ESSAYS IN DEVELOPEMENT ECONOMICS

By

Sambojyoti Biswas

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics – Doctor of Philosophy

2019

ABSTRACT

ESSAYS IN DEVELOPEMENT ECONOMICS

By

Sambojyoti Biswas

In chapter one I use a large nationally representative data set from India to analyze the causal effect of age at marriage of a woman on her post marriage health and fertility outcomes. To look at this effect I propose a new instrumental variable in the sparse causal literature. The instrumental variable strategy stems from two social practices within the Indian society; minimum age targeting at marriage and seasonality of marriage dates. I find that delaying marriage causally decreases the probability of a women being diabetic and having elevated blood pressure post marriage, increases her age at first birth and has a zero effect on the number of children she has. I also find that delayed marriage causally increases the women's educational attainment, improves her bargaining power within the household and improves her spousal quality. The paper also contrasts its findings with the existing causal literature by producing results using both the new and the existing instrumental variable in the literature. I find that both the IV's produce similar estimates in terms of health, education and spousal quality but the estimates differ for fertility outcomes.

In the absence of universal social security, parents in India depend heavily on their off springs for post-retirement consumption. The patriarchal nature of the Indian society combined with low labor market returns for women skews this dependency towards sons. This skewedness in dependency for old age support, towards the son, might be a potential reason for gender gap in human capital investments of the children. In chapter two I use the 2004 New Pension Scheme reform in India as a quasi-experiment in a Difference in Difference in Difference framework to identify the causal effect of post retirement security on gender gap in human capital investment. I find that with the decrease in post-retirement security for the parents the gender gap against the female child increases. Compared to the male child, a girl child is less likely to be enrolled in a private school or a school where the medium of instruction is English, both these effects are statistically significant.

I also find that the gender gap in test scores is larger post reform for the children of parents whose retirement security decreases, however these results are not statistically significant.

Chapter three looks at the relationship between risk correlation in borrower outcomes and micro finance tools such as monitoring and audit. To look at these relationships I introduce risk correlation in existing theoretical models of moral hazard and costly state verification. I find that as risk correlation in borrower outcomes increases auditing by the lender increases whereas the relation with monitoring is conditional on the probability of success of the borrower and the monitor. I use the Townsend Thai data base to check whether the empirical findings are consistent with the theoretical predictions. I find that the relationship between risk correlation and monitoring depends on the measure of risk correlation I use and for auditing I find that the empirical findings are opposite to the theoretical prediction.

This thesis is dedicated to all the women in developing countries, majority of whom are still strugling for equality and autonomy.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CAUSAL EFFECT OF AGE AT MARRIAGE ON HEALTH AND FERTILITY

1.1 Introduction

In developing countries the practice of women getting married at a young age is still prevalent. In India the legal age of marriage for women is eighteen but we see that almost 41 percent of women are getting married before turning the age of eighteen and 82 percent of women are getting married by the age of 21 (DHS VII, 2016). The numbers are also troubling for other developing countries such as Bangladesh and Niger where 51 percent and 74 percent of the women are married before turning the age of 18 (UNFPA, 2007). Women getting married at a young age is negatively correlated with the women's educational, health and marital outcomes. We also see negative correlation in terms of their child's health and educational outcomes (Bruce 2003, Clark 2004, Nour 2006, Santhya et al. 2010). Early marriage also has negative impact on national development by stunting educational and vocational opportunities for a large sector of the population (Raj et al. 2009). In this paper I try to estimate the causal effect of age at marriage of a women on her post marriage health and fertility outcomes.

There has been a lot of correlational studies looking at the adverse effects of early marriage, but the causal literature on age at marriage is sparse. Field and Ambrus, (2008) was one of the first papers to establish a causal relationship. They found that delayed marriage is causally associated with higher educational attainment and literacy. They also found that delayed marriage is causally associated with an increase in use of preventative health services. Sekhri and Debnath, (2014) and Chari et al., (2017) look at the causal effect of age at marriage of the mother on her children's wellbeing. They find that delayed marriage is causally associated with better health and educational outcomes for the children. Chari et al, (2017) also found that the gain in child outcomes due to delayed marriage is partially driven by a decrease in mother's fertility. A recent study by Dhamuja and Roychowdhury, (2018) also find that delayed marriage is causally associated with a decline in

1

physical violence against the women.

Introduced by Field and Ambrus, (2008), all the aforementioned causal papers uses age at menarche as an instrumental variable for age at marriage to establish causality. They argue that a significant portion in the variation of age at menarche is random, making it a good instrument for age at marriage. There are two issues in using age at menarche as an instrumental variable to establish a causal relationship on heath and fertility in this paper. First, if we look at graph 1 which plots the distribution of age at menarche and age at marriage, we see that 98.7 percent of the women have gone through menarche by the age of sixteen whereas only 30.2 percent of the women are married by the age of sixteen. Since an instrumental variable strategy estimates LATE, the complier group from using age at menarche as an IV is small and might not be representative of the current population of women. A second and more pressing issue is that age at menarche might not be an exogenous variation when we look at women's health and fertility outcomes, and hence fail the exogenity condition of an instrument. In the medical literature we see that both early and late onset of menarche is often associated with adverse pregnancy outcomes such as ectopic pregnancies and spontaneous abortions (Liestol, 1980; Martin et al., 1983; Wyshak, 1983; Sandler et al., 1984). Later onset of menarche is also associated with a slightly higher risk of sub fecundity and infertility (Gulbrandsen et al., 2014). Hence my biggest concern is that age at menarche could affect health and fertility outcomes through channels other than age at marriage, such as underlying reproductive health. For example if late onset of menarche increases the risk of sub fecundity and infertility, then that might result in lower number of children. At the same time late onset of menarche delays marriage age, thus the causal effect of delayed age at marriage on number of children obtained from using age at menarche as an instrument might be biased.

In this paper I propose a new instrumental variable to look at the causal effect of age at marriage of a women on her post marriage health and fertility outcomes. The exogenous variation for the instrument comes from two social practices within Indian society; Seasonality of marriages and parents preference of getting their daughter married off early and after they turn a minimum marriage eligible age. First, the seasonality of marriages partly stems from the deep rooted belief

in astrology by the Hindu population in India. When it comes to choosing marriage dates certain dates in the Hindu calendar year are considered more auspicious for getting married compared to others. If we look at graph 2 which plots the distribution of marriages by month we see that certain months of the year on average have more marriages compared to others. Second, parents in India on average prefer to marry off their girl child at an early age. The legal age for marriage in India is eighteen for women. This creates a weak lower bound on how early parents can get their daughters married off. If we look at graph 3 which pots the distribution of marriage age in months, we see a spike in the number of marriages at the exact age of eighteen (216 months). From the data we see these spikes also at whole number ages from fifteen through twenty one. I call this phenomenon minimum age targeting at marriage. Where parents prefer to get their daughters married off early and after they turn a minimum marriage eligible age. The minimum age after which parents can get their daughters married off should be eighteen due legal reasons but we see such targeting at the aforementioned ages of fifteen to twenty one also. This could be driven by localized social and cultural norms about minimum marriage age of a woman. If the minimum target age is binding then the women who turns the minimum target age in the month which has low number of marriage dates i.e. women who are born in the month with low number of marriage dates, has to wait longer in expectation to get married. If being born in a month with low number of marriages versus being born in a month with high number of marriages is random, then that produces an exogenous shock to their age at marriage. I exploit this variation using an instrumental variable strategy.

I use a nationally representative data set from India to estimate the results. First I look at the overall causal effect of age at marriage of a women on her post marriage health and fertility outcomes. The effect of age at marriage on health and fertility are possibly driven by other intermediary variables such as educational attainment, spousal quality, spousal family quality and bargaining power of the women within the marriage. From the previous literature (Chari et al., 2017; Field and Ambrus, 2008; Shekhri and Debnath, 2014; Dhamuja and Roychowdhury, 2018) we know that the intermediary variables itself are also affected by age at marriage. So in the next step I look at the effect of age at marriage on these intermediary variables. Finally, I try to

3

understand the mechanisms and potential channels through which age at marriage affects health and fertility. Also to bench mark my findings to the previous causal literature, I re-estimate and compare the results obtained from using both the instrumental variables (age at menarche IV and my IV) separately. The prior is that both the IV's should produce similar results for the intermediary variables and the estimates might differ when we look at health and fertility outcomes.

For post marriage health outcomes, I find that one month delay in marriage reduces the probability of the woman being diabetic (2.8 percent), being pre-diabetic (3.5 percent) and having elevated blood pressure (0.8 percent). The results are statistically and economically significant. Diabetes and blood pressure are strongly driven by genetics but they are also affected by lifestyle choices such as lack of exercise, unhealthy meal plans and obesity (Nahas et al., 2009). I also look at the effect of age at marriage on anemia (0.3 percent increase), I do not find a statistically significant effect. For fertility outcomes, I find that, one month delay in marriage has no effect on the total number of children (0.02 percent decrease), decreases the ideal number of children reported by mother (0.09 percent), increases the age at first birth (0.3 percent) and decreases birth gap by (0.05 percent). The effect on age at first birth is highly statistically significant whereas the effect on birth gap and number of children is not.

For intermediary variables, I find that one month delay in marriage increases the number of years of education of the woman by (0.5 percent) and having at least primary level of education (0.04 percent). The effect on years of education is statistically and economically significant. One month delay in marriage increases husband's years of education (0.4 percent), decreases husbands age[1] (0.05 percent), increases the wealth of the in laws household (0.07 percent) and decreases the probability that the in laws household belongs to a rural area (0.9 percent). The effects are statistically significant. One month delay in marriage decreases the probability of the husband limiting the women to contact her natal (3.6 percent), decreases the probability of the husband insist of knowing her whereabouts (3.3 percent) and decreases the number of control questions

---

[1]A large age gap between men and women at time of marriage would generally indicate that the younger partner has less power and less say in the relationship (Carmichael, 2011).

answered yes[2] (0.2 percent). The effects are statistically significant. Overall we see that women who get married later have higher education, better spousal and spousal family quality, and higher bargaining power in the marriage.

To systematically unpack the channels for the effect of age at marriage on health and fertility I look at three different specifications. I look at the results controlling for spousal and spousal family characteristics, controlling for spousal, spousal family and bargaining power and including spousal and spousal family controls for the zero education population. For the health outcomes I see that adding spousal controls and bargaining power controls does not change the estimates for the health outcomes, for the zero education I find that the gains are even larger this rules out the possibility that the gains in health are completely driven by gains in education and spousal quality. To conclude, I see that delaying marriage causally results in better health outcomes for the women and the gains are not completely driven by gains in education, spousal quality and bargaining power.

Next, I unpack the effect of delaying age at marriage on fertility. Previously we saw that delaying marriage age had no effect (0.02 percent decrease) on the total number of children but the effect was not statistically significant. One month delay in marriage increases the number of children by 0.009 (0.3 percent) when I include spousal and spousal family controls, by 0.008 (0.3 percent) when I include spousal controls and bargaining controls, and by 0.016 (0.5 percent) for the zero education sample with spousal controls. The effects are highly statistically significant but small in magnitude. For age at first birth and birth gap I see that adding spousal and bargaining controls do not change the estimate. Overall I conclude that delayed age at marriage has zero or a very small positive effects on the number of children and it increases the age at first birth.

When I compare the results of the effect of age at marriage on the intermediary variables across the two IV's (age at menarche and my IV), I find that the estimates are equal in direction and magnitude. I also compare these results with estimates from the previous literature and find that they are equal in direction and mostly similar in magnitude. When I compare the results on health outcomes across the two IV's, I find that the estimates are similar in direction and magnitude. My

---

[2]Lower number of control questions answered yes to implies greater bargaining power for the women.

prior was that, both the IV's should produce similar estimates for the intermediary variables but the estimates on health and fertility might differ. A possible explanation for why I see similar results across the two IVs for health outcomes could be that the health factors which are influenced by early or late onset of menarche might not affect health outcomes such as diabetes, blood pressure and anemia. The main divergence in results is when I look at the effect of age at marriage on number of children. My IV estimates a zero or small positive effect on the number of children. Whereas age at menarche IV estimates a negative, statistically significant effect on the number of children. The previous literature which uses age at menarche as an IV also found negative and significant effect. There are two plausible explanations for this, the first one which has already been stated, age at menarche is associated with reproductive health using it as an IV to measure total number of children might downward bias the results. Another explanation could be that the compliers from both the IV's are different and since an IV estimates LATE, the results are different.

There are three major contributions of this paper, first this is the first paper in the literature to look at the causal effect of age at marriage on diabetes, blood pressure and anemia. Second, the paper introduces a new instrumental variable strategy in the literature to look at the causal effects of age at marriage, the IV is less demanding in terms of data compared to age at menarche IV and can easily be applied to other countries which has minimum marriage age laws and seasonality of marriage dates. Finally, the results from the paper on intermediary variables such as education, spousal quality and bargaining power, concretizes the findings from the previous causal literature and further justifies their use of age at menarche as an IV.

The rest of the paper is structured as follows, section 2 formally introduces the instrument, section 3 discusses the data, section 4 discusses the empirical strategy, section 5 presents the results, section 6 compares my IV with the age at menarche IV, section 7 presents robustness checks and finally section 8 is conclusion.

1.2   Instrumental Variable

The instrumental variable strategy stems from two social phenomenon within the Hindu population in India, preference to get their daughter married off early and after turning a particular minimum age, and preference to get married on auspicious dates. Parents in India prefer to get their daughter married off at an early age but how early they can get their daughter married off depends on laws and social norms. I call the preference of getting your daughter married off early and after they turn a minimum marriage eligible age "minimum age targeting at marriage". The girls who turn these minimum marriage eligible age in a month of low auspicious marriage dates in expectation has to wait longer to get married compared to girls who turn the minimum marriage eligible age in a month of high auspicious marriage dates. For example imagine a girl born in January (an unpopular month for getting married), 1980. If there were no preference for getting married on auspicious marriage dates and no minimum age targeting (parents do not have a preference to get their daughter married off early and after turning a particular minimum age), then she would get married in any month/year. In presence of only minimum age targeting (say at the legal age of 18) she would get married in January, 1998. If there is preference for getting married on auspicious marriage month and minimum age targeting then she would likely get married in June (popular month for marriage), 1998 at the age of 18 years and 6 months. Compare this to a girl born in June, 1980. In absence of minimum age targeting and preference for getting married on auspicious marriage dates she could get married whenever. In presence of minimum marriage age targeting (again at the legal age of 18) she would get married in June, 1988. In presence of age targeting and preference for getting married on auspicious dates she would still get married in June,1980 at the age of 18. Hence the girl who was born in January in expectation has to wait longer to get married. If being born in a popular versus unpopular marriage month is random then we get an exogenous variation for age at marriage.

Before constructing the IV we look at the practice of minimum age targeting and preference to get married on auspicious dates. This paper is the first to introduce the phenomenon of minimum age targeting in the literature. I present suggestive empirical evidence to establish it. The second

phenomenon comes from Hindu religious beliefs and astrology where people prefer to get married on certain auspicious dates.

1.2.1   Minimum age targeting at marriage

Parents in Indian tend to marry off their daughters at an early age but how early the daughter is married off depends on various social, cultural and economic reasons. One candidate for the minimum age which parents can target their daughters to get married after is the minimum legal marriage age of eighteen. Graph 3 plots the age at marriage in months. We see a spike at 216 months (Exactly 18 years) i.e. there is a jump in the number of marriages at the exact age of 18 compared to getting married at 17 years 11 months or 18 year 1 month. This is suggestive that parents are getting their daughters married off as soon as they turn eighteen. In India 41 percent of women get married before turning the age of eighteen. Thus for a majority of the population the legal minimum age constrain is not binding. When we look at graph 3 we see spikes similar to 216 months at 180, 192, 204, 228 and 240 months. I cite this as evidence for parents targeting ages apart from the minimum legal age as a minimum age for their daughters to get married after. I call this phenomenon minimum age targeting at marriage. This phenomenon is comprised of two social preference i) Parents want their daughter to get married early ii) Parents want their daughter to get married after turning a minimum marriage eligible[3] age. In the existing literature there is plenty of evidence of the first preference. In this section I will present some evidence that both of them holds together.

Let us define the variable Target = mod (age at marriage in months, 12). Graph 4 shows the density of the variable target for our population. The maximum frequency for target is at the value zero implying maximum number of girls are getting married in the same month as their birth month i.e. as soon as they turn a particular minimum marriage eligible whole number age. Then we see a monotonic decrease in density as the variable target increases, reaching a minimum at six[4],

---

[3]The age itself could be driven by social and cultural norm, in certain communities it might be frowned upon to get your daughter married of before turning sixteen

[4]This could be because of both targeting and the seasonality of birth and marriage months, in

8

implying least number of girls are waiting six months to get married after turning a particular whole number age. We then see the variable monotonically rising towards zero. In Graph 5 I plot the average number of marriages by the absolute distance in months from a whole number age. We see that as the distance from the whole number ages decreases the average number of marriage decreases.

In Appendix Section I, I do a back of the envelope calculation for the variable target. Given the population distribution of marriages by month and the population distribution of births by month I calculate the distribution of the variable target, under the assumption that month of birth is not correlated to the month of marriage i.e. no age targeting. Graph 6 plots the actual (in presence of targeting) and calculated (absence of targeting) distribution of the variable target. We see that at the value zero there is gap between the two distributions implying that if women were not coordinating their month of marriage with their month of birth then less number of women would get married in their birth month. This gap is indicative of minimum age targeting at marriage.

In graphs 7-9 I plot the mean values of household characteristics by the distance from a whole number age. We see that women who gets married closer to a whole number age are on average less educated, gets married into families belonging to rural areas, and gets married into families with lower wealth. Thus families which practice minimum age targeting usually belongs to the lower socio-economic strata of the society. In graphs 10-11 I plot the distribution of the variable target by wealth and education. We see that the women who gets married into the top wealth quintile families do not practice minimum age targeting at marriage whereas women who gets married into the bottom quintile practice age targeting at marriage. We see similar trends for education and other women and family characteristics. This gives us an idea about the compliers for the IV. The variables looked at in the above graphs are outcomes rather than covariates as they are determined after marriage. In graph 14 in I plot the age at menarche[5] by the distance from a whole number age

appendix Section I do a back of the envelope calculation for the variable target there also we see a similar shape for the variable target at 6.

[5]For women who get married after menarche.

we see that the average age at menarche does not vary by the variable target[6] , this is indicative of the fact that targeting affects women outcomes through marriage.[7]

1.2.2 Auspicious Marriage Dates

Hinduism is one of the most prominent religion in the world, it is the most dominant religion in India. Indian people belonging to the Hindu religion have a deep rooted believe in astrology. They depend on astrology to make many life decisions. So when it comes to setting a marriage date Hindus normally consult a horoscope interpreted by the pundits (Narlikar 2013, Lal 2015). Now when it comes to marriage dates certain months of the year are considered more auspicious than others this results in the auspicious months getting more marriage dates which in turn translates into more marriages occurring in certain months of the year than others. Graph 2 shows the distribution of number of marriages by month. Graphs 12 and 13 show the distribution for other religions and over years. We see that the distributions looks different by religion but similar over the years. Graph 18 plots the distribution of month of marriage by month of birth, we see that the distribution looks similar across month of birth.

The distribution of marriage months might not solely be driven by auspicious dates it could also be due to other reason such as overlap with the festive season which is usually post-harvest, climate, availability of wedding services etc. Thus in equilibrium some we see some periods of the year are more favorable for getting married compared to other hence we see a seasonality in marriage.

To understand who gets married on these auspicious dates I plot the mean values of household and woman characteristics by month of marriage, graph 15-17 presents these. We see that similar to the population that practice age targeting people who gets married in the auspicious months are on average less educated, gets married into families belonging to rural areas, and gets married into

---

[6]Menarche is an outcome realized by the women prior to their marriage and hence should not vary by whether their families practice minimum age targeting at marriage.

[7]Ideally I would like to show more pre-marriage outcomes but unfortunately the data doesn't allow that.

families with lower wealth. Since the IV is relevant to the population that practice age targeting and prefer getting married on auspicious dates this gives us an idea about the compliers for the IV.

1.2.3 Constructing the IV

To get an exogenous variation for age at marriage I make use of the practice of minimum age targeting at marriage and seasonality of marriage dates. From the data we see that a lot of women are getting married as soon as they turn a particular whole number age. Now if the women is born in a period of low number of auspicious marriage dates then she turns the minimum marriage eligible whole number age for getting married in that period of the year which has low number of auspicious marriage dates. So in expectation she has to wait longer to find a suitable auspicious date to get married. Whereas if the individual is born in a period of high number of auspicious marriage dates then the individual turns the minimum marriage age in that period of the year which has high number of marriage dates. Then she has to wait less to find a suitable auspicious marriage date. Since the distribution of auspicious dates comes from astrology, I assume that is has no real effect on marriage outcomes. Thus if being born in a low marriage frequency period versus high marriage frequency period is random[8] then in presence of the two social phenomenon minimum age targeting at marriage and the desire to get married at an auspicious date; we get an exogenous shock to age of marriage. Where women born in low marriage frequency period waits longer in expectation to get married.

I construct two IVs, IV1 and IV2. To capture the aforementioned effect, the IVs should not only take into account whether the woman was born in a month of low marriage frequency or not, it should also take into account whether the period following her birth month has a high number of marriages or not. For example the month of January and July have similar amount of marriages but January is followed by months which have high marriage frequency and July is followed by months with low marriage frequency. Hence someone born in January might have to wait less in

---

[8]I present evidence for this in section 4.2.

expectation[9] To capture this I create a relative score for each birth months (IV2). The score should capture two things, firstly months which have high number of marriage dates should have higher relative score and secondly the months which are followed by months with high number of marriage dates should also have higher relative score. I capture these two factors using the following equation

IV2=Relative Frequency of Birth Month + Relative Frequency of the Period after Birth Month

$$IV2_i t = M_i t + (\beta^1 m_1 it + \beta^2 m_2 it + \beta^3 m_3 it + ...) \tag{1.1}$$

Let t be the whole number age at which the individual i gets married then $M_i t$ is the relative frequency of marriages of the month she was born in, this term captures the first factor that months with high marriage number of marriage dates have higher relative score. Let $m_1 it$ be the relative frequency of the month after her birth month, $m_2 it$ is the relative frequency of the month two months after birth month and so on. Beta is a discounting factor with a value less than one. The power of beta increases as the distance from the birth month increases implying lower wait on months which are further away. The term $(\beta^1 m_1 it + \beta^2 m_2 it + \beta^3 m_3 it + ...)$ captures the second factor that months immediately followed by period of high number of marriages have a higher relative score. Thus IV2 gives us a relative score for each birth month which captures how favorable the month is for getting married. In Appendix Section II I give an example of IV2 for various values of beta.

I also construct a binary IV, IV1. IV1 is a binary variable which is equal to 1 if you are born in a month which belongs to a period where there is low number of marriages and 0 otherwise. IV1 takes the value 1 for the months (7,8,9,10) and 0 for the rest. These months were chosen as they have the lowest relative frequency according to the ranking generated by the relative score computed in equation 1[10].

---

[9]This rules out using some common instruments such as quarter of birth or quartile of birth by marriage frequency.

[10]See Appendix section II for example

1.3 Data

The data used for the analysis comes from the 2015-16 National Family Health Survey (NFHS-4). This is the fourth data set in the NFHS series[11]. The data set provides information on population, health and nutrition for India and each state and union territory. The survey has four different questionnaire – woman's, man's, biomarker and household. All women age 15-49 and men age 15-54 in the selected sample of households were eligible for individual interviews. The household questionnaire covers basic information on all household members along with information on socioeconomic, sanitation, water, health and deaths in past three years. For the women's interview two sets of questionnaire were used. The first version (district module) collected information on women's basic characteristics, marriage, fertility, reproductive health, children characteristics. This first version was collected on all women in the household survey. The second version (state module) was collected for sub sample of the households. It has additional four topics on sexual behavior, HIV/AIDS, husband's background and woman's work, and domestic violence. The man's questionnaire covers information on man's characteristics, life style, sexual behavior, marriage, fertility, nutrition, attitude towards gender roles, contraception and HIV/AIDS. The biomarker questionnaire covered information on height, weight and hemoglobin levels for children; height, weight, hemoglobin levels, blood pressure and random glucose for women and men.

For my analyses I restrict the sample to Hindu women who gets married before the age of 22, this is 87 percent of the population. The minimum age targeting at marriage and auspicious dates phenomenon is most relevant for this group, I call this the "full sample"[12]. I also analyze the results on a sub-sample of women with zero years of education. I call this the "zero education sample". The purpose is mainly to understand that if we block the channel of education gains by delayed marriage, do we still see an effect[13]. The Age at Menarche question was asked to a sub sample

---

[11]I use this round as it is the latest round in the series with higher number of observation for my population

[12]For women getting married after the age of 21 it is unlikely that her parents were waiting for some minimum marriage eligible age.

[13]More on this in Section 3.3

of the women. To compare the results from my IV and Age at Menarche IV I use the sample of women who were asked about their age at menarche. I call this the "Menarche Sample". In all the samples I drop the women for whom the information on month/year of birth or month/year of marriage was imputed (18 percent of the population was dropped).

Table 1 provides the descriptive statics for all three samples. The uneducated sample as expected performs worse than the full sample on dimensions of wealth, education, family size among others. The menarche sample is much younger than the full sample, this is because the age at menarche question was asked to women age less than 25.

The main outcomes includes measures of health and fertility. The intermediary outcomes includes measures of educational attainment, spousal quality, spousal family quality and bargaining power.

1. Health: It includes measures of diabetes, blood pressure and anemia. For diabetes I look at two outcomes. a) A binary variable indicating blood glucose level greater than 200mg/dl, when the blood sample is collected randomly. This indicates whether the person is diabetic. b) A binary variable indicating blood glucose level greater than 100 mg/dl, when the blood sample is collected at least after 8 hours of fasting. This indicates whether the person is pre diabetic. For high blood pressure I look at two outcomes. c) A binary variable indicating systolic blood pressure level greater than 120. This indicates high blood pressure. d) A binary variable indicating diastolic blood pressure level greater than 80. This indicates high blood pressure. For anemia I look at two outcomes. e) A binary variable indicating a hemoglobin level (g/dl) less than 6.9, indicating severe anemia. b) A binary variable indicating a hemoglobin level (g/dl) less than 9.9, indicating moderate or severe anemia.

2. Fertility: It includes four measures a) Number of children. b) Age at first birth. c) Desired/Ideal number of children reported by the ever married woman d) Birth Spacing[14].

3. Women's Educational attainment: It includes two measures of women's education a) highest grade attained by the woman. b) A binary indicator for whether the woman has at least primary

---

[14]Number of months between the birth of first and second child.

level of education.

4. Husband and Husband's Family Characteristics: [15] : It includes four measures a) Highest grade attained by husband. b) Husband's age in years. c) An index variable indicating wealth of husband's family. One being the poorest and five being the richest. d) A binary variable indicating whether the family belongs to a rural or urban area.

5. Proxies for Bargaining Power: These are reported by an ever married women to whom the domestic violence questionnaire was given. The questionnaire included questions to access control issues within marriage. There are several control issue questions in the data, I report the results from a few. a) A binary variable indicating whether the husband limits the woman from contacting natal family b) A binary variable indicating whether the husband insists on know where she is. c) Number of control issue questions answered yes.[16]

1.4 Identification Strategy

I want to find the causal effect of age at marriage for a woman on her health and fertility. The main empirical challenge is that age at marriage might be endogenous due to omitted variables. Age at marriage might be correlated with unobserved family and woman characteristics which will bias the results from simple correlational studies. To overcome this challenge I use an Instrumental variable approach. I use functions of month of birth as an IV for age at marriage. The instrumental variable approach is detailed in section 2. The exogenity assumption of this IV approach is that women being born in low marriage frequency months are not systematically different from women who are born in high marriage frequency months.

---

[15]Ideally I should at other husband outcomes such as husband's income and occupation. The NFHS-4 data has a men's questionnaire but it does not report income. It reports occupation but the way it is reported makes it difficult to create a relative ranking over the various occupations reported, hence I do not include them

[16]The higher number of control issue questions answered yes indicates lower bargaining power of the women in the marriage.

1.4.1 Instrumental Variable Approach

I consider the following regression model where Y_i denotes outcome for women i.

$$Y_i = \beta Age of Marriage_i + \gamma WomenControls_i + \epsilon_i \tag{1.2}$$

Women controls include age and month of marriage FE[17]. I report robust standard errors.

Now the major concern with estimating the above equation is that age of marriage might be endogenous. To overcome this I use IV1 and IV2 as my instrument for age at marriage. The motivation for the IVs are discussed in section 2.

1.4.2 Validity of the Instrument

1.4.2.1 First Stage

First I examine whether the IVs predict the age at marriage. Table 2 reports the first stage estimates. I control for women's age and include month of marriage FE. We see that being born in a low frequency month delays age at marriage by 3 months in the full sample and by 3.4 months in the zero education sample. One unit increase in IV2 delays age at marriage by 0.15 months in the full sample and 0.16 months in the zero education sample. All the estimates are statistically and economically significant.

The first stage results using both the IV's are highly significant, which deals away with the weak instrument issue. Stock and Yogo (2002) formalized the definition of "weak instruments" and it suggested that a first stage fstat of greater than 10 is sufficient to nullify concerns about weak IV bias. Stock and Yogo (2005) go into more detail and provide useful rules of thumb regarding the weakness of instruments based on a statistic due to Cragg and Donald (1993). To test for this I use the Weak IV test in Stata which reports the Cragg-Donald Wald F statistic and reports the critical values complied by Stock and Yogo (2005). Appendix 1 presents the results. We see a very high

---

[17]I include month of marriage FE to account for the possibility that women getting married in different months might vary systematically

fstat for IV1 and IV2 which rejects the null hypothesis that instruments are weak at the highest critical value.

1.4.2.2 Exogenity of the Instrument

The instruments uses being born in a low marriage frequency month versus being born in a high marriage frequency month as the primary source of exogenous variation for age at marriage. There are two separate but related threats to the exogenity of the instruments. First if month of birth is not randomly assigned between the low and high marriage frequency groups. This is problematic if for example wealthy girls are born in the low frequency months and wealthy girls marry wealthy boys (assortative matching). Then it might appear that girls born in low frequency marriage months marry later and have better spousal quality, but this is not causal. To tackle this issue I look at a population of unmarried girl child less than age 10[18] and see whether the IV is predictive of their background mother and family characteristics. For mother characteristics I look at mother's education, age at marriage, age at first birth and number of children. For family characteristics I look at a measure of wealth, whether the household belongs to a rural versus urban area, household size and whether the household has electricity.

The second threat to the validity is if being born in low marriage frequency months versus high marriage frequency months affects outcomes for the girls through channels other than age at marriage, e.g if girl child born in low marriage frequency months are less subjected to early childhood diseases and are therefore healthier independent of age at marriage. Then the estimates we get from this causal analysis would be biased. To refute this I show that the IV is not be predictive of early childhood health outcomes of the girls. I look at a population of girl child between ages 0-4, for this population the NFHS-4 data set provide measures of natal health such

---

[18]The reason I chose this population is because in the data we have mother and family characteristics only for girls whose mothers are sampled, i.e who live with their natal family which only happens if they are unmarried and at older ages (e.g 14,15,16) the sample of unmarried girls living with their natal families becomes more and more selected (because girls start getting married and move out).

as size at birth; weight at birth and one year mortality. I see whether being born in low versus high marriage frequency month is predictive of these early childhood health outcomes.

To show these I run an OLS regression model (Equation 3) without including any controls. The purpose is to see whether the IV is correlated with early childhood health and family characteristics. The main coefficient of interest is Beta which captures the effect of the IVs on background characteristics and natal health.

$$Y_i = \beta IV_i + \epsilon_i \tag{1.3}$$

Table 3 panel C presents the results for the effect on early childhood health outcomes. The outcomes I look at are i) size at birth this is a categorical variable[19] reported by the mother with 1 being the largest and 5 being the smallest ii) weight at birth in pounds iii) one year mortality. Being born in a low frequency marriage months is associated with a 0.01 points (0.3 percent) increase in log odds for being at a higher birth size, 0.001 pounds (0.03) percent decreases in child's weight at birth and 0.0003 percentage points (0.6 percent) increase in the probability of one year mortality. None the effects are statistically or economically significant.

Table 3 panel A presents the results on mother characteristics. Being born in a low marriage frequency month is associated with 0.008 (0.2 percent) increase in mother's education, decreases mother's age at marriage by 0.097 years (0.5 percent), decreases mother's age at first birth by 0.002 years (0.01 percent) and increases total number of children by 0.02 (0.05 percent). None of the results except mother's age at marriage is statistically or economically significant. From previous literature (Chari et al, 2017); we see that a 0.097 years decrease in mother's age at marriage is causally associated with 0.01 years (0.19 percent) decrease in number of years of education for the child, 0.25 percentage points (0.31 percent) decrease in probability of being enrolled and 0.3 percent decrease in test scores, these effects are small in magnitude and not economically significant. The mechanism is that women who are born in low marriage frequency months gets married later and

---

[19]Here I run an ordered logit regression without any controls

have better health and fertility outcomes. The fact that women who are born in low frequency months have a lower mother age might slightly downward bias the benefits we see from delayed marriage.

Table 3 panel B presents the results for the effect on household characteristics. Being born in low frequency month increases household members by 0.008 (0.1 percent), increases the probability of having electricity by 0.3 percentage points (0.4 percent), decreases the probability of being from a rural area by 0.5 percentage points (0.8 percent) and increases wealth by 0.03(0.1 percent). None of these effects except the effect on wealth is statistically or economically significant. The effect on wealth is very small and the direction of the effect is opposite to what I found for mother's age at marriage. Overall I conclude that being born in low frequency month does not systematically affect the woman's early childhood health, mother and family characteristics.

1.4.2.3 Mechanisms

Conceptually the effect of age at marriage on fertility and health should occur in steps. In the first step age at marriage influences decisions regarding intermediary variables such as educational attainment, spousal quality, spousal family quality. In the next step each of these decisions affects more tertiary outcomes such as bargaining power of the women in the household and finally all these channel combined affects her fertility decision and health. Note that these steps might not always be sequential, for example bargaining power of the women can be directly affected by the age at which she enters the household at the same time the bargaining power of the women can be influenced by her health and fertility outcomes, hence it's hard to judge whether bargaining power is an intermediary variable for final outcomes such as health and fertility. Broadly categorizing, three potential channels through which age at marriage might affect health and fertility are educational attainment, spouse and spousal quality and bargaining power. I first establish the overall causal effect of age at marriage on health and fertility. Then I try to disentangle the effects of the potential channels. First I control for spouse and spousal quality then the effect we see are thorough the channels of education and bargaining power. Next I control for bargaining power along with spouse

and spousal family characteristics. Finally following Chari et al, (2017) and Field and Ambrus, (2008) I restrict the sample to women with zero years of education and include the controls for bargaining power and spousal quality. In this sample the effect we see should be driven by factors other than the above three channel. Since education itself is an outcome variable, restricting the sample on it makes the results biased[20]. Also the zero education population itself might be very different form the rest of the sample and hence the estimates obtained on them is not reflective of the true population effect. The purpose of restricting the sample to zero education is to show that age at marriage affects health and fertility through channels other than gains in education.

1.5 Results

To understand the effect of age at marriage of a women on her health and fertility outcomes I start by analyzing the full sample. First I look at the effect of age at marriage on her health outcomes such as diabetes, blood pressure and anemia. Then I look at the effect on her fertility outcomes such as total number of children, age at first birth and birth spacing. These results are estimated using regression equation 2, I include minimal amount of controls for the above set of regressions such as age at survey and month of marriage FE. Next I look at the effect of age at marriage on more intermediary variables such as education, spousal and spousal family characteristics and bargaining power. Here I control for women[21] and family characteristics.[22] For the above set of outcomes I present the results for both the instruments, IV1 and IV2[23]. I also present the OLS results, though the discussions would mainly focus on the estimates using IV1. In appendix table 4 I present the Wald IV estimates for some of the outcomes Finally I present the results on fertility and health under

[20]Yes years of education is affected by age at marriage but having zero versus some year of education is not strongly affected by age at marriage. The decision to give the female child some education versus no education is taken at a much earlier age compared to the decision of getting them married. When I do a simple OLS controlling for household char we see a very weak correlation (0.03 percent) between age at marriage and having zero years of education versus some.

[21]Women characteristics includes age, and year of birth FE.

[22]Family characteristics includes number of household members, whether the family lives in an urban versus rural area, age and sex of household head, whether the household has access to electricity and wealth.

[23]With 0.7 as the value of Beta.

four specifications; specification 1 presents the result for the full sample with minimal controls[24], specification 2 presents the results including spousal[25] and spousal family controls[26], specification 3 presents the results including spousal, spousal family and proxies for bargaining power controls and specification 4 presents the results including spousal and spousal family controls for the zero education sample. This will help us understand the channels through which age at marriage for a women affects her health and fertility.

1.5.1 Health outcomes

Table 4 reports these results. One month delay in marriage decreases the probability of a woman being diabetic by 0.07 percentage points (2.8 percent) and decreases the probability of being pre-diabetic by 1.3 percentage points (3.5 percent). One month delay in marriage decreases the probability of a woman having high blood pressure under the diastolic measure by 0.27 percentage points (0.8 percent) and decreases the probability by 0.18 percentage points (0.6 percent) under the systolic measure. These results are highly statistically and economically significant. The magnitude of some of these results are small but we need to keep in mind that these are the effects of one month delay in marriage and if we linearly[27] transform the effects into one year delay then the effects are pretty substantial. Next I look at anemia, one month delay in marriage increases the probability of being severely anemic by 0.01 percentage points (0.1 percent) and increases the probability of being moderately anemic by 0.05 percentage points (0.3 percent). The results for anemia are not statistically or economically significant. Overall we see that as age at marriage increases women have better health outcomes.

---

[24]This is what I already report.

[25]Husband's age and education.

[26]Family characteristics includes number of household members, whether the family lives in an urban versus rural area, age and sex of household head, whether the household has access to electricity and wealth.

[27]This might not be completely accurate but it helps us to understand the magnitude of these effects

1.5.2 Fertility Outcomes

Table 5 reports these results. One month delay in marriage decreases the total number of children by 0.0005 (0.02 percent). The effect is very small and statistically not significant. When we look at the ideal number of children the woman wants to have, we see that one month delay in marriage decreases the ideal number of children by 0.002 (0.09 percent). Again the effect is very small and statistically not significant. One month delay in marriage increases the age at first birth by 0.06 months (0.3 percent) and decreases birth gap by 0.019 months (0.05 percent). The effect on age at first birth is highly statistically significant whereas the effect on birth gap[28] is not.

1.5.3 Intermediary Outcomes

Table 7 Panel A reports these results. Now we look at the effect of age at marriage on some of the intermediary outcomes such as women's education, spouse and spousal family quality and proxies for bargaining power within the marriage. One month delay in marriage increases the number of years of education of the woman by 0.03 years (0.5 percent) and the effect is statistically significant. One month delay in marriage increases the probability of the women having at least primary level of education by 0.03 percentage points (0.04 percent) and the effect is statistically not significant.

Table 6 reports these results. One month delay in marriage increases husband's years of education by 0.03 years (0.4 percent), decreases husbands age by 0.024 years (0.05 percent), increases the wealth of the in laws household by 0.02 (0.07 percent) and decreases the probability that the in laws household belongs to a rural area by 0.7 percentage points (0.9 percent)[29]. Overall we see that delayed marriage age improves spousal and spousal family quality. One thing to think about here is the timing of the spousal match. Usually first, the spousal match occurs and then the search for the marriage date begins. Now being born in a low marriage frequency month delays marriage age through the search for the marriage date process. In that scenario the spousal quality

---

[28]In appendix table X we see that one month delay in marriage delays the age at first and second birth by the same magnitude, hence we do not see an effect on birth gap.

[29]Appendix table 2 presents some more results on in laws family characteristics

should not be affected. One possibility is that the delayed marriage gives the husband more time to increases his educational attainment which in turn increases the family's wealth. This can explain why we see a gain in spousal education and family characteristics but it doesn't explain the decrease in husband's age. The decrease in husband's age is small (0.05 percent) but still significant. The only other possible explanation is that, since women born in low marriage frequency months have a longer wait time between spousal match and the marriage there might be a higher chance for the match breaking before marriage and if bad matches (higher age differences) break more frequently, then that might explain why we see an effect on spousal age.

Table 7 Panel B reports these results. One month delay in marriage decreases the probability of the husband limiting the women to contact her natal family by 0.6 percentage points (3.6 percent), decreases the probability of the husband insist of knowing her whereabouts by 0.7 percentage points (3.3 percent) and decreases the number of control questions answered yes by 0.01 (0.2 percent). The effects are statistically significant.

1.5.4 Discussion and Mechanism

1.5.4.1 Health

I find that delaying age at marriage decreases the probability of a woman having diabetes and high blood pressure. Though partially genetic, high blood pressure and diabetes can also be driven by lifestyle choices such as lack of exercise, unhealthy meals and obesity. Lifestyle choices can very well be influenced by the intermediary variables we looked at such as education, spousal quality and bargaining power in the marriage. Hence the health gains we saw could be partially driven by the gains in the intermediary variables we looked at. To systematically unpack the channels I look at three different specifications, table 8 panel B presents the results on health controlling for spousal and spousal family characteristics, panel C presents the results including spousal, spousal family and bargaining controls and panel D presents the results including spousal and spousal family controls for the zero education population.

We see that adding spousal controls (panel B) does not change the estimates for the health outcomes, this rules out the theory that the gains in health is completely driven by the gains in spousal quality. Next I add bargaining controls (panel C), when I control for bargaining power I see that estimates are more negative implying higher gains in health outcomes. One thing to note here is that the bargaining questions are at best, proxies to the actual bargaining dynamics. Also these questions were asked to a sub-population of the sample, on further inspection I find that the higher gains are not driven by the controls itself rather it is driven by the chosen sub population. That is if I run specification 1 on the sub population to which the bargaining question were asked, I get a bigger negative effect. Hence I fail to conclusively state that whether the gains in health are driven by a higher bargaining power within the marriage. Finally in specification 4, which looks at the zero education sample, I find that the gains are even larger this rules out the theory that the gains in health are completely driven by gains in education. To conclude we see that delaying marriage causally results in better health outcomes for the women and the gains are not completely driven by gains in education or spousal quality.

1.5.4.2 Fertility

I unpack the effect of delaying age at marriage on fertility in table 9. Previously we saw that (section 5.2) delaying marriage age had no effect (0.02 percent decrease) on the total number of children but the effect was statistically not significant. One month delay in marriage increases the number of children by 0.009 (0.3 percent) in specification 2, by 0.008 (0.3 percent) in specification 3 and by 0.016 (0.5 percent) in specification 4. The effects are statistically significant but small in magnitude. Conceptually, delayed age at marriage can have opposing effects on the number of children the women has. Getting married later gives the woman less time to bear children whereas getting married early and having children early possess health risks to the woman and child. When I control for spousal quality in specification 2 and block the education channel in specification 4 I see a positive effect on fertility compared to the zero effect in specification 1. This might be driven by the fact that higher education among women is correlated with lower number of children and

24

higher spousal education and wealth can lead to lower fertility through higher use of contraception and lower son preference. These could also be the reason why we see a statistically significant positive effect on ideal number of children under specification 2 and 4 but a zero effect under specification 1. For age at first birth we see that adding spousal and bargaining controls do not change the estimate.

1.6 Comparison with the Menarche IV

The previous causal literature which looks at the effect of age at marriage uses age at menarche as an instrumental variable (Field and Ambrus 2008, Sekhri and Debnath 2014, Chari et al 2017) for age at marriage. In this section I presents the results comparing the two IVs using the menarche sample. The menarche sample is a sub sample of the full sample. I obtain the results from both IV on the menarche sample so that we can draw a direct comparison. The menarche sample is quite different from the full sample, table 1 provides the descriptive statistics for both the sample. The question about age at menarche was asked to ever married women between the age of 15-25 whereas the full sample consists of ever married women between the ages 15-49, hence the menarche sample on average is much younger than the full sample. The menarche sample is also on average more educated, has lower number of children, belongs to larger household and have lower wealth compared to the full sample.

1.6.1 First Stage

Table 10 reports the first stage estimates for both the instrumental variables.[30] Both the instruments are strong predictors of age at marriage. A one year delay in age at menarche delays age at marriage by 2 months whereas for IV1 the first stage estimate is 2.3 months, both the estimates are highly significant. For IV1 we saw a first stage estimate of 3 months for the full sample, the difference is attributed to the aforementioned difference in sample. For age at menarche, previous literature

---

[30]To draw direct comparison with the previous literature I control for age, height and include caste and district FE, following Chari et al, (2017) and Field and Ambrus, (2008)

reports a first stage of around 4.8 months (Chari et al, 2017), 4 months (Sekhri and Debnath, 2014) and 8.8 months (Field and Ambrus, 2008). The difference from the previous literature could be attributed to the fact that the menarche sample on average is much younger and consists of anewer cohort. Both Chari et al, (2017) and Sekhri and Debnath (2014) uses the IHDS 2005 data set from India and report an average age of 33, Field and Ambrus, 2008 uses the 1996 MHSS data set form Bangladesh and report an average age 33. A more recent data set combined with lower average age implies a more recent cohort. The lower magnitude of the first stage implies that over time the predictive power of age at menarche for age at marriage decreases.

1.6.2 Intermediary Variables

Before we compare the final outcomes on health and fertility I look at the intermediary variables such as education, spousal and spousal family quality. My prior was that both the IVs should produce similar estimates for these outcomes. Table 11 reports the results, we see that one month delay in marriage increases years of education by 0.14 months using IV1 and 0.13 months using age at menarche, the effects are statistically significant. The full sample estimate is 0.03 months. Field and Ambrus, 2008 found that delaying marriage by one year increases education by 0.22 years, this is comparable to the baseline results converted to years (0.36 months). The menarche sample estimates are higher when compared to the full sample and previous literature. Menarche sample looks at a more recent cohort compared to the other samples and if the benefits of delaying marriage increases with time then that might explain why we see a larger effect in the menarche sample. One month delay in marriage increases the probability of having at least primary education by 0.5 (0.6 percent) percentage points using IV1 and 0.6 (0.6 percent) percentage points using age at menarche. Again the effect for the menarche sample is bigger than the full sample and statistically significant.

Now we look at spousal and spousal family quality. One month delay in marriage increases husbands education by 0.06 years using IV1 and 0.07 years using age at menarche, the results are statistically significant. Again the estimates are higher than the full sample (0.03). Field and Ambrus, 2008 also looked at husband's education and found that one year delay in marriage

increases husbands age by 0.04 years but the effect is not statistically significant. We see a pattern similar to women's education here, where the gain in husband's education are higher for the more recent cohorts. When I look at husband's age I see that one month delay in marriage decreases husbands age by 0.01 years using IV1 and 0.1 years using age at menarche. The full sample estimate is a decrease of 0.02 years. The results using IV1 are not statistically significant hence I cannot draw a conclusion about why we see a bigger effect for the menarche IV. For spousal household quality we see a similar effect from both the IVs and the full sample. Overall we see my prior that both the IVs should produce similar results for the intermediary variable, holds.

1.6.3   Health

Since early or late onset of menarche is correlated with underlying health factors my prior was that my IV and age at menarche IV will produce different estimates. Table 12 reports the results, I see that delaying age at marriage by one month decreases the probability of being diabetic by 0.01 percentage points (0.6 percent) using IV1 and 0.03 percentage points (1.8 percent) using age at menarche IV. The full sample estimate showed a decrease of 2.8 percent. One month delay in marriage decreases the probability of having high blood pressure by 0.2 percentage points (0.8 percent) using IV1 and 0.7 percentage points (2.8 percent) using age at menarche IV. The full sample estimate is a decrease of 0.8 percent. For anemia we see that one month delay in marriage has no effect on the probability of being severely anemic using IV1 and 0.05 percentage points (5 percent) increase in probability using age at menarche IV. I didn't find any significant effect from the full sample. Over all we see that the health estimates are not very different across the two IVs. The estimates from the menarche IV are slightly higher but overall has the same direction as IV1. A possible explanation for why we see similar results across the IVs could be that the health factors which are influenced by early or late onset of menarche might not affect health outcomes such as diabetes, blood pressure and anemia. These health outcomes are more driven by genetics and lifestyle choices whereas late or early onset of menarche affects the women's reproductive health.

1.6.4 Fertility

Table 12 reports these results. The first divergence in the results from the two IV is when I look at total number of children. One month delay in marriage decreases the number of children by 0.0001 (0.005 percent) using IV1, the effect is not statistically significant. For the full sample we see a 0.02 percent decrease, not statistically significant. When I include spousal controls I find a small (0.4 percent), positive and statistically significant effect. In summation IV1 estimates a zero or small positive effect on the number of children. Whereas age at menarche IV estimates a negative, statistically significant effect on the number of children. Using age at menarche IV, one month delay in marriage decreases the number of children by 0.017 (1 percent) decrease, and the effect is statistically significant. Using age at menarche as an IV, Chari et al, (2017) find that one year delay in marriage decreases the number of children by 0.13 (4.64 percent). Hence both the menarche sample and previous literature estimates a negative, statistically and economically significant effect of age at marriage on number of children, whereas my IV1 predicts a zero or small positive effect.

One reason for the difference in results could be that the set of compliers for both the IVs are different. In the menarche paper the compliers are women who delay marriage because of delay in menarche whereas in my paper the compliers are women who delay marriage to get married at an auspicious marriage month. The other possible reason is that age at menarche is not exogenous when we look at the effect of age at marriage on fertility. In the medical literature we see that both early and late onset of menarche is often associated with adverse pregnancy outcomes such as ectopic pregnancies and spontaneous abortions (Liestol, 1980; Martin et al., 1983; Wyshak, 1983; Sandler et al., 1984). Later onset of menarche is also associated with a slightly higher risk of sub fecundity and infertility (Gulbrandsen et al, 2014). Age of menarche has also been associated with other health issues for women such as breast cancer and cardiovascular diseases (Petridou et al., 1996; Garland et al., 1998). Since age at menarche is associated with reproductive health using it as an IV to measure total number of children might bias the results. Women with late onset of menarche might have lower reproductive health and delayed marriages. In this scenario the lower number of children are driven by both lower reproductive health and delayed age at marriage, hence

the estimates could be biased.

In the data we have a few measure of reproductive health such as irregularities in period and pregnancy termination. I use these variable to see whether age at menarche directly affect these variables even after controlling for age at marriage. I use an OLS regression model with age fixed effects. I ran the model on the Menarche sample excluding woman who are currently pregnant. I look at two variables, the first variable is a binary variable indicating whether the woman had periods in the last six weeks. I view this as an indicator for period irregularities. The second variable I look at is pregnancy termination. Appendix table 3 reports the results, I see that even after controlling for age at marriage, women's education, height, age, spousal and spousal family characteristics age at menarche has a statistically significant effect on period irregularities. As age at menarche increase irregularities decreases. This is again supportive evidence that the age at menarche might directly affect reproductive health and hence might not be a valid instrument for this outcome.

When I look at other measures of fertility like ideal number of children and age at first birth I find that both the IV's produce similar estimates and the results are comparable to the baseline estimates and previous literature.

1.7 Robustness Checks

1.7.1 Month of Birth

One non parametric way to capture the essence of the instrumental variable introduced in this paper is to use month of birth dummies as instruments. Note that to use month of birth dummies as an IV we need to make a stronger assumption for exogenity, compared to the current IV. For the current IV the exogenity assumption is that women who are born in months with low number of marriage dates are not systematically different from women who are born in months with high number of marriage dates, this I show in the paper. The exogenity assumption for using month of birth dummies is, women born is different months are not systematically different from each other. From the previous literature (Weber et al., 1998; Lokshin and Radyakin, 2009) we know that this

might not hold. In table 13 panel A, I present few of the results using month of birth dummies as IV. I see that the estimates are similar to the findings of the paper.

1.7.2 Measurement Error

Graph 19 looks at the distribution of birth month. There is an unnatural jump at January. This might be because of miss reporting. When we look at the distribution of birth months of children between ages 0-4 in appendix graph 20 we do not see such jumps. Since their birth is more recent there is less chance of misreporting. Since the IVs are created using month of births this might have an effect if there is non-random bias in reporting. Like if more people from low socioeconomic background do not know their birth months and report as January. To account for this first I drop all the women for whom month of birth or month of marriage were imputed in the sample. This is done for all the results presented in the paper. Next I present some of the results dropping women who reports their month of birth as January. Table 13 panel B reports the results using IV1 for the full sample dropping the women who reported their month of birth as January, I see that the estimates are slightly bigger in magnitude, statistically significant and have the same sign's as the main results.

Another form of misreporting could be if the women does not remember her month of marriage and when asked, only remembers her age at marriage in years then she might report her month of marriage as her month of birth. For example if the women remembers that she got married when she turned eighteen and does not remember the month of marriage then she might just report the month she turned eighteen in, i.e. her birth month. To see check for this I estimate some of the results dropping women who reports the same month of marriage and birth, i.e I drop the women for whom the variable target takes the value zero. I see that the estimates are statistically significant and similar to the main results.

1.7.3   Non Monotonicity of the IV

Monotonicity is a condition that the IV needs to satisfy for the estimates to be interpretable as LATE (Hahn, Todd, and Van der Klaauw (2001)). Unlike the relevance and exogenity condition this condition in rarely tested in the applied literature (Fiorini et al. 2016). There is no formal test of the monotonicity condition and this is still an active area of research. Fiorini et al., (2016) in their paper suggest that the interpretation of the estimates is not completely lost if the monotonicity condition fails; they also suggest that the monotonicity condition should be tested using economic intuitions and suggestive data patterns.

In my paper the monotonicity assumption might fail if being born in a period of low number of marriages encourages some part of the population to get married earlier. The IV stems from two social practices; seasonality of marriage dates and minimum age targeting at marriage. I defined minimum age targeting as a preference where parents want their daughters to get married off early and after turning a marriage eligible whole number age. Another form of targeting could be if the parents want their daughters to get married before turning a particular age (type II). For example if parents want their daughters to get married before they turn 21. Type II targeting in itself would not affect monotonicity. If the parents still want their daughter to get married early then this kind of targeting would not cause bunching at any particular age. Type II targeting combined with the parent's preference to get their daughter married off as late as possible would challenge the monotonicity condition. For example if the parents want their daughter to get married before turning 21 and they want their daughter to get married as late as possible, now if the daughter turns 21 in a period of low marriage dates then the parents might get her married early. Hence for monotonicity to fail two things must hold i) Parents want their daughter to get married as late as possible ii) Parents want their daughters to get married before turning a particular age (type II). I argue that in context of India there is no evidence that parents have preference for getting their daughters married as late as possible hence the monotonicity condition holds. In the next paragraph I will presents suggestive evidence that the above conditions which invalidates monotonicity does not hold.

Suppose both the conditions which invalidates monotonicity holds then when we look at the distribution of target (Graph 4) we should see a spike at 11 similar to the one we see at 0. We do not see any such spikes. When we plot the fraction of marriages by distance from a whole number age (Graph 5) we see similar fraction of people getting married at 1 and -1. This implies that the targeting is towards 0 and not towards -1 which would not have been case if the conditions for non-monotonicity holds.

In this paragraph I provide another suggestive evidence that type II targeting does not exist. When we look at the back of the envelope calculation for the distribution of the variable target (Graph 6) we see that the actual distribution of the variable target diverges from the calculated one at the value 0 and not at 11. Hence the fraction of people who would get married at 11 is the same as if would have happened if month of birth and month of marriage is uncorrelated. This gives further evidence to the fact that monotonicity conditions holds.

1.7.4   Seasonality versus Targeting: Non-Hindu Sample

When we look at the distribution of marriages by month over religion in graph 13 we see that there is seasonality in marriages within some other religions also. If we look at the Muslim and Sikh population we see that they have a seasonality of marriage similar to the Hindu population. As I mentioned before that seasonality in marriage might be driven by reasons apart from religious beliefs such as overlap with the festive season which is usually post-harvest, climate, availability of wedding services etc. When we look at the distribution of the variable target by religion in graph 21 we see that distribution of the variable target looks similar between the Hindu and Muslim population whereas it is not present in the Sikh population. Hence for the Muslim population we see both seasonality in marriage dates and minimum age targeting at marriage, whereas for the Sikh population we see seasonality but not targeting.

In appendix table 5 I present the first stage estimates of IV1 on the Muslim and Sikh population, we see that for the Muslim population the first stage effect is pretty similar to that of the Hindu population whereas for the Sikh population there is no first stage effect. This is suggestive evidence

that the IV strategy presented in this paper is only valid for populations which practice both seasonality in marriage dates and minimum age targeting. Seasonality of marriage dates in absence of minimum age targeting at marriage does not delay age at marriage for women who are born in the low marriage frequency month.

In appendix table 6 I present some of the IV results for the Muslim population, I see that the effects are similar to the main estimates with the Hindu population in terms of direction but the magnitudes of the effects are higher. We also see that on average the Muslim women has higher diabetes, higher blood pressure, more number of children and lower age at birth. This might be a probable reason of why we see higher magnitudes compared to the Hindu population.

Hence this IV can easily be extended to any population which has seasonality in marriage dates and practices age targeting at marriage.

1.8   Conclusion

In this this paper I use a nationally representative data set from India to look at the causal effect of age at marriage of a women on her health and fertility outcomes. To establish this causal relation I propose a new instrument, the instrument uses two social practices. First parent's preference to get their daughter married off early after they turn a minimum marriage eligible age and second seasonality of marriage dates. I find that delayed age at marriage for a woman improves her health post marriage in terms of being diabetic or having elevated blood pressure. Delayed marriage also increases her age at first birth and has a zero effect on number of children. I also find that delayed marriage age also increases her educational attainment, increases her spousal and spousal family quality and increases her bargaining power within the marriage. When I looked at the potential channels through which delayed marriage improves the women's health I find that the benefits are not completely driven by gains in education and spousal quality.

I also compare my findings with the existing instrumental variable in the literature, age at menarche. I find that the results from both the IVs are similar when I looked at health, education and, spousal and spousal family quality. The results differ when I look at the effect on total number

of children.

The instrumental variable strategy used in this paper can be easily extended to other countries which has minimum age laws and seasonality in marriage dates. Since these features are not uncommon in developing countries I plan to extend this study to those countries. Previous causal literature used age at menarche as an IV but overtime as the average age of marriage increases the IV would become less relevant. The IV introduced in this paper can be used to draw causal inferences in those scenarios.

The paper is also first in the literature to talk about minimum age targeting in marriage. It shows how background characteristics vary for people who practice minimum age targeting at marriage. I belief that there is a lot more to explore about this social phenomenon.

The findings from the paper supports the existing inference in the literature that increasing age at marriage has many benefits for women in developing countries. The paper provides support to any policy which aims at delaying the age at marriage for girl child.

Figure 1.1: Distribution of Age at Menarche and Age at marriage
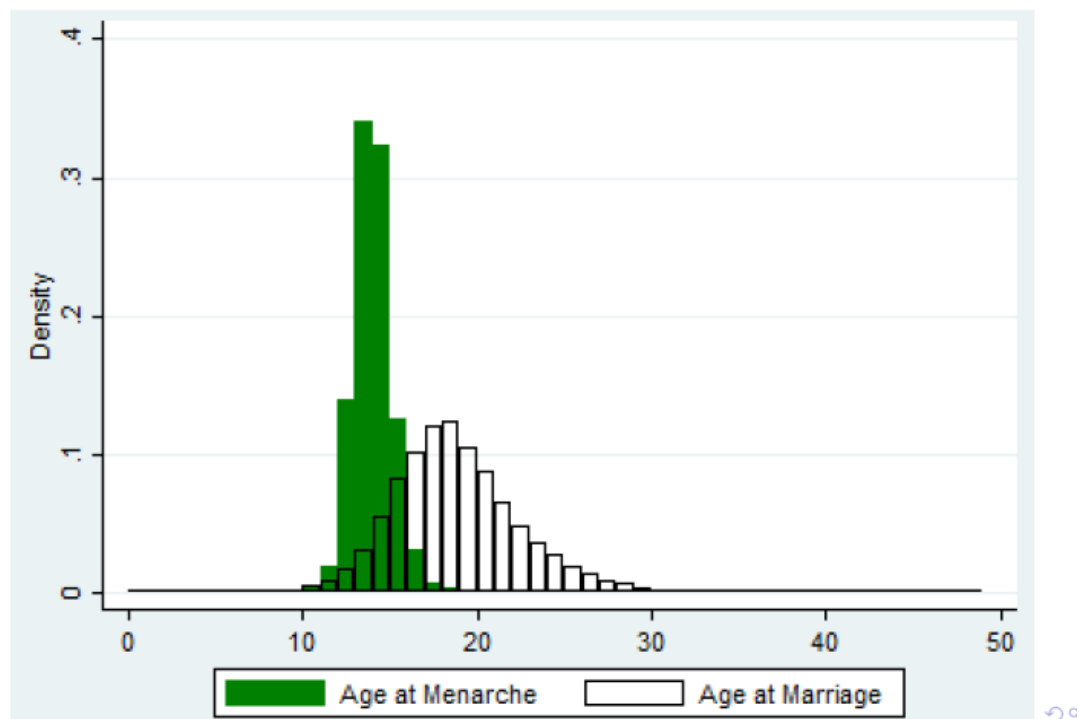
Figure 1.2: Distribution of fraction of marriages by months
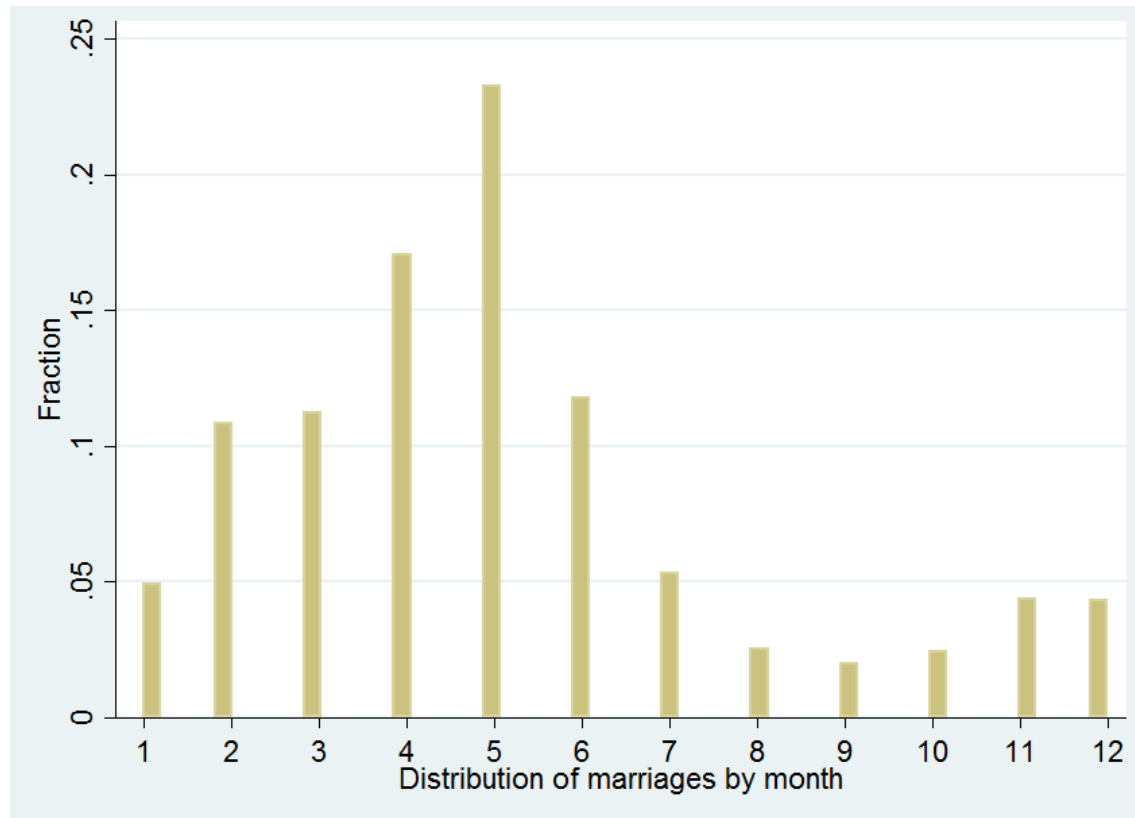
Notes** 1=January, 2=February ....12=December

Figure 1.3: Distribution of Age at Marriage in months
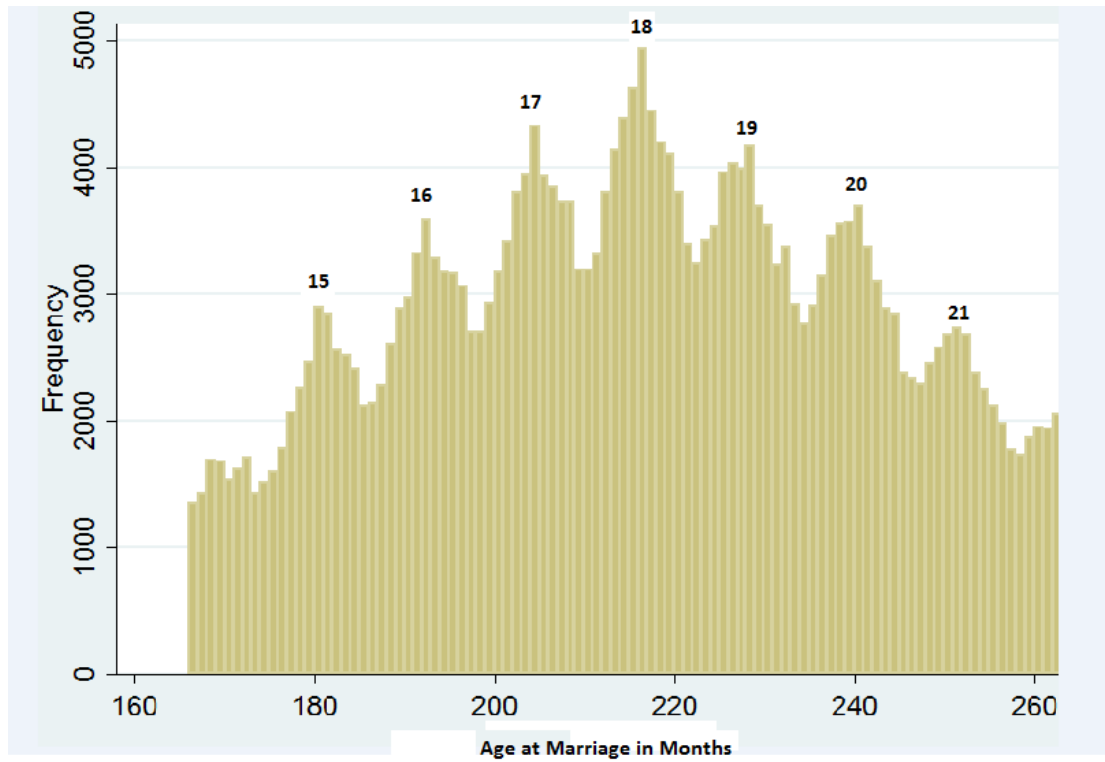
Figure 1.4: Distribution of the variable Target
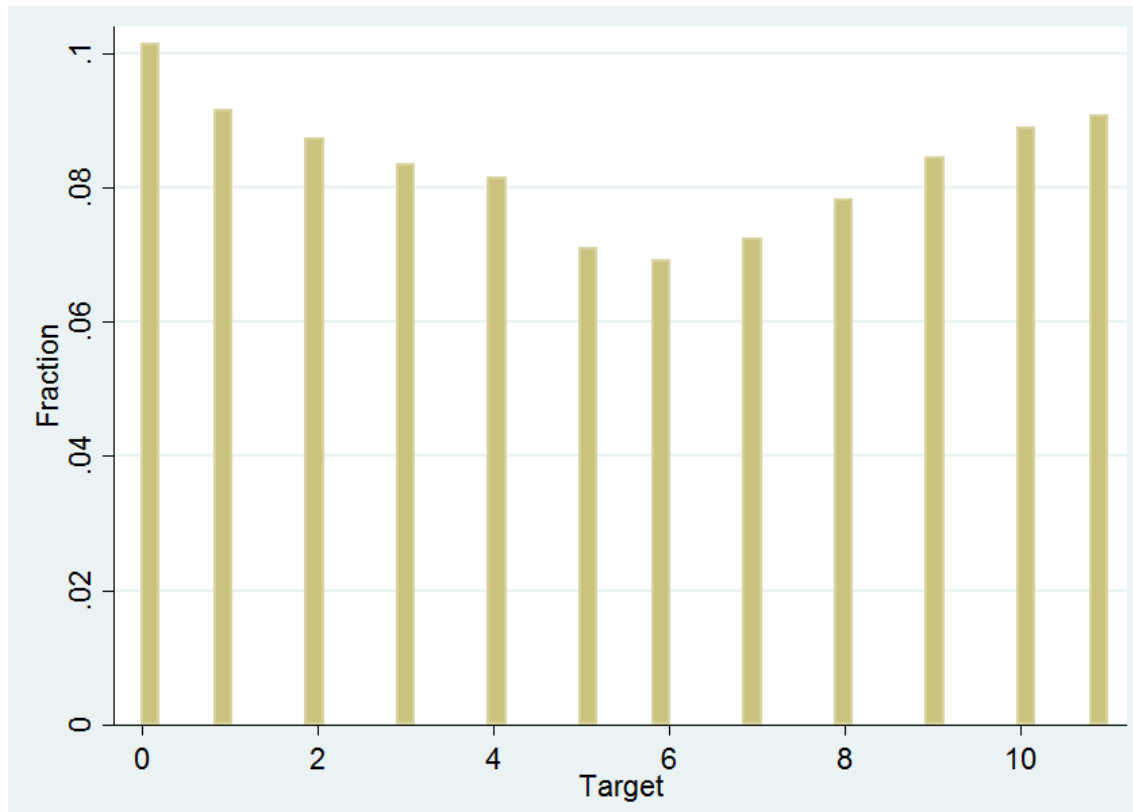
Figure 1.5: Fraction of Marriages by distance from a whole number age
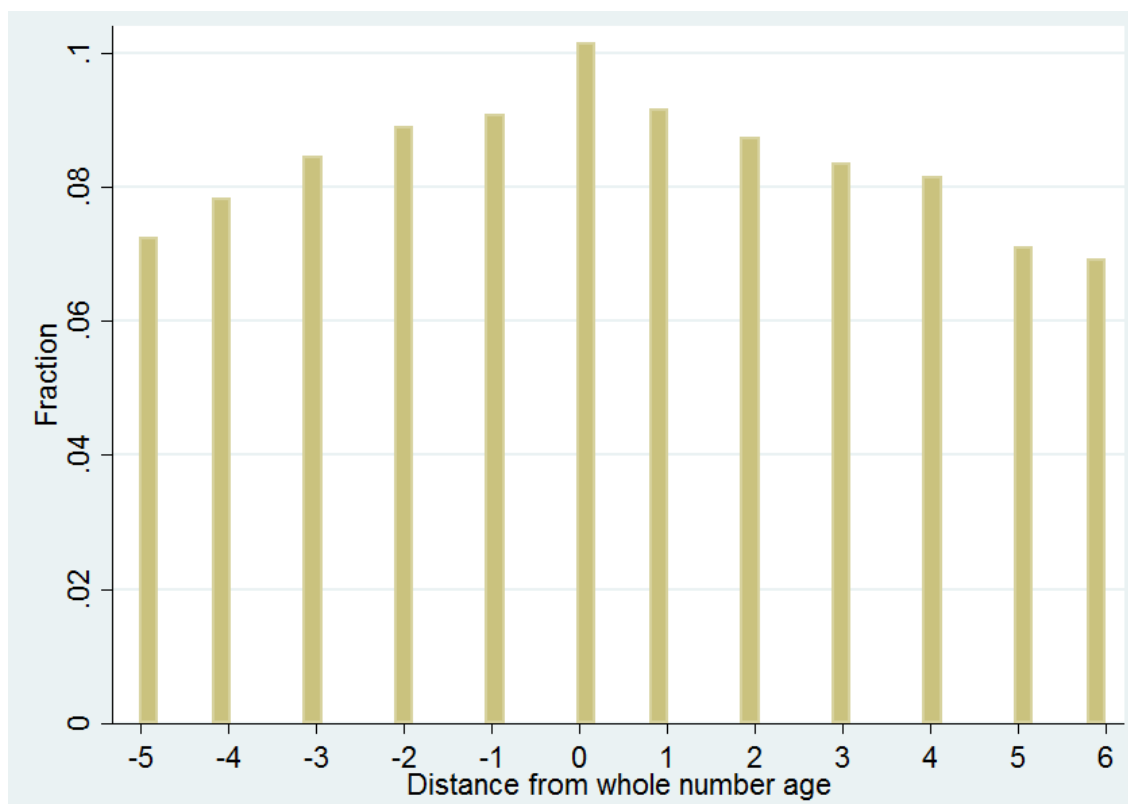
Figure 1.6: Back of the Envelope Distribution of the Variable Target

Distribution of Target
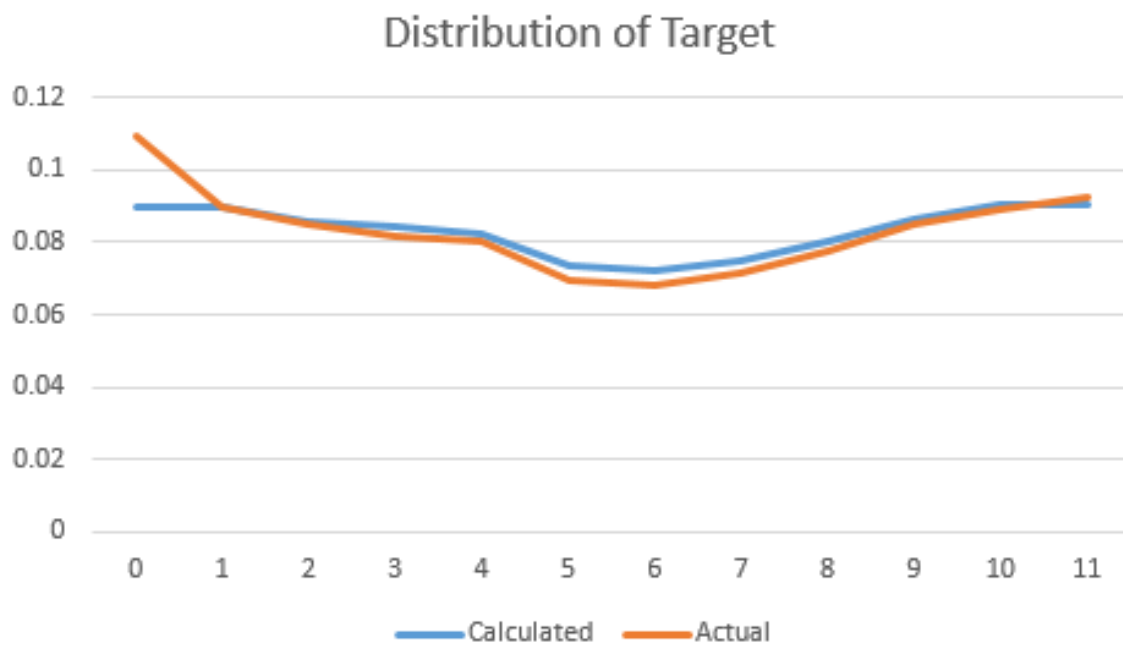
Figure 1.7: Wealth by Distance from a whole number marriage age
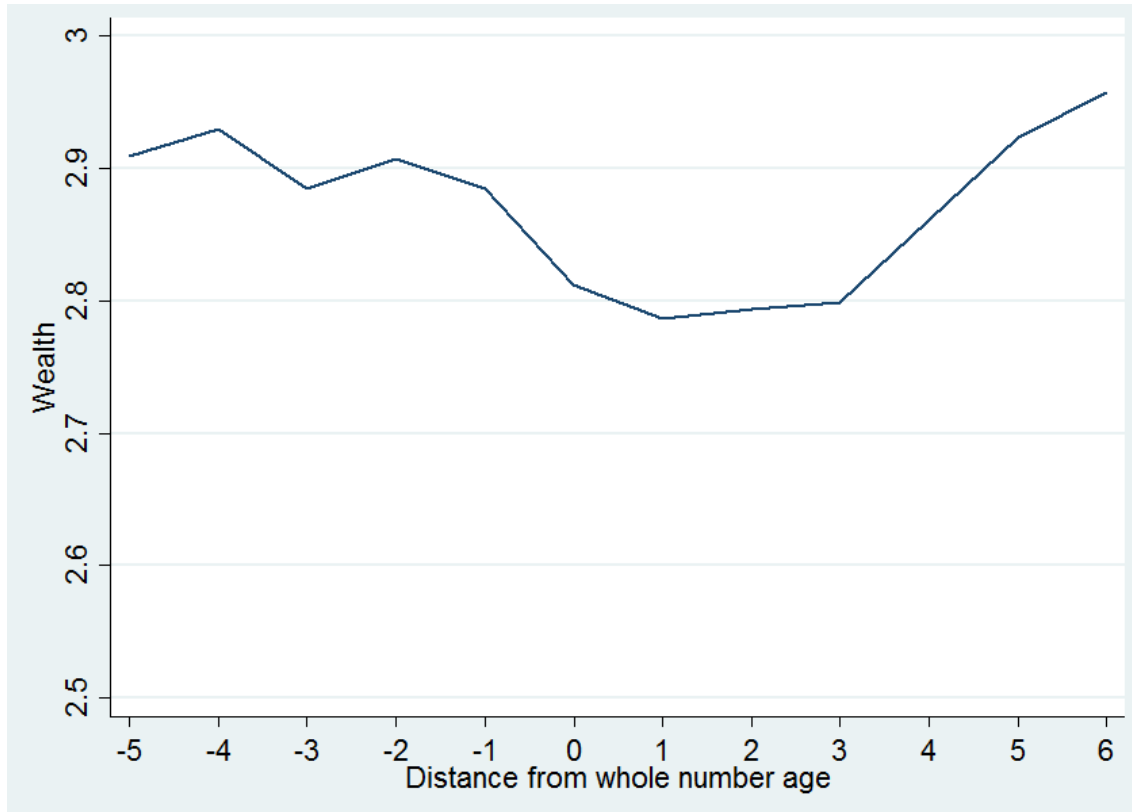
Figure 1.8: Education by Distance from a whole number marriage age
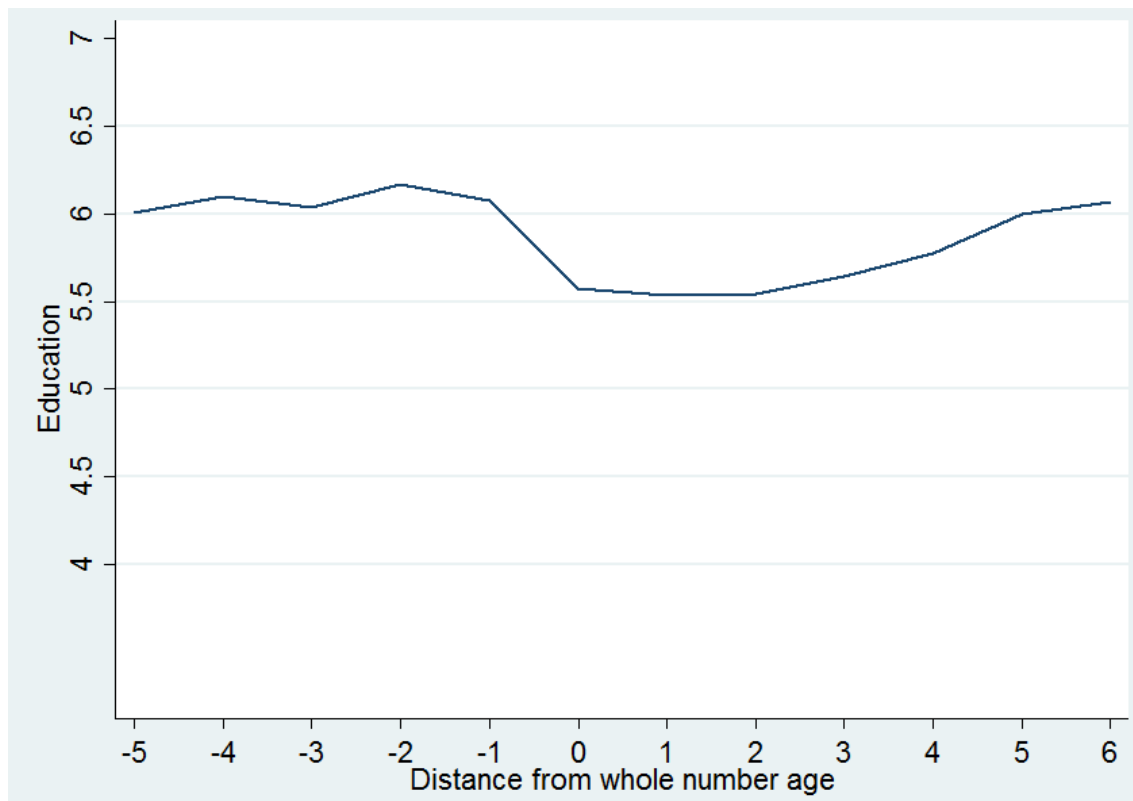
Figure 1.9: Rural by Distance from a whole number marriage age
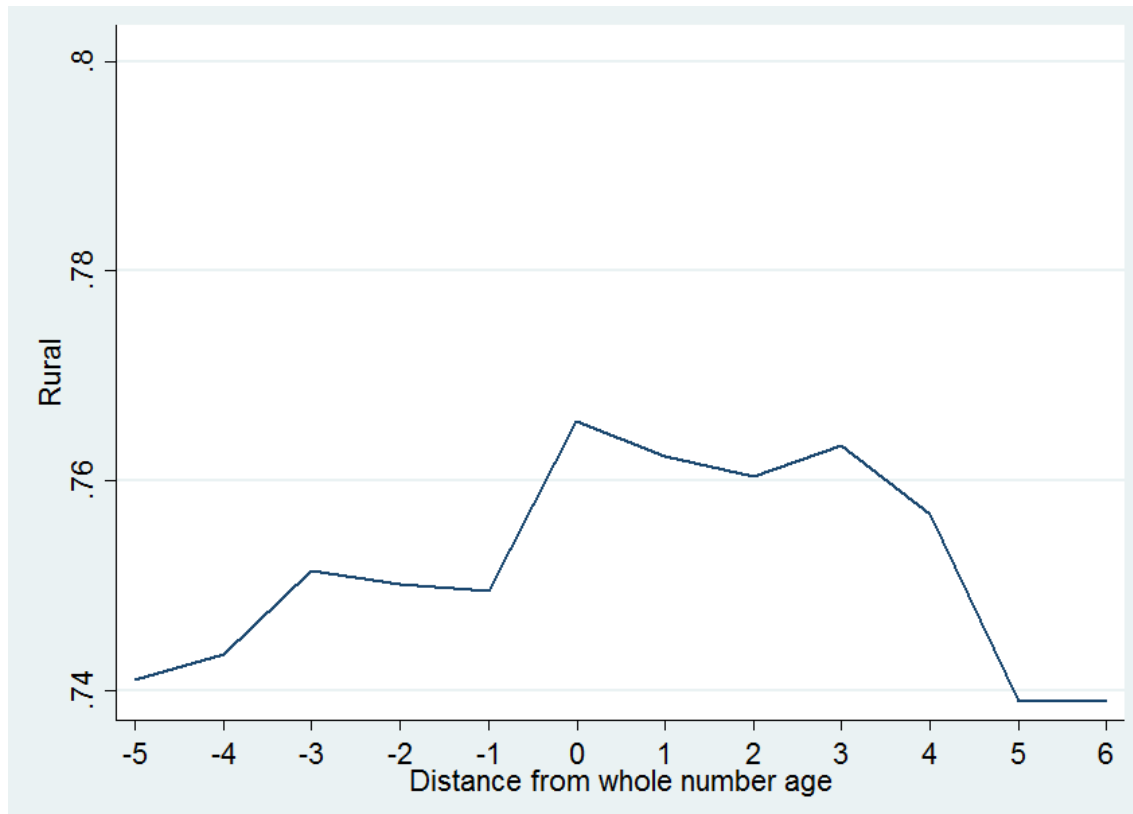
Figure 1.10: Distribution of Target By Wealth

Bottom 20 %   Top 20%
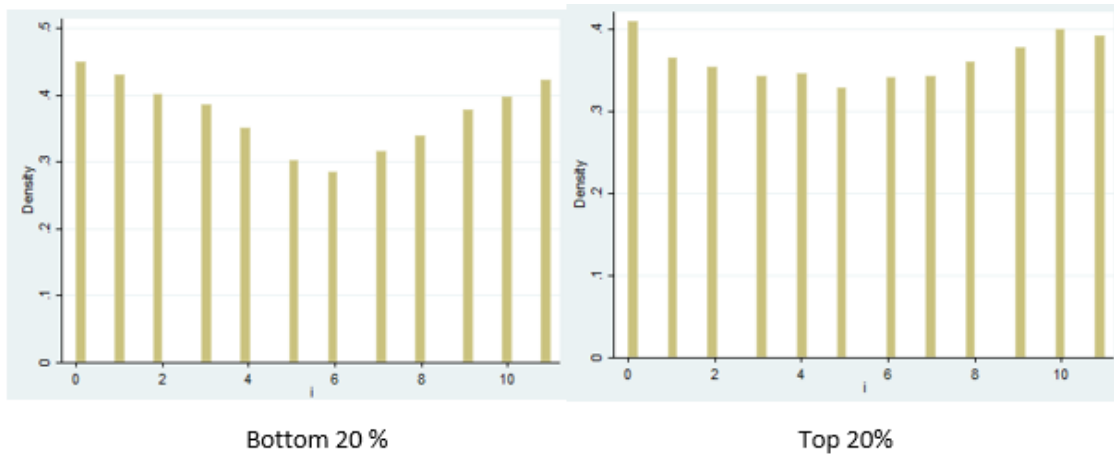
Figure 1.11: Distribution of Target By Education

Zero Education Sample   Completed Secondary Education
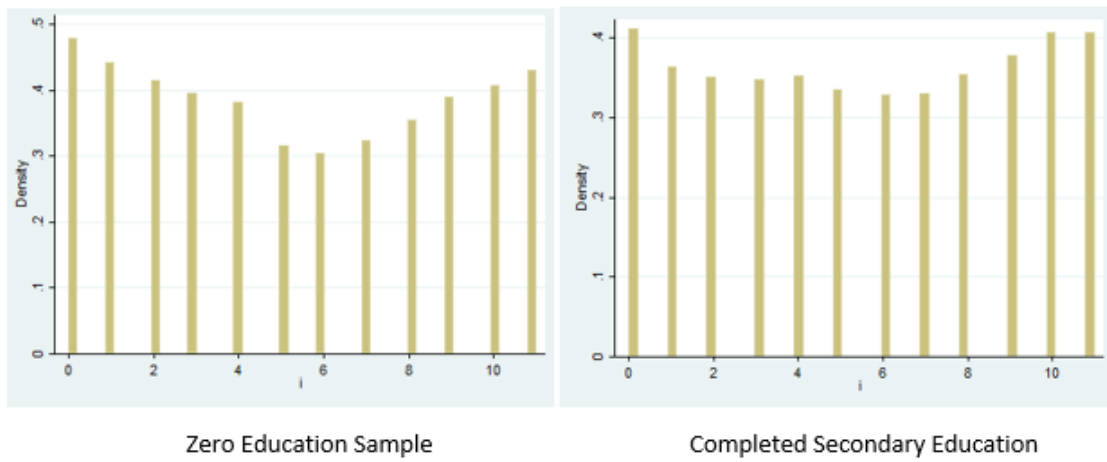
Figure 1.12: Distribution of marriages by month over time
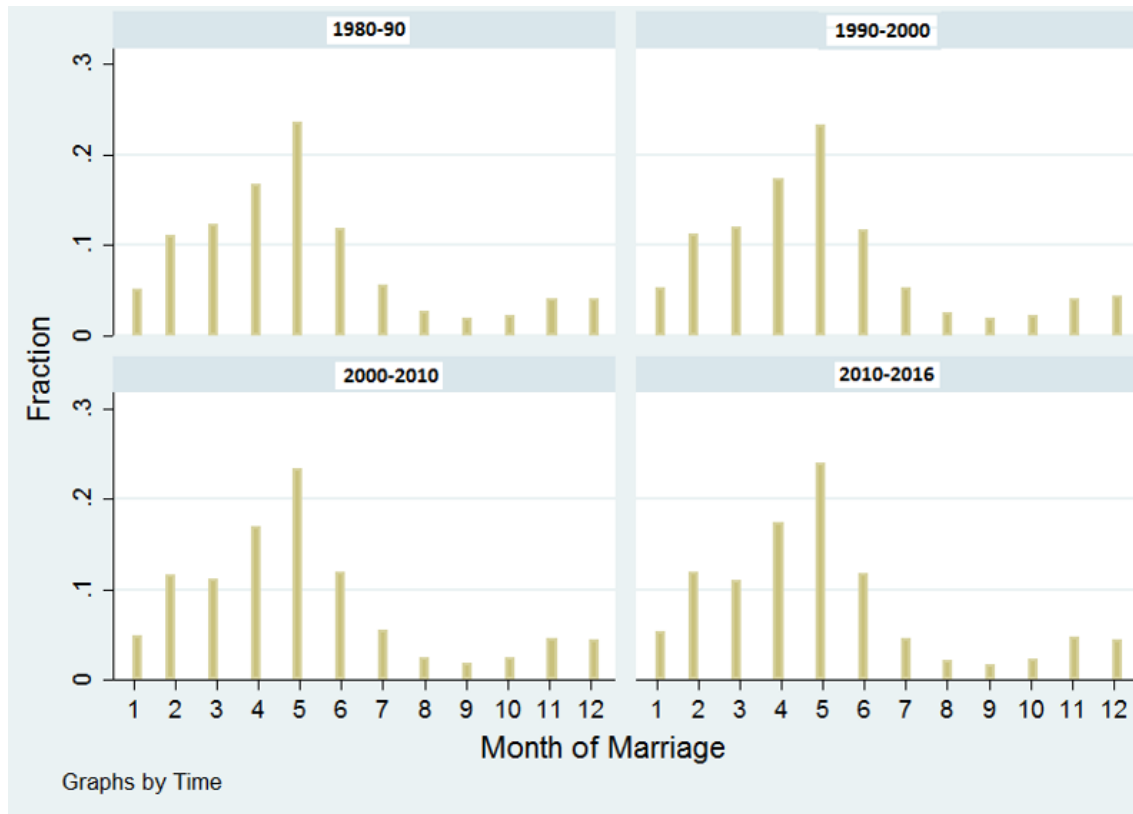
Notes** 1=January, 2=February ....12=December

Figure 1.13: Distribution of marriages by month across religion
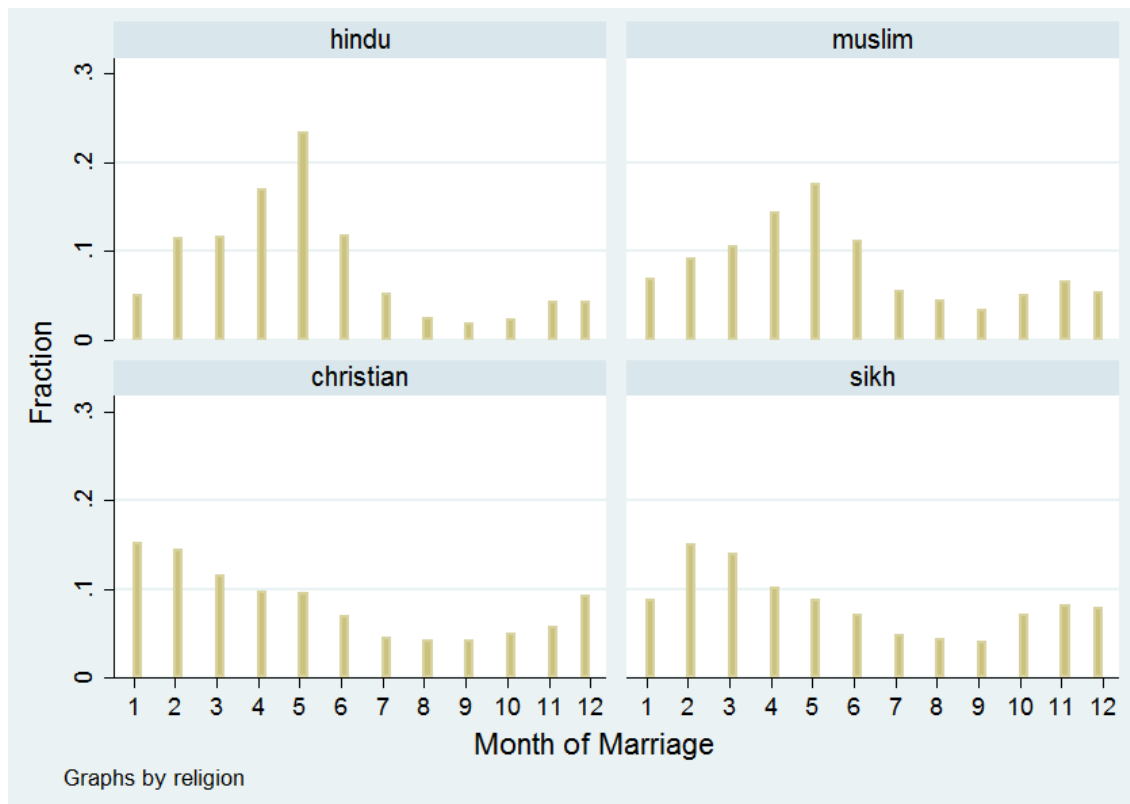
Notes** 1=January, 2=February ....12=December

Figure 1.14: Age at Menarche by distance from a whole number age
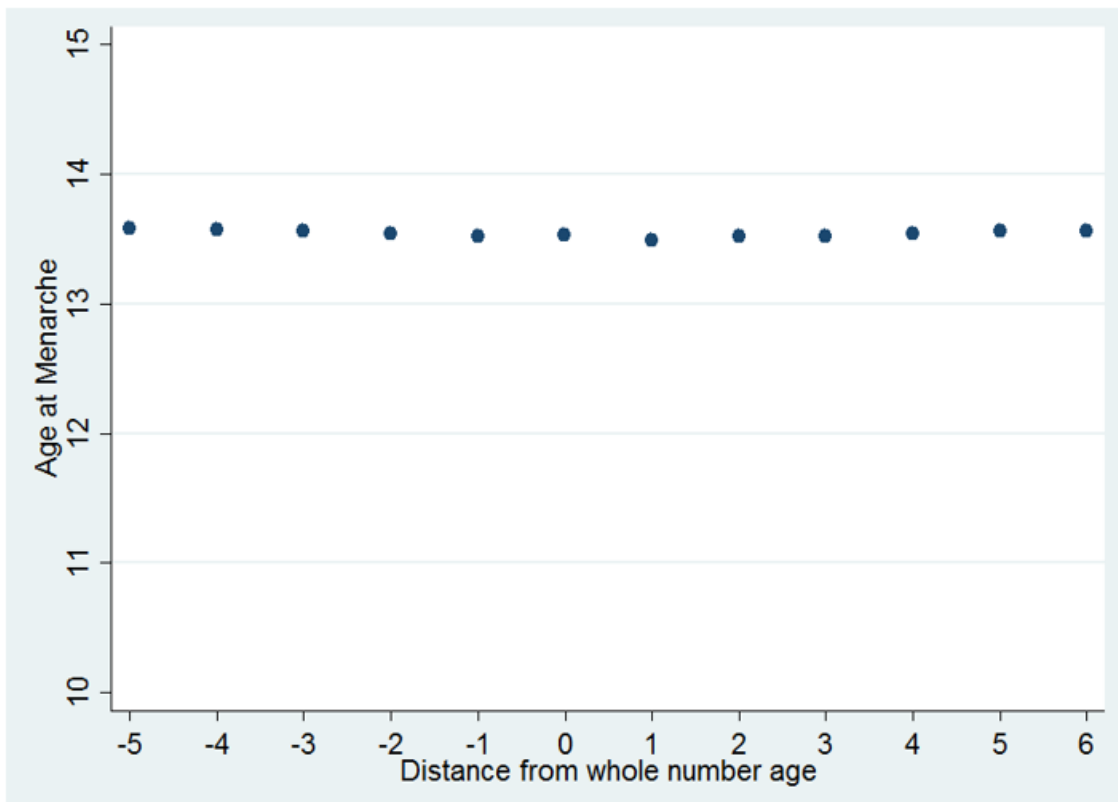
Figure 1.15: Distribution of Education by month of marriage
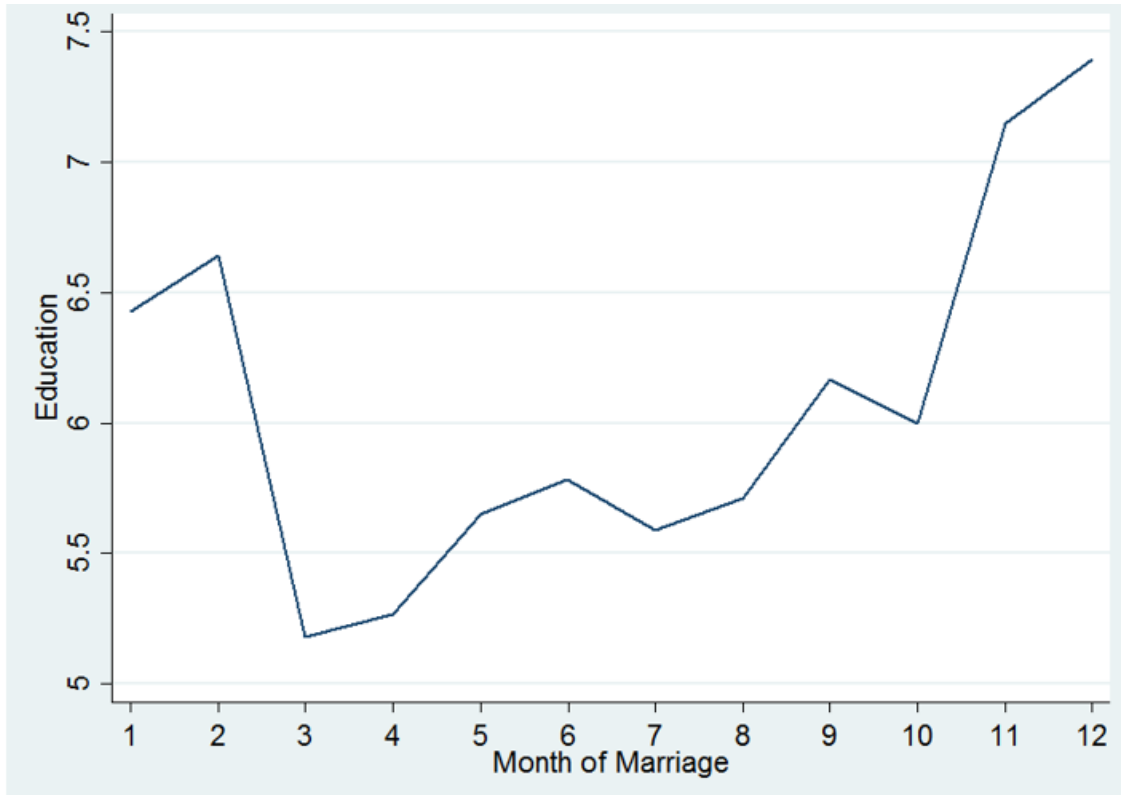
Notes** 1=January, 2=February ....12=December

Figure 1.16: Distribution of Rrural by month of marriage

Notes** 1=January, 2=February ....12=December

Figure 1.17: Distribution of Wealth by month of marriage
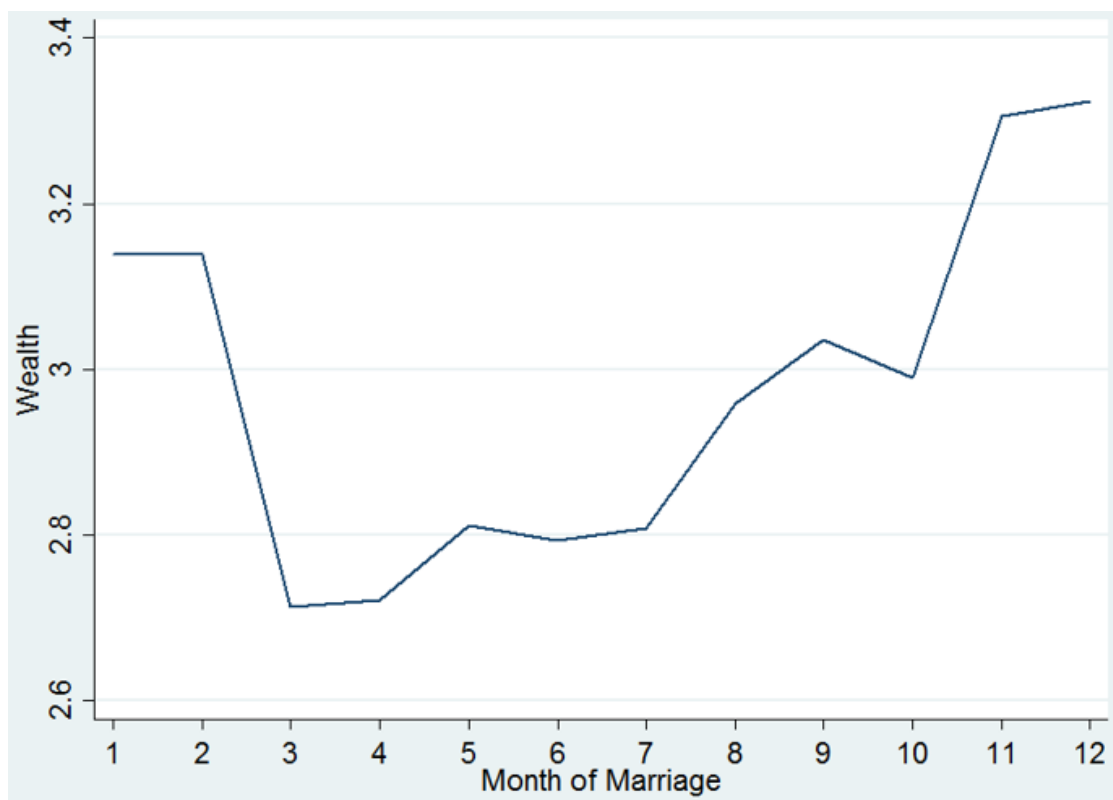
Notes** 1=January, 2=February ....12=December

Figure 1.18: Distribution of month of marriage by month of marriage

Notes** 1=January, 2=February ....12=December

Figure 1.19: Distribution of fraction of births by months for respondents
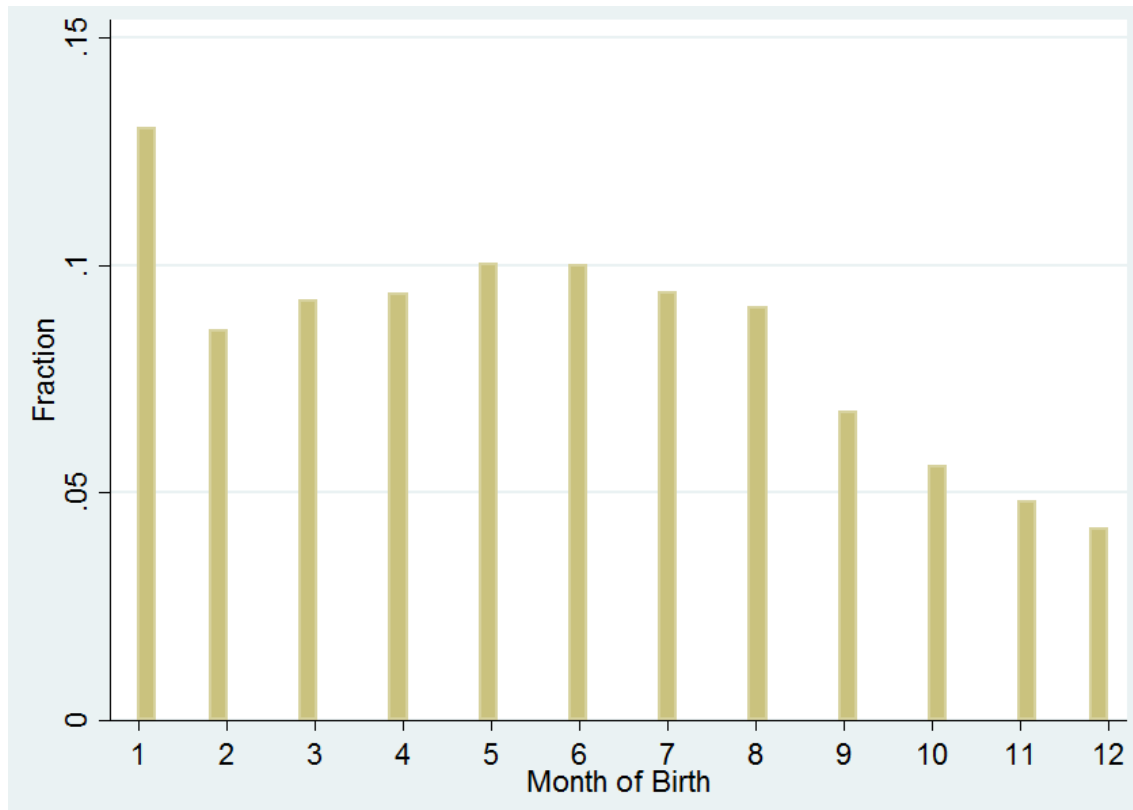
Figure 1.20: Distribution of fraction of births by months for children age 0-4
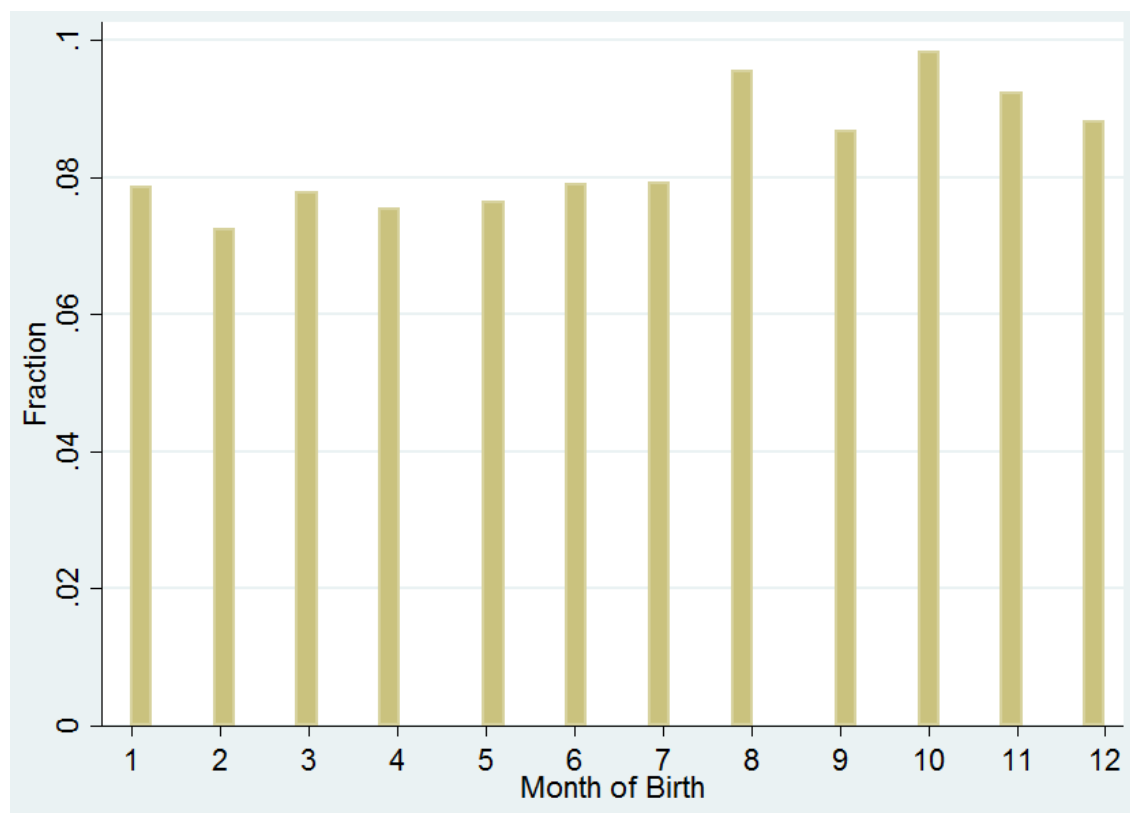
Notes** 1=January, 2=February ....12=December

Figure 1.21: Distribution of Target by Religion

Graphs by religion

Descriptive Statistics						
Women Characteristics						
	Full Sample		Zero Education Sample		Menarche Sample	
	Mean	S.D	Mean	S.D	Mean	S.D
Education(years)	5.69	4.92	0	0	7.58	4.53
Age	31.91	8.57	34.95	8.34	21.32	2.04
Age at Marriage (years)	17.19	2.58	16.44	2.71	17.56	2.28
Age at first birth	19.54	2.80	19.19	3.07	19.16	2.02
Number of children	2.45	1.57	3.23	1.76	1.034	0.90
Height	151.81	6.0	150.95	6.05	151.60	5.98
Household Characteristics						
Number of Member	5.81	2.69	5.86	2.64	6.22	2.95
Wealth	2.85	1.38	2.07	1.15	2.67	1.30
Rural	0.75	0.43	0.85	0.34	0.8	0.39
Electricity	0.87	0.32	0.77	0.41	0.86	0.33
Observations	255397		86977		60639	

Notes**

The table presents the mean and standard deviation of selected household and woman-specific variables for each of the three analysis samples. The household is defined as the persons who live under the same roof. The full sample includes all ever-married women who were administered a women's questionnaire and were married before the age of 21.The zero education sample consists of ever married woman who were administered the women's questionnaire, were married before the age of 21 and have zero years of education. The menarche sample consists of ever married women between ages 15-25 who were asked the age at menarche question.

Table 1.2: First Stage

First Stage Estimates	
Panel A : The effect of IV1 on age at marriage in months for the Full Sample	
IV1 (Being born in low versus high frequency months)	3.07*** (0.013)
Observation	256474
R-Squared	0.009
Panel B :The effect of IV2 on age at marriage in months for the Full Sample	
IV2	-0.15*** (0.005)
Observation	256474
R-Squared	0.01
Panel C :The effect of IV1 on age at marriage in months for the Zero Education Sample	
IV1 (Being born in low versus high frequency months)	3.47*** (0.24)
Observation	86977
R-Squared	0.005
Panel D :The effect of IV2 on age at marriage in months for the Zero Education Sample	
IV2	-0.16*** (0.01)
Observation	86977
R-Squared	0.005

Notes***
The table presents the results of the first stage OLS regression for both the instrumental variables. All regressions control for woman's age at survey and include month of marriage FE. Robust standard errors are reported in parenthesis.
Panel A presents the results of the first stage OLS regression for IV1 for the full sample.
Panel B presents the results of the first stage OLS regression for IV2 for the full sample.
Panel C presents the results of the first stage OLS regression for IV1 for the zero education sample.
Panel D presents the results of the first stage OLS regression for IV2 for the zero education sample.
***=significant at 1%
** =significant at 5%

Table 1.3: Effect of IV1 on Early Childhood Outcomes

Effect of IV1 on Early Childhood Characteristics				
Panel A : Mother Characteristics				
	Years of Education	Age at Marriage	Age at First Birth	Number of Children
IV1	0.008	-0.097**	-0.002	0.0202
	(0.044)	(0.0333)	(0.034)	(0.0179)
Observation	53975	53939	53975	53975
Mean	4.51	17.10	19.69	3.54
Panel B: Household Characteristics				
	Number of Members	Electricity	Rural	Wealth
IV1	0.008	0.003	-0.005	0.03**
	(0.044)	(0.004)	(0.004)	(0.012)
Observation	53975	53975	53975	53975
Mean	4.51	0.74	0.61	3.075
Panel C: Birth Characteristics				
	Size at Birth		Weight at Birth	One Year Mortality
IV1	0.01		-0.005	-0.002
	(0.025)		(0.0487)	(0.002)
Observation	24335		23451	24756
Mean	3.028		6.99	0.048

Notes**

The table presents the OLS regression estimates from equation 2. Panel A presents the estimates for mother characteristics, Panel B presents estimates for household characteristics and Panel C presents the estimates for birth characteristics. Robust standard errors are reported in parenthesis. For size at birth the estimates are reported from an ologit regression model.

***=significant at 1%

** =significant at 5%

Table 1.4: Effect of Age at Marriage on Wealth

Effect of Age at Marriage on Health						
Panel A : Diabetes						
	Diabetic			Pre-diabetic		
	OLS	IV1	IV2	OLS	IV1	IV2
Age at Marriage (in months)	0.0001*** (0.00004)	-0.0007*** (0.0002)	-0.0004*** (0.0002)	0.002*** (0.0004)	-0.013*** (0.003)	-0.012*** (0.0026)
Observation	255397			20835		
Mean	0.025			0.37		
Panel B : Blood Pressure						
	Diastolic			Systolic		
Age at Marriage (in months)	0.0007*** (0.00012)	-0.0027*** (0.00074)	-0.0026*** (0.00067)	0.00003 (0.00012	-0.0018*** (0.0006)	-0.0019*** (0.0006)
Observation	255397			255397		
Mean	0.32			0.28		
Panel C: Anemia						
	Severe Anemia			Moderate Anemia		
Age at Marriage (in months)	0 (0.00002)	0.0001 (0.0001)	0.0001 (0.0001)	0.0002** (0.00009)	0.0005 (0.0005)	0.0004 (0.0004)
Observation	251599			251599		
Mean	0.01			0.13		

Notes**
The table presents the results of OLS and IV regressions in which the dependent variables are health outcomes for the full sample (women who got married before the age of 22). Column with heading OLS presents the ordinary least square estimates. Column with heading IV1 (IV2) presents the estimate from equation 2 using IV1 (IV2) as the instrument. All regressions control for age of the women at survey and includes month of marriage of marriage FE. Robust standard errors are presented in parenthesis.
Panel A presents the result on Diabetes, Panel B presents the results on Blood Pressure, Panel C presents the results on Anemia
***=significant at 1%
** =significant at 5%

Table 1.5: Effect of Age at Marriage on Fertility

	Total number of Children			Ideal number of Children		
Effect of Age at Marriage on Fertility						
Panel A : Number of Children						
	OLS	IV1	IV2	OLS	IV1	IV2
Age at Marriage (in months)	-0.0117*** (0.0003)	-0.0005 (0.002)	-0.0002 (0.0019)	-0.002*** (0.0003)	-0.002 (0.002)	-0.0087*** (0.0015)
Observation	255397			254368		
Mean	2.45			2.1		
Panel B : Timing of Birth						
	Age at First Birth			Birth Spacing		
Age at Marriage (in months)	0.063*** (0.0006)	0.073*** (0.003)	0.074*** (0.003)	0.002 (0.04)	-0.0190 (0.0419)	-0.0616 (0.0506)
Observation	232115			224761		
Mean	19.54			34.188		

Notes**
The table presents the results of OLS and IV regressions in which the dependent variables are fertility outcomes for the full sample (women who got married before the age of 22). Column with heading OLS presents the ordinary least square estimates. Column with heading IV1 (IV2) presents the estimate from equation 2 using IV1 (IV2) as the instrument. All regressions control for age of the women at survey and includes month of marriage FE. Robust standard errors are presented in parenthesis. Panel A presents the results on number of children, Panel B presents the results on ideal number of children reported by mother.
***=significant at 1%
** =significant at 5%

Table 1.6: Effect of Age at Marriage on Spouse and Spousal Family Quality

	Effect of Age at Marriage on Spouse and Spousal Family Quality					
Panel A : Spousal Quality						
	Husband's Years of Education			Husband's Age		
	OLS	IV1	IV2	OLS	IV1	IV2
Age at Marriage (in months)	-0.003 (0.003)	0.031*** (0.009)	0.031*** (0.01)	0.0175*** (0.0011)	-0.024*** (0.0067)	-0.024*** (0.0067)
Observation	219965			219965		
Mean	7.27			38.32		
Panel B : Spousal Family Quality						
	Wealth			Rural		
Age at Marriage (in months)	0.008*** (0.0003)	0.0248*** (0.002)	0.034*** (0.0018)	0.0001 (0.0001)	-0.007*** (0.0007)	-0.007*** (0.0006)
Observation	254539			254539		
Mean	2.85			0.75		

Notes**
The table presents the results of OLS and IV regressions in which the dependent variables are spousal outcomes for the full sample (Women who got married before the age of 22). Column with heading OLS presents the ordinary least square estimates. Column with heading IV1 (IV2) presents the estimate from equation 2 using IV1 (IV2) as the instrument. All regressions control for age of the women at survey and includes month of marriage FE. It also controls for household characteristics. Robust standard errors are presented in parenthesis.
Panel A presents the results for Spousal Quality, Panel B presents the results for Spousal Family Quality.
***=significant at 1%
** =significant at 5%

Table 1.7: Effect of Age at Marriage on Education and Bargaining Power

Effect of Age at Marriage on Education and Bargaining Power									
Panel A : Education									
	Years of Education			Primary					
	OLS	IV1	IV2	OLS	IV1	IV2			
Age at Marriage (in months)	0.034*** (0.0011)	0.032*** (0.0066)	0.0351*** (0.0059)	0.001*** (0.0001)	0.0003 (0.0007)	0.0002 (0.0006)			
Observation	253080			254539					
Mean	5.7			0.6					
Panel B : Bargaining Power									
	Limit women to contact family			Insist on knowing where partner is			Number of control issue questions answered yes		
	OLS	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2
Age at Marriage (in months)	0.003 (0.002)	-0.006** (0.0027)	-0.014** (0.0063)	0.002 (0.003)	-0.007** (0.0029)	-0.01** (0.0040)	0.011*** (0.001)	-0.001 (0.0019)	-0.006 (0.006)
Observation	27421			27421			27421		
Mean	0.1661			0.2142			0.41		

Notes**
The table presents the results of OLS and IV regressions in which the dependent variables are spousal family outcomes for the full sample (Women who got married before the age of 22). Column with heading OLS presents the ordinary least square estimates. Column with heading IV1 presents the estimate from equation 2 using IV1 as the instrument. Column with heading IV1 presents the estimate from equation 2 using IV2 as the instrument. All regressions control for age of the women at survey and includes month of marriage FE. It also controls for household characteristics. Robust standard errors are presented in parenthesis.
***=significant at 1%
** =significant at 5%

Table 1.8: Effect of Age at Marriage on Health

Effect of Age at Marriage on Health			
	Diabetes	Diastolic	Anemia
Panel A: Full Sample(1)			
	IV1	IV1	IV1
Age at Marriage	-0.0007***	-0.0027***	0.0001
	(0.0002)	(0.00074)	(0.0001)
Observation	255397	255397	251599
Mean	0.025	0.32	0.01
Panel B: Full Sample + Spousal Controls(2)			
Age at Marriage	-0.0008***	-0.0028***	0.0002
	(0.0002)	(0.0008)	(0.0003)
Observation	220694	220694	217466
Mean	0.025	0.32	0.01
Panel C: Full Sample + Spousal Controls + Bargaining Controls(3)			
Age at Marriage	-0.0014**	-0.0035	-0.0008
	(0.0007)	(0.0022)	(0.0007)
Observation	31779	31779	31779
Mean	0.028	0.3	0.01
Panel D: Zero Education Sample + Spousal Controls(4)			
Age at Marriage	-0.0012***	-0.0026	0.0002
	(0.0004)	(0.0015)	(0.0004)
Observation	76342	76342	76342
Mean	0.022	0.33	0.01

Notes**

The table presents the results from the IV regression using IV1 as the instrument where the dependent variables are health outcomes for the full sample (Women who got married before the age of 22). Robust standard errors are presented in parenthesis. Panel A presents the result from specification 1 which is the baseline specification. The IV regression controls for women's age at survey and include month of marriage FE. Panel B presents the results from specification 2. The IV regression include controls from specification 1 plus spousal controls. Panel C presents the results from specification 3. The IV regression include controls from specification 2 plus bargaining controls. Panel D presents the results from specification 4. The IV regression include controls from specification 2 and restricts the sample to the zero education population.

***=significant at 1%

** =significant at 5%

Table 1.9: Effect of Age at Marriage on Fertility

	Total Number of Children	Age at First Birth	Ideal Number of Children
Effect of Age at Marriage on Fertility			
Full Sample(1)			
	IV1	IV1	IV1
Age at Marriage	-0.0005 (0.002)	0.073*** (0.003)	-0.002 (0.002)
Observation	255397	232115	254368
Mean	2.45	19.54	2.1
Full Sample + Spousal Controls(2)			
Age at Marriage	0.009*** (0.002)	0.072*** (0.004)	0.005*** (0.001)
Observation	220694	203350	220775
Mean	2.45	19.54	2.1
Full Sample + Spousal Controls + Bargaining Controls (3)			
Age at Marriage	0.008** (0.004)	0.075*** (0.01)	0.007 (0.007)
Observation	27421	25474	27421
Mean	2.43	19.68	2.07
Zero Education Sample + Spousal Controls(4)			
Age at Marriage	0.0167*** (0.0044)	0.044*** (0.0076)	0.013*** (0.003)
Observation	76342	72954	76571
Mean	3.23	19.19	2.48

Notes**

The table presents the results from the IV regression using IV1 as the instrument where the dependent variables are fertility outcomes for the full sample (Women who got married before the age of 22). Robust standard errors are presented in parenthesis. Panel A presents the result from specification 1 which is the baseline specification. The IV regression controls for women's age at survey and include month of marriage FE. Panel B presents the results from specification 2. The IV regression include controls from specification 1 plus spousal controls. Panel C presents the results from specification 3. The IV regression include controls from specification 2 plus bargaining controls. Panel D presents the results from specification 4. The IV regression include controls from specification 2 and restricts the sample to the zero education population.

***=significant at 1%

** =significant at 5%

Table 1.10: First Stage Comparison

First Stage Estimates	
Panel A : The effect of IV1 on age at marriage in months/cohabitation for the Full Sample	
IV1 (Being born in low versus high frequency months)	2.31***
	(0.13)
Observation	251491
R-Squared	0.0675
Panel B :The effect of IV1 on age at marriage in months/cohabitation for the Menarche Sample	
IV1 (Being born in low versus high frequency months)	1.99***
	(0.23)
Observation	59726
R-Squared	0.1222
Panel C :The effect of IV1 on age at marriage in months/cohabitation for the Menarche Sample	
Age at Menarche (in years)	1.4***
	(0.09)
Observation	59726
R-Squared	0.1244

Notes***

The table presents the results of the first stage OLS regression for both the instrumental variables. All regressions control for woman's age at survey and height, and include caste and district FE. Robust standard errors are reported in parenthesis.

Panel A presents the results of the first stage OLS regression for IV1 for the full sample.

Panel B presents the results of the first stage OLS regression for IV1 for the menarche sample.

Panel C presents the results of the first stage OLS regression for age at menarche for the menarche.

***=significant at 1%

** =significant at 5%

Table 1.11: Comparison : Intermediary Variables

Comparing the Effect of Age at Marriage on Intermediary variables				
Panel A : Education				
	Years of Education		Primary	
	IV1	Age at Menarche	IV1	Age at Menarche
Age at Marriage	0.14 *** (0.015)	0.13*** (0.009)	0.005*** (0.001)	0.006*** (0.001)
Observation	59808		59808	
Mean	7.5		0.77	
Baseline Results	0.032*** (0.0066)		0.0003 (0.0007)	
Panel B : Spousal Quality				
	Husband's Years of Education		Husband's Age	
	IV1	Age at Menarche	IV1	Age at Menarche
Age at Marriage	0.06*** (0.001)	0.07*** (0.002)	-0.03 (0.02)	-0.1*** (0.01)
Observation	45911		45911	
Mean	8.34		26.12	
Baseline Results	0.031*** (0.009)		-0.024*** (0.0067)	
Panel C : Spousal Family Quality				
	Wealth		Rural	
	IV1	Age at Menarche	IV1	Age at Menarche
Age at Marriage	0.03*** (0.004)	0.03*** (0.002)	-0.007*** (0.001)	-0.005 (0.007)
Observation	59808		59808	
Mean	2.67		0.8	
Baseline Results	0.0248*** (0.002)		-0.007*** (0.0007)	

Notes**

The table presents the results of IV regressions in which the dependent variables are intermediary outcomes for the menarche sample. Column with heading IV1 (Age at Menarche) presents the estimate from equation 2 using IV1 (Age at Menarche) as the instrument. All regressions control for age of the women at survey and height, and includes caste and district FE. Robust standard errors are presented in parenthesis. Panel A presents the result on Education, Panel B presents the results on Spousal Quality and Panel C presents the results on Spousal Family Quality.

***=significant at 1%
** =significant at 5%

Table 1.12: Comparison : Health and Fertility

Comparing the Effect of Age at Marriage on Health and Fertility						
Panel A : Health						
	Diabetes		Diastolic		Anemia	
	IV1	Age at Menarche	IV1	Age at Menarche	IV1	Age at Menarche
Age at Marriage	-0.0001 (0.0002)	-0.0003** (0.0001)	-0.002 (0.002)	-0.007*** (0.0008)	0 (0.0002)	0.0005*** (0.0001)
Observation	59808		59808		59808	
Mean	0.016		0.25		0.01	
Baseline Result	-0.0007*** (0.0002)		-0.0027*** (0.00074)		0.0001 (0.0001)	
Panel B : Fertility						
	Total number of Children		Age at first birth		Ideal Number of Children	
	IV1	Age at Menarche	IV1	Age at Menarche	IV1	Age at Menarche
Age at Marriage	-0.0001 (0.002)	-0.0178*** (0.0013)	0.083*** (0.006)	0.083*** (0.003)	-0.009** (0.002)	-0.01*** (0.001)
Observation	59808		40896		59623	
Mean	1.76		19.1		1.89	
Baseline Result	-0.0005 (0.002)		0.073*** (0.003)		-0.002 (0.002)	

Notes**
The table presents the results of IV regressions in which the dependent variables are health and fertility outcomes for the menarche sample. Column with heading IV1 (Age at Menarche) presents the estimate from equation 2 using IV1 (Age at Menarche) as the instrument. All regressions control for age of the women at survey and height, and includes caste and district FE. Robust standard errors are presented in parenthesis.
Panel A presents the result on Health, Panel B presents the results on Fertility.
***=significant at 1%
** =significant at 5%

Table 1.13: Comparison : Robustness

Robustness Checks				
Panel A : Health and fertility results from using month of birth as an IV				
	Health		Fertility	
	Diabetic	Diastolic	Total Children	Age at first birth
IV1	-0.0004***	-0.0012**	0.005***	0.073***
	(0.0001)	(0.005)	(0.0014)	(0.002)
Observation	255397	255397	255397	232115
Mean	0.025	0.32	2.45	19.54
Panel B : Health and fertility results using IV1 and dropping women with month of birth reported as January				
	Health		Fertility	
	Diabetic	Diastolic	Total Children	Age at first birth
IV1	-0.001***	-0.004**	0.009***	0.076***
	(0.0001)	(0.0009)	(0.002)	(0.004)
Observation	223065	223065	223065	202058
Mean	0.025	0.32	2.46	19.52
Panel C : Health and fertility results using IV1 and dropping women with target=0				
	Health		Fertility	
	Diabetic	Diastolic	Total Children	Age at first birth
IV1	-0.0008***	-0.002**	0.008***	0.072***
	(0.0002)	(0.0008)	(0.001)	(0.0037)
Observation	230806	230806	230806	209718
Mean	0.025	0.32	2.45	19.55

Notes**
The table reports the results from the robustness checks, all the regressions control for women's age, spousal and spousal family quality, and include month of marriage FE. Robust standard errors are presented in parenthesis. Panel A presents the result on the full sample using month of birth dummies as an IV. Panel B presents the result on the full sample using IV1 dropping the women who reported their month of birth as January. Panel C presents the results on the full sample using IV1 dropping the women who report the same month of birth and marriage, i.e. target=0.
***=significant at 1%
** =significant at 5%

CHAPTER 2

RETIREMENT SECURITY AND GENDER GAP IN PARENTAL INVESTMENT

2.1 Introduction

Gender gaps favoring the male child at the household level in various aspects such as health, education and personal autonomy are systematically higher in developing countries (Jayachandran, 2014). In the Indian context, there have been various studies trying to establish the reasons behind these gaps. Some of the established reasons for these gaps include the patriarchal nature of the Indian society, son preference, poverty, illiteracy, marriage norms and lack of employment opportunities for women. These biases lead to differences in investment of resources between the boy child and the girl child at the household level (Sen and Sengupta 1983, Das Gupta 1987 and Behrman et al. 1988). In this paper, I will try and investigate the old age (post retirement) security motive for the parents as a potential reason for this differential investment between a girl child and a boy child.

With the absence of proper credit markets and retirement security in developing countries parents are incentivized to invest in children as means of old age support (Nugent, 1987). It has also been seen in the existing literature that transfer from children and well-being of the parents depends on the education level of the children (Zimmer et al, 2007). These provide incentives for the parents to invest in their children's education, besides altruistic reasons. In developing countries like India, investment in the male child on average generates higher return to the parents as compared to investment in their daughter, in terms of financial support at an old age (Kishor, 1993). This could be attributed to labor market returns and marriage norms.[1] Also parents in India live with their sons in old age and it is the responsibility of the son to provide for parents whereas the daughter is expected to take care of the husband's family (Das Gupta, 1987). This skewedness

---

[1] When daughters get married they usually live with their husband and his family. She is also expected to take care of her husband's parents and her income goes to the husband's family

in dependency towards the son for old age support can incentivise the parents to invest more in sons compared to daughters. In this paper I will try and investigate this old age support (retirement security) motive for unequal investment between a boy child and a girl child. Now, parents who have other sources of post retirement income such as pension or savings will depend less on their children for financial support in their old age. Whereas parents who do not have any pension income will primarily depend on their children and savings for post-retirement consumption. In this paper I use a pension policy change as an exogenous shock to post retirement security to look at the causal effect of retirement security on gender gap in investment.

In 2004, the Government of India introduced the New Pension Scheme for its public sector employees joining after January 1st, 2004. The public sector employees who joined prior to that were enrolled in the Civil Servant Pension Scheme. This shift from the Civil Servant Pension Scheme to the New Pension Scheme was seen as a decrease in the post retirement security by the public sector employees. While the policy change decreased the post retirement security for the public sector employees it had no effect on the private sector employees. Due to the policy change the parents who joined the public sector post 2004 (after the policy change) had lower retirement security compared to the parents who joined prior to 2004 (before the policy change). This implies that parents who joined the public sector post 2004 will be more dependent on their children for old age support and since the dependency is skewed towards the son we might see a rise in gender gap. In graph 1, I plot the difference in private school enrollment between boys and girls. We see that there is a rise in gender gap in enrollment for children whose parents joined the public sector post 2004 compared to those whose parents joined prior to 2004. We do not see such a rise for children whose parents worked in the private sector or agricultural sector. This is indicative of the fact that gender gap increased for children of public sector employees post policy reform.

To capture the causal effect of retirement security on gender gap in child investment and outcomes, I use the policy change in a triple difference framework. The first difference between the public sector employees who joined post 2004 and prior to 2004 will give us the effect of decrease in post-retirement security and time effects on investment in children. A second difference between

70

parents who joined the private sector post 2004 and prior to 2004 will give us the time effects as there was no major pension plan change between these two time periods for the private sector employees. Assuming that the time effects are same for the two sectors, a difference in difference between sectors (public versus private) and joining time (post versus pre) will give us the effect of decrease in retirement security on investment in children. Finally a third difference by the gender of the child will capture the effect of retirement security on gender gap in child investment. This is captured by running a triple difference model across sector, time and gender. The prior is that, as post retirement security decreases the girl child is worse off compared to the boy child in terms of educational investment and outcomes. Alternatively stated, gender gap goes up for children of public sector employees versus private sector employees, post reform.

I use a nationally representative panel data from India, the Indian Human Development Survey 2004 and 2011. The major outcomes that I look at are schooling choices[2] and standardized test scores for children.[3]. While schooling choice such as public versus private school is indicative of the monetary investment made in the child, test scores reflect the sum total of various investments made in the child. Private schools in India are more expensive than the public schools and are associated with better outcomes for the children, so when a child is enrolled in a private school compared to a public school it is indicative of higher investment in the child. Test scores on the other hand are a reflection of the total investment made in the child.

I find that a decrease in post-retirement security, increases the gender gap against the girl child. We see that the gender gap in private school enrollment is 0.07 (19 percent)[4] higher for children of public versus private sector employees after the reform relative to before and the effect is statistically significant. Also compared to a boy child, girl child is 0.0684 (16 percent) less likely to be enrolled in a school where the medium of instruction is English[5] and the effect is statistically

---

[2]Private versus public.

[3]I do not look at enrollment as 97% of the kids in my sample report being in school (i.e. there is no variation in school enrollment)

[4]This is driven by 15.3% increase in private school enrollment for boy child and 4% decrease for girl child

[5]The magnitudes of change in enrollment for the boy child and girl child is similar to that of

significant. The girl child also performs worse compared to the boy child on standardized test scores in math(-0.07), writing(-0.04) and English ability(-0.06) however these results are not statistically significant. Overall we see that a decrease in post-retirement security for the parents are associated with worse outcomes and lower investment for the girl child compared to the boy child.

The rest of the paper is organized in the following way. In section II I discuss the relevant institutional details to set up the empirical strategy, I discuss the pension system in India, the reform and how the new pension scheme is worse for the public sector employees. I also briefly discuss the discrimination against girl child present in India and finally I discuss private versus public schools in India. In Section III I provide a brief Literature Review where I talk about previous work on this topic. In Section IV I develop a theoretical model to explain the linkage between old age security and differential investment. In section V I discuss the empirical strategy, section VI I discuss the data. Section VII presents the results followed by common trends and falsification tests in section VIII. Section IX presents the conclusion.

2.2 Background

2.2.1 Pension System in India

The Indian pension system suffers from a lot of drawbacks such as the absence of an universal social security for the unorganized sector which constitutes the bulk of the work force, low coverage in general, ever increasing fiscal stress on the government and long term structural instability for the public sector pension system(Sadhak, 2009). In 2012 that the estimated population of elderly people living in India was above 100 million, with an annual growth rate of 3.9 percent during the last decade. Thus for a country with a high rate of growing elderly population it has a very low pension coverage of around 12 percent of the working population. In the absence of a universal pension system the elderly in India depend mostly on their own savings and transfers from relatives for their old age consumption.

---

private school enrollment. This might be driven by the fact that private school and medium of instruction being English is highly correlated (0.7)

Contrary to the present scenario India actually had a long history of pension benefits, some even dating back to the 3rd Century B.C, where the king had to pay half of the wages for people who have completed 40 years of service (Gayithri, 2006). Pension system was also present during the Colonial period of British India. Under the Royal Commission of Civil Establishments 1881, pension benefits were given to the public sector employees. Indian government further consolidated these benefits under the Acts of 1915 and 1935. Post-independence other schemes were created to cover the private sector (Gowswami, 2001). The structure and benefits from the different pension schemes currently available to different sectors vary a lot. So it makes more sense to look at them separately. In the next few paragraphs we will discuss the pension scenario in India prior to 2004, for the three broad sectors.

First, I discuss the most generous pension scheme in India, the Civil Servant Pension scheme. This scheme covered public sector employees under the Central and State government. This is a pay-as-you-go defined benefit system, so the employees do not contribute anything towards their retirement account. Thus the entire pension expenditure is paid from the government revenues. By 2012 the scheme had 2.4 million subscribers (Sanyal, 2013). An Indian government employee is entitled to a monthly pension post retirement or superannuation or invalidation. The pension amount is fixed and is calculated based on the number of years of service and the average salary of the last ten months of service. Full superannuation benefit is 50 percent of the average salary of the last ten months of service. The benefit upon retirement has two components: the above mentioned monthly pension and a lump sum payment called the retirement gratuity. Dearness Pension benefit in the form of family pension is given to the spouse after the death of the employee. This compensation is approximately 30 percent of the monthly pension. The government sector employees also have a commutation provision where they can forgo 40 percent of the pension and take it as a lump sum restored after 15 years of retirement. The pension amount is indexed to the consumer price index to provide real annuities to the retired government servants. To claim the entire benefits of the pension scheme the employee has to retire after the age of 60 or after 33 years of service. The penalty for early retirement varies by state and different branches of the

government.

The second major pension scheme which primarily covers the organized private sector employees is the Employee Provident Fund Scheme 1952 (EPF) and the Employee Pension Scheme 1995 (EPS). Any organization with more than 20 employees are mandatorily covered by EPF. The provident fund is a defined contribution, fully funded pension scheme. The employee and the employer each contribute 12 percent of monthly earnings to the EPF account. This is returned to the employee in lump sum with interest (Current annual interest rate is 11percent). For employees drawing a basic pay of up to Rs 6,500 ($10) per month it is mandatory to make the contribution towards EPF and EPS. Employees drawing more than that have an option of getting PF deducted from the salary. In normal circumstances the employee and the employer contribute 12 percent each of the basic salary of the employee. The contribution of 12 percent from the employee is put into the EPF account. Out of the 12 percent contribution from the employee 8.33percent is invested in the EPS and the rest in EPF. The interest rates on these accounts are set by the government every year. Though the contribution is monthly the interest rate is set annually. In case of death of the employee the money in the EPF account is handed over to the nominee. The employer's contribution is tax free and the employee's contribution is eligible for tax deduction. EPF withdrawals are permitted under certain circumstances.

The government scheme along with employee provident fund schemes covers almost 93 percent of the pension subscribers. There are other tertiary target pension programs which are usually for the poor. One such pension program is the National Old Age Pension Scheme launched in 1995, targeted to people above 65 years old and below the poverty line. The contribution from the government coupled with the state was as low as Rs.200 ($3) per month. Other such programs include micro pension schemes run by MFI and NGOs. The purpose of these pension schemes is to reach the disadvantaged people at an affordable price. Since in our sample we only look at private and public sector employees from white collar jobs these schemes are not relevant for our sample.

2.2.1.1 The Reform

The Civil Servant Pension Scheme was becoming unstable because of the fiscal stress it was creating on the central and state government. With the demography aging and increased lifespan of the elderly, the fiscal stress on the government increased three folds in the past decade. In 2004, 16 percent of the tax revenue from the state government and 12 percent from the central government were spent on making pension payments (Sadhak, 2009). A reform in the government pension scheme was long overdue. The Government of India introduced the New Pension System (NPS) in 2004, new government employees joining after January 1st 2004 will be placed under the NPS instead of the traditional CSP and other government pension schemes. The scheme was announced to the public in December of 2003. The NPS is a defined contribution pension scheme which requires 10 percent contribution from the employee and ten percent contribution from the employer (government). The NPS will have two tier, Tier – I is mandatory for all government employees joining post 2004 and Tier-II is optional. In Tier-I account the employee will contribute 10 percent of his basic pay, deducted directly from his salary and the government will make an equal matching contribution. The government of India established the Pension Fund Regulatory and Development Authority (PFDRA), as the prudential regulator for the NPS. Funds from Tier-I account are mostly invested in government and corporate bonds. Tier-I is mostly non-withdraw able. Tier-II is optional and at discretion of the government employee. Government will not make any contribution to the Tier-II account. There is no limit to the number of withdrawals from the Tier-II account. There is also a choice for the employee to which fund to invest in. Tier-II account can be looked at as a medium term investment tool. Since there is no locking period for the funds there is no tax-benefits also. To make the NPS more effective as a retirement saving tool there is provision for mandatory annuitization. Individuals can normally exit at the age of 60 or at superannuation from Tier-I of the pension system. At the time of exit the individual would be mandatorily required to invest at least 40 percent of the pension wealth to purchase an annuity (from an Insurance Regulatory and Development Authority (irda)-regulated life insurance company). Finally the existing defined benefit CSP scheme would not be available to any government employee joining the government

sector post Jan 1st, 2004.

2.2.1.2   Comparison between the NPS and CSP

In this section I will discuss why the government sector employees who joined the government sector post 2004 and fall under the NPS are worse off in terms of retirement security than government sector employees who joined the sector prior to 2004 and fall under the CSP. The first major difference between the two schemes is the fact that the NPS is defined contribution (DC) whereas CSP is defined benefit (DB). A change from DB to DC schemes shifts the risk of investment to the employee from the employer. The employee is more exposed to variability in retirement income. (Broadbent, Palumbo and Woodman, 2006). This variability in retirement income in itself is viewed as a dis-utility by the employee. DB benefit pension schemes are back loaded and favors the long tenured employees. Government sector jobs are one of the most secured jobs in India, so people joining the government sector end up working in the sector all their lives. Thus with high job security and long tenure the DB scheme was highly beneficial for these employees. Further a secured job with predictable growth and defined income stream post retirement made working in government sector very reliable. The change from DB to DC takes away this security from the employees. Other issues which make the NPS worse than the CSP are the following: the income under CSP was indexed to inflation whereas under NPS it is not and that the contribution rate went from 0 percent to 10 percent for the government employees.

Now a move from DB to DC is not necessarily detrimental to the employee as they have more labor force mobility among others. So what is more important is how this change was perceived by the people. The introduction of the NPS was met by protests all over the country. The people who joined the sector post 2004 viewed this as an act of stripping them of their pension rights. They argued that since the return under NPS is market driven there is no guaranteed return.

Another major indication that the government themselves believed that the CSP is better than the NPS is the fact that the Armed Forces were kept under CSP. Armed Forces are the most valued functionaries of the nation, the fact that the government is keeping the armed forces under CSP was

further seen as a reason that the CSP is better than NPS.

Finally there are reports (K.V Ramesh, 2013) which calculate returns under the two pension scheme, they look at returns under various market interest rate, ranging from 8-12 percent. Taking into account that NPS is not inflation proof they show that at all plausible market return rates the CSP outperforms the NPS.

2.2.2   Unequal Investment in Sons and Daughters in India

There is a large literature which analyses the issue of son preference in India. Pande and Astone (2007) discuss the existence of son preference in rural India and analyze the role of individual versus structural effects that might generate this son preference. This taste based preference among parents has been seen to result in pronounced effects on the sex composition of children in families (Clark, 2000; Mutharayappa, 1997; Arnold et al 1998) as well as in parental investments in child care (Barcellos et al 2014), nutrition (Chen, Huq and D'Souza, 1981; Das Gupta, 1987; Behrman 1988), vaccinations (Borooah 2004) and healthcare (Basu, 1989). There is evidence that girls receive less investment in their human capital as compared to boys in both developing as well as developed nations (Behrman, Pollak and Taubman 1986). There is also evidence that discrimination begins in the womb (Bharadwaj & Lakdawala 2013). They find that mothers visit antenatal clinics and receive tetanus shots more frequently when pregnant with a boy.

The preference based reasons cited for these gender gap in investment includes the patriarchal nature of the Indian society, son preference, poverty, illiteracy and marriage laws. Unfortunately there also exists economic reasons for gender gap in educational investment. Girls face poorer economic incentives to invest in schooling than boys because they reap lower labor market returns to education than boy (Kingdon, 1998). Women also suffer high levels of wage discrimination in the Indian urban labor market, but education contributes little to this discrimination (Kingdon and Unni, 2001). Overall in the literature we see that both economic and preference based reasons drive gender gap against the girl child.

2.2.3 Private versus Public Schools in India

Public schools also knows as government schools have rapidly lost their enrollment to private schools in India (Murlidharan and Kremer, 2006). Due to poor performance of over-crowded government schools there has been a steady surge in private school enrollments (Kingdon, 2007; Wadhwa, 2009). Parents have started to value education more and are sending their children to private schools which are more expensive compared to government schools (Goyal and Pandey, 2009). The teachers in private school are more accountable as they can be fired whereas as government school teachers are mostly tenured and have low accountability (e.g. Probe Report, 1999). There have been a few a studies comparing the performance of private versus public schools in India. These studies find that the private schools out performs the public schools on grounds of test scores, English ability and other child outcomes (Goyal and Pandey, 2009; Kingdon, 2007; Tooley, Dixon, Shamsan, and Schagen, 2010).standardizing for home background and controlling for sample selectivity greatly reduces the raw average achievement advantage of private school students over public school students, but does not wipe it out (Kingdon 1996). Private schools being more expensive are usually attended by children from better off and better informed families (Goyal and Pandey, 2009). Thus enrolling your child in private schools could be viewed as a form investment on the parent's part as private schools are more expensive and associated with better outcomes for the child. Appendix table A3 summarizes private school enrollment in the data, we see that the average enrollment for boys are higher than girls.

2.3 Literature Review

The Old Age Security hypothesis views children as capital goods. They are used as investments to transfer resources from present to future (Srinivasan, 1988). One implication of this statement, which is the core of this paper is that if the male child produces higher return than the female child then by the Old Age Security hypothesis the male child should get higher investment. There is a huge literature which talks about old age security and son preference as potential reasons for increased fertility (Jensen, 2003; De Vos, 1985; Hodinott, 1992; Nugent, 1985; Zhang and

Nishimura, 1993; Mutharrayappa, 1997), but there is a dearth of good empirical papers which establishes a causal link between old age security of parents and differential investment in children by gender. This dearth of papers could be attributed to the lack of good intergenerational data in developing countries.

Ebenstein and Leung (2010) discuss a potential link between old age security and son preference. They use a voluntary pension program introduced in rural China in 1990 to confirm two hypotheses. They establish that parents with sons are less likely to enroll in the program and less likely to save for old age. They also establish that the implementation of rural old age pension programs mitigated the increase in the sex ratio observed in counties that had at least partial adoption of the program. Both the findings support the argument that sons provide higher old age security than daughters and preference for sons could be driven by this old age security motive. What the paper fails to capture is how significant this old age security motive is in explaining the differential treatment of a boy and girl child, which I intend to capture in my paper.

With the access to new data there has been some recent study which looks at the effect of Old Age Pension on inter generational resource allocation. Mu and Du (2015) link educational investment in children and old age security for parents. They find that as old age security increases parents invest more in their children's education. They talk about two reasons why increase in old age security should affect investment in children's education. Firstly, the lifetime income of parents is increasing; so by income effect the investment should increase. Secondly, as old age security increases, the parents are less reliant on their children for old age support so they should decrease their investment. They conclude that the income effect is greater than the second effect so they see a rise in the investment in children's education. This paper again does not look at the differential investment by gender which is the focus of my paper. Another discussion paper, Chen (2015) looks at the effect of old age security on inter generational living arrangements. In the paper the author finds that as old age security increases for the parents, they are less likely to live with their children.

2.4   Model

I construct a very simple two period model to explain how dependency on children for post retirement consumption brings in gender gap in investments.The model is constructed under the assumption that there is no taste based gender discrimination by the parents when it comes to investing in their children. The purpose of this assumption is to highlight the fact that even in the absence of taste based gender discrimination, parents invest deferentially in their children by gender.

I consider a simple two period model, were parents derive utility from their consumption in period one and period two and they also derive utility from their children's well being in period two. It is a time separable concave utility function. In period one parents earn income Y_1, they choose how much to consume C_1, how much to invest in their boy child S_B, how much to invest in their girl child S_G and how much to invest in assets A. In period two parents consume C_2, earn lump sum pension income T , they get return from their son $R_B S_B$, they get return from their daughter $R_G S_G$ and the return from the asset $R_A A$. For interior solution we assume that return from assets is higher than return from children[6].

I prove three propositions :

Proposition 1 : *If the cumulative return from investing in a boy child is higher than the cumulative return from investing in the girl child then parents invest more in the boy child compared to the girl child.*

Proposition 2 : *As post retirement income increases the gender gap in investment decreases.*

Proposition 3 : *For household with children of one gender as pension income T decreases investment in children increases if the cumulative return to investing in children is higher than that of investing in the asset. The increase/decrease in investment is proportional to the difference in return between the asset and the children*

---

[6]If return to assets is lower than return to children then parents would not invest in assets as investing in children produce greater returns and added altruistic utility

2.4.1 Maximization Problem for Households with Children of Both Gender

Agent maximizes

$$\max_{\{C_1,C_2,S_B,S_G,A\}} U(C_1) + \beta\{V(S_B) + V(S_G) + U(C_2)\} \tag{2.1}$$

subject to:

$$C_1 = Y_1 - S_B - S_G - A \tag{2.2}$$

$$C_2 = T + R_B S_B + R_G S_G + R_A A \tag{2.3}$$

Combining the FOC w.r.t S_G and S_B we get:

$$\frac{V'(S_B) - V'(S_G)}{U'(C_2)} = R_G - R_B \tag{2.4}$$

From the existing literature we know that $R_G < R_B$ (Jayachandran 2014, Das Gupta et al. 2003, Kishor 1993), i.e. the return and the dependency are higher for boys. This implies $V'_B(S_B) < V'_G(S_G)$. Since the utility functions are concave, we get $S_B > S_G$.

Proposition 1 : **_Proved_**

From the FOC w.r.t R_A we get:

$$U'(C_1) = \beta R_A U'(C_2) \tag{2.5}$$

Differentiating (5) w.r.t to Pension Income T we get:

$$\frac{\frac{dC_1}{dT}}{\frac{dC_2}{dT}} = \frac{\beta R_A U''(C_2)}{U''(C_1)} > 0 \tag{2.6}$$

When pension income changes consumption in both period moves in the same direction. As consumption in both period cannot decrease with an increase in income we get

$$\frac{dC_2}{dT} > 0 \tag{2.7}$$

Now differentiating (4) w.r.t T and using (6) we get:

$$V''(S_B)\frac{dS_B}{dT} - V''(S_G)\frac{dS_G}{dT} = U''(C_2)\frac{dC_2}{dT}(R_G - R_B) > 0 \tag{2.8}$$

Under the Assumption that V() is concave we get:

$$\frac{dS_B}{dT} < \frac{V''(S_G)}{V''(S_B)}\frac{dS_G}{dT} = k\frac{dS_G}{dT} \qquad (2.9)$$

Where k is less than 1 as:

$$V''(S_B) > V''(S_G) \qquad (2.10)$$

If

$$\frac{dS_G}{dT} \subset (-\infty, \frac{1}{1-k}] \qquad (2.11)$$

We get

$$\frac{dS_B}{dT} < \frac{dS_G}{dT}$$

Proposition 2 : ***Proved under the assumption that*** $\frac{dS_G}{dT} \subset (-\infty, \frac{1}{1-k}]$

2.4.2 Maximization Problem for Households with Children of one Gender

Agent maximizes

$$\max_{\{C_1, C_2, S_I, A\}} U(C_1) + \beta\{V(S_I) + U(C_2)\} \qquad (2.12)$$

Where

$$I \subset (B, G) \qquad (2.13)$$

subject to:

$$C_1 = Y_1 - S_I - A \qquad (2.14)$$

$$C_2 = T + R_I S_I + R_A A \qquad (2.15)$$

Combining the F.O.C w.r.t S_I & A

$$\beta V'(S_I) = U'(C_2)(R_A - R_I) \qquad (2.16)$$

Differentiating w.r.t T

$$\frac{dS_I}{dT} = \frac{1}{\beta}\frac{U''(C_2)}{V''(S_I)}\frac{dC_2}{dT}(R_A - R_I) \qquad (2.17)$$

From the above equation we get

We know that, $U''(C_2) > 0$ & $V''(S_I) > 0$ & $\frac{dC_2}{dT} > 0$

If $R_A < R_I$

$$\frac{dS_I}{dT} < 0 \qquad (2.18)$$

If $R_A > R_I$

$$\frac{dS_I}{dT} > 0 \qquad (2.19)$$

And

$$\frac{dS_I}{dT} \propto (R_A - R_I) \qquad (2.20)$$

This implies that the decrease or increase in investment in children is proportional to the difference between the return to asset and children. Since $R_B > R_G$, and $\frac{dS_I}{dT} \propto (R_A - R_I)$ the decrease in investment would be higher for the girl child in an all girl household compared to a boy in a all boys household.

Proposition 3 : ***Proved***

Thus with a decrease in pension we should see an increase in the Intra-household gender gap between all girls and all boys household if the return to the boy child is greater than that of the girl child.

2.5 Empirical Framework

2.5.1 Main Identification Strategy

The empirical strategy hinges on a differential decrease in retirement benefits across the private and the public sector employees. We use this variation in difference in difference in difference (DDD) strategy. The first difference is over time as public sector employees joining the sector post 2004 are exposed to a different retirement benefit scheme compared to the people joining prior to 2004. The second difference is over the work sector of the father of the child namely private versus public, this difference will account for other time trends which affects both the sector. The third difference is

among the child gender to capture the differential effect of the retirement benefit change by gender. The net differential effect by gender of the pension contraction for the parents in the public sector on various child outcomes is estimated by the following regression equation.

$$Y_i = \beta_1 Sector_i + \beta_2 Post_i + \beta_3 Gender_i$$
$$+ \beta_4 Sector_i \times Post_i + \beta_5 Sector_i \times Gender_i + \beta_6 Post_i \times Gender_i \quad (2.21)$$
$$+ \beta_7 Post_i \times Gender_i \times Sector_i + \gamma X_i + \epsilon_i$$

Sector is a binary variable which takes value one if the parent works in the public sector and is zero if the parent work in the private sector. Post is a binary variable which takes value one if the parent joined the current sector post 2004 and is zero if the parent joined the current sector prior to 2004. Gender is a binary variable for the sex of the child, zero for a boy and one for a girl. We include a full set of interaction terms in the regression equation namely sector-post, sector-gender and post-gender. The gender differential treatment effect of the pension reform is captured by β_7. We hypothesize to see a negative a β_7 which would imply that as the retirement security decreases girls are worse off related to boys. The main identifying assumption is that the evolution of gender gap in investment in child's education, over time follows the same trend for the public and private sector employees. In the above baseline specification we are not explicitly using within household variation i.e we are comparing across all household within children[7]. In section 7.3 we include household fixed effects, this helps us compare inter versus intra-household effect.

We also include a rich set of individual, father, mother and household specific control variables in the above equation. Child characteristics includes age, grade and other variables depending on the outcomes we are looking at. Father characteristics include age, salary[8] and years of education[9].

---

[7]Hence the sample consists of households with only boys, only girls and both boys and girls. This is more representative of the population rather than just looking at household with both boys and girls. Also this gives us a bigger sample size to work with

[8]One might argue that this is endogenous to the policy change, since I am interested in the effects of the policy that work through the dependence on children and not current income, I control for current income. For the same reason I also control for mother's work status.

[9]We are unable to control for father's industry here as the measure for industry is not consistent across the two panels at a micro level.

Mother characteristics include age, years of education and work status[10]. House hold characteristics include caste, income, poverty level, highest education level for adult male and female, number of children and district.

We analyze two different set of outcomes, first set contains outcome variables which are direct indicators of investment in the child. These variables include the type of school the child is enrolled in private or public school, medium of instruction at school. The second set of variables are more derivative as they reflect the returns to the investment in a child. These variables include English speaking ability of the child, mathematics ability, reading ability and writing ability.

2.5.2 Common Trend and Falsification

The identifying assumption for the validity of our results is that the evolution of gender gap in investment in child's education, over time follows the same trend for the public and private sector employees. This might not hold if gender attitudes evolve differently for the employees of the two sector. Besides economic returns there are other social and cultural norms that make parents discriminate against their female child if such taste based gender attitudes evolve differentially over time for the two sectors then that will be a threat to the validity of the identifying assumption. To show that this is not the case I will provide two arguments: first the fathers selected in the sample in private and public sector have a considerable amount of overlap when it comes to the industry[11] they are employed in and second we will use a difference in difference strategy to show that gender attitudes does not evolve differentially across the two sector. We will use the data on gender relation question asked to an ever married woman in the house hold, these gender relation questions capture the gender attitudes of the household. I will use a difference in difference strategy to see that if these gender attitudes have changed differentially for both the sectors over time. The questions we specifically look at are the following: is the women frequently beaten, is she allowed to visit her natal family, does she have her name on important rental documents , does she have

---

[10]I do not have data for years of work experience for mother and father but it is usually collinear with age and years of education.

[11]Table 4 shows this.

to cover her face and does she have to take permission to go out of the house.These questions capture the attitude of the household towards women and would capture the non-economic part of discrimination against women. Another threat to the identification strategy is if selection into private versus public sector changes as a result of the policy change. To show that the people entering the sectors post policy change are not different I estimate equation 22 with background individual and family characteristics as outcomes

$$G_i = \beta_1 Sector_i + \beta_2 Post_i + \beta_3 Sector_i \times Post_i + \gamma X_i + \epsilon_i \tag{2.22}$$

G_i are variables that capture gender attitudes of household, these gender relation question were asked to an ever married woman. We check that whether the time trend in gender attitudes vary differentially across the employees of the two sector. Here sector represents her husbands work sector. A β_3 significantly not different from zero would imply that the time trends in gender attitudes are similar for the two sector. We want to see if the taste based discrimination towards women have changed differentially for the households of the current student.

I also run a falsification test to make sure that what we are picking up is not some residual time trend effect. i set treat equal to one for private sector and treat equal to zero for agricultural sector. I run the baseline specification (equation 21) between the private sector employees and agricultural sector employees instead of public versus private. Since the reform was for the public sector employees we should not see any significant coefficient when we change the sectors.

2.6 Data description

The data I use for this paper is panel data from India Human Development Survey Data (IHDS), Waves I and II, which are two nationally representative surveys conducted in 2005 and 2011. The data set is very rich as it contains data on household and individual level characteristics. The richness of the data regarding gender specific questions makes it ideal for our study. We will have two samples of interest for whom we looks various outcomes. The purpose of this is to better identify the causality between old age security and child investments at various stages for various outcomes. The first sample consists of all current students who are enrolled in school or some

sort of degree program and are less than 18 years old. Table II presents the descriptive statistics for this sample. The outcome variables we look at for this sample is type of school public or private, medium of instruction in school English or regional language and the English ability[12] of the child. I present the descriptive statistics decomposed by parent's sector of work and whether they joined the sector post or prior to 2004. We see that the public sector households does better than the private sector households, they are more educated, wealthier and belongs to higher caste. This trend is common across all three samples. Compared to the private sector household we see that the children in the public sector households are more likely to be enrolled in private schools or schools where the medium of instruction is English. When it comes to comparing between the people who joined post and prior to 2004 we see that the prior people are in general a little older but other outcomes are comparable across the two sectors.

The second sample consists of children age 8-11, IHDS administers reading, writing and math tests to this group of children. Table II presents the descriptive statistics for this sample. The outcome variable we look at for this sample is test score for mathematics, reading and writing. The scores are standardized[13] at a district level. We also construct a variable Score which is the summation of the three scores. Again I present the descriptive statistics decomposed by parent's sector of work and whether they joined the sector post or prior to 2004.

For both the sample the fathers work in either the public or the private sector. I further restrict the sample to fathers who work in white collar jobs. I identified jobs as white collar jobs [14] and then segregated them into private and public sector. This was done to ensure that the time trends are comparable across the two sectors. Table IV presents the distribution of percentage of people employed in various industries across the two sectors, the list is not exhaustive. We see a considerable overlap of industry across the two sectors, private and public.

For analyzing common trends in the two sector I use the data on eligible woman from the IHDS

---

[12]English ability is a categorical variable taking the value 0 if the child has no knowledge of the language, 1 if the child has some knowledge and 2 if the child is fluent in English.

[13]I compute the z score at a district level by child age

[14]see appendix for the full list

survey. In the data gender relation questions were asked to a married woman in the household. These question captures the gender attitudes of the household toward females. Questions such as is the women frequently beaten, is she allowed to visit her natal family, does she have her name on important rental documents and does she have to take permission to go out of the house. The answer to these questions reflect the patriarchal nature of the household.

Appendix table 6 presents the means of the outcome variables by the three difference groups public versus private, post reform versus pre reform and boy versus girl. These could be read as the raw DDD results.

2.7 Results

2.7.1 Current Students

Table V[15] illustrates the overall gender differential effect of contraction of post-retirement security. First we look at the sample of current students. We are looking at three outcomes for this sample, whether the child is enrolled in a public schools or a private school, whether the medium of instruction is English or a regional language and the English ability of the child. Specification one contains full set of controls including father characteristics, mother characteristics, household characteristics and survey year dummy. Specification two does not contain household characteristics, the purpose of the two specifications is to show that the estimates does not vary a lot under different specifications. Under specification one which contains the full set of controls we see that compared to a boy child a girl child is 0.0701[16] points less likely to be enrolled in a private

---

[15]Table V presents the DDD estimates from equation 21. The top panel presents the results from the current student sample and the bottom panel presents the results from the test scores. Each row presents the result from equation 21 with a particular outcome variable. For example row 1 presents the DDD estimate with enrollment in private school as the outcome variable. Column 1 presents the results from specification 1 which includes the full set of controls: child, parent and household characteristics and column 2 presents the result from specification 2 which does not control for household characteristics.

[16]The pre-reform means for private school enrollment are 0.366 and 0.276 for boys and girl respectively. If we look at the DD separately for boys and girls we see that the 7pp increase in gender gap is driven by 1.62 pp decrease in private school enrollment for the girl and 5.61 percentage point increase in private school enrollment for the boys. I think these are reasonable compared to

school post reform in the public sector, the effect is significant. This means that as post retirement security is decreasing the gender gap in enrollments in private schools increases compared to public schools. We see that the estimates do not vary much under the second specification which does not include household controls. The next outcome is whether the child is enrolled in a school where the medium of instruction is English or a regional language. Under specification one we see that compared to a boy child a girl child is 0.0684 points less likely to be enrolled in a school where the medium of instruction is English, post reform in the public sector. The result is significant and is in line with our hypothesis. The estimate does not vary much under the second specification. Finally we look at the effect on the English ability[17] of the child we see that compared to a boy child a girl child is 0.0684 points less fluent in conversing in English. The effect is not significant but in line with our hypothesis.

We report the Difference-in-Difference for the boy child and girl child separately in Appendix table A1. We see that the decrease in retirement security for the parents increases the investment in the boy child whereas it decreases the investment in the girl child. The effects are not significant, but the overall effect on the gender gap is significant.

2.7.2 Test Scores

For the second sample we look at standardized test scores administered by IHDS. The results are presented in Table V. Under specification one, we see that compared to a boy child a girl child scores 0.07 points less in math post reform in the public sector , the estimate is not significant. This implies that compared to boy child a girl child performs worse post reform, thus as old age security decreases the performance of the girl child falls compared to the boy child which in indicative of increasing gender gap in human capital investment. We see similar results for writing and SCORE which is the summation of all the scores. None of the results are significant but the signs are consistent with our hypothesis.

the pre-reform means.

[17]English Ability is reported by the parent or grandparent so there might be some non-standard measurement error, the sibling fixed effect model will account for this

Overall we see that as post retirement security decrease girl child are getting worse of compared to the boy child. They are less likely to be enrolled in good private school compared to boys or in school where the medium of instruction is English. They also perform worse in exams compared to boys and have lower scores in mathematics and writing, they are also less likely to be speaking in English.

2.7.3   Inter-Household Vs Intra-Household

Appendix table A2 reports the results from our DDD estimation (equation 21) including household fixed effect. We see that the significant effect we were getting in our triple difference estimates vanishes after accounting for sibling fixed effect. To analyze this we divide our sample into families with children of only one gender and household with children of both gender. Table VI reports the results from running the triple difference separately for the two samples. We see that the rise in gender gap is primarily driven by families with children of one gender. We further divide the sample of one gender household into families with one child and families with more than one child. We see that the effect is primarily driven by families with more than one child and child of only one gender.

To further deepen our understanding of why the results are being driven by them we look at the characteristics of the two kinds of households. Appendix Table A8 presents the characteristics of household with children of one gender versus household with children of multiple gender. We see that household with children of only one gender looks similar to household with children of multiple gender in terms of age, education and wealth.

There could be number of reasons why we see this, in the theoretical model in proposition 3 we saw that for families with children of only one gender, as retirement security decreases there will be an increase in investment in children if the return from investing in a child is greater than that of investing in an asset. If the return from the asset lies between the return from a boy child and the return from a girl child ($R_B > R_A > R_G$) then that explains the finding. Proposition 2 states that for household with both girl child and boy child gender gap increases as pension income decreases

but proposition 3 holds under less stricter assumption than proposition 2. Proposition 2 holds if $dS_G/d_T \subset (-\infty, \frac{1}{1-k}$ whereas proportion 3 holds for all values of S_G. This might explain why we see a inter-household effect rather and not a intra-household effect.

Other possible reason could be that parents of families with only girl children are more worried about their retirement security and they shift resources from their girl children into other assets. In appendix table A7 I present the result from running a DD with sector(public versus private) and reform (post 2004 versus pre) to look at the effect of the policy change on investment in other assets[18] we see that families with only girl child invest more in retirement saving and other long term savings instrument compared to families with only boys but the effect is not significant.

If we look at table A6 we see that for both pre and post reform and public and private sector, the girl child on an average is less likely to be send to a private school. In table A8 we look at private school enrollment for a girl child and boy child by family composition. For the boy child we see that the average private school enrollment is similar for household with one child, household with children of all gender (all boys household) and household with children of all gender (mixed). Whereas for the girl child we see that the private school enrollment varies by family composition. Girls from household with only one child on average has the highest private school enrollment. Followed by girls from household with children of both gender (mixed). Finally the girls from household with children of one gender (all girls household) has the lowest average enrollment. Hence in the entire population, irrespective of the policy reform and working sector of the parents, we see that girls from all girls household has the lowest private school enrollment whereas for boys the sibling composition does not matter. This supports the finding that the impact from the policy reform were mainly driven by household who has only children of one gender. The girls from the all-girls household in general receive less investment compared to girls in mixed household, the policy reform might further magnify this discrimination.

---

[18]A binary variable indicating whether the household has invested in some other pension plans or LIC policy

2.8   Common Trends and Falsification Test

2.8.1   Trends in Gender Attitude

In table VI I present the results from equation twenty two. Since the people joining the sectors post 2004 are relatively younger than people joining prior to 2004 the gender relation might evolve differentially over time for the two sector. The first outcome I look at is whether the woman is made to practice "ghunghat" or "purdah", this is an old patriarchal Indian tradition where the married woman is made to cover her face at all times. We see that the coefficient on the interaction term is 0.0011, and it is not significant. Since the coefficient is not significantly different from zero it implies that the trend does not vary differentially for the two groups. The next outcome we look at is if the women is beaten frequently by her husband, the coefficient on the interaction term is -0.0164 and is not significant. We find very similar results for the other indicators of gender attitude all of them have small coefficients which are not significantly different from zero. Overall we can summarize that the gender attitudes does not evolve differentially over time for the two sectors.

2.8.2   Falsification Test

The results are presented in the lower panel of table VI. To further establish that the effects that we are picking up in our main estimation are not some spurious effect driven by the private sector, we look at the outcomes for current students across the private[19] and agricultural sector. We see that the coefficients are smaller, of the opposite sign and insignificant.

2.8.3   Selection into Private and Public Sector

Another threat to the identification strategy is if selection into public versus private sector changes as a result of the policy change. To see if people entering the public sector versus private sector post policy change are different from the people who joined prior to the policy change I use a difference in difference model across Sector (Public versus Private) and Post (joined post or prior to policy

---

[19]The sector dummy takes the value 1 for private and 0 for agriculture

change). The outcome variables I look at are individual characteristics and family characteristics such as education, father's education, caste, religion, urban or rural, district and below poverty line. Table VII reports the results from the difference in difference. I see that these characteristics do not vary between the people who entered the two sector over time. Variables such as fertility, individual income, household income, assets and mother's income might be affected by the policy change and they in turn might affect gender gap. To account for this I control for them in specification1 for equation 21.

2.9 Discussion and Conclusion

We see that the reform of 2004 for the public sector employees is associated with worse outcomes for the girl child compared to the boy child. We established that the 2004 reform was viewed as a reduction in post-retirement security by the public sector employees, which implies that when the perceived retirement security decreases, gender gap in various outcomes increases. As we have seen this change in gender gap is not driven by change in gender attitudes as gender attitudes remain the same for the two sets of household. This would further imply that the effect we see is solely driven by economic incentives rather than taste based change in gender attitudes. What we are capturing is the effect of the contraction of post-retirement benefit and not some interaction effect between taste and incentive. This result has good implication for the future, if people react to economic incentives to change their behavior towards gender then with the right incentives the gender gap between male and female could be reduced in India. The New Pension System has also extended its coverage to the private sector and unorganized sector in the recent years. The take up is slow but it is expected to expand the safety net. This will give us another chance to look into this topic in the future.

Table 2.1: Pension System in India

Government Sector	Private Sector	Unorganized	Government Sector Post 2004
• Civil Servant Pension • Defined Benefit • For employees of Central and State Government • Contribution Rate 0 • Monthly pension • 50% of wages in the last 10 months of employment • Consumer price index	• Employees Provident Fund (EPF) • Defined Contribution • Any Company with more than 20 employees. • 12% from both employee and employer. • Plagued with early withdrawals. • Accumulation plus interest upon withdrawals.	• National Old Age Pension Scheme • Below poverty line (BPL) who were aged 65 and above. • Micro-pensions and Other Alternatives • Very low coverage	• New Pension Scheme • Defined contribution • Involuntary Tier I account 10% contribution by Employee and Government. • Non-withdrawal account. • Can exit at or after the age of 60 years • Mandatory invest 40% pension for annuity. • Tier II account, voluntary, withdrawal able, no contribution from government.

94

Table 2.2: Descriptive Statistics for Current Students

	Post 2004		Prior 2004	
	Public	Private	Public	Private
English Ability	0.6534	0.4527	0.5378	0.3562
English Medium	0.4221	0.3196	0.3669	0.3138
Private School	0.5254	0.496	0.4481	0.4831
Ever Repeated	0.0539	0.0387	0.063	0.0624
Child Gender	0.47	0.47	0.46	0.48
Child Age	11.81	10.71	12.07	10.66
Father's education	11.85	10.51	11.896	10.49
Father's age	40.33	34.51	43.99	40.72
Father's salary	15557	6914	8006	2875
Mother's education	8.57	7.59	8.143	7.42
Mother's Age	37.27	33.51	37.21	34.99
Mother's work status	0.16	0.19	0.1213	0.1319
Household				
Highest adult education	13.13	11.62	12.59	11.35
Highest female education	10.50	8.75	9.08	8.1583
Highest Male education	12.52	10.92	12.27	10.88
Number of children	1.37	1.33	1.8	2.07
Caste	3.39	3.36	3.337	3.27
Below poverty Line	0.038	0.96	0.0275	0.117
N	1336	2171	4500	3439

Table 2.3: Descriptive Statistics for the Test Score Sample

| | Post 2004 | | Prior 2004 | |
	Public	Private	Public	Private
Math	1.98	1.82	2.16	1.91
Reading	3.19	2.91	3.24	3.07
Writing	1.48	1.36	1.01	0.95
Child Gender	0.49	0.46	0.45	0.49
Child Age	9.55	9.57	9.56	9.46
Father's education	11.91	10.77	12.12	10.66
Father's age	39.87	33.47	44.6	40.08
Father's salary	15360	6958	12868	4694
Mother's education	9.42	7.53	8.258	7.09
Mother's age	35.24	33.89	35.12	33.52
Mother's work status	0.2	0.22	0.13	0.14
Household				
Highest adult education	13.26	11.79	13.01	11.65
Highest female education	10.88	8.97	9.8	8.63
Highest Male education	12.62	11.07	12.64	11.11
Number of children	1.21	1.16	1.73	2.03
Caste	3.366	3.35	3.26	3.2
Below poverty Line	0.04	0.08	0.03	0.12
N	320	480	1032	864

Table 2.4: Distribution of Industry Across the Public and Private Sector

Industry	Public Sector Employees	Private Sector Employees
Clerical	25.8	13.77
Teachers	17.33	9.77
Police	16.49	7.18
Government officials	5.12	1.11
Village officials	3.53	0.52
Clerical supervisor	2.9	1.48
Service	2.72	4.77
Transportation conductors	2.33	2.04
Accountants	2.07	1.92
Engineers	1.89	1.7
Mail distributors	1.68	0.48
Book-keepers	1.56	3.33
Trans/communication	1.53	0.78
Nursing	1.44	1.78
Manager finance	1.41	0.26
Physicians	1.32	0.96
Telephone op	1.32	0.96
Eng. tech	1.29	0.85
Typists	1.11	1.22
Managerial	0.93	2.48
Computing op	0.84	0.59
Sales, shop	0.84	19.02
Manager manufacturing	0.63	1.41
Other scientific	0.57	0.26
Lawyers	0.48	1.44
FIRE sales	0.33	2.41
Physical science tech	0.3	0.22
Social scientists	0.27	0.33

	(1)	(2)
Current Students		
Private School	-0.0711**	-0.0717**
	(0.0342)	(0.0359)
English Medium	-0.0684**	-0.0683**
	(0.0340)	(0.0333)
English Ability	-0.0667	-0.0691
	(0.0517)	(0.0530)
N	11462	11336
Test Scores		
Mathematics	-0.0765	-0.0794
	(0.1217)	(0.1383)
Reading	0.029	0.0429
	(0.1169)	(0.1250)
Writing	-0.044	-0.0387
	(0.1588)	(0.1687)
SCORE	-0.1034	-0.0843
	(0.3349)	(0.3721)
N	2632	2603

*Table V contains estimates of β_7 from equation (8) for various outcome variables.

*The top panel presents the results from the current student sample and the bottom panel presents the results from the test scores.

*Each row presents the DDD estimate from equation 21 with a particular outcome variable.

*Standard errors are reported in brackets and are clustered at the district level.

*Specification (1) includes full set of controls : child characteristics, father characteristics, mother characteristics and household characteristics.

*Specification (2) does not include household characteristics.

Table 2.6: Inter versus Intra-Household

Current Students			
	DDD	DDD (Mixed child sex HH)	DDD (All boys or all girls HH)
Private School	-0.0711**	-0.0083	-0.1370**
	(0.0342)	(0.0372)	(0.0646)
English Medium School	-0.0684**	-0.0326	-0.1204**
	(0.0340)	(0.0321)	(0.0625)
English Ability	-0.0667	0.0088	-0.1515**
	(0.0517)	(0.0470)	(0.0776)
Observations	11462	5866	5697

Current Students		
	DDD (one child HH)	DDD (All boys or all girls HH with more than one child)
Private School	0.0134	-0.2547**
	(0.0490)	(0.1052)
English Medium School	-0.0078	-0.2345**
	(0.0739)	(0.1035)
English Ability	-0.1805**	-0.1404
	(0.0935)	(0.1337)
Observations	2794	2838

*Table V contains estimates of β_7 from equation (8) for various outcome variables.

*Each row presents the DDD estimate from equation 21 with a particular outcome variable.

*Standard errors are reported in brackets and are clustered at the district level.

*Column 1 of the top panel shows the DDD estimate from equation 21 for the entire sample.

*Column 2 of the top panel shows the DDD estimate from equation 21 for household which has children of both gender.

*Column 3 of the top panel shows the DDD estimate from equation 21 for the household which has only girl child or boy child.

*Column 1 of the bottom panel shows the DDD estimate from equation 21 for household which has only one girl child or boy child.

*Column 2 of the bottom panel shows the DDD estimate from equation 21 for household which has only more than one girl child or boy child.

Table 2.7: Trends & Falsification

Trends in Gender Attitude	(1)	(2)
Cover Face	0.0011	-0.0038
	(0.0256)	(0.0299)
Beaten Frequently	-0.0164	-0.0040
	(0.0264)	(0.0282)
Permission to go outside	0.0101	0.0224
	(0.0228)	(0.0225)
Visit Natal Family	-0.0166	-0.0127
	(0.0423)	(0.0441)
Name on Rental Paper	0.0102	0.0191
	(0.0532)	(0.0517)
N	9521	8867
Falsification	(1)	
English Medium	0.0413	
	(0.0513)	
Private School	0.0259	
	(0.0235)	
English Ability	0.0363	
	(0.0348)	
N	17326	

*Table VII top panel presents the estimates of β_3 from equation (22) for various outcome variables.

*Each row presents the DD estimate from equation 22 with a particular outcome variable.

*Standard errors are reported in brackets and are clustered at the district level.

*Specification one includes all women who were interviewed and is married to someone in the private sector or public sector. Specification one includes all women who were interviewed and is married to someone in the private sector or public sector and has at least one child.

*Table VII bottom panel presents the estimates of β_7 from equation (21) for various outcome variables, where sector is equal to 1 for private sector employees and 0 for Agricultural sector employees.

Table 2.8: Selection into Sector

Education	-0.0321
	(0.1197)
Father's Education	-0.0609
	(0.4877)
Caste	-0.0922
	(0.0710)
Religion	-0.0731
	(0.0513)
Urban/Rural	-0.0063
	(0.0289)
District	-0.0241
	(0.7408)
Below Poverty Line	0.0028
	(0.0160)

*Table VIII presents the estimates of β_3 from equation (22) for various outcome variables.

*Each row presents the DD estimate from equation 22 with a particular outcome variable.

*Standard errors are reported in brackets and are clustered at the district level.

Figure 2.1: Difference in Means

Gender Gap in Private School Enrollment
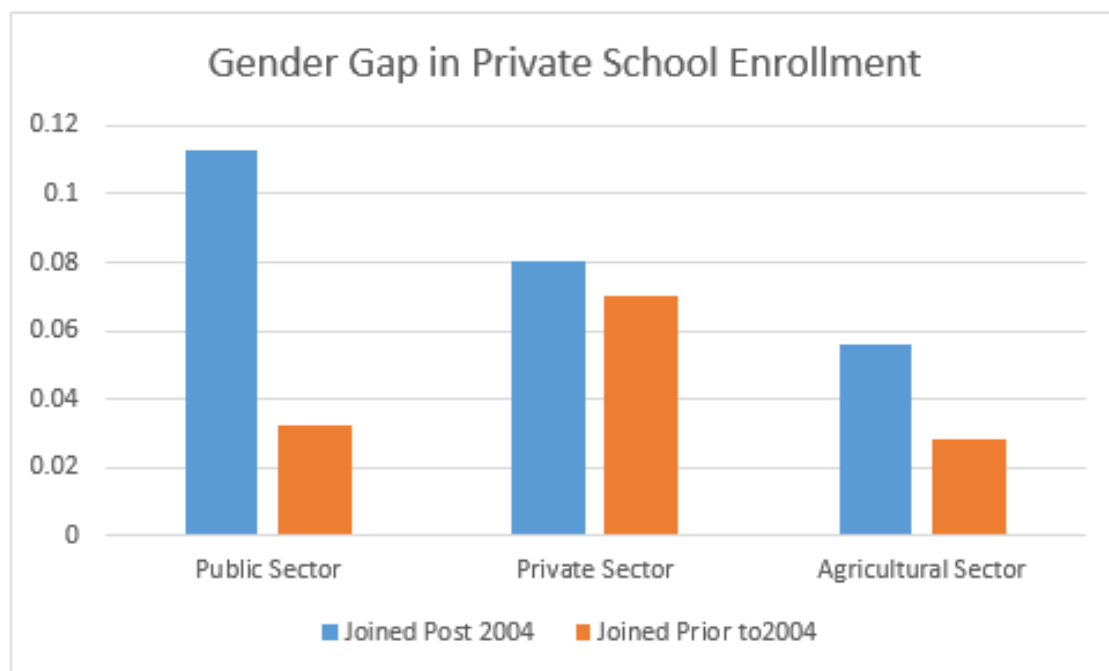
*Graph 1 plots the raw difference in means for the boy child versus girl child from the current student sample.

*Fathers characteristics is used to assign sector and joining date.

()

CHAPTER 3

EFFECT OF RISK CORRELATION ON MONITORING AND AUDITING

3.1 Introduction

Micro finance has been looked upon by policy makers as a viable means of alleviating poverty and providing credit to the poor in developing countries. Bangladesh, Brazil, India, Bolivia, Uganda, Morocco and several other countries have developed such institutions of their own to cater to the financial needs of the poor. For the past forty years, micro finance institutions (also known as MFIs) have developed mechanisms to maintain financial sustainability so as to serve the poor customers, who were deemed unprofitable by the commercial banks, by providing small loans without collateral, selling insurance, all in an effort to bring about poverty reduction.

Group loans have been the dominant model for lending in such institutions. There are numerous papers in the literature that investigates the reasons for why group loans outperform individual loans in terms of repayment. In this paper I am interested in looking at the effect of risk correlation in outcomes among group members, on various aspects of group lending. Risk correlation among group members is an important factor for repayment as most forms of group lending rely on joint liability for repayment. The intuition is that if one of the group members is unable to repay then the other members pay for that member, ensuring future credit. Higher risk correlation increases the probability of multiple group members failing to repay thus breaking the credit line.

Ahlin and Townsend (EJ 2007) is one of the first paper to look at the effect of risk correlation on repayment in group lending. They introduced risk correlation in some of the most cited models in the literature; Stiglitz (1990) and Banerjee et al (1994) both of which focuses on moral hazard problem which joint liability lending can mitigate; Besley and Coate (1995) which focuses on limited liability contracts which village sanctions can mitigate; Ghatak (1999) which focuses on how adverse selection can be dealt with using joint liability contracts. When they introduced risk correlation in the limited enforcement model Stiglitz (1990), they find that higher risk correlation

lowers repayment from the borrower. Whereas in the Beasley and Coates (1995) and Ghatak (1999) models they find higher risk correlation increases repayment by the borrower. They conclude that risk correlation might have counter intuitive positive effect on repayment. Empirically they find a weak positive relation between risk correlation and repayment.

Since the overall relation between risk correlation and repayment is ambiguous, I look at the effect of risk correlation on more intermediary variables which influences repayment. For example in the moral hazard model by Banerjee et al (1994) the adverse effect of moral hazards are mitigated through monitoring within the group members. So rather than directly looking at the effect of risk correlation on repayment, I look at the relationship between risk correlation and monitoring by group members. In the costly state verification model in Ghatak and Guinnane (1999), the high auditing cost on the part of the lenders are mitigated by joint liability contracts. In this model I look at the relationship between risk correlation and probability/cost of auditing by the lender. Looking at the relation between risk correlation and these intermediary variables will give us a better understanding on how risk correlation affects the mechanisms of group loans, as we are unable to understand its effects on more observable outcomes such as repayment.

The structure of the paper is very similar to Ahlin and Townsend (EJ 2007). I first introduce risk correlation to some of the most cited theoretical models in the existing literature. This gives us a theoretical understanding of the effects of risk correlation. Then I test these theories empirically in a non-structural way using the Townsend Thai data base.

To look at the relationship between risk correlation and monitoring I introduce risk correlation in the moral hazard model by Banerjee et al (1994). I find that the relationship between risk correlation and monitoring depends on the probability of success of the borrower and cosigner[1]. If the probability of success for the borrower is lower than that of the cosigner then monitoring decreases as risk correlation increases. If the probability of success for the borrower is higher than that of the cosigner then monitoring increases as risk correlation increases. Hence from the theoretical model the effect is ambiguous. On the empirical side we see that the results are also

---

[1]The cosigner is the group member who is liable to repay the borrower's loan if the borrower is unable to repay

ambiguous. The empirical relation between risk correlation and monitoring also depends on the proxies for risk correlation I am using. When I proxy for risk correlation using a village level measure for risk correlation[2] I find that as risk correlation increases monitoring decreases whereas when I use occupational homogeneity as a measure of risk correlation I find that as risk correlation increases monitoring decreases. These results are not causal rather correlations.

To look at the relation between risk correlation and audit[3] I introduce risk correlation in the costly state verification model in Ghatak & Guinnane (1999). I find that as risk correlation increases the probability/cost of audit increases. Since in the model the lender only audits when all group members fail to pay; a higher risk correlation increases the probability of that state of the world occurring and hence increases the cost/probability of audit. Empirically I find a negative correlation between audit and risk correlation but the results are not statistically significant.

The paper is structured as follows. Section 2 introduces the theoretical model. Section 3 introduces the data. Section 4 talks about the empirical strategy. Section 5 discusses the results. Section 6 compares the theoretical and empirical results. Section 7 presents the conclusion.

3.2 Model

3.2.1 Moral Hazard; BBG

To look at the effect of risk correlation on Monitoring I introduce risk correlation in the Moral hazard model by Banerjee et al (1994). Moral hazard in this context is the temptation by borrowers to gamble with risky projects when they have a limited liability. I use a modified version of the original model as used in Ahlin and Townsend (EJ 2007).

In this model the groups are asymmetric; it consists of one borrower and one cosigner (monitor). The borrower receives one unit of capital from the lender and chooses to invest in a project P_b. Where $P_b \in (\underline{P_b}, 1)$ and $P_b > 0$. The project generates a return of $Y(P_b)$ with probability P_b and 0 with probability $(1-P_b)$. The lender collects a return r from the borrower if the project is successful.

---

[2] defined in section 3

[3] When the lender has to verify the borrowers income, which is costly to do.

In the original model the lender collects q from the cosigner if the borrower fails to repay. This is the point I slightly diverge from the original model. In the original model there was no uncertainty in the cosigner's ability to pay the liability q but in my model the cosigner pays the liability only if he is also successful in his own project[4]. Hence in my model the lender collects q from the cosigner if the cosigner's project is successful and gets 0 otherwise. The cosigner is successful in his project with P_m and fails with probability $(1-P_m)$.

The risk neutral borrowers pay off from the project without any penalties[5] is $E(P_b)-P_br$; where $E(P_b)=Y(P_b)P_b$. By assumption $E(P_b)$ is increasing in P_b. It is also assumed that the interest rate and loan size is such that given limilted liability the borrower would always prefer risky projects to safer ones. So in the absence of a cosigner/monitor he will gamble with the riskiest project $P_b=\underline{P_b}$. I assume that;

$$0<E'(P_b)<r \text{ and } E''(P_b)< 0$$

The assumption implies that the borrowers expected payoff is increasing as riskiness of the project increases but the expected output is decreasing in risk.

The cosigner/monitor can penalize the borrower based on his project choice. To impose a penalty c on the borrower the monitor has to incur a cost M(c) where the function M() is increasing and convex in c.

$$M_c() > 0; M_{cc}() > 0$$

If the monitor wants to enforce a project P_b then he should penalise the borrower in such a way that the borrower's gains from deviating to the riskiest project is exactly equal to the gains from choosing the project P_b.

$$c(P_b,r) = E(\underline{P}_b) - \underline{P}_br - (E(P_b) - P_br)$$

---

[4]This is done to introduce risk correlation among group members
[5]Punishment from the monitor

The monitor then chooses P_b to maximize his payoff; which includes paying the joint liability fee q when the borrower is unsuccessful and the monitor is successful in his project. It also includes the cost of imposing a penalty c, which the monitor has to pay in all states of the world.

$$((1 - P_b)(P_m) - \epsilon)(-q) - M(c)$$

ϵ is the correlation in project outcomes between the borrower and cosigner. I introduce risk correlation in exactly the same way as Ahlin and Townsend (JoE 2007). They introduce risk correlation in a unique way to preserve the individual probabilities of success for the borrower and monitor.

	Monitor Succeeds	Monitor Fails
Borrower Succeeds	$P_bP_m+\epsilon$	$P_b(1-P_m)-\epsilon$
Borrower Fails	$(1-P_b)P_m-\epsilon$	$(1-P_b)(1-P_m)+\epsilon$

$\epsilon = 0$ is the case where there is no correlation in the projects outcomes of the borrower and monitor, $\epsilon > 0$ is the positive correlation case which I focus on in this paper. In general ϵ may depend on P_b and P_m. Since no general structure is available Ahlin and Townsend (JoE 2007) assume two specifications.

Specification I : The constant correlation case.

$$\epsilon(P_b, P_m) = \epsilon$$

Specification II :

$$\epsilon(P_b, P_m) = \mu \times min\{P_b(1 - P_m), P_m(1 - P_b)\}$$

Under specification II if the groups are homogeneous then the correlation coefficient over project returns is equal to μ. For non-homogeneous groups, that is for whom $P_b \neq P_m$, μ is not the correlation coefficient, but something closely related: it is the correlation, expressed as a fraction of the maximum correlation possible given individual probabilities of P_b and P_m. When $P_b = P_m$, the maximum possible correlation coefficient is one, so μ equals the correlation

coefficient. This formulation is the unique way of affecting each potential group's (appropriately normalized) correlation coefficient symmetrically. I will operate under specification II in this paper.

The monitor chooses the project P_b to maximize,

$$((1 - P_b)(P_m) - \epsilon)(-q) - M(c)$$

Where,

$$c(P_b, r) = E(\underline{P}_b) - \underline{P}_b r - (E(P_b) - P_b r)$$

Now I am interested in finding the effect of risk correlation on monitoring $dM/d\mu$; the risk correlation structure under specification II takes two forms given the value of P_b and P_m.

Case I,

$$P_b < P_m \implies \epsilon(P_b, P_m) = \mu \times P_b(1 - P_m)$$

Case II,

$$P_b > P_m \implies \epsilon(P_b, P_m) = \mu \times P_m(1 - P_b)$$

3.2.1.1 Case I

First I look at the effect of risk correlation on monitoring under case I.

The first order condition (F.O.C) w.r.t P_b;

$$q(P_m + \mu(1 - P_m)) = M_C(c)(r - E'(P_b) \text{ (2.1.1)}$$

The lender's break even condition is given by;

$$P_b r + ((1 - P_b)P_m - \mu \times P_b(1 - P_m))q = \rho \text{ (2.1.2)}$$

Where ρ is the outside option for the lender. To analyze the effect of risk correlation (μ) on monitoring ($M(P_b, r)$) we will use the first order condition. Here we are assuming that the lender

is not changing the liability q or the interest rate r to account for the change in risk correlation. [6]

Since M() is a function of P_B I first find the sign of $\frac{dP_b}{d\mu}$. To look at the effect of risk correlation on P_b I differentiate the F.O.C wrt μ.

The F.O.C when differentiated w.r.t μ gives;

$$\frac{dP_b}{d\mu}(M_c E''(P_b) - M_{cc}(r - E'(P_b))^2) = -q(1 - P_m) \ (2.1.3)$$

Rearranging equation 2.1.3 we get;

$$\frac{dP_b}{d\mu} = \frac{-q(1-P_m)}{A} > 0$$

Where;

$$(M_c E''(P_b) - M_{cc}(r - E'(P_b))^2) = A$$

$$A < 0 \text{ as M() is convex}$$

$$-q(1 - P_m) < 0,$$

Hence,

$$\frac{dP_b}{d\mu} > 0$$

Now we look at the effect of risk correlation on monitoring;

$$\frac{dc}{d\mu} = (r - E'(P_b))\frac{dP_b}{d\mu} > 0$$

$$\frac{dM}{d\mu} = M_c \frac{dc}{d\mu} > 0$$

Thus if $P_b < P_m$ then monitoring increases as risk correlation increases.

---

[6]Since we are working with cross sectional data where liability and interest doesn't change, this assumption fits.

3.2.1.2 Case II

$$P_b > P_m \implies \epsilon(P_b, P_m) = \mu \times P_m(1 - P_b)$$

The first order condition (F.O.C) w.r.t P_b;

$$qP_m(1 - \mu) = M_C(c)(r - E'(P_b)) \text{ (2.2.1)}$$

The lender's break even condition is given by;

$$P_b r + (1 - P_b)P_m(1 - \mu)q = \rho \text{ (2.2.2)}$$

Again we will use the first order condition. Since M() is a function of P_B I first find the sign of $\frac{dP_b}{d\mu}$. To look at the effect of risk correlation on P_b I differentiate the F.O.C wrt μ.

The F.O.C when differentiated w.r.t μ gives;

$$\frac{dP_b}{d\mu} \frac{A}{P_m(1-\mu)} = \frac{q}{(1-\mu)} \text{ (2.2.3)}$$

From equation 2.2.3,

$$\frac{dP_b}{d\mu} = \frac{qP_m}{A}$$

Hence,

$$\frac{dP_b}{d\mu} < 0$$

This implies,

$$\frac{dc}{d\mu} = (r - E'(P_b))\frac{dP_b}{d\mu} < 0$$

and,

$$\frac{dM}{d\mu} = M_c \frac{dc}{d\mu} < 0$$

110

If $P_b > P_m$ then monitoring decreases as risk correlation increases.

We see that the effect of risk correlation on monitoring depends on the values of P_b and P_m. If $P_b < P_m$ then monitoring increases as risk correlation increases. Whereas if $P_b > P_m$ then monitoring decreases as risk correlation increases. The result is driven by the structure of risk correlation used in the model. Hence form the theoretical model we find that the effect of risk correlation on monitoring is ambiguous.

3.2.2 Costly state Verification; Maitreesh Ghatak , Timothy W. Guinnane 2009

In this section I want to look at effect of risk correlation on the probability and hence the cost of audit. I will introduce risk correlation in the model of State Verification used in (Maitreesh Ghatak et al, 1999). Formal lenders sometimes cannot lend to poor borrowers because such lenders cannot cost effectively verify the borrowers state. In such a scenario joint liability contracts reduce expected audit cost. The intuition is that suppose group members have a lower cost of verifying each other and they are liable for each other's loan then the bank has to verify only when the entire group defaults. In their paper they introduce a small theoretical model to show how joint liability contracts improve efficiency by lowering audit costs. I will use the same model to show that introducing risk correlation into the model decreases the gains from moving to joint liability contract ie it increases the audit cost.

There are two borrowers and one lender. There are two outcomes from the borrowers $\{Y_H, 0\}$ with probability $(p, (1-p))$ and $(p', (1-p'))$. The lender has to incur a cost γ to verify the individual borrowers output, whereas the borrowers can do it costlessly. Each borrower has an exogenous outside option U. Assume that the borrowers can write side-contracts with each other costlessly and that there is no cost for a borrower to observer her partner's project returns. There are no problems of moral hazard, adverse selection or enforcement of contracts. The financial contract specifies the interest rate r and the probability of an audit λ when both the members default.

The Optimal contract is given by;

$$\max \text{pp'}(Y_H - r) + p(1 - p')(Y_H - 2r) - U$$

The truth telling constrain is given by[7] ;

$$Y^H - 2r \geq max\{0, (1 - \lambda)Y^H\}$$

The break even constrain is given by[8];

$$PP'r + P(1 - P')r + (1 - P)P'r - \lambda(1 - P)(1 - P')\gamma = \rho$$

Since p, Y_H and U are exogenously given the optimal contract is given by r and λ which satisfies both the constrain. The equilibrium value of λ and r is given by;

$$r = \lambda\frac{Y^H}{2}$$

$$\lambda = \frac{\rho}{\frac{Y^H}{2}(PP' + P(1-P') + (1-P)P' - (1-P)(1-P')\gamma)}$$

Now, I add risk correlation among borrower outcomes in the model. I will show the proof under specification I where $\epsilon(P, P') = \epsilon$ but the proof can be easily extended for specification II.

Let the new equilibrium probability of audit and interest rate be λ' and r'.

The truth telling constrain remains the same after adding risk correlation and is given by;

$$Y^H - 2r' \geq max0, (1 - \lambda')Y^H$$

The break even constrain changes, the new break even condition is given by,

$$(PP' + \epsilon)r + (P(1 - P') - \epsilon)r + ((1 - P)P' - \epsilon)r - \lambda'((1 - P)(1 - P') + \epsilon)\gamma = \rho$$

The new equilibrium value of r and λ is given by,

$$r' = \lambda'\frac{Y^H}{2}$$

$$\lambda' = \frac{\rho}{\frac{Y^H}{2}(PP' + P(1-P') + (1-P)P' - (1-P)(1-P')\gamma - \epsilon(\frac{Y^H}{2} + \gamma))}$$

Now we compare the equilibrium value of r and λ,

---

[7]This constraint ensures that it is more profitable for the lender to reveal his true state.
[8]This ensures that the lender is operating at zero profit

$$\lambda' - \lambda = \frac{\epsilon(Y^H/2+\lambda)}{Z} > 0$$

$$r' - r = Y^H/2(\lambda' - \lambda) > 0$$

As long as $\epsilon > 0$, $\lambda' > \lambda$ and $r' > r$. Also $\lambda' - \lambda$ increases as ϵ increases. Thus we see introducing risk correlation increases the probability/cost of monitoring. The increase in cost increases in ϵ. This result can readily be extended to specification II of ϵ, as long as $\epsilon > 0$ the above results hold.

3.3 Data

The data used in this project comes from the Townsend Thai data base. To be more precise the data comes from a large cross section survey conducted in May of 1997 covering 192 villages in two regions of Thailand. The regions covered in the survey includes the central and northeast region. Each of these regions in Thailand consists of provinces and the provinces are further divided into sub-counties. To conduct the survey first sub-counties were selected in a province and then the survey was conducted on a cluster of four villages within the sub-counties. Borrowing groups in the villages belonging to the Bank of Agriculture and Agricultural Cooperatives (BAAC) were interviewed. The BAAC is a government operated development bank in Thailand[9]. Within the selected villages they interviewed as many groups as possible up to two. In the data there are 262 groups, 62 of which are the only groups in their village. Each group had a designated leader who responded to the questions on behalf of the group. The analyses is at the group level.

The dependent variable of interest are measures of risk correlation. Following Ahlin and Townsend (EJ 2007) I proxy for risk correlation by two measures of correlation; Occupational homogeneity within a group and a village-level measure taken from the household survey. Occupational Homogeneity is taken from the BAAC survey and equals to the probability of two randomly chosen group members to have the same occupation. For the village level measure I use

---

[9]The BAAC requires some kind of collateral for all loans but it allows smaller loans to be backed with social collateral in the form of joint liability. The nature of the BAAC loans makes it ideal for me to test my theoretical prediction on.

the household survey where villagers were asked which of the previous five years were the best and worst for income, respectively. The variable is constructed as the probability that two randomly selected respondents from the same village reported the same year as worst. Table 1 summarizes these variables.

The independent variable of interest includes measures of monitoring and auditing. For measures of monitoring within the group I use four questions asked in the BAAC instrument

i) Leader Monitoring 1 : In a typical week how many group members does the leader talks to more than five times; ii) Leader Monitoring 2 : Number of people in the group the leader talks to less than once a week; iii) Peer Monitoring : Are the group members aware of the work of the other members; iv) Meetings : How many meetings did the group had in the past month. These variables are reported in Table 2.

I use two measures of auditing i) Verify Income : Does anyone verifies the members income ii) Officer visit: How many times does the BAAC officer visits. The summary of these variables are reported in Table 2.

I also include group and village level control. I control for group size and group age. I control for village level risk, percent in village claiming production credit group membership and percent in village claiming to be clients of commercial banks.

3.4 Identification Strategy

To empirically see whether risk correlation affects monitoring within a group and audit cost I use a simple OLS framework (Equation1).

$$Y_i = \alpha + \beta X_i + \gamma GroupCharacteristics_i + \delta VillageCharacteristics_i + \epsilon_i \text{ (1)}$$

The coefficient of interests is β, which is the coefficient on the two measures of risk correlation described in section 3 and the dependent variables include various measures monitoring and audit costs. The results are not causal rather correlations. I control for group and village characteristics.

I partial out the effect of group size and group age by controlling for them as they are usually strong predictors of group behavior. I also run the OLS including region fixed effects.

3.5 Results

In this section I present the results from equation 1, the regressions results are estimated for both proxies of risk correlation; occupational homogeneity and village level measure of risk correlation. I also presents the results from including region fixed effects.

First, I look at the effect of risk correlation on monitoring. Table 3, panel A presents the effect of occupational homogeneity on how many group members the leader talk to more than five times in a week (Leader Monitoring 1). The more number of group members the leader talks to more than five times a week implies greater monitoring by the leader on the group members. We see that as occupational homogeneity increases leader monitoring 1 increases and the effect is statistically significant. This implies that even after controlling for group and village characteristics, occupational homogeneity among the group members is positively correlated with leader monitoring. When I include region fixed effect the estimate become smaller and is still statistically significant at a lower level of confidence. Table 4, panel A presents the effect of village level measure of risk correlation on how many group members the leader talk to more than five times in a week (Leader Monitoring 1). We see that as village level measure of risk correlation increases leader monitoring 1 decreases and the effect is not statistically significant.

Table 3, panel B presents the effect of occupational homogeneity on number of people the leader talks to less than once a week (Leader Monitoring 2). The more number of group members the leader talks less than once a week implies lower monitoring by the leader. We see that as occupational homogeneity increases leader monitoring 2 decreases and the effect is statistically significant. This implies that after controlling for group and village characteristics occupational homogeneity among the group members is positively correlated with leader monitoring. When I include region fixed effect the estimate become smaller and is still statistically significant. Table 4, panel B presents the effect of village level measure of risk correlation on number of people the

leader talks to less than once a week (Leader Monitoring 2). We see that as village level measure of risk correlation increases leader monitoring 1 decreases and the effect is statistically significant.

Table 3, panel C presents the effect of occupational homogeneity on group awareness of each other's work (Peer Monitoring). The higher awareness of the group members about each other's work implies higher peer monitoring. We see that as occupational homogeneity increases peer monitoring decreases and the effect is not statistically significant. When I include region fixed effect the estimate become smaller and is still not statistically significant. Table 4, panel C presents the effect of village level measure of risk correlation on peer monitoring. We see that as village level measure of risk correlation increases peer monitoring decreases and the effect is not statistically significant. When we include region FE the sign changes but the effect is still not statistically significant.

Table 3, panel D presents the effect of occupational homogeneity on how many meetings the group had in the past month. We see that as occupational homogeneity increases number of meetings decreases and the effect is statistically significant. This implies that after controlling for group and village characteristics occupational homogeneity among the group members is negatively correlated with number of meetings. When I include region fixed effect the estimate become more negative and is still statistically significant. Table 4, panel D presents the effect of village level measure of risk correlation on number of meetings. We see that as village level measure of risk correlation increases number of meetings increases and the effect is statistically significant.

Overall we saw that as occupational homogeneity increases monitoring by both the leader and among group members increases whereas as village level measure of risk correlation increases monitoring decreases. In section 6 we compare the results from both measures in further details.

Second, I look at the effect of risk correlation on cost/probability of audit. Table 5, panel A presents the effect of occupational homogeneity on the probability of someone verifying the borrower's income. The higher number of borrowers incomes verified implies higher auditing cost. We see that as occupational homogeneity increases the probability of borrower's income being verified decreases and the effect is statistically significant. This implies that after controlling for

group and village characteristics occupational homogeneity among the group members is negatively correlated with audit cost. When I include region fixed effect the estimate is similar and statistically significant. Table 6, panel A presents the effect of village level measure of risk correlation on the probability of someone verifying the borrower's income. We see that as village level measure of risk correlation increases the probability of someone verifying the borrower's income decrease and the effect is not statistically significant.

Table 5, panel B presents the effect of occupational homogeneity on the probability on number of officer visits. The higher number of officer visits implies higher auditing cost. We see that as occupational homogeneity increases the number of officer visit decreases and the effect is statistically not significant. When I include region fixed effect the estimate is more negative but still statistically not significant. Table 6, panel B presents the effect of village level measure of risk correlation on the number of officer visits. We see that as village level measure of risk correlation increases the number of officer visits decreases and the effect is not statistically significant.

3.6   Comparing the theoretical results with the empirical findings

When I introduce risk correlation in the BBG model I find that the relation between risk correlation and monitoring depends on the probability of success of the borrower and cosigner. If the probability of success for the borrower is less than that of the cosigner then as risk correlation increases monitoring decreases. Whereas if the probability of success for the borrower is greater than that of the monitor then as risk correlation increases monitoring decreases. It would have been ideal if we had some individual level data which measures the relative success of a borrower and cosigner then we could have tested the theoretical predictions more meaningfully but unfortunately we do not have such individual level measures. When we look at the empirical result of risk correlation on monitoring we see that the two measures of risk correlation consistently delivers opposite results. The occupational homogeneity measure predicts a positive correlation between monitoring and risk correlation whereas the village level measure of risk correlation predicts a negative correlation.

117

Firstly I acknowledge that both the measures of risk correlation are at best proxies of the true risk correlation. Occupational homogeneity captures the risk co-relatedness that comes from having group members in the same occupation whereas the other measure captures a more macro risk correlation at the village level. In the context of the BBG model I think the occupational measure does a better job of capturing correlation than the village level measure. If the cosigner and borrower are in the same occupation they are more likely to have correlated outcomes. The issue here could be that the increase in monitoring we see from the data might be coming from channels which the model doesn't capture such as information and marginal cost of monitoring. For example if the borrower are cosigner are in the same profession then they might have similar professional circles in the village and the cosigner gets information from the circles and has a reduced cost of monitoring or higher capability to punish. Being in the same profession might also affect the marginal cost of monitoring. For example it might be easier for a peach farmer to verify the output of another peach farmer compared to a shop keeper. The effect from these channels might over/under estimate the effect of risk correlation on monitoring. On the other hand the village level measure of risk correlation might not be representative of the risk correlation at the group level as the BAAC borrowers constitute a very small portion of the total population. Hence the effect we capture from this measure of risk correlation might just be driven by other confounding macro level village variables.

To get the relation between risk correlation and the cost/probability of audit I introduced risk correlation in the Ghatak and Guinnane (2009) model. I found that as risk correlation increases the cost/probability of audit increases. The intuition being that the lender only has to audit when both borrowers fail to repay and as risk correlation increases the probability of that happening increases. From the empirical side we see that risk correlation is negatively correlated with audit using both measures of risk correlation. Again the measures of risk correlation we are using are proxies at best and the divergence in the theoretical and empirical results could be driven by that. To understand the contribution of this paper and importance of the correlational results, we should first try and think about how the ideal causal inference would look like. The ideal causal inference would require an

exogenous variation in the risk correlation of project outcomes among the group members. Since groups are form endogenously and the lending mechanism heavily depends on this endogenous matching process, such a variation might not be feasible. If we think of an experimental setting then this will require us to randomly assign borrowers to groups in order to create some groups with high risk correlation and some groups with low risk correlation. But again such randomization would take away the very essence of group lending which heavily depends on the endogenous matching by the group members which reflects network strength. For example one of the reasons why group lending succeeds is because the borrowers are liable for each other's loan and they are able to effectively punish each other in case someone defaults. If the groups are formed randomly then the borrower's ability to effectively punish or monitor each other would substantially decrease and hence the results we see from such an experiment would capture effects apart from variation in risk correlation. To sum up, an experiment to identify the causal relation would require some sort of exogenous intervention in group formation but the endogenous group formation itself is an important driver of group lending so the intervention would affect a lot of other variables apart from risk correlation. This is why a causal inference might not be feasible in this context.

3.7   Conclusion

In this paper I look at the effect of risk correlation on monitoring and auditing cost in the context of group lending. To identify this I take an approach very similar to Ahlin and Townsend (EJ 2007). First I introduce risk correlation to some of the well established models in the literature which look at monitoring and auditing. These models gives us a theoretical understanding of the relationship between risk correlation and monitoring and auditing costs. Next I use the Thai Townsend data base to get the empirical results. The empirical results are correlations and not causal.

To look at the effect of risk correlation in monitoring I introduce risk correlation in the moral hazard model by Banerjee et al (1994). I find that the effect of risk correlation on monitoring depends on the probability of success of the borrower and cosigner. If the probability of success for the borrower is lower than that of the cosigner then monitoring decreases as risk correlation

increases. If the the probability of success for the borrower is higher than that of the cosigner then monitoring increases as risk correlation increases. Hence from the theoretical model the effect is ambiguous. On the empirical side we see that the results are also ambiguous. When we use occupational homogenity as a measure of risk correlation we find that monitoring increases as risk correlation increases whereas when we use the village level measure of risk correlation we find that monitoring decreases.

To look at the effect of risk correlation on auditing cost I introduce risk correlation in the model of State Verification used in (Maitreesh Ghatak et al, 1999). I find that as risk correlation increases the probability/cost of audit increases. When we look at the empirical results we see that as risk correlation increases audit cost decreases.

The empirical findings of the paper are not causal but correlation. It suffers from the usual demerits of correlation results such as omitted variable bias and reverse causality. The measures of risk correlation are also not perfect as they are at best proxies for the true risk correlations. In the future I plan to dig further into the empirical side and try and establish more robust causal results. The theory model in the paper specifically the BBG model used to look at the effect of risk correlation on monitoring has its drawbacks as the results from the theoretical model is partially driven by the structure of risk correlation. Overall the study should not be treated as a final answer to question of how risk correlation among group members affect group lending. Rather it should be treated as a staring point to motivate research in the sparse literature.

Table 3.1: Descriptive Statistics

Variable	Description	Mean	SD
Controls			
Group Size	Number of members in the borrowing group.	12.3	5.1
Group Age	Age of the Group	11.4	8.5
Village Risk	Village average coefficient of variation for next year's expected income	0.67	0.8
PCG_Membership	Percent in village claiming Production Credit Group membership	0.09	0.15
Commercial Bank	Percent in village claiming to be clients of a commercial bank	0.26	0.17
Measures of Risk Correlation			
Village level Risk Correlation	Measure of coincidence of economically bad years across villagers	0.28	0.16
Homogeneous Occupation	Measure of occupational homogeneity within the group	0.87	0.24

Table 3.2: Descriptive Statistics

Monitoring			
Leader Monitoring 1	In a typical week, how many group members do you talk to more than 5 times?	3.98	3.98
Leader Monitoring 2	No of people you talk to less than once a week?	0.93	0.24
Peer Monitoring 1	Are group Members aware of the work of other members?	1.5	2.5
Peer Monitoring 2	Are group Members aware of the debt of other members?	0.8	0.37
Meetings	How many meetings has the group had in the past 12 months?	1.97	1.16
Auditing Cost			
Verify Income	Does Anyone Verify the member's income?	0.45	0.49
Verify Asset	Does Anyone Verify the member's asset?	0.49	0.5
No of Office visit	How many times does the BAAC officer visit?	3.8	6.3
Screening			
Screen	Do someone wants to join the group but cannot?	0.39	0.49
Punishment			
Punish	Percent of village loans where default is punishable by informal sanctions.	0.9	0.11

Table 3.3: Leader Monitorig 1 RC

Effect of Occupational Homogeneity on Leader Monitoring 1		
	OLS	OLS W/region FE
Occupational Homogeneity	2.8517** (1.1439)	2.2173* (1.2011)
N	244	232
Mean of Dependent Variable	3.98	3.98
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.4: Leader Monitorig 2 OH

Effect of Village level measure of risk correlation on Leader Monitoring 1		
	OLS	OLS W/region FE
Village level risk correlation	-1.5809 (1.7558)	-2.2124 (1.7893)
N	244	232
Mean of Dependent Variable	3.98	3.98
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.5: Leader Monitorig 1 RC

	OLS	OLS W/region FE
Effect of Occupational Homogeneity on Leader Monitoring 2		
Occupational Homogeneity	-1.8776** (0.7105)	-1.7152** (0.7594)
N	244	232
Mean of Dependent Variable	0.93	0.93

Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.

Table 3.6: Leader Monitorig 2 OH

Effect of Village level measure of risk correlation on Leader Monitoring 2		
	OLS	OLS W/region FE
Village level risk correlation	2.106** (1.02)	1.9812* (1.1075
N	244	232
Mean of Dependent Variable	0.93	0.93
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.7: Peer Monitoring RC

Effect of Occupational Homogeneity on Peer monitoring		
	OLS	OLS W/region FE
Occupational Homogeneity	0.0794 (0.0719)	0.0627 (0.0842)
N	244	232
Mean of Dependent Variable	1.5	1.5
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.8: Peer Monitoring OH

Effect of Village level measure of risk correlation on Peer Monitoring		
	OLS	OLS W/region FE
Village level risk correlation	-0.1424 (0.1888)	0.0915 (0.1105)
N	244	232
Mean of Dependent Variable	1.5	1.5
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.9: Monthly Meetings RC

Effect of Occupational Homogeneity on Monthly Meetings		
	OLS	OLS W/region FE
Occupational Homogeneity	-0.7421** (0.3202)	-0.8273** (0.3537)
N	244	232
Mean of Dependent Variable	1.97	1.97
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.10: Monthly Meetings OH

Effect of Village level measure of risk correlation on Peer Monitoring		
	OLS	OLS W/region FE
Village level risk correlation	0.9974** (0.4660)	0.8839 (0.5249)
N	244	232
Mean of Dependent Variable	1.97	1.97
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.11: Borrower Income RC

Effect of Occupational Homogeneity on Verifying Borrower Income		
	OLS	OLS W/region FE
Occupational Homogeneity	-0.2759** (0.1415)	-0.2967** (0.1515)
N	244	232
Mean of Dependent Variable	0.45	0.45
Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.12: Borrower Income OH

	OLS	OLS W/region FE
Effect of Village level measure of risk correlation on Verifying Borrower Income
Village level risk correlation	0.1355 (0.2043)	0.1267 (0.2279)
N	244	232
Mean of Dependent Variable	0.45	0.45

Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.

Table 3.13: Officer Visit RC

	OLS	OLS W/region FE
Effect of Occupational Homogeneity on Number of Officer Visits		
Occupational Homogeneity	-1.3367 (1.7781)	-1.526 (1.9021)
N	244	232
Mean of Dependent Variable	3.8	3.8
Notes: Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

Table 3.14: Officer Visit OH

	OLS	OLS W/region FE
Effect of Village level measure of risk correlation on Number of Officer Visits		
Village level risk correlation	-0.9535 (2.5830)	-1.3818 (2.9125)
N	244	232
Mean of Dependent Variable	3.8	3.8
Notes: Notes: Column 1 presents the results from the OLS. Column 2 presents the results from the OLS including region fixed effects.		

APPENDICES

APPENDIX A

CHAPTER I : APPENDIX

A.1 Back of the Envelope Calculation

In this section I will calculate the distribution of the variable target given the sample distribution of births by month and the sample distribution of marriages by month. Let m1, m2, m3, m4. . . m12 be the number of marriages by month. Where m1 is the number of marriages in January, m2 is the number of marriages in February and so on. Similarly let b1,b2,b3...b12 be the number of births by month. Where b1 is the number of births in January, b2 is the number of births in February and so on.

Similar to section 2 we define the variable target=mod (age at marriage in months,12). Hence the variable target takes the value zero when month of birth is the same as month of marriage. That is if the individual is born in January and gets married in January or the individual is born in February and gets married in February and so on. Takes the value 1 if the month of marriage is one month later than the month of birth and so on.

Prob(target=0)=Prob(b1=m1) + Prob(b2=m2) . . . . + Prob(b12=m12)

Prob(target=1)=Prob(b1=m2) + Prob(b2=m3). . . . . + Prob(b12=m1)

....

Prob(t=11)=Prob(b1=m12) + Prob(b2=m1) + ... + Prob(b12=m11)

Now if there was no age targeting at marriage then; that is people had no preference of matching their month of birth to their month of marriage; that is month of marriage is independent of month of birth then;

Prob(bi=mi) = Prob(bi)Prob(mi)

Where Prob(bi)=bi/B and Prob(mi)=mi/M

$$M = \sum mi$$

136

$$B = \sum bi$$

We obtain the distribution of birth by month and the distribution of marriages by month from our sample. Given the sample distribution we calculate the distribution of the variable target. Graph 6 plots the actual and calculated distribution of the variable target. The difference between the actual and calculated distribution at zero is what I define as age targeting at marriage, where parents target certain whole number age for their daughters to get married after.

A.2 Example for IV2

IV2 captures the one year future discounted number of marriage dates from the month of birth of the woman. IV2 to not only captures the fact that certain months have more marriages than other but it also captures the fact that certain months are followed by a period of high number of marriage dates whereas certain months are followed by a period of low number of marriage dates. For example January and July has similar number of marriage dates but January is followed by a period of high number of marriage dates and July is followed by a period of low number of marriage dates. IV2 would be able to capture this essence whereas IV1 does not. IV2 could be thought of as a relative ranking of a month compared to other months. I provide an example with beta equal to 0 and 0.5.

Example for IV2

Month	Number of Marriages	Percent	IV2 (β=0)	IV2(β=0.5)
1	15,203	4.91	4.91	17.21
2	33,623	10.85	10.85	24.61
3	34,799	11.23	11.23	27.54
4	52,852	17.05	17.05	32.63
5	72,164	23.29	23.29	31.17
6	36,481	11.77	11.77	15.78
7	16,514	5.33	5.33	8.05
8	7,837	2.53	2.53	5.46
9	6,142	1.98	1.98	5.886
10	7,466	2.41	2.41	7.81
11	13,527	4.37	4.37	10.81
12	13,288	4.29	4.29	12.89

A.3   Compliers for the IVs

To better understand why we get different results on fertility outcomes across the two IV's, I take a closer look at the compliers for both the IVs in this section. The complier for the age at menarche IV are women who delays marriage due to delay in age at menarche. Since 94 percent of the women reach menarche by the age of 15, the compliers for the menarche IV primarily consists of young girls who get married at an early age of sixteen or lower. For my IV the compliers are women belonging to families who practice minimum age targetting at marriage and exhibit a preference for getting married on auspicious marriage months. The preference for getting married in an auspicious marriage months is relevant for Hindu women getting married at any age but the age targetting at marriage is relevant for women getting married before the age of 22. Hence the compliers for my IV belongs to the group of women who gets married before the age of 22. When we look at the socio-economic background of the compliers across the two IV. Both minimum age targeting at marriage and preference for getting married on auspicious dates are exhibited by women who are less educated and belonging to families with lower wealth and education [1]. Families who get their daughters married at an early age of less than sixteen (population for the menarche complier) also

---

[1]Graphs 7-9 and graphs 15-17.

belongs to families with lower wealth and education. Hence the differential results are most likely not driven by the fact that the compliers belong to different socio-economic backgrounds.

One potential difference between the compliers for the two IVs is that the compliers for the age at menarche IV are younger than the compliers for my IV. Thus one possible explanation for why we see different results for fertility across the two IVs is that how delayed age at marriage affects fertility differs by age at which the women gets married. A three month delay in marriage at the young age of 12-16 might have a different effect on fertility compared to a three month delay at a higher age of 17-21. To check that whether the difference in results across the two IVs is driven by non linearity in the response function by age, I estimate the results from my IV on a set of age windows. Appendix table 7 presents the results. We see that for women who gets married before the age of 16 one month delay in marriage increases the number of children by 0.01 (0.3 percent) and the effect is statistically significant. For women who got married after the age of 16 one month delay in age at marriage increases the number of children by 0.006 (0.27 percent) and the effect is statistically significant. We also find similar results for delay in age at first birth and ideal number of children. Thus I conclude that the differential fertility results we get for both IV's is not driven by the fact that the menarche IV compliers are younger than the compliers for my IV.

Table A.1: First Stage Summary Stat (Appendix Table 1)

First Stage Regression Summary Statistics					
Panel A: First Stage Regression Summary Statistics for IV1					
	R-Square	Adjusted R-Square	Partial R-Square	Robust F(1,256460)	Probability>F
Age at Marriage (months)	0.0166	0.0166	0.0021	572.329	0
Panel B: First Stage Regression Summary Statistics for IV2					
	R-Square	Adjusted R-Square	Partial R-Square	Robust F(1,256460)	Probability>F
Age at Marriage (months)	0.0141	0.0141	0.0026	671.547	0

Table A.2: Effect of Age at Marriage on Spousal Characteristics (Appendix Table 2)

	Electricity			Car		
	OLS	IV1	IV2	OLS	IV1	IV2
Age at Marriage (in months)	-0.0003*** (0.00009)	0.003*** (0.0005)	0.004*** (0.0005)	-0.0006*** (0.00006)	0.0006** (0.0003)	0.0007** (0.0003)
Observation	244345			244345		
Mean	0.87			0.05		

Effect of Age at Marriage on Husband's Family Characteristics

Notes**
The table presents the results of OLS and IV regressions in which the dependent variables are spousal family outcomes for the full sample (Women who got married before the age of 22. All regressions control for age of the women at survey and includes month of marriage FE. It also controls for household characteristics. Robust standard errors are presented in parenthesis. Electricity is a binary variable taking the value 1 if the household has electricity or 0 otherwise. Car is a binary variable taking the value 1 if the household has electricity or 0 otherwise.
***=significant at 1%
** =significant at 5%

Table A.3: Effect of Age at Marriage on Reproductive Health (Appendix Table 3)

Effect of Age at Menarche on Reproductive Health		
	Period Irregularities	Pregnancy Termination
Age at Menarche	-0.006**	-0.001
	(0.0015)	(0.0011)
Observation	55271	55271
Mean	0.8	0.1588

Notes***
The table presents the results from an OLS regression with age FE. I control for household and individual characteristics.
***=significant at 1%
** =significant at 5%

Table A.4: Wald IV Estimates (Appendix Table 4)

	Health		Fertility	
	Diabetic	Diastolic	Total Children	Age at first birth
Age at marriage (months)	-0.0005*** (0.0002)	-0.003** (0.0006)	-0.001 (0.0014)	0.065*** (0.002)
Observation	255397	255397	255397	232115
Mean	0.025	0.32	2.45	19.54

Wald IV Estimates

Notes***
The table presents the Wald IV estimates.
***=significant at 1%
** =significant at 5%

Table A.5: First Stage Estimates (Appendix Table 5)

First Stage Estimates	
Panel A : The effect of IV1 on age at marriage in months for Muslim Population	
IV1 (Being born in low versus high frequency months)	3.24*** (0.34)
Observation	41472
R-Squared	0.016
Panel B :The effect of IV1 on age at marriage in months for the	
IV1 (Being born in low versus high frequency months)	0.42 (0.58)
Observation	6953
R-Squared	0.011
Notes*** The table presents the results of the first stage OLS regression for the instrumental variable. All regressions control for woman's age at survey and include month of marriage FE. Robust standard errors are reported in parenthesis. Panel A presents the results of the first stage OLS regression for IV1 for the Muslim Population. Panel B presents the results of the first stage OLS regression for IV1 for the Sikh Population. ***=significant at 1% ** =significant at 5%	

Table A.6: Muslim Population (Appendix Table 6)

Effect of Age at Marriage on the Muslim Population				
Panel A : Health and fertility results using IV1 for Muslim Population				
	Health		Fertility	
	Diabetic	Diastolic	Total Children	Age at first birth
IV1	-0.002	-0.006**	0.02***	0.055***
	(0.0011)	(0.002)	(0.007)	(0.001)
Observation	41472	41472	41472	37765
Mean	0.04	0.35	2.92	19.23

Notes**
The table presents the IV results on the Muslim sample (women who got married before the age of 22). All regressions control for age of the women at survey and includes month of marriage FE. It also controls for household characteristics. Robust standard errors are presented in parenthesis.
***=significant at 1%
** =significant at 5%

Table A.7: Effect of Age at Marriage on Fertility (Appendix Table 7)

Effect of Age at Marriage on Fertility			
	Total Number of Children	Age at First Birth	Ideal Number of Children
Panel A : Women who got married before the age of sixteen			
	IV1	IV1	IV1
Age at Marriage	0.01*** (0.002)	0.04*** (0.006)	0.005** (0.002)
Observation	61275	58914	61275
Mean	3.06	16.99	2.18
Panel B : Women who got married after the age of sixteen			
Age at Marriage	0.006** (0.003)	0.08*** (0.005)	0.005** (0.002)
Observation	163917	145141	163917
Mean	2.18	20.76	2.02
Notes** The table presents the results from the IV regression using IV1 as the instrument where the dependent variables are fertility outcomes. Robust standard errors are presented in parenthesis. Panel A presents the results for women who got married before the age of sixteen and Panel B presents the results for women who got married after the age of sixteen. ***=significant at 1% ** =significant at 5%			

CHAPTER II : APPENDIX

B.1   Accounting for age

One of the major difference between people who joined the sectors prior and post 2004 is that the people who joined prior are older. As a robustness check i include interaction terms with age such as age*sector, age*gender and age*sector*gender. Appendix table A4 reports the results of the triple difference estimates including the interaction terms. We see that the estimates do not vary from the original estimates.

B.2   Definition of White Collar Jobs

There are two variable which helps in identifying my desired population, first variable directly asks whether the person works in public sector or private sector or other. The second variable asks the specific occupation of the person. They jobs are grouped into 99 occupations. I separate the occupations into white collar and non-white collar jobs based on my knowledge of occupation in India, like teachers are white collar while bus drivers are not. Now in my analysis I use the population which is the intersection of white collar and public sector or private sector. To see if my analysis is robust to the definition is white collar I use the following specification. I just use the variable which asks whether the person works in a public or private sector. Hence I include both white collar and non-white collar in my analysis. Table A5 reports the estimate from the triple difference with the alternate specification. We see that the estimates do not vary from the original estimates. Hence the estimates are not sensitive to the definition of white collar.

B.2.1   Including Mothers Sector

Sector is defined by fathers occupation as the percentage of female in the public sector is 14% in my data and usually fathers have a higher say in matters of financial decision in the household. I

present the results including mother's sector, now sector is defined as public if either mother or father is in public sector and private if both mother and father are not in public and one of them is in private. Table A6 reports the estimate from the triple difference with the alternate specification. We see that the estimates do not vary from the original estimates. Hence the estimates are not sensitive to the definition of white collar.

B.3 Anthropometric Measures of Early Childhood Investments

In this section I look at some anthropometric measures of early child childhood investment such as height for age, weight for age, weight for height and BMI for age. To compute these scores I use the zscore06 function in STATA. It calculates anthropometric z-scores using the WHO child growth standards in developing countries. Height-for-age, weight-for-height, BMI-for-age and weight-for-age Z-scores are calculated for children 0 to 5 years of age. Appendix table I presents the results. The sample consists of children age 0-5. Overall I do not find any statistically significant or economically meaningful results for this sample. The most likely reason for this is that the data is plagued with measurement errors and missing values. Also since the birth months of a lot of children were imputed I had to use age in years to calculate these scores which might also drive the findings.

Table B.1: A1 Difference-in-Difference Results for Boy child and Girl child

Current Students			
	DD For Male Child	DD For Female Child	DDD
Private School	0.0561	-0.0162	-0.0711**
	(0.0315)	(0.0323)	(0.0342)
English Medium	0.0407	-0.0384	-0.0684**
	(0.0306)	(0.0305)	(0.0340)
English Ability	0.0281	-0.0529	-0.0667
	(0.0381)	(0.0398)	(0.0517)
Observation	6012	5324	11462
Test Scores			
Mathematics	-0.1057	-0.2134	-0.0765
	(0.1049)	(0.1435)	(0.1217)
Reading	0.0156	0.0557	0.029
	(0.0851)	(0.1090)	(0.1169)
Writing	0.0758	-0.0312	-0.044
	(0.1109)	(0.1532)	(0.1588)
Score	0.0061	-0.1697	-0.1034
	(0.2522)	(0.3677)	(0.3349)
Observation	1374	1236	2632

Table B.2: A2 Triple Difference Estimate with sibling fixed effect

Current Students		
	DD with Sibling Fixed Effect	DDD with Sibling Fixed Effect
Private School	0.0207	0.0224
	(0.0145)	(0.0324)
English Medium	0.0054	0.0171
	(0.0108)	(0.0241)
English Ability	-0.0005	0.0080
	(0.0177)	(0.0393)
Observation	11563	11563
Group	6434	6434
Test Scores		
Mathematics	0.2184	-0.1967
	(0.1306)	(0.2898)
Reading	0.1391	.2654
	(0.1295)	(.2966)
Writing	0.0443	.3158
	(0.1282)	(.2981)
Score	0.3211	.0732
	(0.2914)	(.6917)
Observation	2597	455
Group	2187	245

Table B.3: A3 Private School Enrollment

Private School	Percentage	No. of Observation
Percentage of kids in private school	30.8	83939
Percentage of boys in private school	33.4	44931
Percentage of girls in private school	27.9	39008
Percentage of boys in private school	33.4	44931
Percentage of boys in private school from all boys HH	33.6	23028
Percentage of boys in private school from mixed HH	33.1	21903
Percentage of girls in private school	27.9	39008
Percentage of girls in private school from all girls HH	28.7	16421
Percentage of girls in private school from mixed HH	27.4	22587

Table B.4: A4 Robustness Check for Age

	(1)	(2)
Current Students		
Private School	-0.0640*	-0.0681**
	(0.0372)	(0.0332)
English Medium	-0.0702**	-0.0765**
	(0.0353)	(0.0353)
English Ability	-0.0605	-0.0568
	(0.0464)	(0.0463)
N	11336	11336

Table B.5: A5 Alternate definition of Public Sector

	(1)	(2)
Current Students		
Private School	-0.0683**	-0.0729**
	(0.0348)	(0.0348)
English Medium	-0.0765**	-0.0816**
	(0.0354)	(0.0353)
English Ability	-0.0620	-0.0615
	(0.0431)	(0.0422)
N	13669	13669

Table B.6: A5 Including Mother's Sector

	(1)	(2)
Current Students		
Private School	-0.0648*	-0.0685**
	(0.0363)	(0.0341)
English Medium	-0.0675**	-0.0743**
	(0.0351)	(0.0350)
English Ability	-0.0571	-0.0534
	(0.0492)	(0.0486)
N	12079	12166

Table B.7: A6 Means of the Outcome Variables

	Public Sector				Private Sector			
	Post Reform		Pre Reform		Post Reform		Pre Reform	
	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl
Private School	0.56	0.44	0.44	0.4	0.53	0.45	0.51	0.44
English Medium	0.44	0.36	0.34	0.31	0.32	0.3	0.33	0.38
English Ability	0.67	0.63	0.48	0.48	0.44	0.46	0.55	0.25
Reading	3.12	3.2	3.21	3.29	2.89	2.93	3.07	3.07
Math	1.96	2	2.17	2.14	1.81	1.83	1.99	1.83
Writing	1.45	1.5	1.02	1	1.32	1.42	0.95	0.96

Means of the Outcome Variables

Table B.8: A7 Differential Effect of the reform on other forms of Investment

	Family with only Girl Child
Pension, LIC and Others	0.008
	(0.067)
N	1390
Mean of Dependent Variable	0.18

	Family with only Boy Child
Pension, LIC and Others	-0.003
	(0.055)
N	1817
Mean of Dependent Variable	0.19

Table B.9: A8 Household Characteristics by Household Composition

	HH with children of only one gender	HH with children of both gender
Father's Age	40.34	40.64
Father's Education	6.82	6.61
Mother's Age	35.31	35.66
Mother's Education	4.59	4.35
Urban	0.35	0.34
Number of HH members	6.37	6.73
HH Asset Score	14.21	14.11

Table B.10: A9 Private School Enrollment by Household Composition

	Private School Enrollment by Gender and Family Composition		
	HH with one Child	HH with more than One Child (Mixed Gender)	HH with more than One Child (One Gender)
Girl Child	0.31	0.28	0.26
Boy Child	0.34	0.33	0.33

Table B.11: A10 Results on Anthropocentric Measures

Children Age 0-5		
	(1)	(2)
Height for Age	0.35	0.36
	(0.24)	(0.24)
Mean of Dependent Variable	-1.36	
N	3219	
Weight for Age	0.19	0.2
	(0.21)	(0.21)
Mean of Dependent Variable	-1.04	
N	3219	
Weight for Height	0.1	0.08
	(1.68)	(1.1)
Mean of Dependent Variable	0.73	
N	3219	
BMI for Age	-0.2	-0.1
	(0.36)	(0.35)
Mean of Dependent Variable	-0.19	
N	3219	

*Table A10 contains estimates of β_7 from equation (8) for various outcome variables.

*The sample consists of children age 0-5.

*Each row presents the DDD estimate from equation 21 with a particular outcome variable.

*Standard errors are reported in brackets and are clustered at the district level.

*Specification (1) includes full set of controls : child characteristics, father characteristics, mother characteristics and household characteristics.

*Specification (2) does not include household characteristics.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bruce, Judith. "Married adolescent girls: human rights, health, and developmental needs of a neglected majority." Economic and Political Weekly (2003): 4378-4380.

Clark, Shelley. "Early marriage and HIV risks in sub-Saharan Africa." Studies in family planning 35.3 (2004): 149-160.

Nour, Nawal M. "Health consequences of child marriage in Africa." Emerging infectious diseases 12.11 (2006): 1644.

Santhya, Kidangamparampil G., et al. "Associations between early marriage and young women's marital and reproductive health outcomes: evidence from India." International perspectives on sexual and reproductive health (2010): 132-139.

Raj, Anita, et al. "Prevalence of child marriage and its effect on fertility and fertility-control outcomes of young women in India: a cross-sectional, observational study." The Lancet 373.9678 (2009): 1883-1889.

Field, Erica, and Attila Ambrus. "Early marriage, age of menarche, and female schooling attainment in Bangladesh." Journal of political Economy 116.5 (2008): 881-930.

Sekhri, Sheetal, and Sisir Debnath. "Intergenerational consequences of early age marriages of girls: Effect on children's human capital." The Journal of Development Studies 50.12 (2014): 1670-1686.

Chari, A. V., et al. "The causal effect of maternal age at marriage on child wellbeing: Evidence from India." Journal of Development Economics 127 (2017): 42-55.

Dhamija, Gaurav, and Punarjit Roychowdhury. "The Causal Impact of Women's Age at Marriage on Domestic Violence in India." (2018).

Liestøl, Knut. "Menarcheal age and spontaneous abortion: a causal connection?." American Journal of Epidemiology 111.6 (1980): 753-758.

MARTIN, ELIZABETH J., LOUISE A. BRINTON, and ROBERT HOOVER. "Menarcheal age and miscarriage." American journal of epidemiology 117.5 (1983): 634-636.

Wyshak, Grace. "Age at menarche and unsuccessful pregnancy outcome." Annals of human biology 10.1 (1983): 69-73.

SANDLER, DALE P., ALLEN J. WILCOX, and LOUISE F. HORNEY. "Age at menarche and subsequent reproductive events." American Journal of Epidemiology 119.5 (1984): 765-774.

Guldbrandsen, Karen, et al. "Age of menarche and time to pregnancy." Human Reproduction 29.9 (2014): 2058-2064.

Hossain, Parvez, Bisher Kawar, and Meguid El Nahas. "Obesity and diabetes in the developing world—a growing challenge." (2009).

Carmichael, Sarah. "Marriage and power: Age at first marriage and spousal age gap in lesser developed countries." The History of the Family 16.4 (2011): 416-436.

Stock, James H., and Motohiro Yogo. "Testing for weak instruments in linear IV regression." (2002).

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. "A survey of weak instruments and weak identification in generalized method of moments." Journal of Business & Economic Statistics 20.4 (2002): 518-529.

Hausman, Jerry, James H. Stock, and Motohiro Yogo. "Asymptotic properties of the Hahn–Hausman test for weak-instruments." Economics Letters 89.3 (2005): 333-342.

Cragg, John G., and Stephen G. Donald. "Testing identifiability and specification in instrumental variable models." Econometric Theory 9.2 (1993): 222-240.

Petridou, Eleni, et al. "Determinants of age at menarche as early life predictors of breast cancer risk." International journal of cancer 68.2 (1996): 193-198.

Garland, Miriam, et al. "Menstrual cycle characteristics and history of ovulatory infertility in relation to breast cancer risk in a large cohort of US women." American journal of epidemiology 147.7 (1998): 636-643.

Weber, Gerhard W., Hermann Prossinger, and Horst Seidler. "Height depends on month of birth." Nature 391.6669 (1998): 754.

Lokshin, Michael, and Sergiy Radyakin. Month of birth and children's health in India. The World Bank, 2009.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. "Identification and estimation of treatment effects with a regression-discontinuity design." Econometrica 69.1 (2001): 201-209.

Edwards, Ben, et al. "Is Monotonicity in an IV and RD design testable? No, but you can still check it." (2013).

Jayachandran, Seema. "The roots of gender inequality in developing countries." economics 7.1 (2015): 63-88.

Sen, Amartya. "More Than 100 Million Women Are Missing." New York Review of Books (2010).

Behrman, Jere R., and Anil B. Deolalikar. "Health and nutrition." Handbook of development economics 1 (1988): 631-711.

Zimmer, Zachary, et al. "Education of adult children and mortality of their elderly parents in Taiwan." Demography 44.2 (2007): 289-305.

Sadhak, H. "Does Not india Need a Default Option in the New pension system?." Economic and Political Weekly (2009): 59-68.

Gayathri, S. R. I. N. I. V. A. S. A. N. "The elderly in an Indian retirement community: the Influence of the caste system." Bold 17 (2006): 2-10.

Goswami, Ranadev. "Indian Pension System: Problems and Prognosis." Indian Institute of Management, Bangalore, January 262001 (2001).

Sanyal, Ayanendu. Civil Service Pension Reforms in India. Diss. INSTITUTE FOR SOCIAL AND ECONOMIC CHANGE BANGALORE, 2013.

Broadbent, John, Michael Palumbo, and Elizabeth Woodman. "The shift from defined benefit to defined contribution pension plans—Implications for asset allocation and risk management. Paper prepared for a working group on institutional investors, global savings, and asset allocation established by the Committee on the Global Financial System." (2006).

Pande, Rohini P., and Nan Marie Astone. "Explaining son preference in rural India: the independent role of structural versus individual factors." Population Research and Policy Review 26.1 (2007): 1-29.

Mutharayappa, Rangamuthia. "Son preference and its effect on fertility in India." (1997).

Arnold, Fred, Minja Kim Choe, and Tarun K. Roy. "Son preference, the family-building process and child mortality in India." Population studies 52.3 (1998): 301-315.

Barcellos, Silvia Helena, Leandro S. Carvalho, and Adriana Lleras-Muney. "Child gender and parental investments in India: are boys and girls treated differently?." American Economic Journal: Applied Economics 6.1 (2014): 157-189.

Bharadwaj, Prashant, and Leah K. Lakdawala. "Discrimination begins in the womb: Evidence of sex-selective prenatal investments." Journal of Human Resources 48.1 (2013): 71-113.

Chen, Lincoln C., Emdadul Huq, and Stan d'Souza. "Sex bias in the family allocation of food and health care in rural Bangladesh." Population and development review (1981): 55-70.

Gupta, Monica Das. "Selective discrimination against female children in rural Punjab, India." Population and development review (1987): 77-100.

Behrman, Jere R., and Anil B. Deolalikar. "Health and nutrition." Handbook of development economics 1 (1988): 631-711.

Borooah, Vani K. "Gender bias among children in India in their diet and immunisation against disease." Social science & medicine 58.9 (2004): 1719-1731.

Behrman, Jere R., Robert A. Pollak, and Paul Taubman. "Do parents favor boys?." International Economic Review (1986): 33-54.

Chaudhury, Nazmul, et al. "Missing in action: teacher and health worker absence in developing countries." The Journal of Economic Perspectives 20.1 (2006): 91-116.

Wadhwa, Wilima. "Are private schools really performing better than government schools." Annual Status of Education Report (Rural) New Delhi (2009).

Kingdon, Geeta Gandhi. "The progress of school education in India." Oxford Review of Economic Policy 23.2 (2007): 168-195.

Kingdon, Geeta Gandhi. "Does the labour market explain lower female schooling in India?." The Journal of Development Studies 35.1 (1998): 39-65.

Kingdon, Geeta Gandhi, and Jeemol Unni. "Education and women's labour market outcomes in India." Education Economics 9.2 (2001): 173-195

Kingdon, Geeta. "The quality and efficiency of private and public education: a case-study of urban India." Oxford Bulletin of Economics and Statistics 58.1 (1996): 57-82.

Tooley, James, et al. "The relative quality and cost-effectiveness of private and public schools for low-income families: a case study in a developing country." School Effectiveness and School Improvement 21.2 (2010): 117-144.

Pandey, Priyanka, Sangeeta Goyal, and Venkatesh Sundararaman. "Community participation in public schools: impact of information campaigns in three Indian states." Education Economics 17.3 (2009): 355-375.

Srinivasan, T. N. "Population growth and economic development." Journal of Policy Modeling 10.1 (1988): 7-28.

Jensen, Robert. 2003. "Equal Treatment, Unequal Outcomes? Generating Sex Inequality Through Fertility Behavior" Mimeo, Harvard University

De Vos, Susan. "An old-age security incentive for children in the Philippines and Taiwan." Economic Development and Cultural Change 33.4 (1985): 793-814.

Zhang, Junsen, and Kazuo Nishimura. "The old-age security hypothesis revisited." Journal of Development Economics 41.1 (1993): 191-202.

Ebenstein, Avraham, and Steven Leung. "Son preference and access to social insurance: evidence from China's rural pension program." Population and Development Review 36.1 (2010): 47-70.

Mu, Ren, and Yang Du. "Pension coverage for parents and educational investment in children: Evidence from urban China." The World Bank Economic Review (2015): lhv060.

Samanta, Tannistha, Feinian Chen, and Reeve Vanneman. "Living arrangements and health of older adults in India." The Journals of Gerontology Series B: Psychological Sciences and Social Sciences (2014): gbu164.

Kishor, Sunita. "" May God Give Sons to All": Gender and Child Mortality in India." American Sociological Review (1993): 247-265.

Ahlin, Christian, and Robert M. Townsend. "Using repayment data to test across models of joint liability lending." The Economic Journal 117.517 (2007): F11-F51.

Stiglitz, Joseph E. "Peer monitoring and credit markets." The world bank economic review 4.3 (1990): 351-366.

Banerjee, Abhijit V., Timothy Besley, and Timothy W. Guinnane. "Thy neighbor's keeper: The design of a credit cooperative with theory and a test." The Quarterly Journal of Economics 109.2 (1994): 491-515.

Besley, Timothy, and Stephen Coate. "Group lending, repayment incentives and social collateral." Journal of development economics 46.1 (1995): 1-18.

Ghatak, M. "Group Lending, Local Information and Peer Selection ', Journal of Development Economics, vol. 60." (1999).

Ghatak, Maitreesh, and Timothy W. Guinnane. "The economics of lending with joint liability: theory and practice1." Journal of development economics 60.1 (1999): 195-228.