Remember When It Rained: The Elusiveness of Gender Discrimination in Indian School Enrollment

Laura Zimmermann[1]

University of Michigan

September 2011

Abstract

Indian girls have significantly lower school enrollment rates than boys. Anecdotal evidence suggests that gender-differential treatment is the main explanation, but empirical support is often weak and seemingly inconsistent. I propose that this may be due to age-specific forms of discrimination. I analyze school enrollment using rainfall shocks, a plausibly exogenous source of income variation. Girls' school enrollment is more vulnerable to rainfall shocks than that of boys, with 6-10 year olds driving these effects. This is consistent with young girls being out of school because boys' education is prioritized, whereas older girls suffer from low perceived benefits of education.

JEL codes: D13, I21, J16, O12, O15, Q54

1   Introduction

Empirical research documents that parents in some developing countries systematically treat their daughters differently than their sons in a number of areas, including mothers' time allocation, breastfeeding, child health, and educational decisions (see for example Barcellos et al. 2011, Jayachandran and Kuziemko 2011, Rose 1999, Rose 2000, Qian 2008). Issues surrounding gender discrimination in the intra-household allocation of resources have received particular attention in the area of educational investments.[2] Research on this topic is extensive in the Indian context, for example, where female school enrollment and attainment significantly lag behind male educational achievements (see for example Jayachandran 2002, Kingdon 2007).

Anecdotal evidence and predominantly descriptive work have identified gender discrimination in school enrollment and expenditure decisions as a major contributor to this disparity: Households allocate limited resources to the education of boys rather than girls across geographic areas and income levels, and school expenditures on girls tend to be lower than those for boys, especially in private schools (see for example Filmer and Pritchett 1999, Jejeebhoy 1993, Ramachandran 2002, Tilak 1996, Tilak 2002). The underlying reasons for this pattern are typically believed to be higher opportunity costs and lower returns to education for girls than for boys (Burra 2001, Caldwell et al. 1985, Bhatty 1998, Chanana 1990, Dreze and Saran 1995).

Research in economics, however, has only had limited success in verifying this importance in a consistent way: Empirical analyses of gender discrmination often find much weaker evidence of gender-differential treatment within the household, with results typically suggesting gender bias in educational investments only in the 10-16 age range rather than as a pervasive phenomenon (see for example Kingdon 2005 for an overview). In contrast, a growing literature on the impact of exogenous shocks like rainfall or price shocks documents the existence of gender differences in a number of

outcomes already for much younger children. While only few papers focus directly on the impact of these shocks on children's education, the usual explanation given for the found patterns is that resources are allocated differently to sons and daughters (see e.g. Maccini and Yang 2009, Qian 2008).

Taken together, the empirical results in the existing literature therefore seem puzzling: One strand of literature documents anti-girl bias in education among older children, whereas another strand of literature shows that gender discrimination in the intra-household allocation of resources already seems to play an important role for younger children. With education being one of the major parental investments in their children, this suggests that we should already find evidence of gender discrimination in education decisions in the 6-10 year age group. These apparent conflicts raise the question whether our research techniques are ill-suited to detect gender bias empirically, or whether there is an underlying explanation for these discrepancies.

This paper contributes to the gender bias puzzle in educational investments in two ways. First, I set up a simple latent-variable model of school enrollment that demonstrates that the vulnerability of a child's enrollment status to shocks depends both on how close a child is to the cutoff at which the school enrollment status changes, and on how parents reallocate resources after the shock has occurred. This combined effect has often been ignored in the literature, but crucially affects the interpretation of empirical results. The framework also allows me to reconcile the seemingly conflicting empirical patterns typically found in the literature.

Second, I analyze gender discrimination in educational decisions empirically by focusing on the impact of rainfall shocks on school enrollment for 6-14 year old children in India. In rural areas, deviations of rainfall from long-term averages are usually conceptualized as income shocks for households, as agricultural output is influenced by the weather. As household income changes in unexpected ways, parents need to make new

resource allocation decisions about their children's school enrollment, which allows me to analyze whether boys and girls are treated differently in this process.[3]

I find that girls are more vulnerable to rainfall shocks than boys with respect to school enrollment. The results are driven by 8-10 year olds, whereas there are no significant differences for older children. Higher than average rainfall leads to an increase in school enrollment for both boys and girls, but the effect is monotonically decreasing in age. There is little evidence that this is driven by extreme weather phenomena like floods or droughts, and the findings are robust to a number of alternative specifications.

Overall, this paper supports the idea that intra-household gender discrimination is a driving factor in school enrollment differences. The considerations of the theoretical framework together with the empirical results of this paper and the previous literature suggest that households practice different forms of gender discrimination at different ages: Whereas younger girls suffer predominantly from the biased reallocation of resources after shocks have occurred, older girls seem to be much more affected by being far away from the enrollment status changing cutoff.

The rest of this article is structured as follows: Section 2 provides some background on education in India, sets up a simple latent-variable model of school enrolment, and discusses how this framework helps reconcile the different empirical patterns in the literature. Section 3 discusses rainfall in India, and introduces the data and the empirical strategy. Section 4 presents the main results as well as a number of robustness checks. Section 5 concludes.

2 Education in India and a Simple Latent-Variable Model of School Enrollment

2.1 Education in India

Schooling in India is compulsory for all children between 6 and 14 years and spans their elementary education years in school, which consist of primary and upper primary school.[4] This requirement is often not binding, however. Most Indian states have not nearly achieved universal school enrollment. Table 1 demonstrates that school enrollment in rural areas in India is significantly lower for girls than for boys at all ages and that the disparity is increasing in age. It also shows, however, that while enrollment rates for both boys and girls have increased substantially over time, they have improved faster for girls, leading to a narrowing in the gender gap over time. Similar patterns hold at the sub-national level, even though levels and gender gaps differ markedly between states: In the 1991 Indian census, for example, primary school enrollment for Indian children was 35 percent in Bihar but 85 percent in Kerala (Jayachandran 2002, Kingdon 2007). Significant gender differences also hold up for Hindus and Muslims as well as for Scheduled Castes (SC) and Scheduled Tribes (ST) (see for example Jayachandran 2002, Kingdon 2007, Nambissan 1996). [5] Ramachandran (2002) estimates that about 70 percent of out-of-school children are girls.

Table 1 also reveals an inverse U-shaped pattern of school enrollment with age. While many children are enrolled in school when they are six years old, enrollment rises until about age nine, and then starts to decline as children drop out of school.

Lower school enrollment for girls than for boys is driven by two phenomena: Girls are more likely to never attend school at all, and are also more likely to drop out at any age than boys, especially when moving to upper primary and secondary levels of

schooling. While the number of children who have never been enrolled in school has decreased markedly over the years, dropouts still remain high. Many children drop out as early as grades I and II, especially among low-caste children (Caldwell et al. 1985, Nambissan 1996, Ramachandran 2002, Seetharamu and Devi 2007). This pattern suggests that both entry and exit are important margins at which school enrollment for girls and boys differ. [6]

Since the second half of the 1980s, the Indian government has shown efforts to actively increase enrollment rates for children, and girls specifically, through a variety of programs. The national government's New National Policy of Education (1986) and the Programme of Action (1992) were meant to target learning achievements, access and dropouts of 6-14 year olds. A number of specific interventions were made, like distributing free textbooks and uniforms, giving scholarships to SC and ST students, or providing mid-day meal schemes in various Indian states (Government of India 2006). The meal programs were especially successful in raising enrollment rates, particularly among girls. [7]

Public elementary school education in India is supposed to be free. Some evidence suggests, however, that this may not always be the case. Looking at data from the second half of the 1980s which distinguishes between private and public school enrollment, Tilak (1996) finds that less than half of the students in rural India receive free primary school education in government schools. Expenditures can include tuition or examination fees or a variety of other fees. Newer evidence suggests that school expenditures for a child, including tuition fees as well as expenditures on uniforms or stationary, are around 340 rupees in rural India per year. This implies that for an agricultural laborer living in Bihar with three children, the complete earnings from more than 40 days of work would go towards sending his children to primary school (Dreze 2003, Kingdon 2005). Furthermore, there is some evidence that even relatively poor families prefer

sending their children to private schools since government schools often suffer from low levels of school quality (Dreze and Kingdon 2001, Kingdon 1996).

Overall, most households that send their children to school therefore incur some form of education expenditures, and at least some households are likely to be credit constrained because they find themselves unable to borrow against future income to enroll their children in school.

2.2   A Simple Latent-Variable Model of School Enrollment

When faced with credit constraints and scarce resources, parents in India may often be unable to send all their children to school. In deciding whether or not to enroll a child in school, parents therefore take into account the perceived benefits and costs of school enrollment for this particular child. The benefits of school enrollment include broadly defined returns to schooling for the child, for example in the form of increased expected wages and potential marriage partners with a more desirable socio-economic background. Parents themselves can also benefit from their child's education both in monetary terms, since a well-educated child may be better able to take care of them when they are old, and in non-monetary terms through increased prestige in their neighborhood.

Parents will also calculate the costs of the child's education, including both the monetary costs of paying for school fees, supplies and transportation, and the opportunity costs of the child's time. Older children increasingly help at home in India, in farm work, with working for pay, or with doing household chores (Burra 2001, Caldwell et al. 1985, Jejeebhoy 1993). Being enrolled in school therefore reduces the productivity of children in these alternative activities. The costs of education can also include other factors like the cost of borrowing money if a family is credit constrained.

The difference between benefits and costs then reflects whether parents find it op-

timal to send their child to school. If the difference is positive, the child will attend school, whereas it will not be enrolled if the difference is negative. If we convert all benefits and costs into their monetary equivalent, then the difference of benefits and costs of school enrollment will be a continuous variable that reflects parents' net willingness to pay (NWTP) for their child's education.

We cannot observe parents' NWTP directly empirically, although the child's school enrollment status gives us some information about the sign of this latent variable. Nevertheless, imposing some structure on the determinants of the latent variable proves to be insightful with respect to the correct interpretation of empirically estimated effects. I model the net willingness to pay of parents in the following reduced form, where it depends both on observed characteristics like a child's age and gender, and on unobservables like parental preferences or child ability:

$$NWTP_i = \beta_0 + \beta_1 female_i + \beta_2\ age_i + \beta_3 age_i \cdot female_i + \beta_4\ shock_i +$$
$$\beta_5 shock_i \cdot female_i + \beta_6 shock_i \cdot age_i + \beta_7 shock_i \cdot age_i \cdot female_i + \epsilon_i$$

shock refers to an exogenous shock that is experienced by the household, like rainfall shocks. The model therefore does not impose any assumptions on the channel through which shocks affect parents' net willingness to pay for education, but just focuses on the reduced form impact. For simplicity, the error term ϵ is assumed to be normally distributed with mean 0 and variance σ^2.

Empirically, I analyze the change in the probability of being enrolled in school for a change in the magnitude of the rainfall shock. In the model, this effect is given by

$$\frac{\partial P(school_i=1)}{\partial shock_i} = \frac{\partial P(NWTP_i>0)}{\partial shock_i} = \phi[\frac{-A}{\sigma}]\ (\beta_4 + \beta_5 female_i + \beta_6 age_i + \beta_7 age_i \cdot female_i)$$

where A equals the expectation of $NWTP$ conditional on the observable characteristics. The first term is the density of the normal distribution and therefore positive.

This is the density of children at the current school enrollment cutoff, and therefore indicates how many children would be affected by an increase in the rainfall shock. These children are right on the margin of being affected by a shock, and therefore have the highest risk of changing their school enrollment status. Because this is a form of vulnerability that exists even before a shock actually occurs, I refer to this as the ex ante vulnerability to a shock. [8]

The second term's sign depends on the sign and magnitude of the coefficients so that the overall sign of the effect of an increased shock on the probability of school enrollment is ambiguous without imposing more structure on the model. This is therefore an empirical question. Intuitively, this second term is the change in parents' willingness to pay in response to the increased shock. Since this is a form of vulnerability that sets in after the shock has actually occurred, I refer to this as the ex post vulnerability to a shock.

This simple theoretical framework suggests that the empirical effects we analyze are a combination of ex ante and ex post vulnerabilities, which is often neglected in the literature where many papers implicitly interpret the estimated effect of a shock as a measure of ex post vulnerability. While I cannot separate the two effects empirically, this has important consequences for the correct interpretation of empirical patterns. Specifically, it seems plausible that an important part of the seemingly conflicting empirical results found in previous research can be explained by making use of the simple model above.

One type of empirical patterns is found in the growing literature on the gender differential impacts of exogenous shocks for young children. Rose (1999), for example, finds that positive deviations of rainfall from the average significantly increase the survival chances of girls in India. Maccini and Yang (2009) find that higher rainfall in a person's birth year in Indonesia has positive effects on women's long-run socioeconomic

outcomes, whereas there is no effect for men. Jacoby and Skoufias (1997) use rainfall shocks to analyze the impact of income shocks on school attendance in six Indian villages and find that households seem to use school attendance as a form of self-insurance to cope with these shocks. Bjorkman (2008) looks at the impact of income shocks on school enrollment and cognitive skills of children in Uganda. She finds that negative rainfall shocks lead to a significant decrease in school enrollment and test scores for girls, but have a smaller impact on boys.

Similar patterns arise for other types of shocks. Qian (2008) finds that in China sex-specific agricultural income shocks occuring to women raise girls' survival chances whereas a positive male-specific income shock worsens them. Baird et al. (2011) document that negative GDP deviations from trend have a larger impact on female infant mortality than on that of males in a large sample of births from 59 developing countries. Duflo (2003) finds that a grandmother's pension eligibility in South Africa has an impact on the anthropometric status of young girls, but not that of boys.

What most of this research has in common is the conclusion that gender differences in how shocks affect the intra-household distribution of resources are already important for young children. These results may seem inconsistent, however, with the empirical patterns in the intra-household allocation literature. Research in this literature usually analyzes whether household education expenditures respond differently to the addition of a boy to the household than to an additional girl by using an Engel curve or hurdle model approach. Results often suggest much weaker gender discrimination than anecdotal evidence, and find that gender-differential effects are concentrated among children of older ages: Subramanian and Deaton (1991), for example, only find weak evidence of gender discrimination in rural Maharashtra for 10-14 year olds, but not for other age groups, while Lancaster, Maitra and Ray (2008) detect significant gender bias in their sample of rural Bihar and rural Maharashtra only for the age group of 10-16 year olds.

Other papers like Kingdon (2005), Azam and Kingdon (2011), and Zimmermann (forthcoming) have shown that aggregation of the data from the individual to the household level, and the estimation technique used may explain some of this empirical pattern, but still find that significant gender differences are concentrated among older children.

Taken together, the empirical results in the two strands of literature therefore seemingly present a puzzle, since gender discrimination appears to be concentrated in mutually exclusive age groups. The simple theoretical framework in this paper allows these different patterns to be consistent with each other, however: As demonstrated in Table 1, school enrollment decreases at older ages, and this pattern is especially pronounced for girls. At the same time, school expenditures for older children are typically higher than those for younger ones. The gender gap in education expenditures will therefore be most pronounced at older ages because few girls are still attending school, so an additional girl of that age group will raise household education expenditures by significantly less than another boy.

This is consistent with the results in the gender discrimination literature. Especially more recent papers on this issue, for example Kingdon (2005), Azam and Kingdon (2011) and Zimmermann (forthcoming), document that the discrimination against girls at the margin of zero versus positive education expenditures is especially pronounced at older ages. In terms of the model above, this suggests that parents' net willingness to pay for education is much lower for a girl than a boy in this age group. The typical girl will therefore be further away from the school enrollment changing cutoff than the typical boy, implying that girls have a lower ex ante vulnerability to shocks at this age than boys.

By contrast, the gender gap in school enrollment tends to be narrower at younger ages, suggesting that male and female ex ante vulnerability may be relatively similar for this age group. This implies that the pattern of gender differential impacts for younger

children in the shock literature is driven by the higher ex post vulnerability of girls to shocks.

Overall, these considerations show that previous empirical patterns are consistent with each other in a world where different forms of gender discrimination are being practiced at different ages: Young girls are mostly discriminated against in the re-allocation of resources after unanticipated shocks, but for older girls the lower ex ante vulnerability dominates. Put differently, both young boys and girls attend school if possible, but girls may be the first to be taken out of school when an unexpected shock hits. If this type of discrimination is present, we should therefore expect that girls benefit more from a positive rainfall shock than boys at this age since boys are already attending school, so that additional resources are used to send the girl to school. By the same token, girls are more likely to be taken out of school after a bad rainfall shock than their brothers.

At older ages, on the other hand, parents are much less likely to see the benefits from educating their girls further, leading to few girls being at the margin of being affected by a shock. This will at least partially balance out any higher female ex post vulnerability to shocks that is still likely to be present at these ages. If this form of discrimination exists, we would therefore expect to find no gender differences or even a higher effect of shocks for boys in the empirical analysis. The following section sets up the empirical strategy to examine these hypothesized patterns of intra-household gender discrimination in India by using rainfall shocks.

3 Weather in India, Data, and Empirical Strategy

3.1 Weather in India and the Effect of Rainfall Shocks on Agricultural Output

Weather in India varies widely by region and often also at the sub-regional level. India has six different climatic regions, ranging from humid tropical areas to desert-like dry regions. While there are local differences, India in general experiences four seasons: winter (January to February), summer (March to May), the monsoon season (June to September) and the post-monsoon season (October to December) (De et al. 2005, Ribot et al. 1996). About 80 percent of annual rainfall occurs during the monsoon months (Department of Agriculture and Cooperation 2007, Webster and Hoyos 2004).

As about 60 percent of the agricultural sector depends on rain as the only source of water, agricultural output should be heavily influenced by precipitation levels and timing (Directorate of Economics and Statistics India). Unfortunately, the dataset that will be used for the analysis does not ask households about their crop yield or household income, so we cannot verify directly whether rainfall shocks do indeed feed through to changes in household income. [9] Figures 1 and 2 therefore show the relationship between standardized rainfall shocks (actual rainfall - mean rainfall divided by standard deviation) and analogously standardized district output of rice and wheat, two major food grains, with data from the Indian Department of Agriculture and Cooperation for Indian districts for the period 1999-2007. [10]

Positive numbers signify more rainfall or more agricultural output than usual, whereas negative numbers mean less rainfall or output than usual. The figures show the non-parametrically estimated relationship between rainfall and output shocks with the corresponding point-wise confidence intervals. The estimation in Figure 1 suggests that the relationship between rainfall shocks and output of rice is close to concave. [11]

While more rainfall is always beneficial for rice output, the impact of rainfall shocks on output shocks is lower for positive shocks than for negative shocks. This suggests that in my analysis the impact of negative rainfall shocks should be especially pronounced.

Previous literature has usually ignored the potential issue that the impact of rainfall shocks on agricultural output may vary by crop. Papers such as Levine and Yang (2006) focus on the relationship between rainfall and rice to motivate the use of rainfall shocks as income shocks. Rice is a very special kind of crop, however, as rice fields are often flooded during the planting process. In dry parts of India without extensive irrigation systems, for example, rice is usually sown immediately after the onset of the monsoon. Other crops like wheat, on the other hand, are much more vulnerable to excessive rainfall, so that non-linearities of the impact of rainfall on wheat output are an important potential concern (Sreenivasan and Banerjee 1973). Figure 2 presents the results for the relationship between rainfall and wheat output shocks and demonstrates that at least in India even relatively large positive rainfall shocks do not seem to have a negative impact on wheat output. There is a positive relationship between more rain and agricultural output for negative rainfall shocks, but the non-parametrically estimated curve quickly flatens out for positive rainfall shocks. [12]

Overall, the figures suggest that rainfall shocks affect household income through agricultural yield and that therefore one channel through which rainfall affects school enrollment is through the change in household income. The impact of rainfall shocks on agricultural output is always positive but may be non-linear, especially for negative rainfall shocks.

3.2 Data and Variable Construction

The primary datasets used in this paper are the National Fertility and Health Surveys (NFHS) of 1992/93 and 1998/99. The NFHS is a cross-sectional dataset that is

nationally representative of India's population. It includes current school enrollment for children and other family background characteristics as well as information about the district in which households currently live, which will be used to match household survey information to rainfall shocks. [13] Unfortunately, a district variable is missing for the most recent NFHS of 2005/06, so this data cannot be exploited in this article. [14]

Rainfall data comes from the Center of Climatic Research at the University of Delaware, which includes monthly precipitation values on 0.5 intervals in longitude and latitude. This grid is achieved by using data from various weather stations and a number of different sources of rainfall data to interpolate rainfall in these intervals. [15] Rainfall data can therefore be matched to districts by using longitude and latitude coordinates. Rainfall shocks are constructed by subtracting the long-term average rainfall from rainfall in a given period and then dividing this term by the standard deviation of rainfall in that district. A positive number therefore signifies higher than average rainfall, whereas a negative number implies less rainfall than usual. The long-term average includes rainfall from 1970 to 2008, but does not include the current period. Such a construction of the rainfall shock has the advantage of being readily interpretable in terms of standard deviations and ensures that geographic differences in rainfall are taken into account: 200mm of additional rainfall, for example, will be a major positive deviation in very dry parts of the country like Rajasthan, but may not be a large deviation from the average in wetter regions like Assam.

In the main specification of this paper, a rainfall shock is defined as the deviation of rainfall in the 12 months before the interview month and year from typical rainfall in this 12 month interval. For a household interviewed in August 1998, for example, rainfall in the previous 12 months includes all rainfall in the appropriate district from August 1997 to July 1998. The expected rainfall in this period is calculated as average

rainfall in a August to July interval between August 1970 and July 2008, excluding August 1997 to July 1998. Analogously, the standard deviation of rainfall in such a period is constructed. The rainfall shock for a household interviewed in August 1998 is therefore the deviation of rainfall in the previous 12 months from expected rainfall in such a period, divided by the standard deviation. In consequence, a rainfall shock is district-, month- and year-specific, where the important date is the interview date. [16]

The moving time window, which depends on the household's interview month and year, ensures that rainfall shocks for all households really only refer to past rainfall and capture the same time span before the interview date, which would not be the case with the use of the same calendar months for all households in the sample. In both NFHS surveys, households were interviewed over the span of almost two whole years, depending on residence within states and districts. [17] In many Indian states, households were interviewed over the span of a couple of months. This feature makes the use of rainfall shocks by calendar year a less sensible choice.

The caveat of this specification is that shocks in the previous 12 months are implicitly taken to have the same impact on school enrollment regardless of the interview month. Yet, a household that reports the enrollment of children at the beginning of the school year may be different from a household where children attend school near the end of the school year: Resources could for example be relatively abundant when children start the school year, so that shocks in the previous 12 months have a limited impact on enrollment, but later on resources become scarcer and shocks may have a larger effect on education outcomes. I will take account of this problem by including month and year fixed effects into my empirical specifications, which control for seasonality effects as well as general time trends. I also include district fixed effects.

Time and district fixed effects take account of time-invariant regional effects like cultural norms, history or educational priorities, and general time trends. Given that I

16

only have two rounds of data, I am unable to look at changes over time. Therefore, the effects that I am identifying here will be those that remain important over the whole time span. As I only know school enrollment in general, but not the school attended, I am also unable to pick up school transfers in reaction to rainfall shocks, for example from government schools to private schools. My analysis thus focuses on the extensive margin of school enrollment.

3.3   Empirical Strategy and Summary Statistics

My main empirical specification is given by the regression equation

$$school_{ijkl} = \beta_0 + \beta_1 shock_{jkl} + \beta_2\, age_{ijkl} + \theta_k + \eta_l + \gamma_j + \epsilon_{ijkl}$$

where the subscripts refer to individual i in district j in month k and year l. *school* is an indicator variable equal to 1 if a child is currently reported to be enrolled in school and *shock* is the standardized rainfall shock variable. *age* is the child's age. In all empirical specifications, I include age dummies for every age between 6 and 14 years, with age 6 being the reference category. θ are month fixed effects, η are year fixed effects and γ are district fixed effects. Standard errors are clustered at the district level. The coefficient of interest is β_1. I run this specification separately for girls and boys but also test for gender differences by fully interacting the model with gender.

In alternative specifications, I interact the rainfall shock with age dummies to get at non-linear effects in age. I also include rainfall shock lags and look at non-linear effects of rainfall shocks on school enrollment by splitting up the rainfall shock variable into a number of indicator variables depending on the magnitude and sign of the rainfall shock. Lags are constructed by moving the time window backwards: The main rainfall shock variable takes into account rainfall deviations in the 12 months prior to the interview months. Lag 1 of this variable, for example, therefore constructs analogously

17

standardized rainfall shocks by focusing on rainfall deviations 13-24 months prior to the interview date.

My identification strategy therefore uses the variation in rainfall shocks within districts across months and years to identify the effect on school enrollment. Summary statistics of the constructed rainfall shocks are presented in Table 2. If deviations of rainfall from long-term averages were perfectly random and we had an infinite sample size, we would expect the mean of the rainfall shock variables to be zero, and the distribution of the variables to be symmetric around the mean since the probability of a positive shock of a given magnitude occurring is the same as the probability of getting a negative rainfall shock of the same size. In reality, however, the mean rainfall shock will often be different from zero. Table 2 shows that the mean of the rainfall shock variables tends to be relatively small. All means, however, are statistically significantly different from zero. Except for the two constructed leads of the rainfall shock variable, rainfall shocks are relatively symmetric around zero.

Table 2 also splits up the main rainfall shock variable into magnitude categories to give an overview of the distribution of the rainfall shock. As we can see, the distribution is skewed towards negative deviations of rainfall from long-term averages: About 27 percent of rainfall shocks fall into the category of 0.5 to 1.5 standard deviations less rainfall than normal, whereas only about 18 percent of rainfall shocks in my sample are 0.5 to 1.5 standard deviations higher rainfall than expected. About 40 percent of rainfall shocks are relatively small deviations from long-term averages within 0.5 positive or negative standard deviations. Extreme rainfall shocks that are larger than 2.5 standard deviations in absolute value are very rare.

Table 3 reports the break-down of children in my sample by age and gender. It also shows averages for household size, age of the household head, wealth index, and whether the household owns their house and any agricultural land. As we can see, the

sample contains slightly more boys than girls, but the breakdown within gender by age is similar: 12 percent of girls are six years old, for example, whereas six-year old boys make up 13 percent of boys in my sample. Averages for the household variables are also similar between boys and girls, suggesting that the living conditions for girls and boys are close to identical. Both girls and boys live in households where the average age of the household head is 45 years. 95 percent of households own their house and two thirds own any agricultural land. The typical household size of a household with a boy is 7.6, which is only slightly smaller than the average household size for households with a girl, which is 7.8, although this difference is statistically significant. Households with boys and girls also perform very similar for the wealth index. Overall, these summary statistics suggest that at least in my sample there is little room for sample selection concerns of households with boys and girls. Any results presented below are therefore not driven by differences in household characteristics. [18]

4 Results

4.1 Main Results

The main results are presented in tables 4, 5, 6 and 7. Table 4 reports results with just rainfall shocks in the past 12 months as an explanatory variable, controlling for age and including month, year and district fixed effects, while Table 5 interacts the shock variable with age to look at non-linear effects of rainfall shocks with respect to age. Table 6 splits the rainfall shock variable up into indicator variables of different magnitude categories to analyze whether rainfall shocks of different sizes have non-linear effects on school enrollment. Table 7 presents the results of these magnitude categories interacted with age categories.

In Table 4, a rainfall shock of one standard deviation in the previous 12 months

translates into an increase in the probability of being enrolled in school of 2.3 percentage points for girls and 1.3 percentage points for boys. As mean school enrollment in my sample is 0.66 for girls and 0.8 for boys, this translates into a 3.5 percent increase in school enrollment for girls and a 1.6 percent increase in male school enrollment. Both coefficients are highly statistically significant and the effect for girls is statistically significantly greater than the effect for boys at the 5 percent level. Results are similar for the subsample of children aged 6-10 years, where a shock of one standard deviation increases school enrollment by about 2.5 percentage points (3.7 percent) for girls and by 1.2 percentage points (1.5 percent) for boys. For 11-14 year olds, there is no longer a statistically different effect of rainfall shocks on school enrollment by gender. School enrollment for girls improves by 1.7 percentage points (2.9 percent) and that of boys increases by 1.3 percentage points (1.6 percent).

Table 5 looks at the non-linear effects of rainfall shocks with respect to age by interacting dummy variables for each age with the shock variable. The table shows that the effect of the rainfall shock is monotonically decreasing in age: 6 year old children experience the largest increases in school enrollment with close to five percentage points, and the effect decreases as children become older. The table also presents tests for the gender equality of the magnitude of the rainfall shock for every age. Significant gender differences occur for 8-10 year olds, where the rainfall shock has significantly larger effects on school enrollment for girls than for boys. The interaction effects for 8 and 10 year olds are also significantly different from each other for girls and boys, indicating that the impact of the rainfall shock declines more sharply for boys than girls at these ages. The rainfall shock remains highly statistically significant in this specification as well. This confirms that gender differences in the impact of rainfall shocks are driven by younger children.

Tables 6 and 7 focus on non-linearities in the magnitude of the rainfall shock by

splitting up the rainfall shock variable into a number of dummy variables, depending on the magnitude of the shock. The reference category in these regressions are relatively small shocks of a magnitude between -0.5 and 0.5 standard deviations. As Table 6 shows, the impact of rainfall shocks on school enrollment is concentrated among negative shocks between -2.5 and -0.5 standard deviations: A rainfall shock between -2.5 and -1.5 standard deviations, for example, leads to a 6.2 percentage points decrease in school enrollment for 6-14 year old girls and to a decrease of 3.6 percentage points for boys of the same age. Except for 6-10 year old girls, where the category of shocks between 1.5 and 2.5 standard deviations is significant at the 10 percent level, no positive rainfall shock category is statistically significantly different from zero. Significant gender differences between girls and boys are also concentrated on negative shocks and are again driven by 6-10 year old children, where girls experience a significantly greater decrease in school enrollment after a negative rainfall shock than boys. There is also little evidence that large absolute shocks of more than 2.5 standard deviations, which we might think of as floods or droughts, have particularly important effects on children's school enrollment. This may be partly due to the fact that, as Table 2 showed, such extreme rainfall shocks are very rare in my sample and their impact may therefore be imprecisely estimated. But these results are also consistent with the agricultural output graph from Figure 2, which showed that large absolute shocks may have no detrimental effects for some crops such as wheat so that the income shocks of large rainfall shocks may actually not be that different from income shocks caused by smaller rainfall deviations.

One big concern with using rainfall shocks is that rainfall shocks may not just act as income shocks, but may also change the opportunity costs for children. Children may for example need to help on the farm when the harvest is good. This would work in the opposite direction of the effect of alleviating credit constraints since more rainfall and

thus a better harvest would now be associated with lower school enrollment. The same pattern would occur if roads are closed during heavy rainfall or schools inaccessible. Therefore, the positive relationship between rainfall shocks and school enrollment that we find even in this more detailed analysis here suggests that these factors cannot be dominant.

Table 7 interacts the magnitude categories of the rainfall shock variable with the age dummies. For an easier overview over the age and magnitude patterns, Table 7 reports the estimated overall effect for a given magnitude and a given age of a child rather than the full regression table with main and interaction effects. Black borders show statistically significant gender differences at at least the 10 percent significance level. The table confirms the results from the previous tables by showing that most significant results as well as most significant gender differences are concentrated among 6-10 year olds. Consistent with Table 5, Table 7 demonstrates that rainfall shocks matter most for 6 and 7 year old children, but in a quite symmetric fashion: Negative rainfall shocks lead to significant decreases in school enrollment for both boys and girls, whereas positive rainfall shocks increase school enrollment, especially for girls who experience significantly bigger increases in school enrollment than boys. Significant gender differences are also present for 8-10 year olds, where girls are significantly more affected by rainfall shocks than boys. The effects are also more concentrated on the impacts of negative shocks than positive ones, suggesting that girls are more likely to be taken out of school after bad shocks rather than being enrolled after positive rainfall shocks at these ages. As in previous tables, children over 10 years old are relatively little affected by rainfall shocks.

Overall, the analysis shows that girls tend to be more affected by rainfall shocks than boys, with results being driven by the primary school aged children rather than children that are older. If we interpret rainfall shocks as income shocks for these rural households,

then the results suggest that younger children are more vulnerable to income shocks than older children and that gender differences of income shocks are driven by children in the 8-10 year age group. Furthermore, younger children in general are more vulnerable to rainfall shocks than older children, especially at ages 6 and 7.

4.2   Robustness Checks

While the results in the previous section already suggest a consistent pattern where the impact of rainfall shocks is decreasing in age and shocks have larger effects on girls than boys in the 8-10 year range, a number of checks can be performed to ensure that they are robust to different empirical specifications.

One potentially important concern is the chosen interval of the shock variable. The constructed rainfall shock takes into account rainfall in the 12 months prior to the interview month. If there are significant lags in the translation of agricultural output into household income, however, it may be that what matters more for school enrollment is not rainfall in the previous 12 months, but rainfall that occurred prior to this. In order to check whether this is an important concern, Table 8 extends the analysis of Table 4 by also including two lags in addition to the main rainfall shock variable. Lag 1 is constructed by taking into account deviations of rainfall from long-term averages between 13 and 24 months prior to the interview months, and lag 3 analoguously focuses on rainfall in the interval of 25 to 36 months prior to the interview date. If these lags have important effects on school enrollment of children, we should see significant estimated coefficients in Table 8 or at least changes in the estimated coefficient of our main rainfall shock variable. Table 8 shows that neither is the case: The estimated coefficients for the lag variables are in general small and are never statistically significant; furthermore, the estimated coefficients for the main rainfall shock variable are very robust to the inclusion of lags and hardly change from the estimates reported in

Table 4. This justifies just focusing on rainfall in the previous 12 months in this paper.

Another potentially important concern is measurement error. Matching rainfall data to the closest districts necessarily introduces measurement error since local conditions will vary in different parts of a district. In their paper, Maccini and Yang (2009) instrument for rainfall shocks from the closest source (in their case weather stations) with rainfall from the next nearest weather stations since rainfall should be correlated, but measurement error uncorrelated between rainfall data from different weather stations. I use a similar technique in Table 9 where I instrument for the closest grid point of rainfall of a district with the 5 next closest points. As the table shows, measurement error is not much of a concern in my case, since IV results are basically identical to the main results presented in Table 4 of this paper. [19] The first stage IV results are documented in appendix table A1.

A third problem that might be driving the results is selection into my sample, for example through death or migration correlated with rainfall shocks. If selection is not an issue, then regressing cohort sizes of girls and boys on the rainfall shocks should not produce any effects: Whether rainfall in the previous 12 months was particularly high or low should not affect how many children of a particular age and gender interviewers were able to locate. The results of this analysis are shown in Table 10 and demonstrate that selection indeed is not an issue since none of the estimated coefficients is close to being statistically significant. [20]

The hypothesis that the effects of rainfall on school enrollment work through the household income channel, rather than through any other variables correlated with rainfall, also lends itself to being tested empirically by focusing on a falsification test. If rainfall shocks lead to household income shocks through the channel of agricultural output and then influence school enrollment, we should find that the impact of rainfall shocks on school enrollment is significantly lower in urban areas, where agricultural

production is less important for household income. Table 11 presents the results for urban areas, focusing on households living in large cities. [21] The results show that the impact of rainfall shocks on school enrollment is different in urban areas: A one standard deviation of additional rainfall here has no statistically significant effect on school enrollment for girls and older boys, and is even associated with a large negative effect of 5.5 percentage points for 6-10 year old boys. [22] Table 11 also reports the p-value of testing the equality of coefficients in urban areas with the rural coefficients estimated in Table 4; the urban coefficients are significantly lower than those in rural areas for boys of all ages and for the overall sample of girls, although not for the subgroups of 6-10 and 11-14 year old girls. Overall, however, Table 11 supports the idea that rainfall shocks matter more in rural areas and are thus working through household income shocks.

5 Conclusion

The preceding analysis has demonstrated that rainfall shocks have a robust impact on school enrollment for 6-14 year old children in India. The effect is concentrated among young children, with the magnitude of the effect decreasing in age. 6 and 7 year old children experience large increases in school enrollment after positive rainfall shocks, and important decreases in the probability of attending school after a negative rainfall shock. Girls in this age category benefit significantly more from a positive rainfall shock than boys, although there are no gender differences with respect to negative rainfall shocks. Rainfall shocks also affect girls significantly more than boys in the 8-10 year age range, where girls seem to be especially vulnerable to being taken out of school after adverse rainfall shocks. There are no significant gender differences in the impact of the rainfall shocks at older ages, and 11-14 year old children in general are affected

25

relatively little by these shocks.

This pattern of results is consistent with different forms of intra-household gender discrimination being dominant at different ages, as suggested in the discussion of the theoretical framework. Young girls appear to have a higher ex post vulnerability to shocks as parents allocate scarce resources to boys first. As children get older, however, girls reduce their ex ante vulnerability to shocks more quickly than boys by being more likely to drop out of school permanently. The summary statistics reveal that in my sample girls and boys live in similar households, suggesting that sample selection of households that have girls and households that have boys is not driving these results. Such a story also reconciles the seemingly conflicting empirical patterns found in the existing gender discrimination and shock literatures. While this does not mean that there is no other channel that could generate these empirical patterns, such a story needs to match the distinct age and gender patterns of school enrollment.

There is room for important future research in looking more closely at why parents discriminate against girls in education decisions in more detail, since intra-household gender discrimination may be driven by a broad variety of factors, including gender differences in returns to education, opportunity costs and taste-based discrimination. [23] Only if parents' motives are better understood can policies be targeted successfully at improving school enrollment for older children.

The results in this paper have two important implications for economic research as well as policymaking. First, transitory shocks like deviations in rainfall can potentially have important long-run impacts on socio-economic outcomes in two respects: Especially for young children, these household-level shocks influence the acquisition of human capital directly. A child that experienced positive income shocks early in the school career may be able to accumulate more human capital than a similar child that did not benefit from higher than average rainfall. This may lead to long-run differences

between otherwise identical children. Additionally, however, just the possibility of being affected by future household shocks may impact a child's educational career, with more vulnerable children likely to permanently drop out of school earlier than others. This suggests that reducing a household's exposure to transitory shocks like rainfall shocks will benefit even children who do not seem to be affected by these shocks in empirical analyses.

Second, the considerations on gender discrimination in education in India suggest that policies may need to target girls and their families age-specifically in order to close the gender gap in school enrollment. While younger girls are most vulnerable to unanticipated transitory shocks in this paper, the results suggest that school enrollment for girls in this age group may improve as household income rises permanently. Parents do seem to see the benefits of sending young girls to school for at least some time, but choose to prioritize male school enrollment in the face of scarce financial resources. For older children in general, and girls especially, however, school enrollment decisions may be a complicated combination of credit constraints, expectations about future shocks and opportunity cost considerations, so that just alleviating credit constraints without changing children's and parents' situation and perceptions in the other two respects may have only limited beneficial effects on children's school enrollment.

Notes

[2]Intra-household gender discrimination is usually defined broadly as parents systematically treating boys and girls differently, rather than looking narrowly at particular potential explanations like differences in returns to education or opportunity costs.

[3]My analysis focuses on the reduced-form impact of rainfall shocks, however, since rainfall shocks may also have additional impacts on schooling that are not directly related to household income: Heavy rainfalls could also make schools inaccessible or change children's opportunity costs of school attendance.

[4]Elementary education is roughly equivalent to primary and middle school. In general, primary schooling includes classes I through V, and upper primary school classes VI through VIII, although there is state variation in the exact classes which fall under primary and upper primary education (Tilak 2002).

[5]Scheduled Castes (SCs) is the official name for untouchables or dalits. Scheduled Tribes (STs), also often called adivasis, are not part of the usual caste hierarchy and often live in secluded hilly or forest areas. Both SCs and STs are traditionally disadvantaged groups within Indian society and still tend to have much poorer outcomes than members of other castes.

[6]It also suggests that while there is a role for sheepskin effects occurring when children for example complete primary school, dropout decisions especially at young

ages are not driven by such considerations.

[7]Problems with a number of these programs persist, however, with textbooks and uniforms for example not reaching beneficiaries, or not on time. There is also large inter-state variation in the scale of the programs: In a number of states, beneficiaries amount to less than 7 percent of students, while in states like Tamil Nadu or Karnataka about 70 percent of children receive free textbooks (Bhatty 1998, Tilak 2002).

[8]My use of the term ex ante vulnerability is an extension of the concept of ex ante risk used in a number of papers like Fitzsimons (2007) and Kazianga (2006), which is usually conceptualized as living in an area that is more prone to shocks than other regions. While in these papers ex ante risk is identical for all members of a household, ex ante vulnerability in this paper can differ between individuals living in the same household, for example between boys and girls.

[9]Most collected information is related to assset ownership, which is a better proxy for permanent income rather than transitory income changes induced by rainfall shocks.

[10]In India, food grain production accounts for about two thirds of total agricultural production and cropped area. About 37 percent of food grain area was irrigated in the early 1990s, which the time of the first household survey that I am using (Directorate of Economics and Statistics India). This percentage is increasing: In 2007-08, about 47 percent of the food grain area was irrigated. To the extent that irrigation insulates a household from rainfall shocks, this will reduce the observed impact of rainfall shocks.

[11]This pattern differs somewhat from Levine and Yang (2006) who estimate an approximately linear relationship of rainfall shocks and rice output in Indonesia.

[12]The overall less steep relationship between rainfall and wheat is probably heavily influenced by the prevalence of irrigation. Wheat tends to be much more heavily grown in highly irrigated areas than rice, which insulates output from rainfall shocks to some degree.

[13]India's state and district boundaries change dramatically over time. Between 1971 and 2001, for example, the number of districts rose from 356 to 593. New districts are not only created by partitioning existing districts, but may also arise from more complicated rearrangements (Kumar and Somanathan 2009). This often makes district comparisons across time difficult. Both rounds of NFHS surveys that I use base the district boundaries on the boundaries of the Indian Census of 1991, however, and are thus comparable between surveys.

[14]More detailed geographical information is excluded in the NFHS of 2005/06 to ensure confidentiality since HIV test results were collected simultaneously. See NFHS Supplemental Documentation (2010) for more details

[15]For more details, see the University of Delaware Center of Climatic Research website at $http://climate.geog.udel.edu/\ climate/html_pages/Global2_Ts_2009/README.global_{pt}s_2009.html$.

[16]Lags and leads of these rainfall shocks are also constructed, and the construction of these shocks will be detailed further in the robustness check section.

[17]In the 1992/93 survey only the Jammu districts of the state Jammu and Kashmir were interviewed. Households in the state of Tripura were not interviewed in 1992/93 but were interviewed in 2000 as part of the 1998/99 survey. There is no evidence, however, that in general the interview date is correlated with state characteristics.

[18]These patterns also hold up when comparing summary statistics for children within the age groups of 6-10 and 11-14 years. For 11-14 year olds, the gender difference in the wealth index is just significant at the 10 percent level, with girls living in slightly wealthier households than boys. That boys and girls seem to live in very similar circumstances is somewhat surprising given typically practiced differential stopping rules to childbearing in India, which depend on the sex composition of already born children (see for example Jensen 2002).

[19]This is probably more driven by the rainfall data that I use, which has already

been smoothed by using data availlable from various rainfall stations rather than due to measurement error not being an issue in my case.

[20]This conclusion is robust to the use of the alternative rainfall specifications that include lags of the rainfall shock or split the rainfall shock variable up into magnitude indicator variables.

[21]The NFHS datatsets distinguish between the countryside, towns, small cities and large cities. Only the countryside gets coded as a rural area, whereas the rest is coded as urban even though towns may lie in major agricultural areas and have close economic connections to the surrounding rural areas. Agricultural production may therefore still play an important role in towns and small cities, but should have much less of an impact on household income in large cities. Consistent with this idea, in results not shown the impact of rainfall on school enrollment is higher the more rural an area becomes: The impact of rainfall shocks on school enrollment for both boys and girls is lowest in large cities and then increases for small cities and towns.

[22]The large negative effect of rainfall on school enrollment for young urban boys is somewhat puzzling. Maccini and Yang (2009) find a similar effect in their paper and hypothesize that this may be due to water-borne illnesses, since male health is typically more vulnerable than that of females. Alternatively, high rainfall in cities could have other effects like the flooding of roads, making school inaccessible. It is unclear, however, why such factors should not have a similar effect on female school enrollment.

[23]Data limitations do not allow me to convincingly separate different potential channels empirically in this paper.

References

[1] Aslam, Monazza, and Geeta Gandhi Kingdon, 'Gender and Household Education Expenditure in Pakistan,' *Applied Economics* 40(2008), 2573-2591.

[2] Azam, Mehtabul, and Geeta Gandhi Kingdon, 'Are Girls the Fairer Sex in India? Revisiting Intra-Household Allocation of Education Expenditure,' IZA Discussion Paper no.5706 (2011).

[3] Baird, Sarah, Jed Friedman, and Norbert Schady, 'Aggregate Income Shocks and Infant Mortality in the Developing World,' *The Review of Economics and Statistics* 93(2011), 847-856.

[4] Barcellos, Silvia Helena, Leandro Carvalho, and Adriana Lleras-Muney, 'Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?,' RAND Working Paper (2011).

[5] Bhatty, Kiran, 'Educational Deprivation in India: A Survey of Field Investigations,' *Economic and Political Weekly* 33(1998), 1731-1740.

[6] Bjorkman, Martina, 'Income Shocks and Gender Gaps in Education: Evidence from Uganda,' Working Paper (2008).

[7] Burra, Neera, 'Cultural Stereotypes and Household Behaviour: Girl Child Labour in India,' *Economic and Political Weekly* 36(2001), 481-488.

[8] Caldwell, John C., P.H. Reddy, and Pat Caldwell, 'Educational Transition in Rural South India,' *Population and Development Review* 11(1985), 29-51.

[9] Chanana, Karuna, 'Structures and Ideologies: Socialisation and Education of the Girl Child in South Asia,' *The Indian Journal of Social Science* 3(1990), 53-71.

[10] De, U.S., R.K. Dube, and G.S. Prakasa Rao, 'Extreme Weather Events over India in the Last 100 Years,' *Journal of Indian Geophysical Union* 9(2005), 173-187.

[11] Department of Agriculture and Cooperation, Government of India, *Annual Report 06-07* (2007).

[12] Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Government of India, *Agricultural Statistics at a Glance* (2010).

[13] Dreze, Jean, 'Patterns of Literacy and Their Social Context,' (pp. 974-997), in Veena Das (Ed.), *The Oxford India Companion of Sociology and Social Anthropology Vol II* (New Delhi: Oxford University Press, 2003).

[14] Dreze, Jean and Mrinalini Saran, 'Primary Education and Economic Development in China and India: Overview and Two Case Studies,' in K.P. Pattanaik Basu and K. Suzumura (Eds.), *Choice, Welfare, and Development: Essays in Honour of Amartya Sen* (Oxford: Clarendon Press, 1995).

[15] Dreze, Jean and Geeta Gandhi Kingdon, 'School Participation in Rural India,' *Review of Development Economics* 5(2001), 1-24.

[16] Duflo, Esther, 'Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa,' *World Bank Economic Review* 17(2003), 1-25.

[17] Filmer, Deon and Lant Pritchett, 'Educational Enrollment and Attainment in India: Household, Wealth, Gender, Village and State Effects,' *Journal of Educational Planning and Administration* 13(1999), 135-164.

[18] Fitzsimons, Emla, 'The Effects of Risk on Education in Indonesia,' *Economic Development and Cultural Change* 56(2007), 1-25.

[19] Government of India, Department of Elementary Education and Literacy, Department of Secondary and Higher Education, Ministry of Human Resource Development, *Annual Report 2005-06* (2006).

[20] Jacoby, Hanan G. and Emmanuel Skoufias, 'Testing Theories of Consumption Behavior Using Information on Aggregate Shocks: Income Seasonality and Rainfall in Rural India,' *American Journal of Agricultural Economics* 80(1998), 1-14.

[21] Jayachandran, Usha, 'Socio-Economic Determinants of School Attendance in India,' Delhi School of Economics Working Paper no. 103 (2002).

[22] Jayachandran, Seema, and Ilyana Kuziemko, 'Why Do Mothers Breastfeed Girls Less Than Boys? Evidence and Implications for Child Health in India,' *Quarterly Journal of Economics* 126(2011), 1485-1538.

[23] Jejeebhoy, Shireen J., 'Family Size, Outcomes for Children, and Gender Disparities,' *Economic and Political Weekly* 28(1993), 1811-1821.

[24] Jensen, Robert, 'Equal Treatment, Unequal Outcomes? Generating Gender Inequality through Fertility Behavior,' mimeo (John F. Kennedy School of Government, Harvard University, 2002).

[25] Kazianga, Harounan, 'Income Risk and Schooling Outcomes in Burkina Faso,' Working Paper (2006).

[26] Kingdon, Geeta Gandhi, 'The Quality and Efficiency of Private and Public Education: A Case-Study of Urban India,' *Oxford Bulletin of Economics and Statistics* 58(1996), 55-80.

[27] Kingdon, Geeta Gandhi, 'Where Has All the Bias Gone? Detecting Gender Bias in the Intrahousehold Allocation of Educational Expenditure,' *Economic Development and Cultural Change* 53(2005), 409-451.

[28] Kingdon, Geeta Gandhi, 'The Progress of School Education in India,' *Oxford Review of Economic Policy* 23(2007), 168-195.

[29] Kumar, Hemanshu, and Rohini Somanathan, 'Mapping Indian Districts Across Census Years, 1971-2001,' Centre for Development Economics, Delhi School of Economics Working Paper Series no. 176 (2009).

[30] Lancaster, Geoffrey, Pushkar Maitra, and Ranjan Ray, 'Household Expenditure Patterns and Gender Bias: Evidence from Selected Indian States,' *Oxford Development Studies* 36(2008), 133-157.

[31] Levine, David I., and Dean Yang, 'A Note on the Impact of Local Rainfall on Rice Output in Indonesian Districts,' Working Paper (2006).

[32] Maccini, Sharon, and Dean Yang, 'Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall,' *American Economic Review* 99(2009), 1006-1026.

[33] Nambissan, Geetha B., 'Equity in Education? The Schooling of Dalit Children in India,' *Economic and Political Weekly* 31(1996), 1011-1024.

[34] NFHS, *NFHS Supplemental Documentation: Data Set Version 52* (2010).

[35] Qian, Nancy, 'Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance,' *Quarterly Journal of Economics* 123(2008), 1251-1285.

[36] Ramachandran, Vimala, *Gender and Social Equity in Primary Education: Hierarchies of Access* (New Delhi: Sage Publications, 2004).

[37] Ribot, Jesse C., Adil Najam, and Gabrielle Watson, 'Climate Variation, Vulnerability and Sustainable Development in the Semi-Arid Tropics,' (pp.1-10), in Jesse C. Ribot, Antonio Rocha Magalhaes, and Stahis Panagides (Eds.), *Climate Variability, Climate Change and Social Vulnerability in the Semi-Arid Tropics* (Cambridge: Cambridge University Press, 1996).

[38] Rose, Elaina, 'Consumption Smoothing and Excess Female Mortality in Rural India,' *Review of Economics and Statistics* 81(1999), 41-49.

[39] Rose, Elaina, 'Gender Bias, Credit Constraints and Time Allocation in Rural India,' *The Economic Journal* 110(2000), 738-758.

[40] Seetharamu, A S and Usha Devi, *Education in Rural Areas - Constraints and Prospects* (New Delhi: A.P.H. Publishing Corporation, 2007).

[41] Sreenivasan, P.S., and J.R. Banerjee, 'The Influence of Rainfall on the Yield of Rain-Fed Rice at Karjat (Colaba District),' *Agricultural Meteorology* 11(1973), 285-292.

[42] Subramanian, Shankar, and Angus Deaton, 'Gender Effects in Indian Consumption Patterns,' *Sarvekshana* 14(1991), 1-12.

[43] Tilak, Jandhyala B.J., 'How Free is Free Primary Education in India?,' *Economic and Political Weekly* 31(1996), 275-282.

[44] Tilak, Jandhyala B.G., 'Determinants of Household Expenditure on Education in Rural India,' NCAER Working Paper no.88 (2002).

36

[45] Webster, Peter J., and Carlos Hoyos, 'Prediction of Monsoon Rainfall and River Discharge on 15-30-Day Time Scales,' *Bulletin of the American Meteorological Society*, 85(2004), 1745-1765.

[46] Zimmermann, Laura, 'Reconsidering Gender Bias In Intra-Household Allocation in India,' *Journal of Development Studies*, forthcoming.

Table 1: School Enrollment in Rural Areas by Age and Gender

	1992/1993		1998/1999		2005/2006	
	girls	boys	girls	boys	girls	boys
age						
6	0.47	0.57	0.71	0.76	0.51	0.53
7	0.57	0.71	0.78	0.86	0.72	0.75
8	0.56	0.76	0.78	0.85	0.77	0.82
9	0.61	0.78	0.80	0.89	0.85	0.89
10	0.56	0.76	0.72	0.85	0.79	0.87
11	0.58	0.81	0.74	0.87	0.84	0.91
12	0.50	0.73	0.63	0.79	0.74	0.84
13	0.46	0.74	0.60	0.78	0.70	0.81
14	0.39	0.68	0.52	0.72	0.66	0.74
6 - 10	0.55	0.72	0.76	0.84	0.73	0.77
11 - 14	0.48	0.74	0.62	0.79	0.74	0.82

Notes: Author's calculation using NFHS 1992/93, 1998/99 and 2005/06. Observations are weighted.

Table 2: Summary Statistics of Rainfall Shocks

	mean	std.dev	min	max
shock	-0.2396	0.9919	-3.0739	2.8243
lag 1	-0.0182	0.7901	-2.3098	2.1448
lag 2	0.0901	0.9483	-3.2978	3.7539
lead 1	-0.2054	0.8823	-2.8513	4.0243
lead 2	0.1059	1.1071	-2.7740	4.1786

distribution of shock variable by magnitude (percent of shocks):

shock<-2.5	2.1
-2.5 < shock < -1.5	7.7
-1.5 <shock < -0.5	27.4
-0.5 < shock < 0.5	40.8
0.5 < shock < 1.5	17.9
1.5 < shock < 2.5	3.7
shock > 2.5	0.5

N	158532

Table 3: Summary Statistics of Children by Age and Gender

	girl	boy	p-value
age 6	0.12	0.13	0.60
age 7	0.11	0.11	0.15
age 8	0.13	0.13	0.77
age 9	0.10	0.10	0.81
age 10	0.13	0.14	0.38
age 11	0.08	0.08	0.22
age 12	0.13	0.13	0.08
age 13	0.09	0.09	0.00
age 14	0.10	0.10	0.63
household size	7.78	7.57	0.00
age of household head	45.28	45.19	0.16
own house	0.95	0.95	0.68
own agricultural land	0.66	0.66	0.54
wealth index	0.38	0.35	0.12
N	76541	81965	

Note: Wealth index was created by principal component analysis of asset and living condition questions for rural households. Own house and own agricultural land are available only for children in the 98/99 survey. The p-value column reports the results from a test of equality of female and male characteristics.

Table 4: The Impact of Rainfall Shocks on School Enrollment in Rural Areas

	(1)	(2)	(3)	(4)	(5)	(6)
	girls	boys	girls	boys	girls	boys
	6 - 14	6 - 14	6 - 10	6 - 10	11 - 14	11 - 14
shock	0.0230***	0.0131**	0.0252***	0.0123*	0.0174**	0.0129*
	(0.0070)	(0.0058)	(0.0074)	(0.0064)	(0.0085)	(0.0068)
age dummies	Yes	Yes	Yes	Yes	Yes	Yes
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
p-value of gender equality test	0.047		0.025		0.549	
Observations	76228	81549	45700	48920	30528	32629
R-squared	0.191	0.1063	0.1917	0.1299	0.2076	0.1001
mean of dependent variable	0.6556	0.7949	0.6881	0.7985	0.6070	0.7895
percent increase	3.5	1.6	3.7	1.5	2.9	1.6

Robust standard errors for clustering at district level in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects here include month, year and district fixed effects

Table 5: Age Non-linearities in the Impact of Rainfall Shocks

	(1)	(2)
	girls	boys
	6 - 14	6 - 14
shock	0.0492***	0.0474***
	(0.0088)	(0.0077)
shock-age interactions		
shock*age 7	-0.0064	-0.0093
	(0.0072)	(0.0067)
shock*age 8	-0.0169***	-0.0317***
	(0.0059)	(0.0065)
shock*age 9	-0.0190**	-0.0309***
	(0.0073)	(0.0067)
shock*age 10	-0.0244***	-0.0432***
	(0.0066)	(0.0071)
shock*age 11	-0.0370***	-0.0443***
	(0.0080)	(0.0074)
shock*age 12	-0.0445***	-0.0476***
	(0.0077)	(0.0080)
shock*age 13	-0.0473***	-0.0588***
	(0.0082)	(0.0081)
shock*age 14	-0.0526***	-0.0552***
	(0.0088)	(0.0076)
age dummies	Yes	Yes
fixed effects	Yes	Yes
p-value of F test of joint significance	0	0
p-value of gender equality test		
shock	0.831	
shock at age 7	0.5453	
shock at age 8	0.0244	
shock at age 9	0.0726	
shock at age 10	0.0024	
shock at age 11	0.2825	
shock at age 12	0.5238	
shock at age 13	0.1303	
shock at age 14	0.6009	
Observations	76228	81549
R-squared	0.1923	0.1084

Robust standard errors for clustering at district level in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Fixed effects here include month, year and district fixed effects

Table 6: The Non-linear Impact of Rainfall Shocks on School Enrollment in Rural Areas

	(1)	(2)	(3)	(4)	(5)	(6)
	girls	boys	girls	boys	girls	boys
	6 - 14	6 - 14	6 - 10	6 - 10	11 - 14	11 - 14
shock<-2.5	0.0195	-0.0269	0.0390	-0.0101	-0.0033	-0.0396
	(0.0285)	(0.0467)	(0.0277)	(0.0457)	(0.0373)	(0.0633)
-2.5<shock<-1.5	-0.0623***	-0.0363**	-0.0603**	-0.0235	-0.0581*	-0.0542***
	(0.0237)	(0.0183)	(0.0248)	(0.0199)	(0.0297)	(0.0207)
-1.5<shock<-0.5	-0.0352**	-0.0207*	-0.0353**	-0.0160	-0.0327*	-0.0285**
	(0.0151)	(0.0118)	(0.0164)	(0.0138)	(0.0180)	(0.0128)
0.5<shock<1.5	0.0044	-0.0016	0.0109	0.0085	-0.0080	-0.0151
	(0.0175)	(0.0132)	(0.0185)	(0.0151)	(0.0216)	(0.0157)
1.5<shock<2.5	0.0423	0.0083	0.0644*	0.0289	0.0128	-0.0260
	(0.0295)	(0.0238)	(0.0350)	(0.0273)	(0.0361)	(0.0243)
shock>2.5	0.0516	0.0212	0.0974	0.0100	-0.0256	0.0230
	(0.0731)	(0.0554)	(0.0671)	(0.0680)	(0.0819)	(0.0469)
age dummies	Yes	Yes	Yes	Yes	Yes	Yes
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
p-value of F test of joint significance	0.0104	0.2992	0.003	0.545	0.355	0.0577
p-value of gender equality test						
shock<-2.5	0.227		0.2520		0.46	
-2.5<shock<-1.5	0.076		0.037		0.866	
-1.5<shock<-0.5	0.148		0.098		0.789	
0.5<shock<1.5	0.571		0.848		0.681	
1.5<shock<2.5	0.106		0.146		0.213	
shock>2.5	0.257		0.003		0.315	
Observations	76228	81549	45700	48920	30528	32629
R-squared	0.1913	0.1063	0.1922	0.1299	0.2078	0.1005

Robust standard errors for clustering at district level in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

Table 7: The Non-linear Impact of Rainfall Shocks and Age

age	shock<-2.5 girl	shock<-2.5 boy	-2.5<shock<-1.5 girl	-2.5<shock<-1.5 boy	-1.5<shock<-0.5 girl	-1.5<shock<-0.5 boy
6	-0.0139	-0.0922*	-0.0858***	-0.0950***	-0.0810***	-0.0720***
7	-0.0389	-0.1026**	-0.0975***	-0.0878***	-0.0556***	-0.0723***
8	0.0339	-0.0378	-0.101***	-0.0179	-0.0441**	-0.0104
9	0.0090	-0.0367	-0.0703**	-0.0371	-0.0241	-0.0182
10	-0.0197	-0.0118	-0.0675**	-0.0091	-0.0305*	-0.0019
11	-0.0006	0.0151	-0.0162	-0.0388*	-0.0254	0.0040
12	0.0646	0.0156	-0.0151	-0.0101	0.0021	0.0035
13	0.0650	-0.0007	-0.0290	0.0124	-0.0253	0.0094
14	0.12**	0.0695	-0.0419	-0.0313	-0.0241	-0.0166

age	0.5<shock<1.5 girl	0.5<shock<1.5 boy	1.5<shock<2.5 girl	1.5<shock<2.5 boy	shock>2.5 girl	shock>2.5 boy
6	0.0358*	0.0252	0.1088**	0.0651**	-0.0182	0.0095
7	0.0010	0.0038	0.1033***	0.0191	0.2008***	0.0826***
8	0.0071	0.0183	0.0552	0.0192	0.0708	0.0663
9	0.0098	0.0060	0.0907**	0.0169	0.1021	0.0760
10	0.0140	0.0011	0.0103	0.0114	0.0074	-0.0449
11	-0.0031	-0.0206	0.0457	0.0013	-0.0075	0.0891*
12	0.0022	0.0013	0.0384	-0.0174	0.0072	-0.0281
13	-0.0097	-0.0274	-0.0320	-0.0221	0.1003	0.0234
14	-0.0242	-0.0371**	-0.0368	-0.0259	0.0036	-0.0664

Robust standard errors for clustering at district level in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects here include month, year and district fixed effects

Table 8: The Impact of Lagged Rainfall Shocks on School Enrollment

	(1)	(2)	(3)	(4)	(5)	(6)
	girls	boys	girls	boys	girls	boys
	6 - 14	6 - 14	6 - 10	6 - 10	11 - 14	11 - 14
shock	0.0233***	0.0133**	0.0253***	0.0120*	0.0178**	0.0136**
	(0.0069)	(0.0058)	(0.0073)	(0.0065)	(0.0084)	(0.0068)
lag 1	0.0109	0.0064	0.0103	0.0043	0.0140	0.0104
	(0.0081)	(0.0063)	(0.0086)	(0.0072)	(0.0096)	(0.0070)
lag 2	-0.0021	-0.0017	-0.0011	0.0019	-0.0034	-0.0057
	(0.0072)	(0.0061)	(0.0080)	(0.0074)	(0.0084)	(0.0061)
age dummies	Yes	Yes	Yes	Yes	Yes	Yes
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
p-value of F test of joint significance	0.004	0.116	0.0037	0.26	0.0847	0.134
p-value of gender equality test						
shock	0.048		0.022		0.582	
lag 1	0.423		0.334		0.657	
lag 2	0.927		0.556		0.722	
Observations	76228	81549	45700	48920	30528	32629
R-squared	0.1911	0.1063	0.1918	0.13	0.2078	0.1003

Robust standard errors for clustering at district level in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects here include month, year and district fixed effects

Table 9: IV Results of the Impact of Rainfall Shocks on School Enrollment

	(1)	(2)	(3)	(4)	(5)	(6)
	girls	boys	girls	boys	girls	boys
	6 - 14	6 - 14	6 - 10	6 - 10	11 - 14	11 - 14
shock	0.0239***	0.0124**	0.0252***	0.0118*	0.0190**	0.0120*
	(0.0075)	(0.0062)	(0.0080)	(0.0071)	(0.0086)	(0.1004)
age dummies	Yes	Yes	Yes	Yes	Yes	Yes
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76043	81365	45583	48803	30460	32562
R-squared	0.1915	0.1066	0.1922	0.1303	0.2081	0.1004

Robust standard errors for clustering at district level in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects here include month, year and district fixed effects

shock was instrumented using rainfall shocks in the 5 closest districts as instruments

Table 10: Test for Selection into the Sample

	(1)	(2)
	girls	boys
	6 - 14	6 - 14
shock	0.1025	0.1432
	(0.3393)	(0.3455)
age dummies	Yes	Yes
fixed effects	Yes	Yes
Observations	8835	8961
R-squared	0.036	0.0368

Robust standard errors for clustering at district level in parenthesis

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects here include month, year and district fixed effects

Table 11: The Impact of Rainfall Shocks on School Enrollment in Urban Areas

|	(1)	(2)	(3)	(4)	(5)	(6)
	girls	boys	girls	boys	girls	boys
	6 - 14	6 - 14	6 - 10	6 - 10	11 - 14	11 - 14
shock	-0.0046	-0.0545**	0.0190	-0.0479*	-0.0340	-0.0637
	(0.0144)	(0.0266)	(0.0231)	(0.0253)	(0.0364)	(0.0397)
age dummies	Yes	Yes	Yes	Yes	Yes	Yes
fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
p-value of gender equality test	0.072		0.002		0.659	
p-value of equality with rural	0.093	0.013	0.795	0.02	0.172	0.057
Observations	9095	9841	5106	5574	3989	4267
R-squared	0.0595	0.0528	0.0561	0.0622	0.0735	0.0545

Robust standard errors for clustering at district level in parenthesis

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Fixed effects here include month, year and district fixed effects

Figure 1: The Relationship between Rainfall and Rice Output in Indian Districts 1999-2007
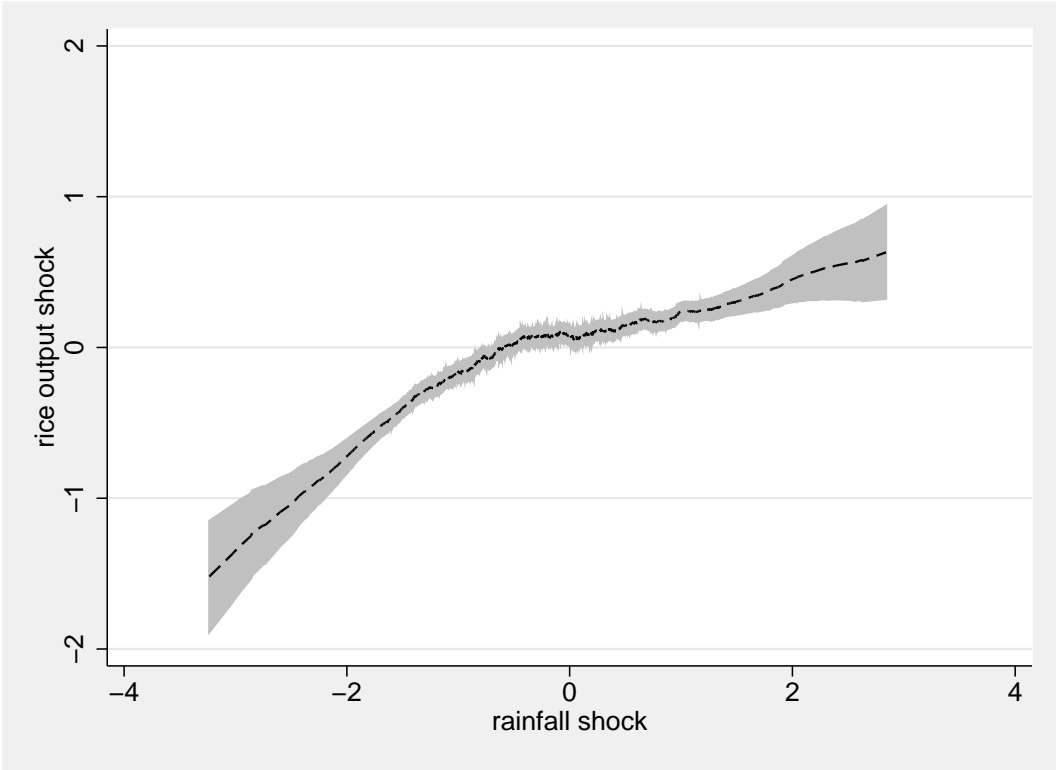
Note: district-level output data comes from the Indian Department of Agriculture and Cooperation; rainfall and output shocks have been standardized, see the text for details; graph shows the non-parametrically estimated relationship between rainfall and output shocks and the corresponding point-wise confidence interval

Figure 2: The Relationship between Rainfall and Wheat Output in Indian Districts 1999-2007
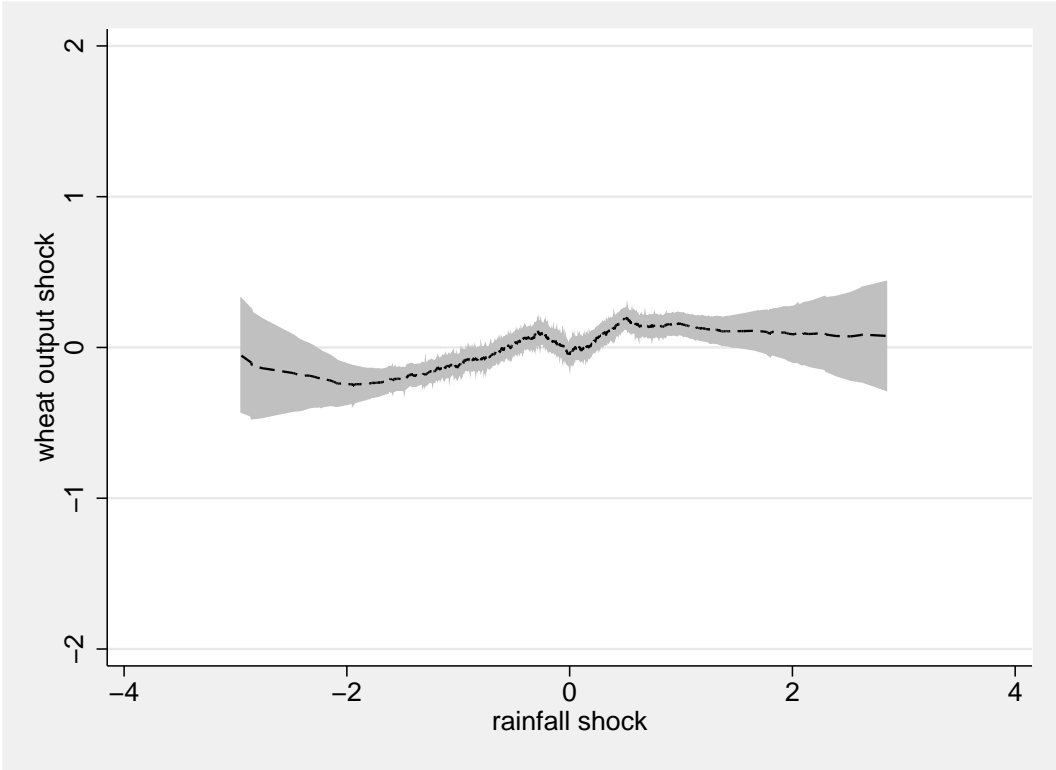
Note: district-level output data comes from the Indian Department of Agriculture and Cooperation; rainfall and output shocks have been standardized, see the text for details; graph shows the non-parametrically estimated relationship between rainfall and output shocks and the corresponding point-wise confidence interval

Appendix Table A1: First Stage IV Results

	(1)	(2)
	girls	boys
	6-14	6-14
IV1	0.5611***	0.5689***
	(0.0700)	(0.0701)
IV2	0.2323***	0.2345***
	(0.0681)	(0.0664)
IV3	0.0586	0.0506
	(0.0537)	(0.0535)
IV4	0.0885	0.0916
	(0.0579)	(0.0567)
IV5	0.0332	0.0308
	(0.0351)	(0.0360)
age dummies	Yes	Yes
fixed effects	Yes	Yes
Observations	76043	81365
F statistic	393.4	386.3
R-squared	0.9767	0.9763

Robust standard errors for clustering at district level in parenthesis

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Fixed effects here include month, year and district fixed effects