The Changing Geography of Gender in India

Scott L. Fulford *

November 2013

Abstract

This paper examines the changing distribution of where women and girls live in India at the smallest possible scale: India's nearly 600,000 villages. Village India is becoming more homogeneous in its preferences for boys even as those preferences becomes more pronounced. A consequence is that more than two thirds of girls now grow up in villages where they are the minority. Most Indian women move on marriage, so parents' sex selection decisions are felt well beyond the village. Yet marriage migration does not have an equalizing influence on sexual imbalances across villages. Linking all villages across two censuses, I show that changes in village infrastructure such as roads, power supplies, or health clinics are not related to changes in child sex. Geographically close villages reinforce each other's preferences. The results suggest that there are no easy policy solutions for addressing the increasing masculinization of Indian society.

JEL classification: O15; J12; J16
Keywords: Sex ratios; Marriage migration; Son preference; Geographic distribution of women; Asia; India

*Boston College Department of Economics and International Studies Program, 140 Commonwealth Ave, Chestnut Hill, MA 02467; email: scott.fulford@bc.edu. Thanks to Mashfiqur Khan who provided research assistance, the Boston College Faculty Workshop on Global Development, and the participants and organizers at the 2013 World Bank/Center for Global Development/Agence Française de Dévelopment International Conference on Migration and Development in Morocco.

1 Introduction

Child sex ratios have become more masculine across Asia over the last three decades, partly driven by the spread of sex selective abortions (Guilmoto, 2009). In India child sex ratios have increased rapidly to 107 boys for every 100 girls (Guilmoto and Depledge, 2008); in some states the ratio is over 120 boys to girls. Overall sex ratios in India are around 108 men for every 100 women. Since without any sex selective abortion or excess female mortality there would typically be more women than men, that has led to the calculation that there are around 40 million "missing women" (Sen, 1992) in India and perhaps 100 million across the world (Klasen and Wink, 2002), although defining the benchmark is not trivial (Anderson and Ray, 2010).

A second, less well studied, phenomenon is that most migration in India is by women. Across India 75% of women over 21 have migrated, while only 15% of men live someplace other than their birth village or town. Female migration is so high because almost all women migrate on marriage in most of the populous states; marriage migration accounts for nearly 90% of female migration. Around 300 million women have migrated for marriage and an additional 20 million are migrating each year.[1]

These two phenomena of a large and growing sex imbalance for children and gendered migration interact, creating complex spatial and gender dynamics. Sex imbalances among children are largely determined by the pre-and post-natal decisions of parents (Arokiasamy, 2007). Parents live in a particular village and have access to sex selection technology and medical care that depends on the local infrastructure. Their preferences for sons and approaches to gender and marriage are shaped locally as well (Desai and Andrist, 2010). Yet with marriage migration so pervasive, parents' decisions are felt well beyond their own village. The combination of diverse sex ratios across India and marriage migration, for example, will determine the spatial distribution and severity of the coming "marriage squeeze" in India (Guilmoto, 2012; Sautmann, 2011).

[1] Known as patrilocal village exogamy, women are married outside of their natal village, joining their husband's family in his village. The statistics in this paragraph are based on calculations by the author from the Indian National Sample Surveys and the India Human Development Survey (Desai, Vanneman, and National Council of Applied Economic Research, 2008). Fulford (2013) provides a full description of marriage migration.

To understand how the the geography of gender is changing in India, this paper examines the spatial distribution of women and girls across all of India's nearly 600,000 villages in 1991 and 2001. Villages are the smallest rural administrative unit and contain two thirds of the Indian population and so are an appropriate place to start in order to understand the large social and demographic changes. This paper is concerned with two kinds of change: how the spatial distribution of where women and girls live has changed over time both within and across villages, and in the changes over the life-cycle as women move on marriage.

Although almost all of the focus on sex ratios in India is on variation at the state or district level (for example, Dyson and Moore (1983); Echávarri and Ezcurra (2010); Guilmoto (2009)), I show that most of the actual variation in where women and girls live is at the village level. The differences between states or districts are masking the much larger variation at the local level. In some ways that is not a surprise—much of the village variation is driven by randomness due to the small size of villages—but it is important to remember that the context in which parents make decisions is far more variable than the aggregates imply.

Decomposing the variance across villages, I find that village India is becoming more homogeneous in its preference for boys at the same time as this preference is becoming more pronounced. Villages are generally closer to the mean so there are fewer villages with a very large female deficit, but also fewer with a female surplus. At the same time, the entire distribution is shifting away from having girls. A consequence of these trends is that equality is much less common; while 36% of girls under seven lived in a village with at least as many girls as boys in 1991, only 30% lived in such a village in 2001 . While there used to be villages that might produce more girls than boys, such villages are increasingly rare. In some states they are practically non-existent. In Haryana and Punjab, for example, less than 10% of girls lived in villages where there were at least as many girls as boys in 2001. The overall masculinization of India is becoming increasingly universal.

The geographic distribution of where girls are born and survive then becomes the distribution of where women live as women migrate on marriage. I show that despite its large size, marriage migration does not have a balancing effect on sex imbalances. That suggests that the consequences of the growing preferences for boys will be felt mostly locally despite the vast movement of women. Since the migration distance is typically not very large (Fulford, 2013), and I show that village preferences for sons are positively correlated with nearby villages, there does not appear to be a stabilizing influence on sex imbalances through the local marriage market. Marriage practices may be changing in response to the growing female deficit (Kaur, 2004, 2013; Mishra, 2013), but the spread of dowries suggests those changes will not necessarily make daughters more desirable for parents (Anderson, 2003).

Finally, this paper examines the determinants of changes in the proportion of girls by linking each village across the 1991 and 2001 censuses to create a panel. Comparing the same village over time allows me to remove fixed village variation so instead of looking merely at cross-sectional correlations it is possible to examine what village changes are associated with changes in child sex.

Except for gaining a primary school, changes in village infrastructure such as gaining access to a road, a bus stop, medical facilities, drinking water or power sources are not linked to changes in the fraction of girls. A recent literature has emphasized excess female mortality as the cause of many of the missing women (Anderson and Ray, 2010; Oster, 2009a,b), but it is not clear that more medical service provision will have much effect. That could be because the quality of health clinics is often poor (Das, Hammer, and Leonard, 2008). On the other hand, there has been growing concern about the increasing penetration of ultrasonography and amniocentesis tests for sex selection before birth (Khanna, 2010). A largely ineffective ban on using ultrasonography to determine sex passed in 1994 and was strengthened in 2003. At least between 1991 and 2001, however, gaining access to roads, a bus stop, or a medical facility are not associated with a fall in the proportion of female children. That suggests that changes in physical access to facilities providing sex tests and abortions may not be primarily responsible for worsening sex ratios. That does not mean sex selective abortions are not widespread, only that villages that gained better physical access did not show larger declines in the proportion of girls.

Instead, it is changes in education that have the most important effects. Increases in female literacy *decrease* the fraction of children female, while increases in male literacy tend to increase the fraction female. These estimates are likely a combination of direct effects of education and other changes that are associated with literacy such as declines in family size and increase in income. I find little evidence of the non-linear education effect hypothesized by Echávarri and Ezcurra (2010).

That changes in village infrastructure have little effect and improvements in female literacy overall are associated with declines in the fraction of girls suggest that it may be difficult to enact policies to halt or reverse the growing sex imbalances. That is in keeping with the weak record of programs to provide financial incentives to have girls (Anukriti, 2013). Instead, the increased homogenization and masculinization of India will likely continue as female literacy spreads. Moreover, villages close together tend to move in the same direction so it seems that the larger cultural changes combined with the marriage markets are reinforcing the trend towards having fewer girls rather than slowing it.

2 The overall trends in village India

Approximately two thirds of all Indians still reside in rural areas. Within rural areas, the village is the smallest administrative unit. Villages need not include an actual village in the sense of a group of dwellings in close proximity although many do. As of 2001 there were 593,000 populated villages in India with an average of approximately 1,250 residents. Migration in the census and in the surveys used in this paper is defined as having left the village of birth and so villages also represent the relevant unit to understand how migration affects the geographic distribution of women. Table 1 gives summary statistics on villages and sex imbalances across the universe of villages in India and within the large Indian states in 1991 and 2001. The 2011 census is not yet available.

The growth in population in villages is substantial, as is evident in table 1. While the population of urban areas has grown somewhat faster than rural, urban migration has not been substantial and the India remains largely rural. Village population varies substantially both within states and between them.

The percent of children that are girls, already low at 48.67% in 1991, has decreased over the decade to 48.28%. Throughout I present the statistics in percents or fractions female rather than the more common sex ratio since the analysis of the variance in later sections makes more sense statistically and is less subject to outliers in fractions rather than ratios. That is because the child sex ratio divides by the number of boys which could be zero in small villages, leaving the ratio undefined. Even when the ratio is defined it can have extreme variability that is difficult to separate from village size and to link to the underlying preferences.

The census at the village level reports ages 0-6, the total population, and no other age groups. So children are age 0-6, and adults are everyone seven and older. Since the average age of marriage for women is under 18 and marriage is nearly universal (Desai and Andrist, 2010), the adult age group includes ages 7-17 who are typically unmarried and living with their parents. It would be interesting to separate age groups more finely and examine how the differential mortality examined by Anderson and Ray (2010) varies with age across villages, but that is not possible with the information at the village level. Between 1991 and 2001 there was an increase in the fraction of the "adult" population that was female. That increase appears to be because of more rapid decreases in female mortality (Guilmoto and Depledge, 2008).

The overall sex imbalances and trends in India mask large differences across regions. In Kerala the percent girls slightly increased, while in some states in the north it fell by a percentage point or more. The states are listed in order of rural population size rather than alphabetically or geographically to give a sense of the importance of different states in the overall trends. While in some states the fall has been precipitous, it is downward for all states except Kerala. On the other hand, almost all states have increased the percent women seven and older.

Similar spatial heterogeneity exists in the extent of marriage migration as well and for other measures of female autonomy (Dyson and Moore, 1983; Fulford, 2013). In the large northern states marriage migration is nearly universal. It is less common in the south, and uncommon in the north-west. Examining how migration rates change by age, Fulford (2013) shows that there have not been any large changes within living memory in migration rates. Women move on average about three and a half hours of travel time from their home village, although there is substantial variance as many move much closer and a few much farther.

3 The changing spatial distribution of girls

Figure 1 shows that there is substantial variation across village India in the proportion of girls and women by plotting the density of the fraction female across villages. I weight villages in the kernel density by village population, so the figure shows the density of women and girls in India living in a village with a given fraction women. Since the population of women is larger than the population of girls, the distribution of the fraction women is more compressed. The distribution of the fraction girls shows a distinct leftward shift from 1991 to 2001 consistent with the overall shift down in the fraction girls from table 1. There has also been a distinct compression of the distribution of the women from 1991 to 2001, largely because the density in the left tale of extreme female deficit has diminished.

The variation across villages has three primary sources: One source is the variation across states as is evident by looking at the fraction of children under six across states in table 1. Some states produce substantially a substantially lower fraction of girls than others, and this tendency became worse between 1991 and 2001 across almost all states. For some states such as Punjab and Haryana it became much worse. A second source is village level variation in the willingness to produce girl children within each state. Finally, village populations are small and so there is random variation in the number of girls born even without differential preferences.

To understand the sources of variation between villages, I perform a simple variance decomposition which can account for random variation, village population size, state or district differences, and underlying unobserved spatial variation in the willingness or ability to produce girls. The simple statistical approach cannot distinguish between preferences for sons, the ability to act on those preferences such as through sex selective abortion, and differential treatment in health or education that results in differential mortality. Bhat and Zavier (2003) have pointed out that the son preferences and the ability to act on those preferences through sex selective abortion are distinct. I will attempt to disentangle them in the regressions in section 5, but for now I continue to refer to underlying spatial differences as preferences to distinguish them from random differences. The decomposition tells three important things: how important district variation is in understanding the overall spatial distribution, how important variation in preferences is in determining the geographic distribution of girls and women, and whether the distribution of preferences has changed over time.

Suppose that in village i the probability of each child being a girl is p_i and there are n_i^c children. Then if n_i^f is the number of girls, the random variable $f_i = n_i^f / n_i^c$ has expected value $E[f_i|p_i, n_i^c] = p_i$ and variance $Var[f_i|p_i, n_i^c] = (1 - p_i)p_i/n_i^c$ since each child is equally likely to be a girl. Note that the variance of f_i falls as the population of children rises. Larger villages are much more likely to be close to their preferred fraction of girls. The simplicity of the statistical model comes from assuming that the probability of the next surviving child being a girl is constant within a village but may vary across villages. That is a useful simplification especially since the limitations of the census data make peering into household decisions impossible, but it clearly sweeps away possibly important family decisions. For example, parents may adapt their choices to the current and predicted sex ratios in their village and community or there may be heterogeneity in preferences within a village as well as across it.

Over a population of villages the distribution of f_i depends on the random variation of multiple draws from a binary distribution, the distribution of village sizes, and the underlying distribution of the unobserved preferences p_i. Denote $Var[f_i]$ as the population variance over villages i. With known child population n_i^C and unknown p_i for each village, the variance can be decomposed using the law of total variance into the portion of the variance that comes from random variation around p_i and preference variation from the distribution for p_i:

$$Var[f_i] = E[Var[f_i|p_i, n_i^c]] + Var[E[f_i|p_i, n_i^c]] = E[(1 - p_i)p_i/n_i^c] + Var[p_i] \quad (1)$$

using the binomial distribution that each village is drawing from.

Then the simplest approach to calculating the decomposition is to assume the independence of n_i^c and p_i. The advantage of this approach is that it makes very clear how changes in population affect the variance. The disadvantage is that it does not constrain the preference variance to be positive because it implies that there is no behavioral relationship between the number of children in a village and its preferences for girls. I relax this assumption in section 5 when I create a panel of villages across censuses to examine changes in preferences.

Assuming the independence of p_i and n_i^c then $E[(1-p_i)p_i/n_i^c] = \omega_c(\bar{p}(1-\bar{p}) - Var[p_i])$ where $\omega_c = E[1/n_i^c]$ is the population mean of the inverse of the number of children and a bar represents the mean. Then solving for $Var[p_i]$:

$$Var[p_i] = (Var[f_i] - \omega_c \bar{p}(1-\bar{p}))/(1-\omega_c). \tag{2}$$

This formula is useful because it emphasizes the importance of village population size. The larger ω_c is, the lower the random variation across villages in the fraction of girls because each village will be (on average) closer to its preferred fraction girls. So larger populations within villages on average imply lower geographic variance and the formula can remove this source of variation.

Table 2 summarizes the results of calculating the village level variances in the percent of girls and decomposing that variance. The variances are calculated in percents for readability. The overall village level variance fell somewhat from 1991 to 2001. The size of the villages also increased by nearly a third, however, and the variance decomposition essentially asks how much variance we should expect given the size of the villages.

First, very little of the total variation comes from differences across states and districts. I show this result by first removing the district mean from the village percent female before calculating the variance. In 2001 only 2% of all variation between villages across India can be explained by the broader state and district differences. Although the demography and economics of sex imbalances has been focused on explaining district or regional level differences (Echávarri and Ezcurra, 2010; Guilmoto and Depledge, 2008; Murthi, Guio, and Drèze, 1995), most of the actual variation takes place at the village level. The extensive local variation matters because explaining the overall declines requires understanding the decisions of parents who see much more variation than the district aggregates imply. The state and district contribution to the variance has doubled since 1991, however, as some states have moved increasingly further from the average.

More important, village India is becoming increasingly homogeneous in its preference for boys. The fraction of the variance explained by the underlying preference variation halved between 1991 and 2001, even as the total variance decreased. In 1991 about 10% of the overall variance came from different preferences among villages across all of India, in 2001 only 5% of a smaller variance came from such variation.

Villages within each state are also becoming more homogeneous in their preferences as well. While there are substantial differences in the level of variance between states, these differences are largely due to differences in average village size. Since the calculation of the preference variance does not constrain it to be greater than zero, negative values suggest that in those states random variation explains nearly all of the spatial variance. For example, Punjab has become almost entirely homogeneous in its strong preferences for boys. But it is the large declines in preference heterogeneity in the biggest states such as Uttar Pradesh and Bihar with a combined rural population of 200 million that drive the overall changes. Kerala is an outlier because of its large village sizes.

Increased homogeneity has the important implication that there is unlikely to be a geographic effect to help stabilize sex imbalances. It seems intuitively appealing to think that some parts of Punjab, for example, might start producing more women due to the sexual imbalance since women are more in demand in the rest of the state. Such a specialization in producing women is the geographic implication of the model introduce by Edlund (1999). Yet the exact opposite is happening: far from there being evidence of increasing geographic specialization, there is instead increased homogenization

The changes in the distribution of girls across villages have combined to make girls the minority almost everywhere. The last three columns of table 2 show the fraction of all girls in village India who live in villages where girls make up 50% or more of children under seven. In 1991 around 36% of girls lived in a village where the children were at least half female. In 2001 only 30%

lived in such a village, a decline of 17% or six percentage points. The experience of the large and growing majority of Indian girls is to grow up in a context where they are a distinct minority.

The decline in the fraction of girls who grow up in villages with at least the same number of girls as boys is driven by two distinct trends. First, the overall downward shift in the fraction female makes equality a less frequent event. That is evident in the fall in percent female under seven in table 1 and the shift left in figure 1. Second, the distribution of girls across villages is becoming more homogeneous so the density in the right tail of the distribution is falling. The decline in the variance is clear from the first columns of table 2 and in figure 1. As table 2 shows, some of the tightening variance comes from an increase in village size and some from the increasingly homogeneous preferences.

The overall Indian statistics again hide some startling differences across states. In Haryana and Punjab less than 10% of all girls grow up in a village where they make up half of the children in 2001. Both states also had precipitous falls from already low levels in 1991. In these and several other states in the north the experience of almost all Indian girls is to grow up as a noticeable minority. Kerala also has a noticeably lower percentage of girls in equal sex villages, but that is caused not by a strong son preference but instead by the large village sizes in Kerala. Large village sizes mean low variance and so there are few villages away from the mean and so very low density in the tails of the distribution.

Village variation does not seem to be systematically related to sex imbalance. The most imbalanced states, Punjab and Haryana, do not have higher village level variation.

4 The changing spatial distribution of women

Where women live in India is determined by two forces. First, there is the the geographic dimension of where girls are born and survive the first several years. Since villages are small, some of that variation is random, but some is determined by preferences of families and their access to sex selection technologies and health care. Second, the geographic distribution of women is determined largely by marriage migration. This section documents that female marriage migration is so extensive that it is not possible to understand the wider effects of declining female to male sex ratios among children without understanding that parents export the effects of their decisions by marrying their daughters outside the village. Yet I show that even with so much migration, the variance across villages looks similar to what it would be if no women moved from their birth villages. That suggests that the influence of marriage migration is primarily through how it changes preferences for girls and not in helping to equalize imbalanced sex ratios across villages. Some recent scholarship has suggested that marriage and migration practices are changing in response to the growing female deficit (Kaur, 2004, 2013; Mishra, 2013), although the focus so far is on the rise in the very small number of cross-region marriages.

Across rural India, about three quarters of all women over 21 have migrated. The first column in table 2 shows the percent of women who have migrated for any reason. 90% of women migrating do so on marriage. There is substantial regional variation in migration rates. Migration exceeds 90% in many of the northern states, is around 60% in much of the south, and around 30% in the north-east. Fulford (2013) provides a more detailed geographic breakdown and examines the causes for this widespread migration.

The previous section documented the extensive spatial variation in where girls live across village India. That suggests that at least some marriage migration would be necessary to equalize the distribution of women. A useful way to characterize the variation across India and within states of the fraction of children female is to calculate how many girls in a given 0-6 cohort would need to move eventually to equalize fraction everywhere. That also helps to understand the role played by marriage migration. The calculation is relatively simple: just add up all of the girls who live in villages with more girls than the national or state average.

Across India, despite the large disparities in villages across and within states only 2.7 percent of girls would need to migrate eventually to exactly equalize the geographic distribution in their cohort across all states and all villages. These calculations are shown in in the first columns of table 2. There would still be too few women but there would be exactly the same proportion of women everywhere. State level variation is relatively unimportant in the number of girls who must move. Although there is more village level variation in some states than others, and some states have worse sex ratios, equalizing migration is similar in most states. The exceptions are Kerala where there is little village level variation since villages are large, and the states Himchal Pradesh and Uttaranchal (now Uttarakhand) where there is much higher village variation.

Since 75% of women over 21 have migrated, the actual movement of women is far larger than necessary to equalize the geographic distribution of girls. That means that the distribution of women across village India is not directly related to the distribution of girls but instead is almost entirely determined by marriage migration. The proportion of women in a cohort that would need to move has been falling in most states and across India because of the increased homogenization.

Yet even though most women in India move, the actual variance across villages is very similar to what would obtain if no women moved. It is possible to calculate what the spatial variance of the adult population would be if there were no migration based on the same variance decomposition in section 3. Then each village simply draws from its underlying preference for girls p_i for a larger population. That should reduce the variance since there are more adults (where adult is over six since that is what the census measures) than children and so the village should be closer to p_i.[2]

The last four columns of table 3 show the variance that would occur if there were no marriage migration of women and the actual village variance of the percent women. Even though most women live some place other than where they were born, the actual sex imbalances are not much different than if no women moved. One reason for this result may be that most women do not move far (Fulford, 2013) and village preferences are strongly shaped by the preferences of the villages close by. I provide some evidence for this hypothesis in the next section.

It is important to note that even if marriage migration does not exacerbate the sex ratio imbalances for the current generation, it may reinforce the value of sons and increase the incentives to spend scarce resources on sons' health and education over daughters' (Jayachandran and

[2] The predicted variance of the fraction female is then $Var^P[F_i] = \omega_A(\bar{p}(1-\bar{p}) + Var[p_i])$ where $\omega_A = E[1/n_i^A]$ replaces ω_c. This calculation applies the variance of p_i to the entire adult population and so it assumes that the variation in the birth and survival of girls is the same as the variation in the survival of women. Given the limited age information in the village census it is not possible to do better, but the evidence does suggest that excess female mortality changes with age (Anderson and Ray, 2010). It seems likely that even if it varies with age, across villages excess mortality at any given age is closely correlated with p_i.

Kuziemko, 2011; Oster, 2009a,b), worsening the imbalances for the next.

5 The village level determinants of child sex imbalances

The previous sections have documented a substantial overall fall in the fraction of children that are female from 1991 to 2001, accompanied by an increase in the homogeneity of preferences across villages. This section examines the sources of these trends at the village level by linking each village across the censuses to examine what has prompted changes within villages. By creating a panel of villages, I can difference out fixed village level unobservables and district trends and examine what changes within each village are associated with changes in the fraction female. Using a panel offers a substantial improvement over using a cross-section since many of the unobservables are likely to be highly correlated with sex ratios and with other determinants of the sex ratio and so lead to heavily biased estimates.

To my knowledge, only one other study uses the village level data to examine sex imbalances although they have been used to study fertility changes (Fulford, 2009; Guilmoto, 2005). Deolalikar, Hasan, and Somanathan (2009) create a panel of villages from the 1991 and 2001 censuses to examine how the juvenile sex ratio changes. This paper matches nearly 200,000 more villages and accounts for the differences in village sizes and so offers much more robust results.[3]

[3]The approach used in this paper differs both methodologically and in the included regressors: First, I am able to match close to 200,000 more villages than Deolalikar, Hasan, and Somanathan (2009) and so have close to the universe of villages (see footnote 4). Second, I account for village level heteroskedasticity—smaller villages must have larger variances. That matters since I find that the effects of literacy vary with village size. Not accounting for heteroskedasticity means that small villages receive disproportionate weight. Third, I use the percent of children that are girls rather than the sex ratio. The sex ratio, while easy to interpret, has a number of properties that make it an unwieldy transformation at the village level. Since villages are small, the sex ratio involves dividing by a random variable that is potentially zero. That shows up in the extreme variability of the 0-6 sex ratio: the standard deviation of the change is 63 even as the mean is a smaller -2.49. Since small villages are more likely to be in the tails, that exacerbates the problem of heteroskedasticity. Finally, I include district level effects. The evidence suggest that there is a strong spatial correlation among districts (Guilmoto and Depledge, 2008). When I follow an estimation approach similar to Deolalikar, Hasan, and Somanathan (2009), I find that these methodological differences matter. Some village infrastructure variables are significant in some specifications but not in others. Altogether, when I include district effects, allow for heteroskedasticity, and limit the estimation to villages with child sex ratio changes less than 60, none of the village infrastructure coefficients are large or significant. That suggests that the results of Deolalikar, Hasan, and Somanathan (2009) are largely driven by the unobserved district level correlations and the large outliers from using the sex ratio.

The Census of India provides demographic information for each village in its Primary Census Abstract and information on "village amenities" in the village directory. The amenities include whether the village is accessible by a footpath, a dirt road, or a paved road, whether it has a power supply, its drinking water source, and whether it has a hospital or clinic, schools, post offices or any transportation facilities like bus stops or rail stations. Starting with the village directory from 1991 census, I match it with the primary census abstract, and then use the village location codes introduced in the 2001 census to match with the primary census abstract and village directory from the 2001 census. Each matching brings attrition and some errors from incorrectly coded location variables. The primary census abstract contains information on 593,000 villages with non-zero population in 2001, and approximately 580,000 thousand villages with non-zero population in 1991. I can match 567,000 inhabited villages with information across both census years. Of these, 556,500 have village directory information for both years for all the relevant variables.[4]

The interactions between the literacy rates for men and women and of village population size are crucial for understanding changes in the fraction female, so I first describe these distributions. Figure 2 shows the distribution across villages of the population size, the change in the fraction female children within a village from 1991 to 2001, the distribution of literacy rates for men and women across villages, and the change in the male and female literacy rates within villages. As was clear from table 1, villages are typically small with a mean population size of about 1,000 in 1991 and 1,250 in 2001, but there is substantial variation in village size both within and across states.

While figure 1 showed how the density of the fraction female across villages changed from 1991 to 2001, by linking the same villages across censuses the top right panel of 2 shows the

[4]The census was not collected in Jammu and Kashmir in 1991, and Meghalaya is missing from the 2001 primary census abstract (although not from the census in general, so this omission may be a problem with the published data). Together these two states account for 13,000 of the non-matching villages between 1991 and 2001. The village is an administrative unit and so there are many villages that do not have any population, such as a national park or for which information is not available, such as a military base. The data are from the village directory and primary census abstracts of the 1991 and 2001 Census of India Registrar General of India (2004), Registrar General of India (2005), and Registrar General of India (1999). The naming conventions for variables varies across states and censuses as does the coding and storage of the data. That makes linking the villages across censuses and comparing them across states arduous and time consuming.

density of changes within villages. The overall downward shift in the fraction girls is nearly impossible to discern within the overall variation over the decade. Some villages have far more girls in 2001 than in 1991 and some far fewer. From the parents' perspective within a village, there may be no obvious trend, and in many villages it may look like girls are becoming more common. That parents are making decisions in a context far more variable than the aggregate is important for understanding those decisions; for almost half of villages there are more girls than there used to be.

Literacy has increased substantially for both men and women (bottom left panel of figure 2). The average village increased the literacy rate by 0.15 for men and 0.18 for women, although again with substantial variance. The distribution of the changes within villages are surprisingly similar for both men and women (bottom right panel), despite the very different initial distributions.

The estimation approach is the regression analog of the variance decomposition in section 3. Suppose the preferences or ability to act on those preferences for the fraction of children that are female in a given village can be expressed as:

$$p_{it} = \theta_i + p_{dt} + X_{it}\beta + \nu_{it} \tag{3}$$

where θ_i represents all of the fixed characteristics of a village such as its location, its immutable preferences, and its level of economic development; p_{dt} is the district level fraction of children female; and X_{it} is a vector of village level characteristics such as transportation, health care centers, and literacy rates that may change. Since we would expect that θ_i is related to both observed and unobserved variables, a single cross-section at the village or district level will not in general produce reliable estimates of β. Including districts effects allows for arbitrary spatial correlations at the district level.

As in the variance decomposition, it is not possible to observe p_{it} directly, but instead we observe the actual fraction of children that are female $f_{it} = p_{it} + \epsilon_{it}$ where ϵ_{it} is the difference between them. Since ϵ_{it} is the result of multiple draws from the same binary distribution it is possible to calculate its variance for each village. Replacing p_{it} with f_{it} in equation (3) does not create a measurement error problem in the standard sense of introducing attenuation bias (Deaton, 1997, p. 99-101) in the coefficients since the measurement error is on the dependent variable. Instead, it creates a problem of heteroskedasticity since $Var[\epsilon_{it}]$ varies with the number of children. I deal with this problem directly by calculating the variance for each village assuming draws from a binary distribution and allowing for it in the regressions. Correctly accounting for this heteroskedasticity is very similar to weighting by village size since for larger villages f_{it} is closer to the p_{it} and so these villages should get higher weight. Weighting is very important in this context since the estimated effects vary with village size and so not correctly weighting means that changes in small villages receive disproportionate importance.

Then with each village measured at each census I take the first difference to remove the village specific fixed effects θ_i and allow for district specific trends:

$$\Delta f_{it} = \Delta p_{dt} + \Delta X_{it}\beta + \Delta\tilde{\nu}_{it} \tag{4}$$

where $\Delta X_{it} = (X_{it} + X_{it-1})$ represents the change in village characteristics from census to census and $\Delta\tilde{\nu}_{it}$ absorbs both the measurement error and unobservable changes.[5] Within the change in village characteristics I include the changes in village infrastructure that the census reports for villages such as roads, medical clinics, schools, access to drinking water, and power supplies, as well as changes in population size and the percent of scheduled tribes and scheduled castes. The full list of regressors is in table 4. Echávarri and Ezcurra (2010) suggest that education may have non-linear effects. To allow for such complex effects the estimates includes the change in the male literacy rate, the change in the female literacy rate, their squares, and all of their interactions, as well as all of their interactions with log population in 1991.[6] Allowing for non-linear interactions

[5] In practice, I estimate equation 4 using weighted least squares and so assume multiplicative heteroskedasticity rather than additive. So $Var[\Delta\tilde{\nu}_{it}|\Delta X_{it}] = \sigma_i^2 = \sigma^2((1-p_{d2001}) * p_{d2001}/n^C_{i2001} + (1-p_{d1991}) * p_{d1991}/n^C_{i1991})$.

[6] So letting ML stands for male literacy rate, FL for the female literacy rate, and VP for the log village population in 1991 the regressions include the following variables: ΔML_{it}, $(\Delta ML_{it})^2$, ΔFL_{it}, $\Delta ML_{it} * \Delta FL_{it}$, $(\Delta ML_{it})^2 * \Delta FL_{it}$, $(\Delta FL_{it})^2$, $\Delta ML_{it} * (\Delta FL_{it})^2$, $(\Delta ML_{it})^2 * (\Delta FL_{it})^2$, $VP_i * \Delta ML_{it}$, $VP_i * (\Delta ML_{it})^2$, $VP_i * \Delta FL_{it}$, $VP_i * \Delta ML_{it} * \Delta FL_{it}$, $VP_i * (\Delta ML_{it})^2 * \Delta FL_{it}$, $VP_i * (\Delta FL_{it})^2$, $VP_i * \Delta ML_{it} * (\Delta FL_{it})^2$, $VP_i * (\Delta ML_{it})^2 * (\Delta FL_{it})^2$, $(VP_i)^2 * \Delta ML_{it}$, $(VP_i)^2 * (\Delta ML_{it})^2$, $(VP_i)^2 * \Delta FL_{it}$, $(VP_i)^2 * \Delta ML_{it} * \Delta FL_{it}$, $(VP_i)^2 * (\Delta ML_{it})^2 *$

allows a small increase in female literacy to have a different effect than a large increase, for those effects to vary with how much male literacy has increased, and to vary with how large the village is.

Table 4 shows the estimated effects for village infrastructure, and I examine the effects of literacy and population size in figures 3 and 4. I show figures rather than report the coefficients for the interaction terms since the individual coefficients are impossible to interpret in isolation. Both figures 3 and 4 evaluate the effects at the mean of all other covariates not shown to isolate changes in literacy. Since the interactions allow the effects to vary with population size, I limit the sample to villages that were larger that 50 and smaller that 7,000 in 1991 to keep the results from being driven by the very large or very small villages; as is clear from figure 2 that includes approximately 96% of all villages.

Changes in village infrastructure are not in general strongly related to changes in the proportion children female. Gaining a primary school reduces the percent female by about one fourth of the mean reduction—a large effect closely related to increases in literacy—but otherwise changes in infrastructure are small in size and not statistically significant. Against the overall negative trend caught by the district changes, higher growth in village population tends to increase the proportion female. That is consistent with recent work suggesting that in at least some parts of India smaller families may lead to more skewed sex ratios (Anukriti, 2013; Khanna, 2010).

Even with the literacy rate interactions and allowing for district specific trends, the explanatory power of the model is low. While the difference between the observed fraction of girls f_{it} and the underlying preferences p_{it} does not necessarily produce bias, it does introduce unexplained variance in the model. As the variance decomposition demonstrated, the fraction of female children within a village is largely and increasingly a matter of chance.

Bhat and Zavier (2003) point out that gender preferences and the ability to act on those preferences are not necessarily the same since the ability to act on them is partly determined by access to sex selective technologies. That means that one potential reason for the increasing homogenization

ΔFL_{it}, $(VP_i)^2 * (\Delta FL_{it})^2$, $(VP_i)^2 * \Delta ML_{it} * (\Delta FL_{it})^2$, $(VP_i)^2 * (\Delta ML_{it})^2 * (\Delta FL_{it})^2$. I do not include the main effect of log population since it is already included in the fixed effects.

is that more villages have access to sex-selective technologies. It does not seem that changes in access to transportation or medical facilities have much of an effect, however. Perhaps the spread is happening outside of villages, or demand for sex-selection technologies is highly inelastic so that those who want it manage to gain access regardless of the transportation costs.

Instead of changes in village infrastructure, it is changes in literacy rates that are more important, but the effects vary by population size. The top panel of figure 3 shows the estimated marginal effects from changes in male and female literacy for different population sizes. Each plot shows the impact of increasing the female or male literacy rate evaluated at different village population sizes, and the different plots show how the marginal effects change as the changes in literacy rates increase. So the dashed line in the top right plot, for example, shows the marginal effect of an increase in male literacy by one unit in villages which had no change in male literacy but increased the female literacy rate by 0.32 (the mean plus a standard deviation) between 1991 and 2001. The shaded areas are 95% confidence intervals.

Changes in male and female literacy have opposing effects: in small villages increases in male literacy are associated with a higher fraction of female children while increases in female literacy tend to lower the fraction female. The marginal effect of male literacy is declining with village size, however, while the marginal effect of female literacy is increasing. For large villages increases in male literacy decrease the fraction of children female, while increases in female literacy have little effect.

Variations in population and literacy combine to produce meaningfully large differences in the fraction female. I show these predicted effects in the bottom panel of figure 3. Each plot shows the predicted change in the fraction female for different size villages when there is a change in male or female literacy holding all of the other variables constant at their means. So the middle plot shows the predicted change in female literacy when the female literacy rate increases by 0.18 (its mean) and the male literacy rate increases by 0.15 (its mean) for villages of different sizes. The leftmost column predicts the effects when there are no changes in female literacy, the rightmost when there is an increase of 0.32 (the mean plus one standard deviation).

The effects of increasing both male and female literacy rates is an overall decrease in the percent of children female. Because male and female literacy rates have marginal effects that are changing in the opposite direction with village size, when they are both increasing together at approximately the same rate the village size does not matter. The negative effects of female literacy tend to predominate, however, so in the bottom panel of figure 3 comparing the plots along the diagonal from top left to bottom right as both female and male literacy increases, the overall fraction of children female falls.

Echávarri and Ezcurra (2010) propose a model in which there is an non-linear relationship between gender bias and eduction. Education may change preferences. It may also change the ability to access or pay for sex-selection technologies, and their model suggests that these two effects may work in opposite directions. Using a cross-section of districts in India, they find an inverse-U relationship between literacy rates and the sex-ratio at birth. Yet literacy rates are likely to be highly correlated with unobservables at the district level such as female autonomy. Figure 4 suggests that the effect they find is likely caused by omitted variable bias. It shows the effect from changes in male and female literacy rates evaluated at the mean of all other covariates. An increase in female literacy has an approximately constant negative effect. An increase in male literacy may have a non-linear effect, but it seems to follow a U shape rather than the inverse-U proposed by Echávarri and Ezcurra (2010).

Finally, I briefly examine the importance of surrounding villages. To examine whether the immediate surroundings matter more than district level trends, columns two and three in table 4 include the average tehsil fraction female. Tehsils are sub-district administrative units and contain an average of approximately 135 villages each. I calculate the tehsil mean for each village excluding the individual village so there is no mechanical correlation between the two. Even including the district level trend, what happens in the surrounding villages is important: a one percentage point increase in the percent female in the surrounding tehsil is associated with a 0.17% village increase. When I exclude the district trends, the surrounding area becomes even more important. Villages seem to act very strongly together. That is very much in keeping with Guilmoto and

Depledge (2008) who find that at the district level sex ratios display a high degree of spatial correlation. Note that by including district level trends, the estimates already allow for arbitrary spatial correlation at the district level.

A positive coefficient for the tehsil average means that villages tend to move together and implies that sex imbalances are amplified geographically. If one village has fewer girls, other villages have fewer girls, which strengthens and reinforces the tendency in the first village.

6 Discussion and Conclusion

The falling female to male ratio in India and other countries in Asia has prompted efforts to slow the trend including banning sex-selective abortions (Arnold, Kishor, and Roy, 2002) and offering monetary incentives to have girls (Anukriti, 2013), but these efforts have proved largely ineffective. Moreover, since skewed sex ratios are often the worst in elite groups (Patel et al., 2013), this problem is not likely to diminish with economic development. The results in this paper suggests that otherwise positive developments such as increased female literacy or the spread of primary schools are associated with declines in the fraction of children female in villages in India. Very little else about village infrastructure matters. That suggests that it will be difficult to halt the trend by providing more services and that some development efforts such as promoting education for women may make the situation worse.

Instead all of the trends seem to be towards homogenization and reinforcing the skewed sex ratios. Rural India has become increasingly homogeneous in its preferences for boys. There appears to be no geographic specialization encouraging some villages to have more girls. Since the pervasive marriage migration of women means that parents spread the effects of their decisions well beyond the village itself, it is not clear that the marriage market provides strong pressure to have more girls. The pressure seems to be working the other way as dowries have spread and perhaps become larger despite being illegal (Anderson, 2003; Sautmann, 2011). Moreover, the positive correlation between changes in nearby villages means that declining female to male sex ratios tend to amplify each other.

The trends seem to be reinforcing and making more universal the low value placed on having daughters. It seems that more parents are adopting the view that “raising a daughter is like watering your neighbor’s garden.” The geography of gender in India is becoming more homogeneous as well as more masculine and there do not appear to be easy ways to change the trends.

References

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

Anderson, Siwan and Debraj Ray. 2010. "Missing Women: Age and Disease." *The Review of Economic Studies* 77 (4):1262–1300.

Anukriti, S. 2013. "The Fertility Sex-Ratio Trade-off: Unintended Consequeces of Financial Incentives." Working Paper 827, Boston College. URL http://fmwww.bc.edu/EC-P/wp827.pdf.

Arnold, Fred, Sunita Kishor, and T. K. Roy. 2002. "Sex-Selective Abortions in India." *Population and Development Review* 28 (4):759–785.

Arokiasamy, Perianayagam. 2007. *Watering the Neighbour's Garden: The Growing Demographic Female Deficit in Asia*, chap. Sex Ratio at Birth and Excess Female Child Mortality in India: Trends, Differentials and Regional Patterns. Committee for International Cooperation in National Research in Demography, 49–72.

Bhat, P. N. Mari and A. J. Francis Zavier. 2003. "Fertility decline and gender bias in Northern India." *Demography* 40 (4):637–657.

Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. 2008. "The Quality of Medical Advice in Low-Income Countries." *The Journal of Economic Perspectives* 22 (2):pp. 93–114.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: Johns Hopkins University Press.

Deolalikar, Anil B., Rana Hasan, and Rohini Somanathan. 2009. "Public Goods Access and Juvenile Sex Ratios in Rural India: Evidence from the 1991 and 2001 Village Census Data." ADB Economics Working Paper 167, Asian Development Bank. URL http://www.adb.org/sites/default/files/pub/2009/Economics-WP167.pdf.

Desai, Sonalde and Lester Andrist. 2010. "Gender scripts and age at marriage in India." *Demography* 47 (3):667–687.

Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research. 2008. "India Human Development Survey (IHDS) 2005." Inter-university Consortium for Political and Social Research (ICPSR). Accessed 12 November 2011.

Dyson, Tim and Mick Moore. 1983. "On Kinship Structure, Female Autonomy, and Demographic Behavior in India." *Population and Development Review* 9 (1):pp. 35–60.

Echávarri, Rebeca A. and Roberto Ezcurra. 2010. "Education and gender bias in the sex ratio at birth: Evidence from India." *Demography* 47 (1):249–268.

Edlund, Lena. 1999. "Son Preference, Sex Ratios, and Marriage Patterns." *Journal of Political Economy* 107 (6):pp. 1275–1304.

Fulford, Scott. 2009. "The tradeoff between children and savings: Banking and fertility in India." In *Essays in Development Economics*. Proquest.

———. 2013. "The Puzzle of Marriage Migration in India." Working Paper 820, Boston College. URL http://fmwww.bc.edu/EC-P/wp820.pdf.

Guilmoto, Christophe Z. 2005. "Fertility Decline in India: Maps, Models and Hypotheses." In *Fertility Transition in South India*. New Delhi: Sage Publications, 385–435.

———. 2009. "The Sex Ratio Transition in Asia." *Population and Development Review* 35 (3):519–549.

———. 2012. "Skewed Sex Ratios at Birth and Future Marriage Squeeze in China and India, 2005–2100." *Demography* 49 (1):77–100.

Guilmoto, Christophe Z. and Roger Depledge. 2008. "Economic, Social and Spatial Dimensions of India's Excess Child Masculinity." *Population (English Edition)* 63 (1):pp. 91–117.

Jayachandran, Seema and Ilyana Kuziemko. 2011. "Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India." *The Quarterly Journal of Economics* 126 (3):1485–1538.

Kaur, Ravinder. 2004. "Across-Region Marriages: Poverty, Female Migration and the Sex Ratio." *Economic and Political Weekly* 39 (25):pp. 2595–2603.

———. 2013. "Mapping the Adverse Consequences of Sex Selection and Gender Imbalance in India and China." *Economic and Political Weekly* 48 (35).

Khanna, Sunil K. 2010. *Fetal/Fatal Knowledge: New Reproductive Technologies and Family-Building Strategies in India*. Belmont, CA: Wadsworth/Cengage Learning.

Klasen, Stephan and Claudia Wink. 2002. "A Turning Point in Gender Bias in Mortality? An Update on the Number of Missing Women." *Population and Development Review* 28 (2):285–312.

Mishra, Paro. 2013. "Sex Ratios, Cross-Region Marriages and the Challenge to Caste Endogamy in Haryana." *Economic and Political Weekly* 48 (35).

Murthi, Mamta, Anne-Catherine Guio, and Jean Drèze. 1995. "Mortality, Fertility, and Gender Bias in India: A District-Level Analysis." *Population and Development Review* 21 (4):745–782.

Oster, Emily. 2009a. "Does increased access increase equality? Gender and child health investments in India." *Journal of Development Economics* 89 (1):62 – 76.

———. 2009b. "Proximate sources of population sex imbalance in India." *Demography* 46 (2):325–339.

Patel, Archana B., Neetu Badhoniya, Manju Mamtani, and Hemant Kulkarni. 2013. "Skewed Sex Ratios in India: Physician, Heal Thyself." *Demography* 50 (3):1129–1134.

Registrar General of India. 1999. “Village Directory and Primary Census Abstract, Census of India 1991.”

———. 2004. “Primary Census Abstract, Census of India 2001.” Volumes 1-12.

———. 2005. “Village Directory and Location Codes, Census of India 2001.”

Sautmann, Anja. 2011. “Partner Search and Demographics: The Marriage Squeeze in India.” Working Papers 2011-12, Brown University, Department of Economics. URL http://ideas.repec.org/p/bro/econwp/2011-12.html.

Sen, Amartya. 1992. “Missing Women: Social inequality outweighs women’s survial advantage in Asia and north Africa.” *British Medical Journal* 304:587–588.

Table 1: Village census statistics

State	Vill. Pop. (millions) 2001	Villages 2001	Mean Village Size 2001	Mean Village Size 1991	Female$\leq$6 (%) 2001	Female$\leq$6 (%) 1991	Female$\leq$6 (%) Change	Female>6 (%) 2001	Female>6 (%) 1991	Female>6 (%) Change
India	742.30	593,622	1,250	991	48.28	48.67	-0.39	48.67	48.35	0.32
Uttar Pradesh	131.66	97,942	1,344	988	47.94	48.09	-0.15	47.36	46.45	0.91
Bihar	74.15	39,020	1,900	1,285	48.56	48.80	-0.24	47.94	47.45	0.49
West Bengal	57.72	37,945	1,521	1,192	49.05	49.22	-0.17	48.65	48.31	0.34
Maharastra	55.78	41,095	1,357	1,106	47.81	48.79	-0.98	49.19	49.48	-0.30
Andhra Pardesh	55.40	26,613	2,082	1,733	49.05	49.48	-0.43	49.66	49.42	0.24
Madhya Pradesh	44.38	52,117	852	662	48.44	48.55	-0.11	48.03	47.80	0.23
Rajasthan	43.29	39,753	1,089	853	47.76	47.88	-0.12	48.31	47.90	0.41
Tamil Nadu	34.92	15,400	2,268	1,764	48.26	48.61	-0.35	50.01	49.65	0.36
Karnataka	34.89	27,481	1,270	1,056	48.69	49.06	-0.38	49.53	49.40	0.14
Gujrat	31.74	18,066	1,757	1,445	47.53	48.37	-0.83	48.79	48.81	-0.02
Orissa	31.29	47,529	658	535	48.86	49.22	-0.36	49.81	49.82	-0.01
Kerala	23.57	1,364	17,283	15,642	49.01	48.94	0.07	51.76	51.22	0.54
Assam	23.22	25,124	924	777	49.16	49.42	-0.26	48.44	48.03	0.41
Jharkhand	20.95	29,354	714	522	49.31	49.64	-0.33	48.96	48.53	0.43
Chhattisgarh	16.65	19,744	843	721	49.54	49.71	-0.16	50.22	50.15	0.07
Punjab	16.10	12,278	1,311	1,096	44.42	46.77	-2.35	47.51	47.15	0.36
Haryana	15.03	6,765	2,222	1,765	45.12	46.73	-1.61	46.66	46.28	0.38
Jammu & Kashmir	7.63	6,417	1,189		48.90			47.65		
Uttaranchal	6.31	15,761	400	320	47.85	48.79	-0.94	50.64	49.83	0.81
Himachal Pradesh	5.48	17,495	313	268	47.37	48.70	-1.33	50.09	50.00	0.09

Notes: Calculations are from the 1991 and 2001 Village Primary Census Abstract and village directories. The table does not show smaller states but these are included in the all India calculations. The states are sorted by village population size.

Table 2: Village variance in fraction girls age 0-6

	Total Village Variance % female $\leq$6			Percent total variance from district level variation			Percent total variance from variation in preferences			Percent female $\leq$6 in villages $\geq$ 50% girls		
State	2001	1991	Change	2001	1991	Change	2001	1991	Change	2001	1991	Change
India	66.3	69.2	-3.0	2.1	1.1	1.0	5.1	10.0	-4.9	30.5	36.2	-5.7
Uttar Pradesh	45.2	51.1	-5.9	1.3	0.6	0.6	6.9	15.2	-8.3	25.6	30.7	-5.1
Bihar	37.1	45.2	-8.1	0.6	0.5	0.1	9.7	23.5	-13.8	27.1	35.3	-8.2
West Bengal	51.6	53.5	-1.9	0.0	0.2	-0.1	1.0	8.3	-7.4	37.5	41.5	-4.0
Maharastra	42.4	44.9	-2.5	1.4	0.3	1.1	5.8	9.6	-3.8	26.8	37.1	-10.3
Andhra Pardesh	48.8	50.6	-1.7	0.6	0.3	0.3	4.7	11.1	-6.4	35.6	43.6	-8.0
Madhya Pradesh	55.9	60.6	-4.7	1.4	1.0	0.4	2.7	7.4	-4.7	35.9	39.1	-3.1
Rajasthan	56.4	67.2	-10.8	1.6	1.4	0.3	3.0	7.4	-4.5	25.3	28.0	-2.7
Tamil Nadu	31.2	34.3	-3.1	3.6	3.1	0.5	6.7	17.0	-10.3	31.1	36.9	-5.8
Karnataka	75.4	68.1	7.3	0.1	0.0	0.0	2.2	5.0	-2.9	35.0	39.7	-4.7
Gujrat	31.2	33.9	-2.7	2.7	2.0	0.7	13.2	13.8	-0.6	23.1	32.1	-9.0
Orissa	99.3	107.4	-8.1	0.4	0.5	-0.1	3.0	5.9	-2.9	43.0	47.2	-4.2
Kerala	3.0	1.9	1.1	0.8	2.9	-2.1	21.6	-1.0	22.6	17.1	15.7	1.3
Assam	72.5	66.4	6.1	0.1	0.0	0.0	7.6	7.5	0.1	43.3	47.3	-3.9
Jharkhand	76.9	88.7	-11.8	0.1	0.0	0.1	5.1	15.4	-10.3	45.4	51.0	-5.7
Chhattisgarh	48.9	50.6	-1.6	0.2	0.1	0.1	6.3	11.8	-5.5	48.2	51.6	-3.4
Punjab	51.9	57.6	-5.8	0.8	0.1	0.6	-2.4	6.5	-8.9	7.8	19.1	-11.3
Haryana	29.2	28.0	1.2	2.9	0.3	2.6	7.0	3.2	3.9	4.3	12.2	-7.9
Jammu & Kashmir	67.5			5.7			21.8			42.9		
Uttaranchal	182.3	165.0	17.3	0.2	0.2	0.0	-1.2	4.1	-5.4	34.4	44.3	-9.8
Himachal Pradesh	218.9	199.3	19.5	0.8	0.1	0.7	-0.8	4.6	-5.4	39.8	46.8	-7.0

Notes: Calculations are from the 1991 and 2001 Village Primary Census Abstract and village directories. The table does not show smaller states but these are included in the all India calculations. The states are sorted by village population size.

Table 3: Marriage migration and the distribution of women

State	Female≤6 need to migrate to equalize (%)			Female migration	Autarchy Variance % female >6		Village Variance % female > 6	
	2001	1991	Change	(%)	2001	1991	2001	1991
India	2.65	2.87	-0.22	75.0	19.1	21.8	17.8	22.6
Uttar Pradesh	2.55	2.98	-0.43	95.1	20.3	21.6	16.8	21.1
Bihar	2.04	2.41	-0.37	69.2	11.3	14.5	12.0	19.8
West Bengal	2.33	2.48	-0.15	80.2	9.4	11.5	10.6	15.7
Maharastra	2.86	2.83	0.03	87.4	15.7	19.8	12.8	17.5
Andhra Pardesh	2.18	2.25	-0.07	67.7	8.9	8.4	10.0	14.3
Madhya Pradesh	3.16	3.12	0.03	93.4	16.6	18.1	17.2	21.9
Rajasthan	2.75	3.02	-0.27	95.5	16.7	18.2	15.5	21.1
Tamil Nadu	2.65	2.83	-0.18	49.9	5.1	5.2	5.9	10.6
Karnataka	2.74	2.71	0.03	70.6	12.3	13.7	15.9	17.4
Gujrat	2.70	2.74	-0.04	92.6	8.7	8.8	9.9	11.0
Orissa	3.77	3.88	-0.11	83.2	18.5	19.8	22.4	28.2
Kerala	0.88	0.88	0.00	62.1	1.9	2.2	0.9	0.2
Assam	2.90	3.05	-0.15	35.6	15.5	20.9	18.4	19.6
Jharkhand	3.13	4.40	-1.27	56.7	16.7	19.1	21.4	32.1
Chhattisgarh	3.16	4.20	-1.04	89.0	7.9	8.1	11.9	15.9
Punjab	3.19	3.05	0.14	91.4	12.0	14.0	8.5	15.3
Haryana	2.35	2.25	0.10	97.1	11.2	12.6	8.8	9.4
Jammu & Kashmir	3.73			57.8	19.1		23.5	
Uttaranchal	4.31	5.35	-1.04	91.8	68.1	64.9	38.3	43.0
Himachal Pradesh	5.78	5.31	0.47	88.9	46.0	53.4	39.4	53.2

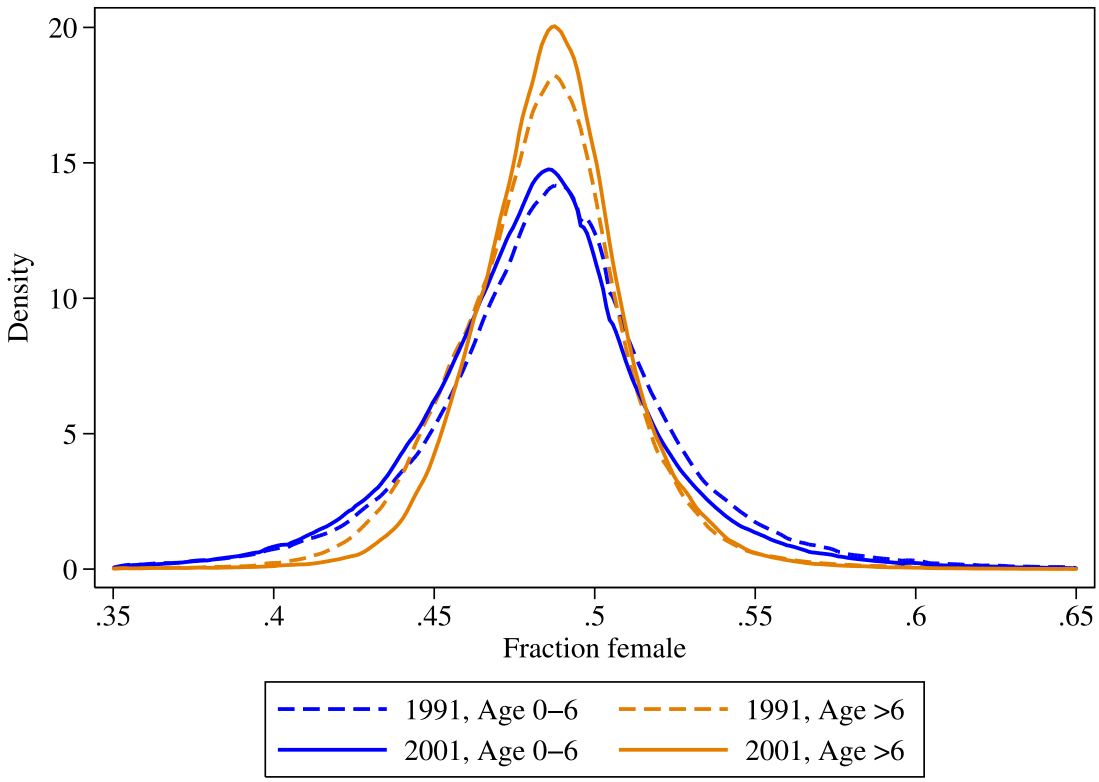
Calculations are from the 1991 and 2001 Village Primary Census Abstract and village directories. The table does not show smaller states but these are included in the all India calculations. Female migration is the percent of women 22 and older in rural areas who have migrated calculated using the 64th round of the National Sample Survey. The predicted variance is what the variance would be if there were no migration. The states are sorted by village population size.

Table 4: Village panel estimation

Dependent variable: Change in percent children 6 and under female			
Village change in:			
Primary school?	-0.0947***	-0.0956***	-0.0730***
	(0.0274)	(0.0274)	(0.0270)
Middle school?	0.0323*	0.0328*	0.0311
	(0.0192)	(0.0192)	(0.0190)
Secondary school?	-0.0314	-0.0311	-0.0376
	(0.0233)	(0.0233)	(0.0229)
Access by footpath?	-0.00872	-0.00790	0.0359***
	(0.0177)	(0.0177)	(0.0131)
Access by paved road?	-0.00116	-0.000339	0.0316*
	(0.0180)	(0.0180)	(0.0164)
Access by dirt road?	0.0199	0.0194	0.0295**
	(0.0171)	(0.0171)	(0.0150)
Bus, rail, taxi	0.00286	0.00276	-0.00803
stop?	(0.0177)	(0.0177)	(0.0172)
Phone connection?	0.000261	0.000412	-0.0506***
	(0.0163)	(0.0163)	(0.0153)
Post office?	-0.0225	-0.0225	-0.0120
	(0.0227)	(0.0227)	(0.0225)
Drinking water	-0.0320	-0.0277	-0.0182
source?	(0.136)	(0.136)	(0.135)
Power supply?	-0.0116	-0.0101	0.0199
	(0.0209)	(0.0209)	(0.0200)
Medical facility?	-0.0153	-0.0147	-0.0283**
	(0.0152)	(0.0152)	(0.0132)
Percent scheduled	0.00833***	0.00829***	0.00789***
caste	(0.00141)	(0.00141)	(0.00140)
Percent scheduled	0.0441***	0.0441***	0.0417***
caste	(0.00154)	(0.00154)	(0.00149)
log total population	0.195***	0.192***	0.181***
	(0.0399)	(0.0399)	(0.0379)
Tehsil percent children		0.172***	0.634***
female		(0.0145)	(0.00978)
Observations	534,204	534,190	534,190
R-squared	0.015	0.015	0.011
District effects	Yes	Yes	No
Interactions	Yes	Yes	Yes

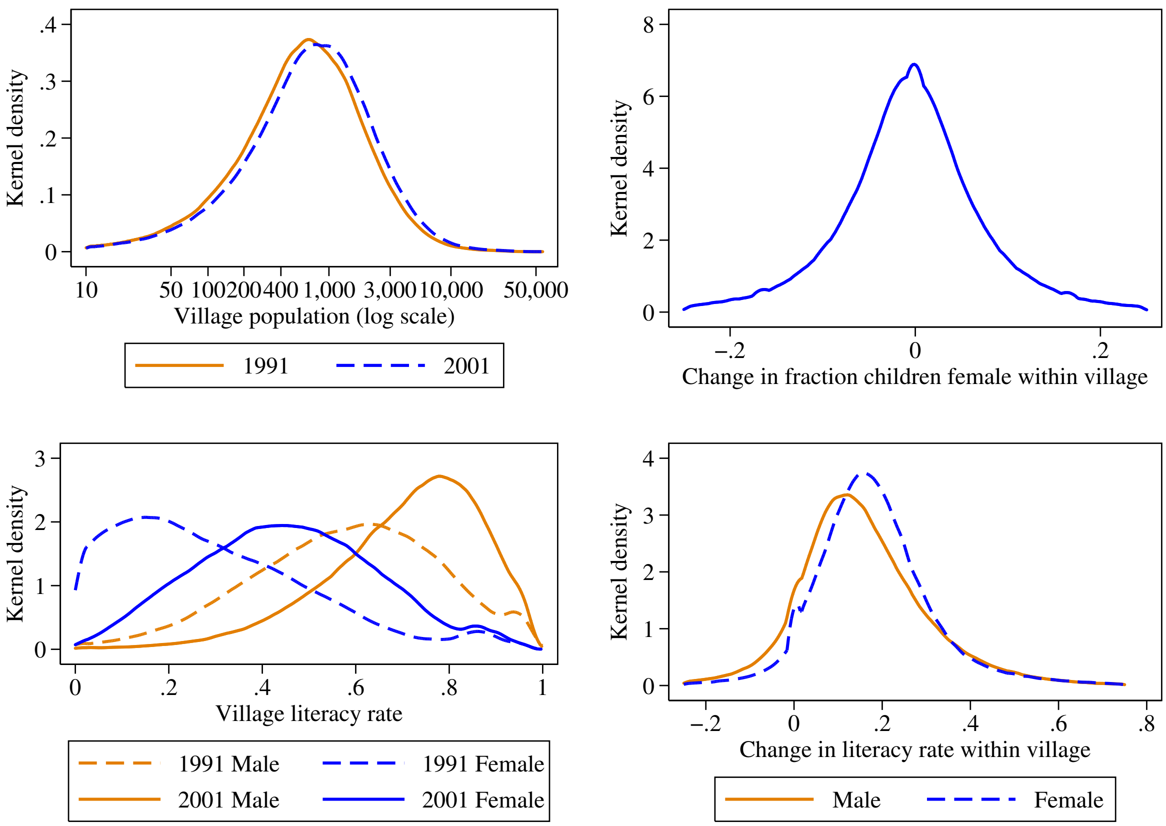
Notes: All columns include the interactions listed in footnote 6 including the change in male and female literacy, their squares, and all their interactions as well as all of their interactions with log population in 1991. These effects are plotted in figure 3 and 4. Village infrastructure enters as 1 if it exists in the village, 0 if not. Calculations are from the 1991 and 2001 Village Primary Census Abstract and village directories. The tehsil mean excludes the individual village.

Figure 1: The distribution of women and girls across village India 1991-2001



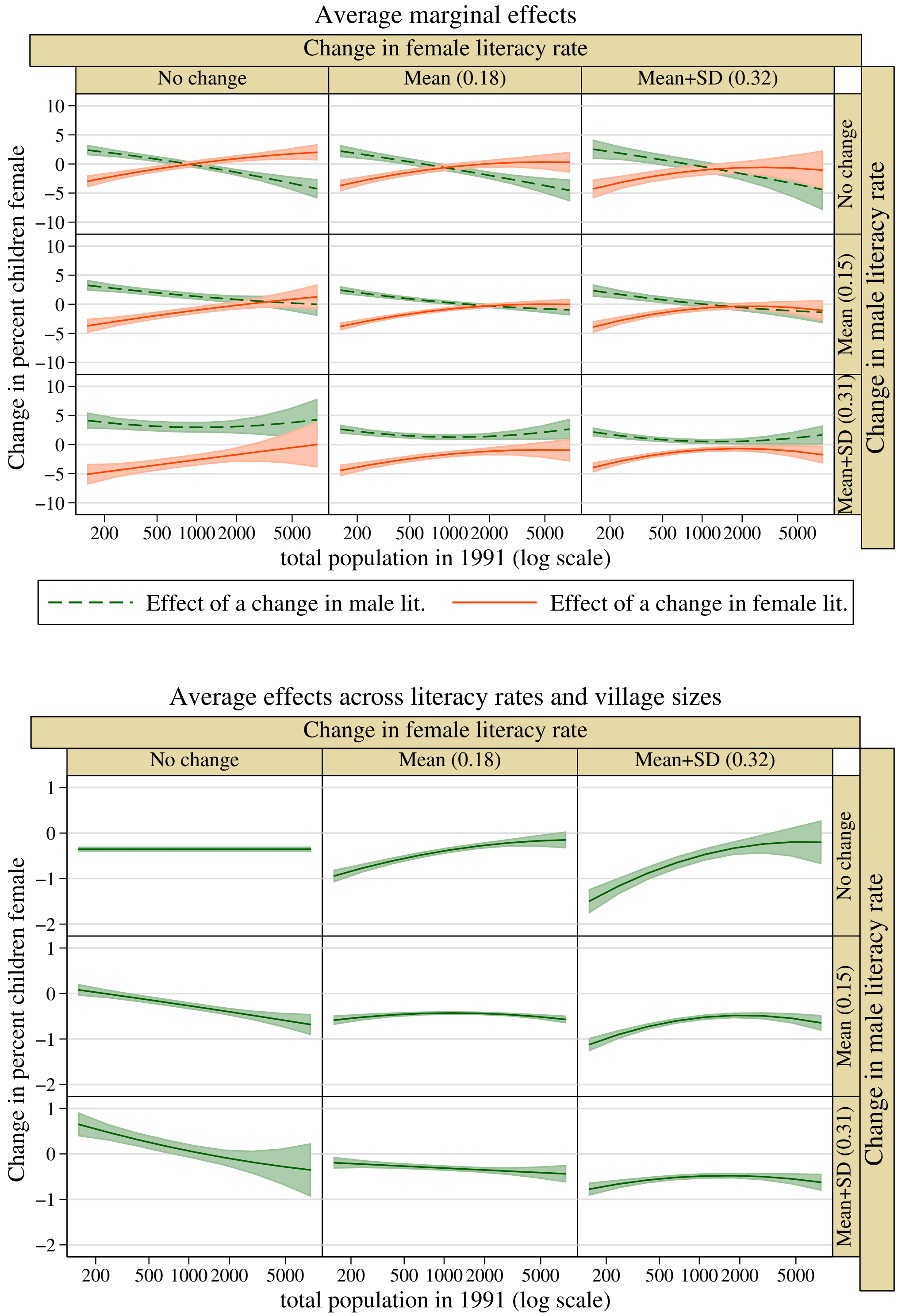
Notes: Uses the village census from 1991 and 2001. The kernel density is weighted by total village population.

Figure 2: Village density of population, change in fraction female, and change in literacy



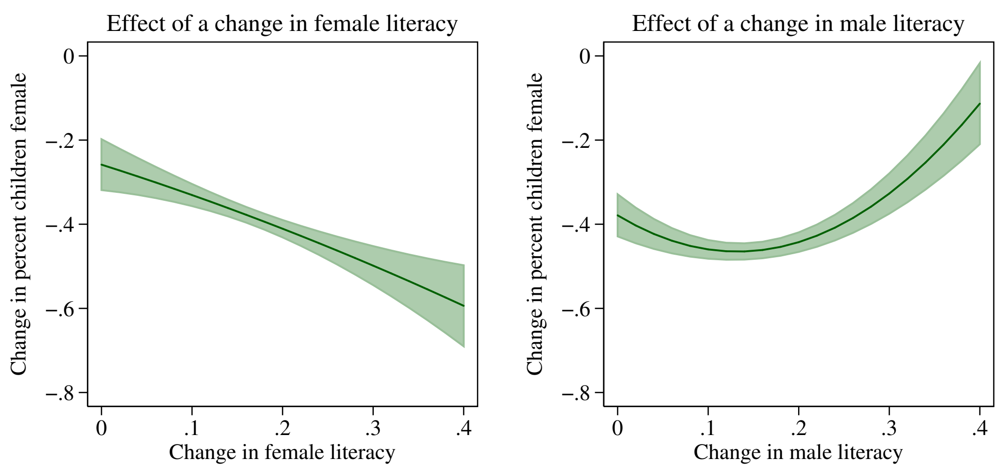
Notes: Uses the village census from 1991 and 2001. The densities are across villages and are not weighted by population size.

Figure 3: Effects of changes in literacy on percent children female across village sizes



Notes: The top plot shows the marginal effect ($E[dy/dx]$) of an additional increase in female or male literacy evaluated at different population sizes and for different changes in female and male literacy. The bottom plot shows the predicted change in the fraction children female across different changes in literacy and population. The interaction effects are estimated but not reported separately in table 4 and the full list of regressors is in footnote 6.

Figure 4: Average changes in percent children female across literacy rates



Notes: Shows the predicted change in the fraction of children (0-6) female for different changes in male and female literacy rates holding all other covariates fixed at their means. The interaction effects are estimated but not reported separately in table 4 and the full list of regressors is in footnote 6.