Developing World. Oxfam GB.

Kalmuss, D. and P. Namerow (1994). Subsequent Childbearing Among Teenage Mothers: The Determinants of a Closely Spaced Second Birth. *Family Planning Perspectives 26*(4), 149–153.

Kiernan, K. (1986). Teenage Marriage and Marital Breakdown: A Longitudinal Study. *Population Studies 40*(1), 35–54.

Klepinger, D., S. Lundberg, and R. Plotnick (1999). How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? *Journal of Human Resources 34*(3), 421–448.

Leon, A. (2004). The Effect of Education on Fertility:Evidence From Compulsory Schooling Laws. Unpublished manuscript, available at `http://www.pitt.edu/~aleon/papers/fertility.pdf`.

Martin, S. P. (2004). *Social Inequality*, Chapter Women's Education and Family Timing: Outcomes and Trends Associated with Age at Marriage and First Birth, pp. 79–118. New York: Russell Sage.

Mullin, C. H. and P. Wang (2002). The Timing of Childbearing Among Heterogeneous Women in Dynamic General Equilibrium. *NBER Working Paper 9231*.

Olivettti, C. (2006). Changes in Women's Hours of Market Work: The Role of Returns to Experience. Unpublished Manuscript available at `http://people.bu.edu/olivetti/papers/returnspap_revision.pdf`.

Oppenheimer, V. K. (1988). A Theory of Marriage Timing. *The American Journal of Sociology 94*(3), 563–591.

Oppenheimer, V. K., M. Kalmijn, and N. Lim (1997). Men's Career Development and Marriage Timing During A Period of Rising Inequality. *Demography 34*(3), 311–330.

Oreopoulos, P. (2007). Do Dropouts Drop Out Too Soon? Wealth, Health, and Happiness from Compulsory Schooling. Unpublished manuscript available at `http://www.chass.utoronto.ca/~oreo/research/dropouts/do%20dropouts%20drop%20out%20too%20soon.pdf`.

Qian, Z. (1998). Changes in Assortative Mating: The Impact of Age and Education, 1970-1990. *Demography 35*(3), 279–292.

Shammas, C. (2002). *A History of Household Government in America.* University of Virginia Press.

3.9 Tables and figures

Table 3.1: **Summary statistics**

	Males				**Females**			
	All		**Currently married**		**All**		**Currently married**	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Demographics								
Age	45.12	15.02	46.54	13.76	46.71	15.81	44.15	13.54
White	0.90	0.30	0.92	0.28	0.89	0.32	0.92	0.28
Black	0.09	0.29	0.08	0.27	0.10	0.31	0.08	0.27
Other race	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.09
Marriage								
Ever married	0.85	0.36	1.00	0.00	0.89	0.31	1.00	1.00
Age at first marriage*	24.27	2.11	24.20	5.62	21.59	1.74	21.40	5.15
Age difference with spouse			3.00	4.87			2.99	4.94
Fertility								
Children ever born					2.37	0.89	2.56	1.97
Number of children in hhd	1.16	0.88	1.52	1.56	1.18	0.94	1.51	1.57
Age of eldest child#	16.06	9.46	13.42	7.74	19.68	12.51	13.69	7.80
Age of youngest child#	12.62	9.95	9.26	8.04	16.58	13.48	9.49	8.16
Other outcomes								
Highest grade attained	11.07	1.82	11.16	3.50	10.99	1.49	11.21	2.86
Employed last week	0.79	0.24	0.83	0.37	0.42	0.20	0.38	0.48
Wage income (1980$)	12525	5776	14097	12770	3620	2248	3067	5225
N	5100514		2491207		4436765		2380419	

All summary statistics are weighted by Census sample-line weights.
*Only available for individuals ever married.
#Only available for individuals with own children in the household

Table 3.2: **Age of consent to marriage, by state, 1885-1980**

State	"No consent laws"						"Consent laws"					
	Male			Female			Male			Female		
	1900	1940	1980	1900	1940	1980	1900	1940	1980	1900	1940	1980
Alabama	21	21	**18**	**18**	**18**	18	17	17	14	**14**	**14**	**14**
Alaska	21	21	18	18	18	18	NA	18	16	NA	16	16
Arizona	21	21	18	18	18	18	14	18	16	12	16	16
Arkansas	**21**	**21**	**21**	**18**	**18**	**18**	17	17	17	14	14	16
California	21	21	18	**18**	**18**	**18**	**18**	**18**	**18**	15	16	18
Colorado	21	21	18	**18**	**18**	**18**	14	16	16	12	16	16
Connecticut	21	21	18	21	21	18	14	16	16	12	16	16
DC	21	21	18	16	18	18	14	18	18	12	16	16
Delaware	21	21	18	18	18	18	**18**	**18**	**18**	**16**	**16**	**16**
Florida	21	21	18	21	21	18	14	18	16	12	16	16
Georgia	17	17	18	18	18	18	17	17	16	14	14	16
Hawaii	NA	20	18	NA	20	18	17	18	16	14	16	16
Idaho	**18**	**18**	**18**	18	18	18	14	14	16	12	12	16
Illinois	21	21	18	**18**	**18**	**18**	17	18	16	14	16	16
Indiana	21	21	18	**18**	**18**	**18**	18	18	17	16	16	17
Iowa	21	21	18	**18**	**18**	**18**	16	16	16	14	14	16
Kansas	15	21	18	12	18	18	15	18	14	12	16	12
Kentucky	21	21	18	21	21	18	14	16	14	12	14	12
Louisiana	**21**	**21**	**21**	**21**	**21**	**21**	14	18	18	12	16	16
Maine	21	21	18	**18**	**18**	**18**	14	14	16	12	16	16
Maryland	21	21	18	16	18	18	14	18	16	12	16	16
Massachusetts	21	21	18	**18**	**18**	**18**	14	18	18	16	16	18
Michigan	**18**	**18**	**18**	18	18	18	**18**	**18**	**18**	**16**	**16**	**16**
Minnesota	21	18	18	18	16	18	18	18	18	15	15	16
Mississippi	**21**	**21**	**21**	18	18	21	14	14	17	12	12	15
Missouri	21	21	18	**18**	**18**	**18**	15	15	15	12	15	15
Montana	21	21	18	18	18	18	18	18	16	16	16	16
Nebraska	21	21	19	18	21	19	18	18	17	16	16	17
Nevada	21	21	18	**18**	**18**	**18**	18	18	16	**16**	**16**	**16**
New Hampshire	14	20	18	13	18	18	**14**	**14**	**14**	13	13	13
New Jersey	21	21	18	**18**	**18**	**18**	14	18	16	12	16	16
New Mexico	21	21	18	**18**	**18**	**18**	18	18	16	15	16	16
New York	18	21	18	18	18	18	18	16	16	18	14	16
North Carolina	**18**	**18**	**18**	**18**	**18**	**18**	**16**	**16**	**16**	14	16	16
North Dakota	21	18	18	18	18	18	18	18	16	15	15	16
Ohio	21	21	18	18	21	18	**18**	**18**	**18**	**16**	**16**	**16**
Oklahoma	21	21	18	18	18	18	18	18	16	15	15	16
Oregon	21	21	18	**18**	**18**	**18**	18	18	17	15	15	17
Pennsylvania	21	21	18	21	21	18	14	16	16	12	16	16
Rhode Island	21	21	18	21	21	18	14	18	18	12	16	16
South Carolina	14	18	18	12	18	18	14	18	16	12	14	14
South Dakota	21	21	18	18	21	18	**18**	**18**	**NA**	15	15	16
Tennessee	16	18	18	16	18	18	14	16	16	12	16	16
Texas	21	21	18	**18**	**18**	**18**	**16**	**16**	**16**	14	14	16
Utah	21	21	18	**18**	**18**	**18**	16	16	14	14	14	14
Vermont	21	21	18	**18**	**18**	**18**	14	16	16	12	16	16
Virginia	21	21	18	21	21	18	14	18	16	12	16	16
Washington	21	21	18	**18**	**18**	**18**	14	14	17	12	15	17
West Virginia	21	21	18	21	21	18	18	18	18	16	16	16
Wisconsin	21	21	18	**18**	**18**	**18**	18	18	16	15	15	16
Wyoming	21	21	18	21	21	18	18	18	16	**16**	**16**	**16**
Average*	20.00	20.51	18.13	18.24	18.72	18.10	16.03	16.98	16.50	14.09	15.23	16.09
Std. Dev.*	3.39	1.30	0.35	3.36	1.84	0.26	3.01	1.38	1.17	3.93	0.99	1.22

Bold entries identify the state, gender and type of laws for which there were no change in the legal age over the entire period 1882 to 1980 or the period for which there are non-missing entries. The full time-series is available from the author upon request.
*Weighted by the population of each state by year.

Table 3.3: **First stage-Basic specification, male respondents**

	Own age at first marriage				Spouse's age at first marriage			
	Full specification		Only own laws		Full specification		Only own laws	
	Own laws	Spouse's laws	Consent	No consent	Own laws	Spouse's laws	Consent	No consent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Dummy regresssions				
Consent minimum age								
13		0.896***			1.410***		1.573***	
		(0.164)			(0.227)		(0.128)	
14		0.192			0.303		-0.009	
		(0.132)			(0.181)		(0.132)	
15	-0.238**	0.397**	-0.011		0.399*	-0.069	0.189*	
	(0.075)	(0.130)	(0.120)		(0.156)	(0.096)	(0.081)	
16	-0.187	0.168	-0.028		0.320	-0.348*	0.079	
	(0.139)	(0.122)	(0.101)		(0.183)	(0.162)	(0.074)	
17	-0.332*	0.528**	-0.085		0.600**	-0.418	0.325*	
	(0.158)	(0.197)	(0.141)		(0.210)	(0.236)	(0.131)	
18	-0.120	0.415*	0.088		0.602**	-0.254	0.391***	
	(0.141)	(0.164)	(0.111)		(0.196)	(0.206)	(0.102)	
No consent minimum age								
13		4.523***			-1.650***			-1.800***
		(0.215)			(0.246)			(0.113)
15		-0.990			0.007			0.284**
		(0.554)			(0.300)			(0.088)
16		-1.272*			-0.136			0.093
		(0.524)			(0.270)			(0.102)
17	0.511	-0.621		-0.434***	0.298	0.028		0.547***
	(0.517)	(0.532)		(0.045)	(0.286)	(0.273)		(0.112)
18	1.099	-1.305*		0.170	-0.290	0.599*		0.151
	(0.573)	(0.530)		(0.279)	(0.257)	(0.284)		(0.077)
19	0.734	-1.440*		-0.198*	-0.359	0.382		0.261
	(0.526)	(0.539)		(0.075)	(0.297)	(0.255)		(0.139)
20	1.142*	-1.400*		0.152	-0.329	0.701**		0.239
	(0.558)	(0.542)		(0.295)	(0.289)	(0.258)		(0.190)
21	0.986	-1.355*		0.030	-0.215	0.610*		0.296
	(0.516)	(0.533)		(0.124)	(0.285)	(0.249)		(0.164)
F (own instruments)	2.97**	77.34***	0.65	35.59***	4.65***	36.55***	1884.01***	55.14***
F (all instruments)	639.34***				901.16***			
Minimum age				**Panel B: Linear regressions**				
Consent	-0.016	0.038	0.021		0.062	-0.043	0.040*	
	(0.040)	(0.024)	(0.029)		(0.033)	(0.052)	(0.017)	
No consent	0.044*	-0.013		0.046**	0.023	0.018		0.043
	(0.021)	(0.028)		(0.013)	(0.040)	(0.035)		(0.029)
N. of cells	19096		19402	19299	19096		19491	19330

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males. All regressions include state, age and year of birth fixed effects. All regressions are weighted by the sum of sample-line weights of individuals in each cell. * significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.4: **First stage-Basic specification, female respondents**

	Own age at first marriage				Spouse's age at first marriage			
	Full specification		Only own laws		Full specification		Only own laws	
	Own laws	Spouse's laws	Consent	No consent	Own laws	Spouse's laws	Consent	No consent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Dummy regresssions				
Consent minimum age								
13			1.310***			0.935**		
			(0.132)			(0.349)		
14	0.331		0.008			0.073		
	(0.169)		(0.140)			(0.149)		
15	0.414**	-0.052	0.191*		-0.038	0.165	0.087	
	(0.142)	(0.095)	(0.072)		(0.197)	(0.142)	(0.173)	
16	0.294	-0.339	0.044		0.043	0.056	0.116	
	(0.175)	(0.175)	(0.075)		(0.147)	(0.118)	(0.095)	
17	0.569**	-0.524*	0.293**		-0.131	0.283*	-0.039	
	(0.186)	(0.258)	(0.105)		(0.149)	(0.122)	(0.107)	
18	0.658**	-0.306	0.364**		0.024	0.114	0.110	
	(0.200)	(0.203)	(0.112)		(0.137)	(0.153)	(0.082)	
No consent minimum age								
15	-2.524***			-2.685***		-0.822*		
	(0.676)			(0.359)		(0.372)		
16	0.080	1.373***		0.114	1.955***	0.746*		1.884***
	(0.454)	(0.196)		(0.373)	(0.218)	(0.327)		(0.157)
17	-0.542	0.361		-0.510	-0.058	0.504		0.377*
	(0.435)	(0.457)		(0.393)	(0.335)	(0.319)		(0.156)
18	0.167	0.738		0.173	0.274	1.564***		0.822***
	(0.440)	(0.457)		(0.387)	(0.318)	(0.327)		(0.151)
19	-0.417	0.485		-0.256	0.094	0.553		0.613***
	(0.445)	(0.467)		(0.371)	(0.329)	(0.319)		(0.159)
20		1.926***		-0.162	1.426***			1.990***
		(0.538)		(0.385)	(0.373)			(0.179)
21	-3.243***	0.293		1.864***	0.367	3.977***		0.843
	(0.488)	(0.609)		(0.390)	(0.405)	(0.364)		(0.444)
F (own instruments)	43.50***	43.07***	17.51***	408.10***	49.51***	155.95***	0.91	287.87***
F (all instruments)	1083.30***				839.68***			
Minimum age				**Panel B: Linear regressions**				
Consent	0.060	-0.058	0.030		0.008	0.016	0.026	
	(0.033)	(0.052)	(0.016)		(0.034)	(0.027)	(0.021)	
No consent	0.026	0.022		0.045	0.046	0.012		0.066***
	(0.028)	(0.029)		(0.026)	(0.024)	(0.024)		(0.019)
N. of cells	18873		19236	19084	18873		19162	19065

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males. All regressions include state, age and year of birth fixed effects. All regressions are weighted by the sum of sample-line weights of individuals in each cell. * significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.5: **First stage robustness checks, male age at first marriage**

	With state-trends		Full sample		With schooling laws	
	Own	Spouse's	Own	Spouse's	Own	Spouse's
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Dummy regresssions			
Consent minimum age						
13		20.471***				
		(5.224)				
14		0.070		0.192		0.224
		(0.064)		(0.133)		(0.137)
15	0.024	0.094	-0.254***	0.373**	-0.202*	0.426**
	(0.057)	(0.092)	(0.072)	(0.139)	(0.092)	(0.127)
16	0.055	0.135*	-0.097	0.141	-0.219	0.189
	(0.067)	(0.062)	(0.136)	(0.116)	(0.125)	(0.123)
17	0.065	0.294*	-0.286	0.322	-0.318*	0.537**
	(0.081)	(0.117)	(0.151)	(0.170)	(0.155)	(0.191)
18	0.076	-0.065	-0.065	0.387*	-0.113	0.435**
	(0.084)	(0.092)	(0.132)	(0.163)	(0.140)	(0.341)
No consent minimum age						
13		4.365***				
		(0.201)				
15		-1.336**		-0.849***		
		(0.454)		(0.235)		
16		-1.289**		-1.100***		-0.287*
		(0.436)		(0.165)		(0.110)
17	0.306	-1.040*	0.597*	-0.556***	0.687*	0.360***
	(0.446)	(0.441)	(0.288)	(0.204)	(0.341)	(0.093)
18	1.262*	-1.412**	0.261*	-1.109***	0.366**	-0.257*
	(0.490)	(0.441)	(0.104)	(0.172)	(0.084)	(0.117)
19	0.908*	-1.477**	0.649*	-1.181***	0.703*	-0.407*
	(0.451)	(0.449)	(0.265)	(0.209)	(0.314)	(0.187)
20	1.085*	-1.412**	0.485***	-1.152***	0.592***	-0.428*
	(0.478)	(0.462)	(0.143)	(0.215)	(0.165)	(0.184)
21	1.089*	-1.412**	0.445**	-1.130***	0.530***	-0.334*
	(0.437)	(0.445)	(0.134)	(0.199)	(0.132)	(0.166)
F (own instruments)	11.63***	106.20***	4.16***	51.78***	4.22***	27.40***
F (all instruments)	214.13***		58.37***		2345.21***	
Minimum age			**Panel B: Linear regressions**			
Consent	0.018	0.021	-0.008	0.036	-0.006	0.032
	(0.022)	(0.018)	(0.038)	(0.024)	(0.041)	(0.024)
No consent	0.048	-0.021	0.048*	-0.012	0.037	-0.004
	(0.027)	(0.024)	(0.018)	(0.026)	(0.024)	(0.029)
N. of cells	19096		20887		16115	

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males, 14. All regressions include state, age and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.6: **First stage robustness checks, female age at first marriage**

	With state-trends		Full sample		With schooling laws	
	Own	Spouse's	Own	Spouse's	Own	Spouse's
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Dummy regresssions			
Consent minimum age						
13	73.726***				1.811*	
	(9.232)				(0.687)	
14	0.065		0.235		0.392*	
	(0.051)		(0.161)		(0.149)	
15	0.018	0.116	0.324*	-0.053	0.478***	-0.049
	(0.056)	(0.065)	(0.126)	(0.135)	(0.132)	(0.089)
16	0.123*	0.093	0.229	-0.190	0.304*	-0.333*
	(0.046)	(0.064)	(0.160)	(0.177)	(0.148)	(0.148)
17	0.088	0.128	0.422*	-0.429	0.595**	-0.515*
	(0.108)	(0.092)	(0.166)	(0.227)	(0.181)	(0.241)
18	-0.042	0.005	0.447*	-0.220	0.668***	-0.271
	(0.073)	(0.060)	(0.181)	(0.186)	(0.171)	(0.173)
No consent minimum age						
13						
15	-3.187***		-1.154***			
	(0.266)		(0.294)			
16	-0.418*	1.379***	0.552**	0.346	-0.600***	0.927**
	(0.193)	(0.117)	(0.173)	(0.198)	(0.072)	(0.311)
17	-0.738***	0.686***	-0.371*	0.111	0.075	1.313***
	(0.194)	(0.192)	(0.149)	(0.142)	(0.075)	(0.323)
18	-0.411*		0.168		-0.446**	1.013**
	(0.186)		(0.176)		(0.140)	(0.311)
19	-0.642***	0.870***	-0.234	1.268***		2.391***
	(0.168)	(0.168)	(0.180)	(0.260)		(0.373)
20		1.915***	1.679	-0.107		0.854
		(0.218)	(0.273)	(0.358)		(0.654)
21	-3.775***	0.858*	-0.178	0.254	-0.380	1.113***
	(0.203)	(0.363)	(0.273)	(0.197)	(0.208)	(0.316)
F (own instruments)	56.64***	219.33***	68.99***	29.91***	29.99***	31.26***
F (all instruments)	7076.55***		4658.29***		1439.20***	
Minimum age			**Panel B: Linear regressions**			
Consent	0.018	0.004	0.056	-0.050	0.051	-0.040
	(0.020)	(0.018)	(0.030)	(0.045)	(0.031)	(0.049)
No consent	0.012	0.004	0.020	0.037	0.028	0.026
	(0.020)	(0.030)	(0.025)	(0.026)	(0.033)	(0.031)
N. of cells	18873		22166		16121	

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males, 14. All regressions include state, age and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell. * significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.7: **Reduced-form estimates, male fertility outcomes**

	Children in hhd		Age of eldest child		Age of youngest child	
	Own	Spouse's	Own	Spouse's	Own	Spouse's
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Dummy regresssions			
Consent minimum age						
13				-0.441		-0.183
				(0.350)		(0.449)
14		0.033		0.503		0.567
		(0.040)		(0.256)		(0.323)
15	-0.020	0.044	-0.355*	0.516*	-0.324	0.489
	(0.034)	(0.028)	(0.175)	(0.247)	(0.189)	(0.281)
16	-0.033	-0.042	-0.667**	0.270	-0.680*	0.437
	(0.029)	(0.026)	(0.236)	(0.291)	(0.299)	(0.345)
17	-0.052	0.024	-0.830*	0.455	-0.734	0.615
	(0.029)	(0.042)	(0.322)	(0.295)	(0.386)	(0.344)
18	-0.013	0.012	-0.453	1.632***	-0.448	2.036***
	(0.027)	(0.035)	(0.279)	(0.417)	(0.319)	(0.530)
No consent minimum age						
13		0.097		-7.401***		-5.989***
		(0.078)		(0.506)		(0.645)
15		0.044		2.540*		2.567*
		(0.090)		(1.024)		(1.085)
16		0.107		3.007**		3.032**
		(0.077)		(1.019)		(1.091)
17	0.030	-0.057	-1.443	2.670*	-2.676*	3.573**
	(0.062)	(0.089)	(1.012)	(1.011)	(1.076)	(1.083)
18	-0.038	-0.014	-1.697	2.675**	-2.383*	3.141**
	(0.073)	(0.075)	(1.069)	(0.974)	(1.078)	(1.032)
19	-0.037	-0.027	-1.393	2.737**	-2.129*	3.380**
	(0.061)	(0.090)	(0.993)	(1.010)	(1.043)	(1.093)
20	0.001	0.052	-1.354	2.766**	-2.081	3.383**
	(0.062)	(0.099)	(1.077)	(1.008)	(1.087)	(1.089)
21	0.037	-0.013	-1.092	2.983**	-2.058	3.607**
	(0.057)	(0.074)	(1.065)	(1.002)	(1.116)	(1.079)
Minimum age			**Panel B: Linear regressions**			
Consent	0.003	-0.014*	-0.045	0.023	-0.051	0.075
	(0.009)	(0.006)	(0.071)	(0.072)	(0.081)	(0.090)
No consent	-0.022	-0.003	-0.047	0.089	-0.020	0.147*
	(0.012)	(0.008)	(0.044)	(0.054)	(0.048)	(0.067)
N. of cells	19096		16845		16845	

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males 14. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.8: **Reduced-form estimates, female fertility outcomes**

	Children in hhd		**Age of eldest child**		**Age of youngest child**		**Children in hhd**	
	Own	Spouse's	Own	Spouse's	Own	Spouse's	Own	Spouse's
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Dummy regresssions				
Consent minimum age								
13				-0.441		-0.183		
				(0.350)		(0.449)		
14		0.033		0.503		0.567	0.048	
		(0.040)		(0.256)		(0.323)	(0.043)	
15	-0.020	0.044	-0.355*	0.516*	-0.324	0.489	0.045	-0.007
	(0.034)	(0.028)	(0.175)	(0.247)	(0.189)	(0.281)	(0.039)	(0.034)
16	-0.033	-0.042	-0.667**	0.270	-0.680*	0.437	-0.034	-0.040
	(0.029)	(0.026)	(0.236)	(0.291)	(0.299)	(0.345)	(0.033)	(0.029)
17	-0.052	0.024	-0.830*	0.455	-0.734	0.615	-0.022	-0.057
	(0.029)	(0.042)	(0.322)	(0.295)	(0.386)	(0.344)	(0.039)	(0.036)
18	-0.013	0.012	-0.453	1.632***	-0.448	2.036***	0.015	0.016
	(0.027)	(0.035)	(0.279)	(0.417)	(0.319)	(0.530)	(0.033)	(0.029)
No consent minimum age								
13		0.097		-7.401***		-5.989***		
		(0.078)		(0.506)		(0.645)		
15		0.044		2.540*		2.567*	-0.074	
		(0.090)		(1.024)		(1.085)	(0.061)	
16		0.107		3.007**		3.032**	-0.141*	-0.016
		(0.077)		(1.019)		(1.091)	(0.066)	(0.062)
17	0.030	-0.057	-1.443	2.670*	-2.676*	3.573**	-0.003	0.092**
	(0.062)	(0.089)	(1.012)	(1.011)	(1.076)	(1.083)	(0.055)	(0.029)
18	-0.038	-0.014	-1.697	2.675**	-2.383*	3.141**	-0.049	0.018
	(0.073)	(0.075)	(1.069)	(0.974)	(1.078)	(1.032)	(0.070)	(0.033)
19	-0.037	-0.027	-1.393	2.737**	-2.129*	3.380**	-0.144*	0.022
	(0.061)	(0.090)	(0.993)	(1.010)	(1.043)	(1.093)	(0.056)	(0.034)
20	0.001	0.052	-1.354	2.766**	-2.081	3.383**		-0.094
	(0.062)	(0.099)	(1.077)	(1.008)	(1.087)	(1.089)		(0.057)
21	0.037	-0.013	-1.092	2.983**	-2.058	3.607**	-0.114	0.087
	(0.057)	(0.074)	(1.065)	(1.002)	(1.116)	(1.079)	(0.068)	(0.043)
Minimum age				**Panel B: Linear regressions**				
Consent	0.003	-0.014*	-0.045	0.023	-0.051	0.075	-0.014*	0.010
	(0.009)	(0.006)	(0.071)	(0.072)	(0.081)	(0.090)	(0.006)	(0.009)
No consent	-0.022	-0.003	-0.047	0.089	-0.020	0.147*	0.005	-0.024
	(0.012)	(0.008)	(0.044)	(0.054)	(0.048)	(0.067)	(0.007)	(0.013)
N. of cells	19096		16845		16845		18873	

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males 14. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.9: **Reduced form robustness checks, children ever born**

	Females									
	With state-trends		**Full sample**		**With schooling laws**		**Married after 25**		**College attendees**	
	Own laws	Spouse's laws	Own laws	Spouse's laws	Own laws	Spouse's laws	Own laws	Spouse's laws	Own laws	Spouse's laws
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Dummy regresssions									
Consent minimum age										
13	-42.374***		-0.476*		-0.183		-0.236			
	(5.051)		(0.184)		(0.192)		(0.135)			
14	-0.025		-0.019		-0.114		0.057		0.073	
	(0.047)		(0.148)		(0.098)		(0.050)		(0.040)	
15	-0.026	0.004	-0.090	0.056	-0.129	0.040	0.019	0.034	0.085*	0.008
	(0.063)	(0.051)	(0.104)	(0.098)	(0.074)	(0.077)	(0.049)	(0.027)	(0.040)	(0.049)
16	-0.070	-0.062	-0.080	0.011	-0.110*	0.023	0.062	-0.045	0.010	-0.024
	(0.046)	(0.051)	(0.077)	(0.084)	(0.048)	(0.067)	(0.032)	(0.036)	(0.033)	(0.036)
17	-0.159***	-0.048	-0.174	0.095	-0.251***	0.118	-0.412	-0.074	0.017	-0.055
	(0.040)	(0.058)	(0.106)	(0.120)	(0.071)	(0.095)	(0.049)	(0.027)	(0.071)	(0.044)
18	-0.263***	0.060	-0.368**	0.108	-0.403***	0.095	-0.130*	0.030	-0.064	0.069*
	(0.062)	(0.051)	(0.122)	(0.084)	(0.109)	(0.057)	(0.050)	(0.273)	(0.042)	(0.027)
No consent minimum age										
13			-0.718**				-1.306***		-1.951***	
			(0.252)				(0.179)		(0.094)	
15	-0.512***		-0.444	-0.734***			-0.564***	-0.795***	0.189	-0.335
	(0.120)		(0.353)	(0.200)			(0.148)	(0.092)	(0.265)	(0.083)
16	-0.089	-0.210*	-0.191	0.123	0.154***	-0.070	-0.501**	0.144	0.160	-0.475
	(0.192)	(0.095)	(0.331)	(0.291)	(0.053)	(0.186)	(0.136)	(0.099)	(0.257)	(0.249)
17	0.053	0.137	-0.351	-0.332	0.054	-0.549**	-0.383***		0.196	
	(0.190)	(0.195)	(0.347)	(0.298)	(0.030)	(0.201)	(0.137)		(0.266)	
18	-0.030		-0.354	-0.180	-0.010	-0.437*	-0.663***	-0.111	0.064	-0.563*
	(0.194)		(0.329)	(0.297)	(0.110)	(0.184)	(0.137)	(0.110)	(0.257)	(0.250)
19	-0.169	-0.106		-0.460		-0.715**		-0.279*		-0.396
	(0.189)	(0.179)		(0.326)		(0.231)		(0.126)		(0.253)
20		-0.301	-1.548***	0.038		-0.302	-0.785***	0.046		-0.360
		(0.193)	(0.314)	(0.310)		(0.278)	(0.169)	(0.143)		(0.222)
21	-0.294	0.046	-0.298	-0.213	-0.032	-0.424*	-0.597***	-0.034	0.154	-0.538*
	(0.226)	(0.245)	(0.327)	(0.297)	(0.142)	(0.196)	(0.141)	(0.103)	(0.260)	(0.251)
Min. age	**Panel B: Linear regressions**									
Consent	-0.029**	0.022	-0.042**	0.029	-0.037**	0.025	-0.004	0.006	-0.012	0.026
	(0.009)	(0.015)	(0.014)	(0.020)	(0.011)	(0.017)	(0.009)	(0.008)	(0.009)	(0.013)
No consent	-0.006	-0.016	-0.007	-0.047	-0.016	-0.050**	-0.009	-0.034	0.023	-0.032
	(0.013)	(0.035)	(0.023)	(0.026)	(0.019)	(0.018)	(0.016)	(0.017)	(0.012)	(0.016)
N. of cells	18873		32810		16121		425768		579914	

Standards errors clustered at the state level in parentheses. Reference category for females is 12 and for males 14. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.10: **OLS and IV results, fertility outcomes**

	Males			Females			
	Children in hhd	Age of eldest child	Age of youngest child	Children in hhd	Age of eldest child	Age of youngest child	Children ever born
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: OLS			
Age at first marriage							
Own	0.210***	-0.181***	-0.812***	0.116***	0.189**	-0.145**	-0.075***
	(0.020)	(0.031)	(0.058)	(0.016)	(0.064)	(0.049)	(0.009)
Spouse's	-0.088***	0.097*	0.339***	0.023*	-0.300***	-0.461***	0.027**
	(0.007)	(0.041)	(0.049)	(0.008)	(0.048)	(0.067)	(0.010)
				Panel B: IV (Dummies as instruments)			
Age at first marriage							
Own	-0.019	-0.268	-0.163	0.063	0.856	1.250	-0.350
	(0.094)	(0.302)	(0.307)	(0.055)	(0.758)	(0.763)	(0.200)
Spouse's	-0.018	0.898*	1.377***	-0.193*	-1.177	-1.252	-0.271
	(0.088)	(0.343)	(0.391)	(0.078)	(0.775)	(0.819)	(0.294)
N. of cells	19096	16845	16845	18873	16096	16096	18873

Standards errors clustered at the state level in parentheses. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.11: **OLS and IV robustness checks, children ever born**

	Females			
	State-specific trends	With education	Whites	Blacks
	(1)	(2)	(3)	(4)
	IV (Dummies as instruments)			
Own age at first marriage	-0.372	-0.291	-0.418	-0.137
	(0.197)	(0.203)	(0.218)	(0.175)
Spouse's age at first marriage	0.085	-0.296	-0.212	-0.151
	(0.171)	(0.322)	(0.317)	(0.181)
Highest grade achieved		-0.034		
		(0.121)		
N. of cells	18873	16121	8387	5241

Standards errors clustered at the state level in parentheses. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.12: **OLS and IV results, own age at first marriage**

	Males			Females			
	Children in hhd (1)	Age of eldest child (2)	Age of youngest child (3)	Children in hhd (4)	Age of eldest child (5)	Age of youngest child (6)	Children ever born (7)
				Panel A: OLS			
Own age at first marriage	0.015***	-0.412***	-0.488***	0.014	-0.338***	-0.382***	-0.101***
	(0.004)	(0.028)	(0.026)	(0.007)	(0.021)	(0.018)	(0.006)
				Panel B: IV (Dummies as instruments)			
Own age at first marriage	-0.231	1.328	2.935	-0.075	0.244	0.838	-0.657**
	(0.153)	(1.362)	(1.673)	(0.093)	(0.890)	(0.811)	(0.195)
N. of cells	19206	16944	16944	18873	16096	16096	18873

Standards errors clustered at the state level in parentheses. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 3.13: **OLS and IV results, other outcomes**

	Males			Females		
	Highest grade	Employed	Wage income	Highest grade	Employed	Wage income
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS			
Own age at first marriage	0.050***	0.030***	614.810***	0.079***	-0.008***	-66.572**
	(0.009)	(0.003)	(78.933)	(0.011)	(0.002)	(21.025)
Spouse's age at first marriage	0.083***	-0.007***	-156.946***	0.034**	0.022***	178.137***
	(0.012)	(0.002)	(39.278)	(0.011)	(0.002)	(19.503)
			Panel B: Double IV (Dummies as instruments)			
Own age at first marriage	-0.065	0.003	-317.695	-0.274	0.039	473.296*
	(0.205)	(0.014)	(490.424)	(0.215)	(0.021)	(183.889)
Spouse's age at first marriage	0.451	0.013	-1300.729*	0.716*	-0.001	-120.868
	(0.265)	(0.010)	(636.832)	(0.329)	(0.034)	(339.544)
			Panel C: Single IV (Dummies as instruments)			
Own age at first marriage	0.809	0.021	-2375.675*	0.148	0.059**	570.409**
	(0.615)	(0.017)	(907.359)	(0.226)	(0.021)	(187.249)
N. of cells	19096	19096	19096	18873	18873	18873

Standards errors clustered at the state level in parentheses. All regressions include state, year of Census and year of birth fixed effects and are weighted by the sum of sample-line weights of individuals in each cell.
* significant at 5%; ** significant at 1%; *** significant at 0.1%

Figure 3-1: **Mean and median age at first marriage, males**

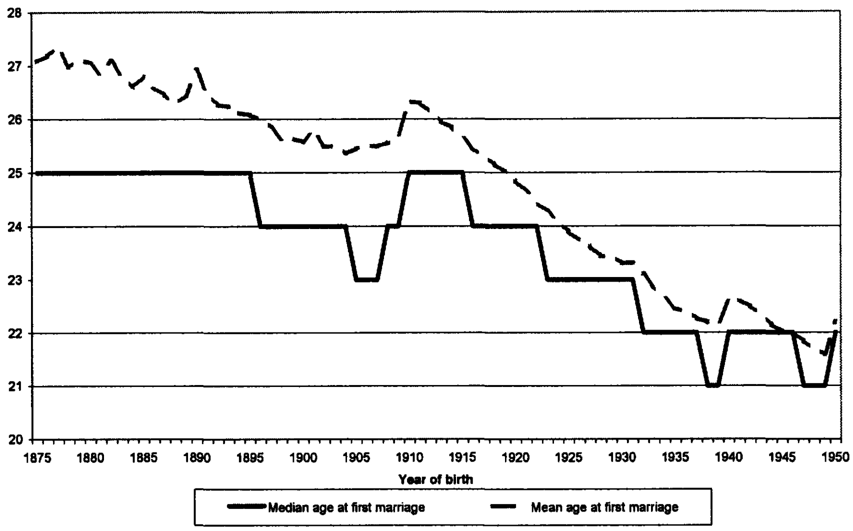


Figure 3-2: **Mean and median age at first marriage, females**

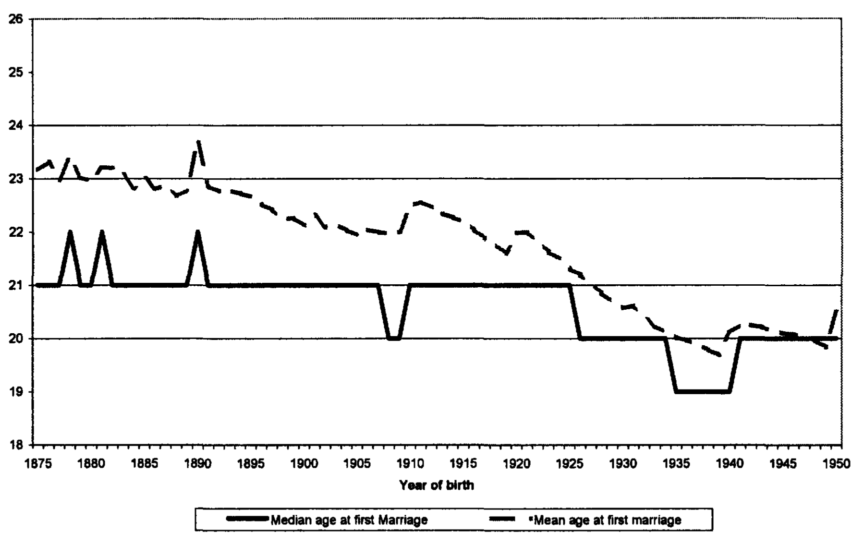


Figure 3-3: **Distribution of age at first marriage, males by no-consent age**

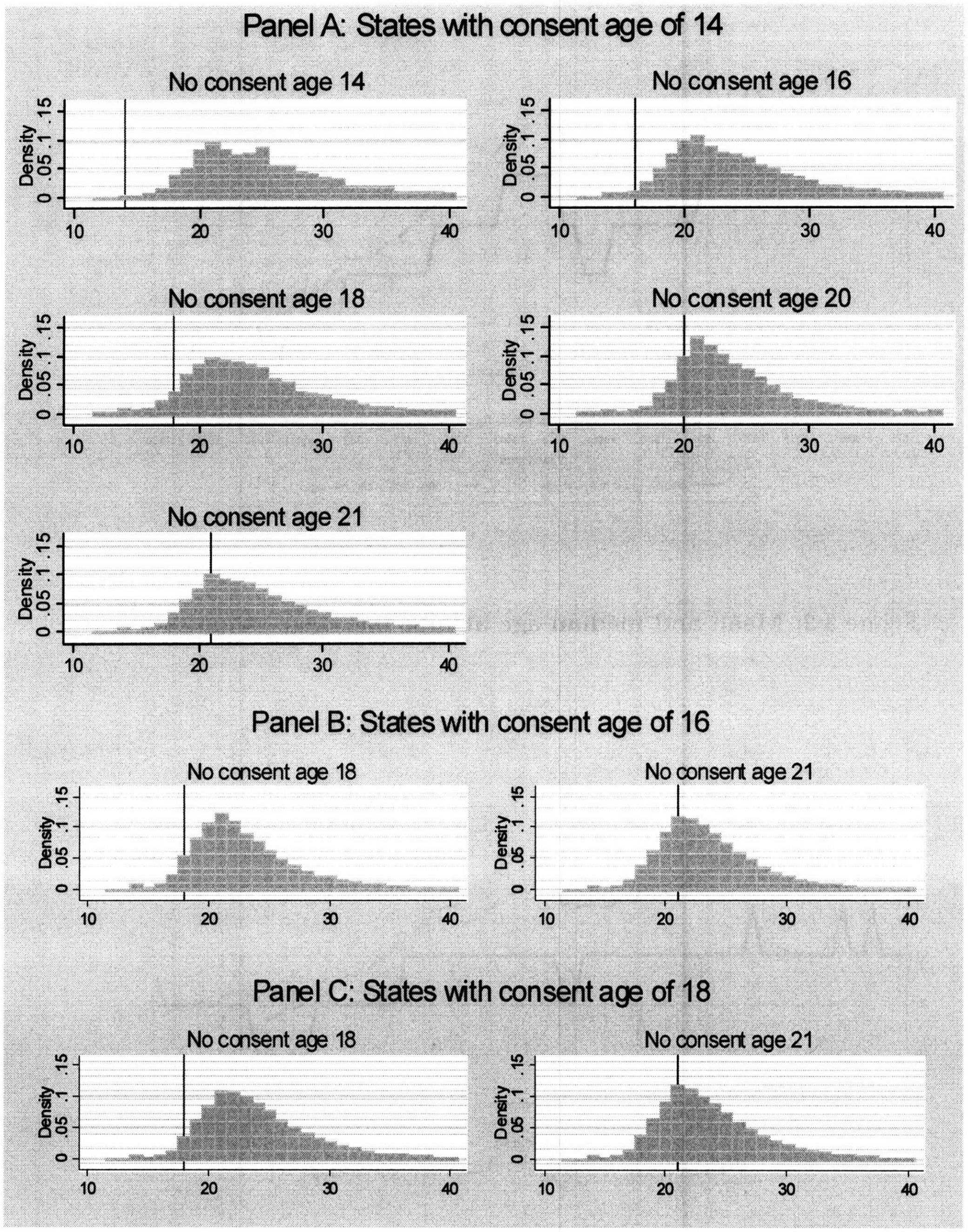


Figure 3-4: **Distribution of age at first marriage, females by no-consent age**

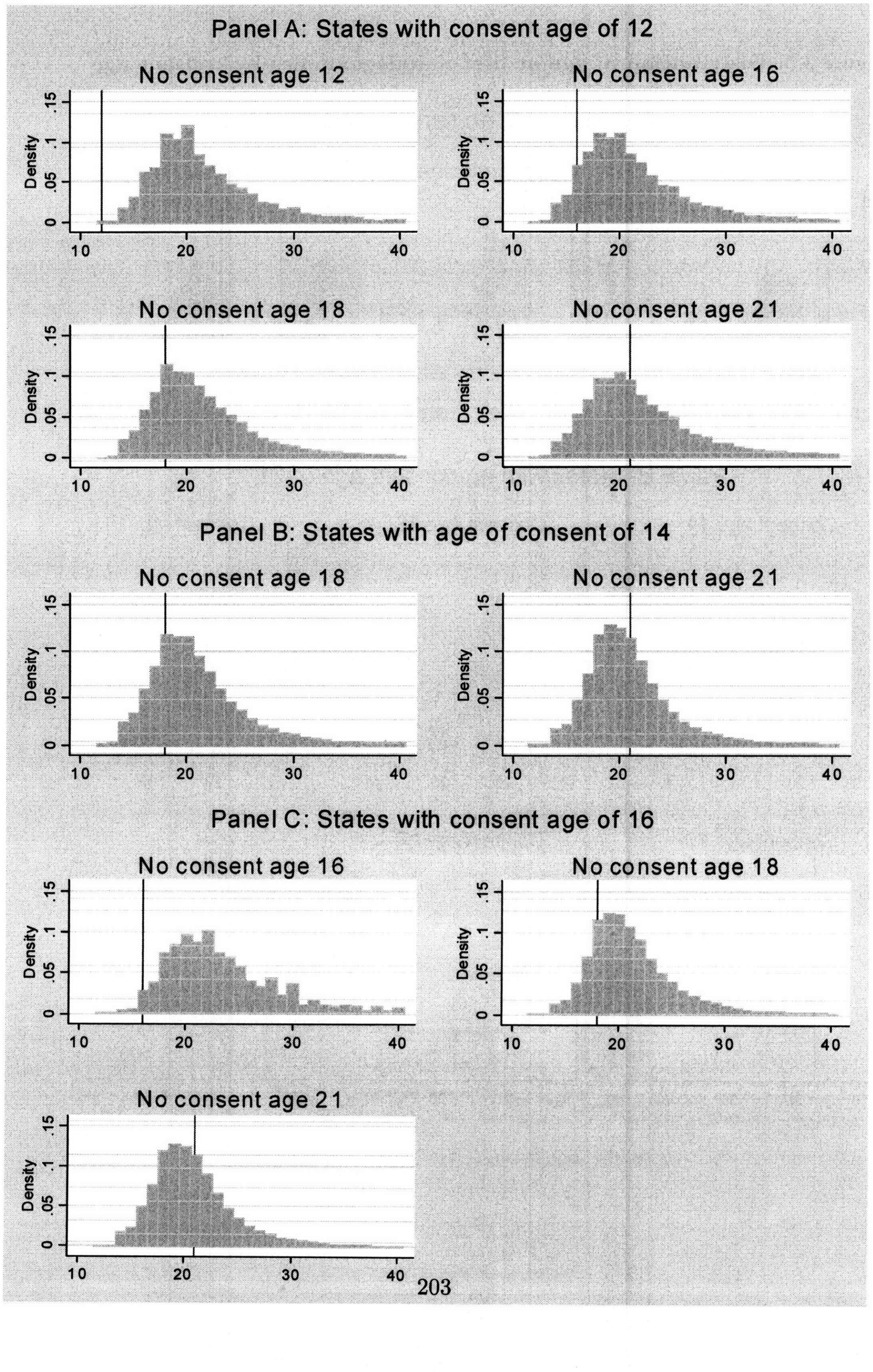


Figure 3-5: **Distribution of age at first marriage, males by consent age**

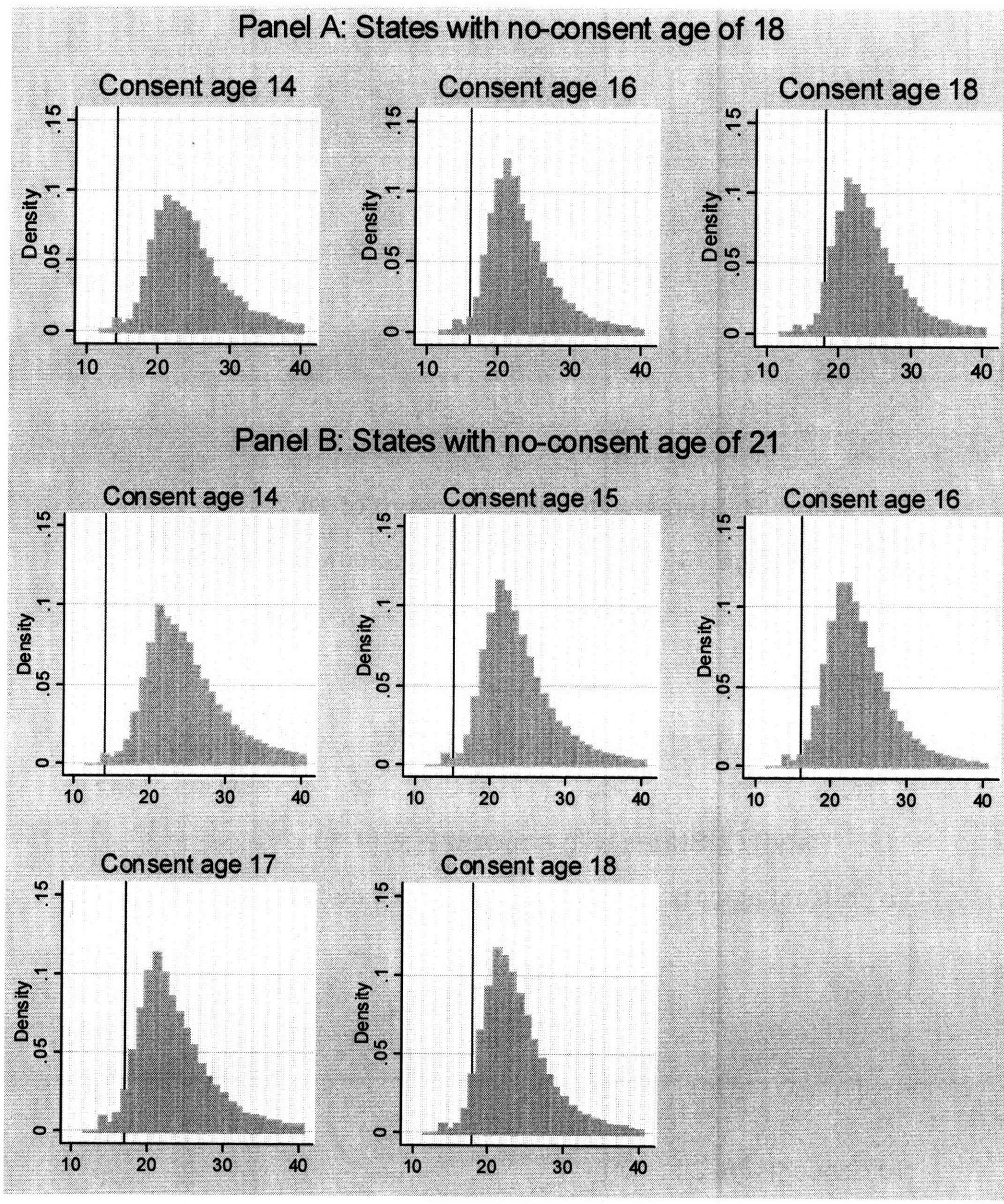
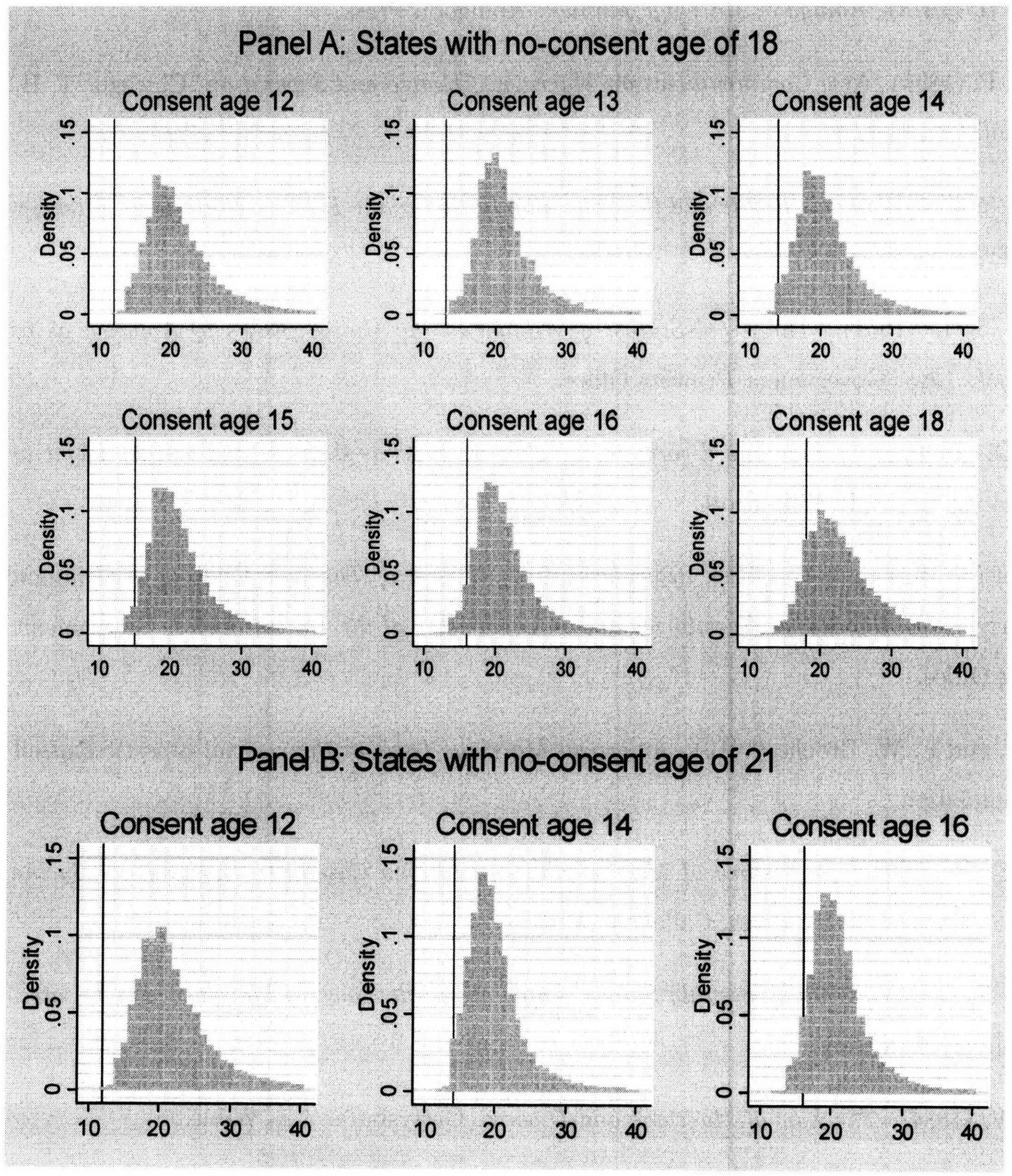


Figure 3-6: **Distribution of age at first marriage, females by consent age**



3.A References used for legal compilation

Bayles, G. J. (1901). *Women and the Law.* New York: Century.

Biddle, P. H. (1974). *Abingdon Marriage Manual.* Abingdon Press.

Bishop, J. P. (1891). *New Commentaries on Marriage, Divorce and Separation.* Chicago: T. H. Flood and Company.

Brown, B. (1977). *Women's Rights and the Law: The Impact of the ERA on State Laws.* Praeger Publishers.

Buchanan, S. L. (1951). *The Legal Status of Women in the United States of America as of January 1, 1948.* Government Printing Office.

Convers, D. (1889). *Marriage and divorce in the United States as they are and as they ought to be.* Philadelphia: J. B. Lippincott.

Director of the Bureau of War Risk Insurance (1919). *Digest of The Law Relating to Common Law Marriage in the States, Territories and Dependencies of the United States.* Government Printing Office.

Hall, F. S. and E. W. Brooke (1919). *American Marriage Laws in their social aspects.* Russell Sage Foundation.

Howard, G. E. (1904). *A History of Matrimonial Institutions Chiefly in England and in the United States.* University Press, Callaghan & Co.

Hudgings, F. (1935). *What Everybody Should Know about the Laws of Marriage and Divorce.* New Century Company.

Hull, R. W. (1884). *The Law of Marriage and Divorce.* Gettysburg: J.E. Wible.

Jacobson, P. H. (1959). *American Marriage and Divorce.* Rinehart & Company, Inc.

Kanowitz (1969). *Women and the Law: The Unfinished Revolution.* Albuquerque: University of New Mexico Press.

Kanowitz, L. (1973). *Sex Roles in Law and Society.* Albuquerque: University of New Mexico Press.

Kaplan, S. (1967). *Marriage and Divorce.* International Institute of Municipal Clerks.

Kuchler, F. (1966). *Law of Engagement and Marriage.* Oceana Publications, Inc.

Kuchler, F. W. (1978). *Law of Engagement and Marriage.* Oceana Publications.

Kuchler, F. W. H. (1961). *Law of Marriage and Divorce Simplified.* Oceana Publications.

Leach, W. H. (1933). *The Cokesbury Marriage Manual.* Cokesbury Press.

Leach, W. H. (1945). *The Cokesbury Marriage Manual.* Cokesbury Press.

Leach, W. H. (1959). *The Cokesbury Marriage Manual.* Cokesbury Press.

Livy, M. I. (1924). *Marriage and Divorce.* New York.

Mackay, R. V. (1946). *Law of Marriage and Divorce Simplified.* Oceana Publications.

Mackay, R. V. and I. Mandell (1959). *Law of Marriage and Divorce Simplified.* Oceana Publications.

Mandell, I. (1951). *Law of Marriage and Divorce Simplified.* Oceana Publications.

Morland, J. W. (1946). *Keezer on the Law of Marriage and Divorce.* The Bobbs-Merrill Company Publishers.

National League of Women Voters (1924). *A Survey of the Legal Status of Women in the Forty-Eight States.* National League of Women Voters.

National League of Women Voters (1930). *A Survey of the Legal Status of Women in the Forty-Eight States.* National League of Women Voters.

Nelson, W. T. (1895). *A Treatise on the Law of Divorce and Annulment of Marriage.* Callaghan and Company.

Noble, C. (1882). *A Compendium and Comparative View of the Thirty-Eight State Laws of Marriage and Divorce in the United States (in 1882).* New York: Baker, Voorhis & Co.

Richmond, M. E. and F. S. Hall (1929). *Marriage and the State.* Russell Sage Foundation.

Speer, O. (1901). *A Treatise on the Law of Married Women in Texas.* The Lawyers Co-operative Publishing Company.

State of Massachusetts (1866). *Laws concerning the Solemnization, Record and Return of Marriages in Massachusetts.* Boston: Wright & Potter, State Printers.

State of Massachusetts (1873). *Laws concerning the Solemnization, Record and Returns of Marriages in Massachusetts.* Boston: Wright & Potter, State Printers.

Vernier, C. G. (1931). *American Family Laws.* Stanford University Press.

Vernier, C. G. (1938). *American Family Laws.* Stanford University Press.

White, F. A. (1910). *Laws on Marriage, Divorce and Property Rights of Married Women of All States and Alaska, Hawaii, Arizona, New Mexico and District of Columbia.* Baumgardt Publishing Co.

Whitney, H. C. (1894). *Marriage and divorce: the effects of each on personal status and property rights with a consideration of fraudulent divorces and the ethics of divorce from popular and professional use.* Philadelphia: J.E. Potter.

Wilson, J. L. (1913). *Woman and the Law.* Chicago: Civics Society.

Women's Bureau (1956). *The Legal Status of Women in the United States of America.* United States Department of Labor.

Wright, C. D. (1897). *Marriage and Divorce in the United States 1867-1886.* Government Press Office.