The changing geography of gender in India

Scott L. Fulford*

September 2013

Abstract

This paper examines the changing distribution of where women and girls live in India at the smallest scale possible: India's nearly 600,000 villages. The village level variation in the proportion female is far larger than the variation across districts. Decomposing the variance, I show that village India is becoming more homogeneous in its preferences for boys even as that preference becomes more pronounced. A consequence is that 70% of girls grow up in villages where they are the distinct minority. Most Indian women move on marriage, yet marriage migration has almost no gender equalizing influence. Further, by linking all villages across censuses, I show that most changes in village infrastructure are not related to changes in child gender. Gaining primary schools and increases in female literacy decrease the proportion of girls. The results suggests that there are no easy policy solutions for addressing the increasing masculinization of Indian society.

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

*Boston College Department of Economics, 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 at the 2013 International Conference on Migration and Development.

Child sex ratios have been worsening all 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). Sex ratios vary substantially by region in India; in some states the the ratio is over 120 boys to girls. Overall sex ratios in India have not increased as fast as sex ratios at birth since there has been a contemporaneous decrease in female mortality but are around 108 men for every 100 women as well. 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. That means that marriage migration in India is the largest migration in the world: around 300 million women have migrated and an additional 20 million are migrating each year.[1]

These two phenomena of gender imbalance for children and gendered migration interact, creating complex spatial and gender dynamics. Gender imbalances among children are largely determined by the pre-and post-natal decisions of parents (Arokiasamy, 2007), who live in a particular village and have preferences for sons and access to sex selection technology and medical care that are determined locally. Yet with marriage migration so pervasive the broader social consequences of parents' decisions are determined by where women eventually move on marriage. With diverse sex ratios across India, how marriage migration changes will determine the distribution of the severity of the coming "marriage squeeze" in India (Guilmoto, 2012; Sautmann, 2011). The effects of the large sexual imbalances in India will differ not just across age groups but by geographic

[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 (2013a) provides a full description of marriage migration.

area as well. Moreover, marriage migration may enhance gender biases and gender imbalances in the next generation; men in rural areas hardly ever move, while the large majority of women have moved to join their husband's family, potentially reinforcing the value of sons and increasing the incentives to spend scarce resources on sons' health and education over daughters' (Jayachandran and Kuziemko, 2011; Oster, 2009a,b).

This paper examines the 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. I present four interlocking results on where women and girls are located in India and how those distributions have changed over time. First, although most of the focus on sex ratios in India is on variation in the state or district level (for example, Dyson and Moore (1983); Echávarri and Ezcurra (2010); Guilmoto (2009)) 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—I show that much of that 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. The variation in the fraction of girls from village to village and decade to decade within each village is so large compared to the overall downward trend that it would be difficult for parents to observe directly.

The second, more important, result is that village India is becoming more homogeneous in its preference for boys. The increasing homogeneity has accompanied the steady increase in preferences for boys across India. A consequence of these two trends is that gender equality is much less common; while 36% of girls lived in a village with at least as many girls as boys in 1991, in 2001 only 30% lived in such a village. The overall masculinization of India is becoming increasingly universal. Most girls in India now grow up in a village where they are the distinct minority. While there used to be villages that produced more women, 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.

One change in Indian society that might have resulted from the increasingly imbalanced gender ratio is that some areas "specialize" by producing more women. This specialization is the geographic implication of the model introduce by Edlund (1999). In that model hypergamy prompts the poorer or lower caste families to produce more girls. Yet far from there being evidence of geographic specialization, there is instead increased homogenization.

Third, I show that the geographic distribution of women in village India is not determined by the village of birth but instead through marriage migration. Despite the large variation in gender ratios at the village level, marriage migration is 30 times larger than necessary to completely equalize sex ratios across all of India. Since most women do not stay in the village of birth to marry, the decisions of parents and their implications for the available spouses create externalities that extend beyond the village through the marriage market. Comparing what the spatial distribution of women would look like if there were no marriage migration with the actual variation, it seems that despite how widespread marriage migration is, it may slightly worsen rather than improve the variation in gender across villages. It is impossible to understand the wider social consequences of the increasing male to female sex ratios without understanding the wider marriage market and marriage migration.

Finally, this paper examines the determinants of child gender imbalances at the village level by linking the villages across censuses. Linking villages 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 gender ratios. Except for gaining a primary school, changes in village infrastructure such as access to a road, medical facilities, drinking water or power sources are not linked to changes in the child gender. While a recent literature which has emphasized excess female mortality as the cause of most of the missing omen (Anderson and Ray, 2010; Oster, 2009a,b) it is not clear that better service provision will change much. Instead, changes in male and female literacy are very important but the effects vary by population size. For small villages increases in female literacy *decrease* the fraction of children female, while increases in male literacy increase it. The effects of male and female literacy move in opposite directions with village size and so tend to offset each other somewhat. The negative effects predominate so that overall increases in male and female literacy together tend to push down the fraction female.

That village infrastructure has almost no effect and improvements in female literacy overall are associated with declines in the fraction of girls suggest that there may be no easy policy solutions to halt or reverse the decline in the proportion of girls and women in India. Instead, these results suggest that the increased masculinization of India will continue and be accompanied by a strong tendency towards homogenization as female literacy spreads.

This paper contributes to a broader trend in understanding spatial demography both in developed countries (Wachter, 2005) and in developing countries (Guilmoto and Rajan, 2005). Understanding the decisions made by parents and how those decisions affect the surrounding areas requires us to understand the very local context of rural India. This paper adds to the literature examining gender imbalances in India that is largely limited to cross-sections or panels of districts (Echávarri and Ezcurra, 2010; Murthi, Guio, and Drèze, 1995) by linking the universe of villages over time. That means it is possible to examine the truly local determinants of gender imbalances. I find that both district trends and changes in surrounding villages are very influential for changes within the village. Such effects are difficult to discern at a district level which also masks the increased homogenization within districts.

1 The overall trends in gender in village India

Approximately two thirds of all Indians still reside in rural areas. Within rural areas, the villages is the smallest administrative unit. Villages do 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 approximately 600,000 villages in India with an average of approximately 1,250 residents. Village population varies substantially both within states and between them. Migration in surveys is generally 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 gender 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.

Several trends are evident in table 1. The growth in population in villages is substantial. While the population of urban areas has grown somewhat faster than rural, urban migration has not been substantial and the country remains largely rural. 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 gender ratio since the analysis of the variance in later sections is substantially more straightforward statistically and less subject to outliers in fractions rather than ratios. Children are defined by the census records as age 0-6, and since that is the only age range reported by the census at the village level, adults are everyone seven and older. Between 1991 and 2001 there was an increase in the fraction of the adult population women. That increase appears to be because of more rapid decreases in female mortality (Guilmoto and Depledge, 2008).

The overall gender 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 side, 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, 2013b). 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 (2013b) 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 further.

2 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, approximately, 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 tighter. 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.

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 fewer 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 ability or 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, but 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 gender 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 but 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 how changes in population affect the variance very clear. The disadvantage is that it does not constrain the preference variance to be positive and 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 4 when I create a panel of villages across censuses to examine changes in preferences. By examining a panel that removes village level preferences that are fixed it is possible to allow for changes in p_i that are impossible to identify in a cross-section.

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 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 gender imbalances has been focused on explaining district or regional level differences (Echávarri and Ezcurra, 2010; Guilmoto and Rajan, 2001; 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 increasing 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. That comparison includes the differences across states, but that contributes a relatively small portion of overall variance.

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 and small number of villages.

Most of the variance is explained by random differences. Across India 95% of the village level variance comes from random variation. Local preference variation plays only a small part and the role it plays is diminishing. That does not mean that preferences for boys are diminishing, but instead that the variation around those preferences is diminishing so that there are fewer villages with preferences different from the average.

Increased homogeneity has the important implication that there is unlikely to be a geographic effect to help stabilize gender 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 the under six population. 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 nearly 6 percentage points or 17%. 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 number 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 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 context 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 be grow up as a noticeable minority. Kerala also has a noticeably lower percentage of girls in equal gender 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 the level of gender imbalance. The most imbalanced states Punjab and Haryana do not have higher village level variation.

3 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.

Across rural India, about three quarters of all women over 22 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 (2013a) provides a more detailed geographic breakdown.

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 given 0-6 cohort would need to move eventually to equalize the distribution whatever the source of the geographical variation. That also helps to understand the role played by marriage migration.

For each village i define f_i as the fraction of children under seven who are female and define f_I as the mean across the entire rural population (not the mean over villages). Then for each village that has more girls than average ($f_i \geq f_I$) and has n_i^C total children a total $(f_i - f_I)n_i^C$ must leave to equalize the distribution ignoring the integer constraint. The limit to only villages with more than the average is to avoid double counting: girls are leaving from some villages but must move to the lower than average female villages. Then the fraction of girls under six who would need to leave to equalize across the state or across India is the sum from each village divided by the total population of girls under six. The calculation is essentially integrating under the the curve in figure 1 starting from the mean and weighting by the size of the village population under six. The same calculation can be done for each state replacing the India mean with the state mean.

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—migration does not create more 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 gender ratios, women would still have to move in similar proportion in most states. The exceptions are Kerala where there is little village level variation, and the states Himchal Pradesh and Uttaranchal (now Uttarakhand) where there is much higher village variation.

The actual movement of women is far larger than necessary to equalize the geographic distribution of girls. Across India 75% of women over 21 have migrated, almost all on marriage, yet only 2.7% would need to migrate to equalize the geographic distribution. The distribution of women across village India is not directly related to the distribution of girls but instead to the villages that women are married into. 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.

It is also 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 2. 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. 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 survival of girls is the same as the variation in the survival of women. While there may be excess female mortality that changes with age (Anderson and Ray, 2010), it seems reasonable to suspect that the geographic variation of the mortality is similar to the geographic variation in excess child mortality. If the village census had more information it would be possible to allow $\bar{p}$ to change with age. Doing so would have little effect on the calculations since $\bar{p}$ is close to 1/2 and so changes in it have only a second order effect on the variance.

The last four columns of table 3 shows the variance that would occur if there were no marriage migration of women and the actual village variance of the percent women. Although the most women in India move, the actual variance across villages is very similar to what would obtain if no women moved. Even though there is a vast movement of women every year and most women live someplace other than where they were born, the actual gender imbalances are not much different than if no women moved. One reason for this result may be that most women do not move far and village preferences are strongly shaped by the preferences of the villages close by. Changes in overall fertility suggest a similarly hyper-local geographic relationship in South India Guilmoto and Rajan (2005). I provide some evidence for this hypothesis in the next section.

4 The micro determinants of child genderimbalance

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 employ a difference-in-difference approach that removes village level unobservables and district trends and so has a potentially causal interpretation.

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 attrition comes because in each census the village directory and the primary census abstract do not match exactly and because the two censuses have some non-matching villages.

From the primary census abstract there are approximately 600 thousand villages with non-zero population in 2001, and approximately 580 thousand villages with non-zero population in 1991. I can match 567 thousand inhabited villages with information across both census years. Of these, 533,500 have village directory information for both years for all the relevant variables.[2]

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 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.

By linking the same villages across 1991 and 2001 is is also possible to look at the distribution of changes within villages (the top right panel). The overall downward shift in the fraction girls is nearly impossible to discern with the the overall variation. 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 village there are more girls than there used to be. While the aggregate trends are clear, from the village level the world looks very different.

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

[2]Census data for Jammu and Kashmir is generally missing for 1991, and the Meghalya data are 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. 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 dates of these sources are somewhat approximate since the data comes on compact discs which do not include much bibliographic information. Different groups within the Census of India were responsible for different parts of the data and for different states and so while the underlying data is collected in the same way across states and districts, 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.

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 the variance decomposition in section 2. Suppose the preferences (or ability to act on those preferences) for the fraction of children that are female of a given village can be written 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 including its location, its immutable preferences, and its wealth or poverty; p_{dt} is the district level fraction of children female; and X_{it} are village level characteristics such as transportation, health care centers, and literacy rates that may change.

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 size of the village. I deal with this problem directly by calculating the variance for each village 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 should get higher weight. Weighting is very important in this context the estimated effects vary with village size.

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:

$$f_{it} - f_{it-1} = (p_{dt} - p_{dt-1}) + (X_{it} + X_{it-1})\beta + (\nu_{it} - \nu_{it-1}) - (\epsilon_{it} - \epsilon_{it-1}). \tag{4}$$

Within the change in village characteristics $\Delta X_{it} = (X_{it} + X_{it-1})$, I include the changes in village infrastructure such as roads, medical clinics, schools, and power supplies, as well as changes in population size. The full list is in table 4. In addition, it seems that the changes in the fraction of children female are closely related to literacy changes, but the effects change with population size and may be non-linear. The model introduced by Echávarri and Ezcurra (2010) suggests 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.[3] Allowing for non-linear interactions 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 does not report the coefficients on the interactions since they are difficult to interpret in isolation. Instead, the top panel of figure 3 shows the marginal effect of an additional increase in female and male literacy evaluated at different population sizes and different changes in literacy. To understand how important these changes in literacy and population size are, the bottom panel of figure 3 shows the predicted change in the fraction female when evaluated at the same changes in literacy and population sizes.

As is clear from table 4, changes in village infrastructure are not in general strongly related to changes in the fraction female. Gaining a primary school reduces the fraction female by about half of the mean change—a large effect closely related to increases in literacy—but otherwise changes are small in size and not statistically significant. The district trends are not shown but are overall negative. Against this overall negative trend, higher growth in village population tends to increase the fraction female. Note that even with the literacy rate interactions and allowing for

[3] 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}, $\Delta ML_{it} * \Delta FL_{it}$, $(\Delta ML_{it})^2 * \Delta FL_{it}$, $(\Delta FL_{it})^2$, $\Delta ML_{it} * (\Delta FL_{it})^2$, $(\Delta ML_{it})^2 * (\Delta FL_{it})^2$, $VP_{i1990} * \Delta ML_{it}$, $VP_{i1990} * (\Delta ML_{it})^2$, $VP_{i1990} * \Delta FL_{it}$, $VP_{i1990} * \Delta ML_{it} * \Delta FL_{it}$, $VP_{i1990} * (\Delta ML_{it})^2 * \Delta FL_{it}$, $VP_{i1990} * (\Delta FL_{it})^2$, $VP_{i1990} * \Delta ML_{it} * (\Delta FL_{it})^2$, $VP_{i1990} * (\Delta ML_{it})^2 * (\Delta FL_{it})^2$, $(VP_{i1990})^2 * \Delta ML_{it}$, $(VP_{i1990})^2 * (\Delta ML_{it})^2$, $(VP_{i1990})^2 * \Delta FL_{it}$, $(VP_{i1990})^2 * \Delta ML_{it} * \Delta FL_{it}$, $(VP_{i1990})^2 * (\Delta ML_{it})^2 * \Delta FL_{it}$, $(VP_{i1990})^2 * (\Delta FL_{it})^2$, $(VP_{i1990})^2 * \Delta ML_{it} * (\Delta FL_{it})^2$, $(VP_{i1990})^2 * (\Delta ML_{it})^2 * (\Delta FL_{it})^2$. I do not include the main effect of log population since with fixed effects and population changes size is already accounted 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.

Instead of changes in village infrastructure, it is changes in literacy rates that are more important. 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 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. Since the interactions allow the effects to vary with population size I limit the sample to villages that are larger that 150 and smaller that 8,100 to keep the results being driven by the very large or very small villages; as is clear from figure 2 that includes approximately 90% of all villages.

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 has either no effect or slightly increases it. The relationship changes with the size of the increase in literacy and with whether men or women are increasing more. Large increases in male literacy make the population size less important. So the bottom row shows that for all village sizes and all changes in female literacy, in villages with large increases in male literacy the marginal effect of additional increases in male literacy are positive and significant and the marginal effect of additional increases in female literacy are negative and significant.

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 evaluated at the mean effect for all of the rest of the variables. 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 left column predicts the effects when there are no changes in female literacy, the right when there is an increase of 0.32 (the mean plus one standard deviation).

The effects of increasing in both male and female literacy rates tend to offset each other across village sizes. Because male and female literacy rates have marginal effects that are changing in the opposite direction, 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 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.

Along the other diagonal the predicted effects are very different. Small villages with no increases in female literacy but large increases in male literacy shown in the bottom left plot have large increases in the fraction female tapering to zero for larger villages. In the top right plot, on the other hand, in small villages large increases in female literacy that are not accompanied by increases in male literacy are associated with large declines in the fraction of children female, but the negative effect goes to zero in larger villages.

These effects are not small. As calculated in table 1 the mean fall in the fraction female is -0.0039. Increasing literacy rates along the diagonal from top left to bottom right drive down the fraction substantially to almost twice the mean. The variation is much larger for villages that had skewed increases in male or female literacy. For villages with the mean increase in female literacy but no increase in male literacy (the top middle plot), for example, the smallest villages had falls in the fraction around four times the mean, while the largest had no fall in the fraction female.

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 exclude 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: about 38% of the change in a tehsil is reflected within changes in the village over and above the district trend. Including the district trend is actually understating the importance of local changes. When I exclude the district trends individual villages have changes that are 139% of the mean tehsil change. Villages seem to act very strongly together.

5 Discussion and Conclusion

The falling female to male ratio in India and other countries in Asia has prompted some 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. Part of the problem is that skewed sex ratios are often the worst in elite groups (Patel et al., 2013). 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 female to male ratio among children. Very little else about village infrastructure matters. That suggests that it will be difficult to halt the trend by providing more services or promoting literacy and that some development efforts may make the situation worse. Instead rural India has become increasingly homogeneous in its preferences for boys with villages strongly positively correlated with villages nearby. There is no geographic specialization encouraging some areas to have more girls. Since the pervasive marriage migration means that villages and 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 (Sautmann, 2011). 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 and Debraj Ray. 2010. "Missing Women: Age and Disease." *The Review of Economic Studies* 77 (4):1262–1300.

Anukriti. 2013. "The Fertility Sex-Ratio Trade-off: Unitended Consequeces of Financial Incentives." Tech. rep., Boston College. BC WP 827 availalbe 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. Available: http://www.cicred.org/Eng/Publications/pdf/BOOK_singapore.pdf, accessed 30 July 2013.

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

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, RebecaA. 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. 2013a. "Marriage Migration in India." Boston College Working Paper 820, Boston College. Available: http://fmwww.bc.edu/EC-P/wp820.pdf, accessed 30 July 2013.

———. 2013b. "Returns to education in India." Working Paper 819, Boston College. URL http://fmwww.bc.edu/EC-P/wp819.pdf. http://fmwww.bc.edu/EC-P/wp819.pdf.

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

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.

Guilmoto, Christophe Z. and S. Irudaya Rajan. 2001. "Spatial Patterns of Fertility Transition in Indian Districts." *Population and Development Review* 27 (4):713–738.

Guilmoto, Christophe Z. and S. Irudaya Rajan, editors. 2005. *Fertility Transition in South India*. New Delhi: Sage Publications.

Guilmoto, ChristopheZ. 2012. "Skewed Sex Ratios at Birth and Future Marriage Squeeze in China and India, 20052100." *Demography* 49 (1):77–100.

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.

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.

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, ArchanaB., 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.

Wachter, Kenneth W. 2005. "Spatial demography." *Proceedings of the National Academy of Sciences of the United States of America* 102 (43):15299–15300.

Table 1: Village census statistics

State	Vill. Pop. (millions)	Villages	Mean Village Size		Female≤6 (%)			Female>6 (%)		
	2001	2001	2001	1991	2001	1991	Change	2001	1991	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 women

	Total Village Variance % female ≤6			Percent total variance from district level variation			Percent total variance from variation in preferences			Percent female ≤6 in villages ≥ 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

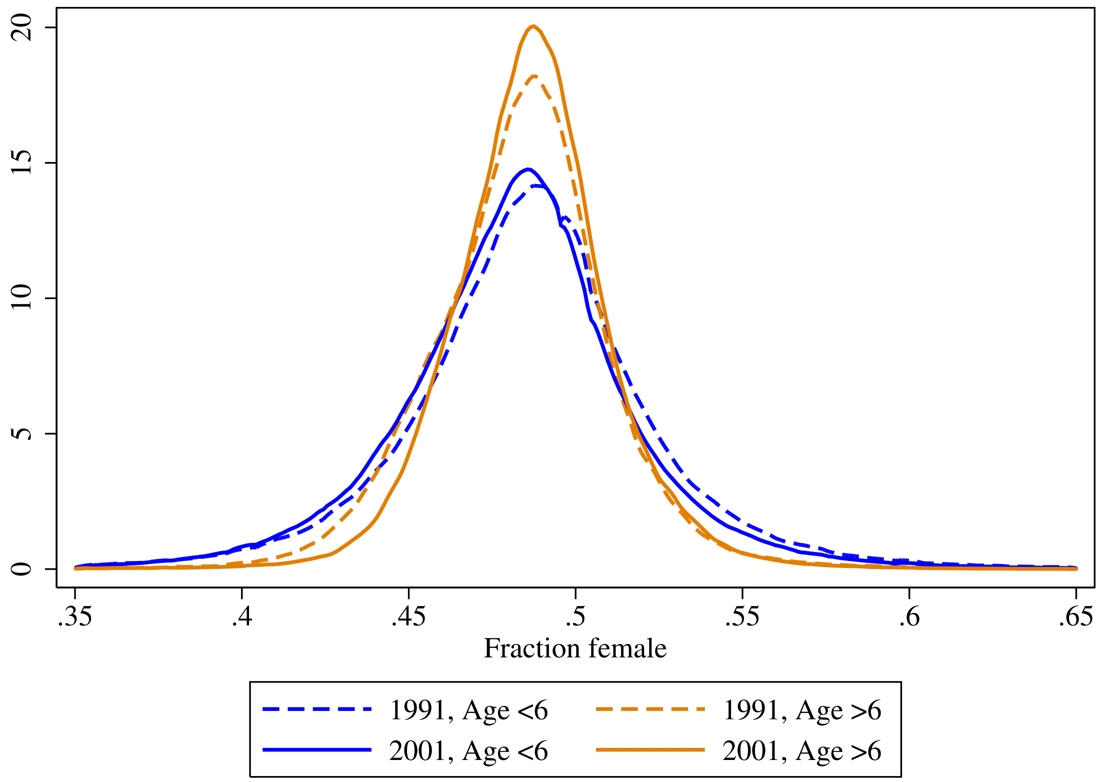
Calculations are from the 1991 and 2001 Village Primary Census Abstract and village directories. The table does not show smaller states but these are included in the all India calculations. Female migration is the percent of women 22 and older in rural areas who have migrated calculated using the 64th 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 fraction children 6 and under female			
Village change in:			
Primary school?	-0.00199***	-0.00200***	-0.00166***
	(0.000586)	(0.000586)	(0.000579)
Middle school?	0.000640	0.000650	0.000644
	(0.000411)	(0.000411)	(0.000406)
Secondary school?	-0.000650	-0.000643	-0.000730
	(0.000499)	(0.000499)	(0.000490)
Access by footpath?	-0.000111	-9.33e-05	0.000849***
	(0.000379)	(0.000379)	(0.000280)
Access by paved road?	-5.34e-05	-3.52e-05	0.000736**
	(0.000386)	(0.000386)	(0.000351)
Access by dirt road?	0.000421	0.000409	0.000687**
	(0.000367)	(0.000366)	(0.000320)
Bus, rail, taxi stop?	0.000148	0.000146	-8.53e-05
	(0.000379)	(0.000379)	(0.000369)
Phone connection?	-1.55e-05	-1.20e-05	-0.00112***
	(0.000349)	(0.000349)	(0.000327)
Post office?	-0.000501	-0.000502	-0.000359
	(0.000486)	(0.000486)	(0.000481)
Drinking water source?	-0.000717	-0.000620	-0.000469
	(0.00290)	(0.00290)	(0.00289)
Power supply?	-0.000362	-0.000328	0.000221
	(0.000446)	(0.000446)	(0.000428)
Medical facility?	-0.000466	-0.000454	-0.000730***
	(0.000326)	(0.000326)	(0.000283)
log total population	0.00812***	0.00806***	0.00774***
	(0.000853)	(0.000853)	(0.000810)
Tehsil fraction children female		0.375***	1.385***
		(0.0311)	(0.0209)
Observations	533,501	533,488	533,488
R-squared	0.014	0.014	0.010
District effects	Yes	Yes	No
Interactions	Yes	Yes	Yes

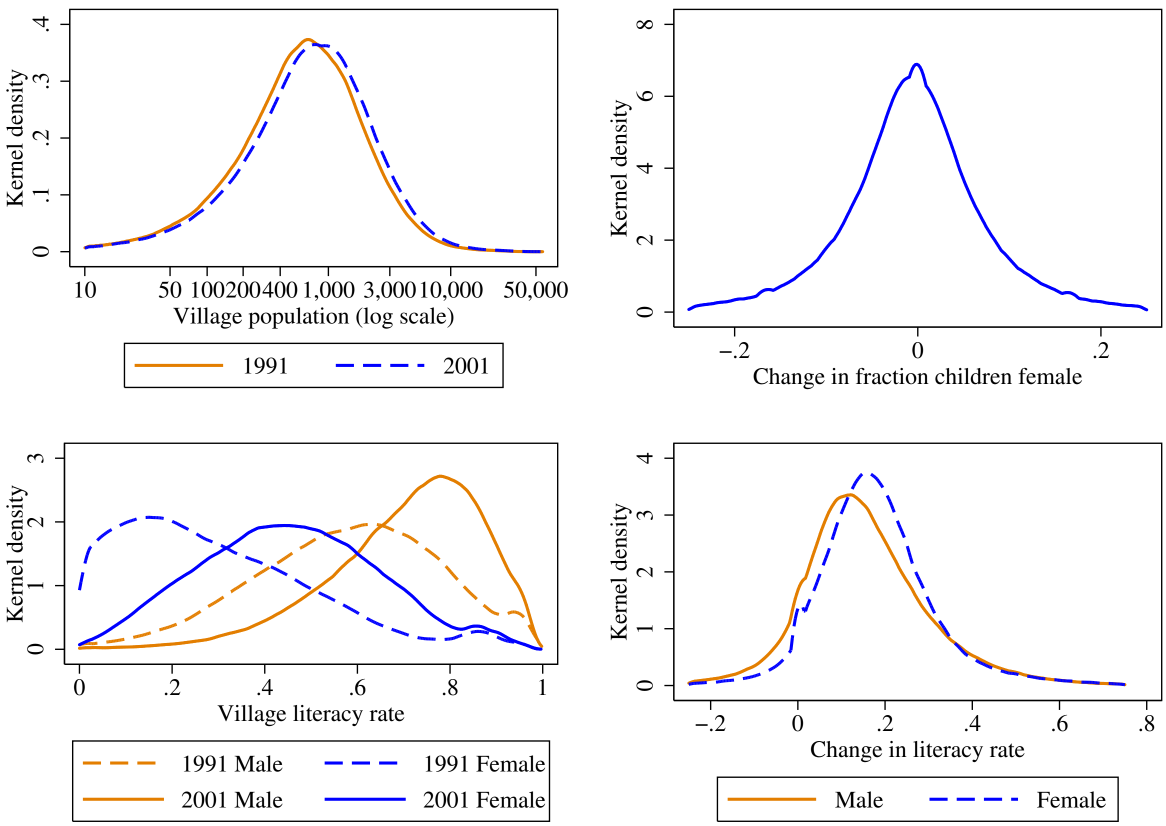
Notes: All columns include the interactions listed in footnote 3 including the change in male and female literacy, their squares, and all their interactions as well as all of their interactions with log population in 1991. These effects are plotted in figure 3. Village infrastructure enters as 1 if it 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



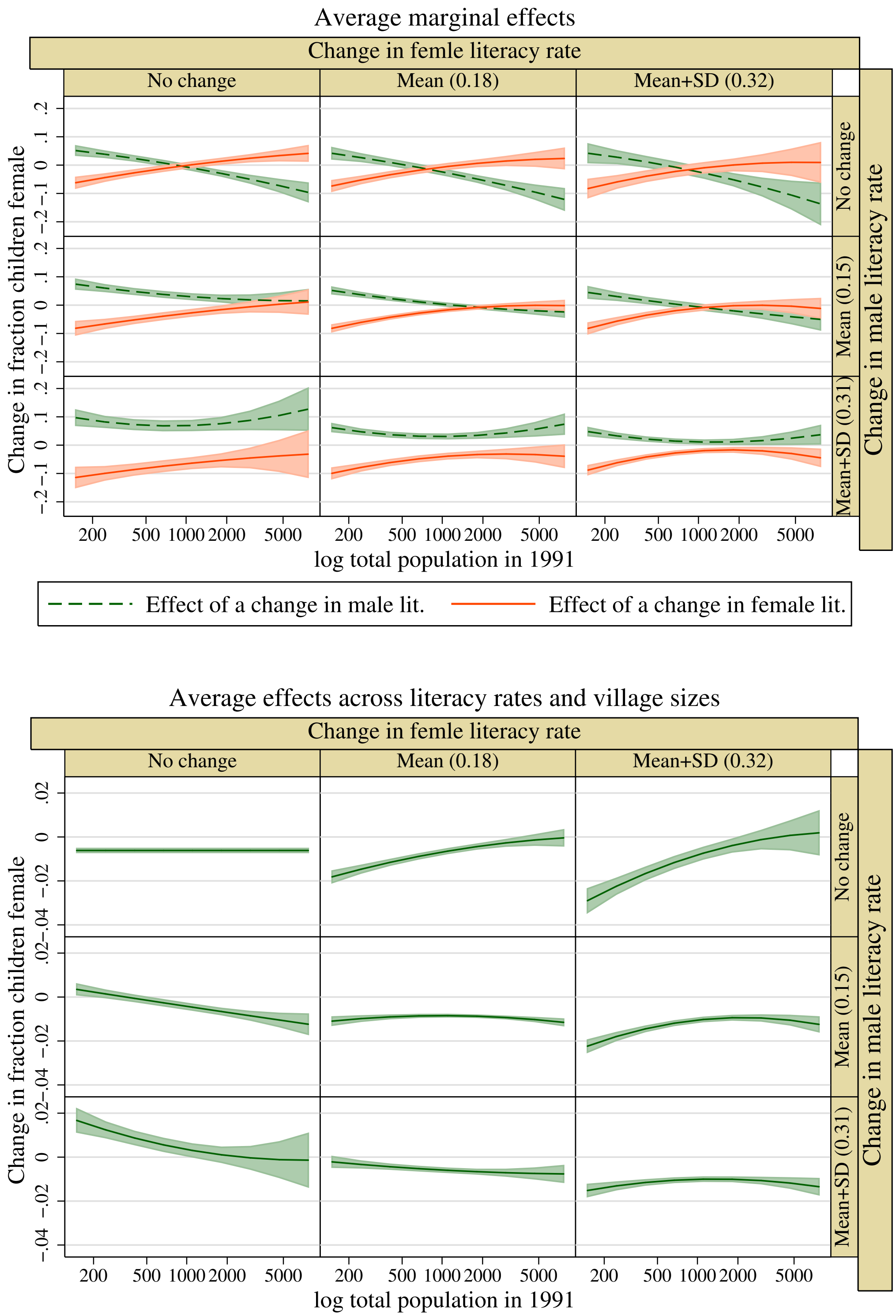
Notes: Uses the village census from 1991 and 2001. The kernel density is weighted by total village population.

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 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 3.