Journal of Development Economics 105 (2013) 237–253

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Income shocks and gender gaps in education: Evidence from Uganda☆☆

Martina Björkman-Nyqvist *

Stockholm School of Economics, Sweden
CEPR, United Kingdom
J-PAL, Europe

ARTICLE INFO

Article history:
Received 22 November 2012
Accepted 30 July 2013

JEL classification:
D1
J16
J24
D13

Keywords:
Rainfall
Education
Test scores
Gender

ABSTRACT

This paper uses exogenous variation in rainfall across districts in Uganda to estimate the causal effects of household income shocks on children's enrollment and academic performance conditional on gender. I find negative deviations in rainfall from the long-term mean to have negative and highly significant effects on female enrollment in primary schools and the effect grows stronger for older girls. I find no effect of rainfall variation on the enrollment of boys and young girls. Moreover, I find that when schooling is free of charge and both marginal boys and girls are enrolled, a negative income shock has an adverse effect on the test scores of female students while boys are not affected. The results imply that households respond to income shocks by varying the amount of schooling and resources provided to girls while boys are to a large extent sheltered — a finding consistent with a model where parents' values of child labor differ across sexes.

1. Introduction

The question of how changes in households' economic conditions differentially affect the treatment of boys and girls in developing countries has long been a concern among development economists and policymakers. Understanding households' decisions regarding their children's education and food consumption conditional on gender in a risky environment is important in order to design sustainable policies to promote gender equality.[1] The importance of this issue has been reemphasized in the last decade since promoting gender equality has been identified as one of the most important goals of the donor community.[2]

Starting with Becker (1981), economists have long argued that households' differential treatment of children conditional on gender can be explained by the underlying economic conditions. For example, Rosenzweig and Schultz (1982) show that in India, households selectively allocate resources to children in response to variations in sex differences in their expected earnings opportunities as adults. Foster (1995) finds that a child's well-being varies with fluctuations in income and prices and that the well-being of girls is more sensitive to these fluctuations than that of boys. Similarly, Behrman (1988) has shown that girls' nutrition suffers more than that of boys in the lean, as opposed to the peak, agricultural season. Differential treatment of boys and girls with regard to intra-household food allocations and long-term consequence on female infanticide and gender imbalance has been

☆☆ I am grateful to the editor and two anonymous referees for their helpful comments. I thank Jakob Svensson for guidance and support; Esther Duflo, Eliana La Ferrara, Andreas Madestam, Edward Miguel, Rohini Pande, and Torsten Persson for their suggestions; participants at the BREAD summer school, the IIES seminar, Bocconi University, University of Copenhagen, Stockholm School of Economics, LSE, and Center for Global Development for constructive comments. I wish to thank the Uganda National Examination Board, the Ministry of Education and Sports, the Uganda Statistical Bureau, the Meterological Department in Uganda, and the Goddard Earth Sciences (GES) Data and Information Services Center (DISC) at NASA for sharing their data with me. I acknowledge financial support from Jan Wallander's and Tom Hedelius' Research Foundation. All errors are my own.

* Tel.: +46 8 7369285.
E-mail address: martina.bjorkman.nyqvist@hhs.se

[1] There are several rationales for gender equality. First, equity is valuable in and of itself and equal participation by both genders at all levels of decision-making is a basic human right. Second, women play a fundamental role in development and gender equality is promoted in order to increase efficiency. For a detailed discussion on gender equality and development, see Duflo (2012).

[2] The Millennium Development Goal.

0304-3878
http://dx.doi.org/10.1016/j.jdeveco.2013.07.013

documented as an evident phenomenon in Asia. In India, Rose (1999) examines the relationship between consumption smoothing and excess female mortality. She finds favorable rainfall shocks in childhood to increase the ratio of the probability that a girl survives to the probability that a boy survives. For China, Qian (2008) finds that increasing total household income has no effect on sex ratios; increasing female income increases the survival rates for girls and the educational attainment for all children whereas increasing male income decreases the survival rates and the educational attainment for girls but has no effect on boys.

One of the more striking (and visible) examples of differential treatment of boys and girls within households in developing countries is the prevalent gender bias in education. In general, girls tend to receive less schooling than boys, particularly so in rural areas, low income countries and South Asia (Alderman et al., 1996; Behrman and Knowles, 1999). The possible causes of this gender gap in schooling have been subject to less study and, similarly, the nature of the relationship between changes in households' economic conditions and differential treatment in children's education is anything but settled.

The main empirical challenge in establishing the link between households' economic conditions and differential investments in boys' and girls' schooling is that economic conditions and intra-household allocations are endogenous to schooling and family structure. The estimated effect of household income on children's developmental outcomes might be spurious, because parental income and outcomes for children may both be driven by an unmeasured factor. Randomized experiments constitute one solution to this omitted-variables problem. In the absence of evidence from such experiments, however, it is necessary to rely on exogenous natural variations in combination with statistical modeling strategies.

This paper develops and implements a strategy capturing the causal effect of changes in households' economic conditions on differential investment in children's primary education. In particular, I exploit the exogenous variation in district income in Uganda over time caused by rainfall shocks (measured as the natural log of annual rainfall minus the natural log of long-term mean annual rainfall in the given district) to study the causal effects of household income shocks on boys' and girls' primary education attainment and achievements. Uganda is an agricultural country where more than 80% of the work force are employed in the mainly rainfed agricultural sector. Agricultural practices depend on natural weather patterns and variations in rainfall levels result in a large variation in total agricultural output and farm incomes.[3] Therefore, rainfall shocks are plausible proxies for income shocks to households in Uganda.[4]

Previous studies on differential treatment in education have primarily been conducted with data from Asia, while this paper attempts at explaining the phenomenon in an African setting.[5] This is also the first paper to study how unexpected changes in households' economic conditions affect students' academic performance and whether there are any differential effects depending on gender.

In this paper, I use district level data covering the full census of children in all primary schools in Uganda for a period of 24 years. Instead of looking at idiosyncratic events (i.e. those that are experienced by the household alone), I use aggregate risk, as reflected by district level rainfall. This might be more critical for rural households, since it is more difficult to insure away the aggregate risk through formal or informal mechanisms.[6] It is also important to understand, across a large number of cohorts and in a representative sample of a national population, whether an unpredictable variation in everyday environmental conditions affects investment in children's education and if it differs across gender. Among other things, these findings are important for guiding policies that aim at helping households cope with the year-to-year variation in economic conditions, as opposed to policies that respond to extreme events.

When a household experiences a rainfall shock, there are two potential effects that could differentially affect investment in boys' and girls' schooling. First, a transitory shock affects crop yields and hence, household income and food availability, and this will force the household to reduce its current consumption. If the reduction in food consumption and/or school expenditures has different effects on boys and girls, then girls' enrollment and ability to perform in school might be affected. Second, in periods of transitory shocks, households are forced to look for alternative income generating activities and food and therefore, the demand for children's participation in home production increases. If the domestic work load is differentially allocated across boys and girls, this will affect both educational attainment and the achievement of girls relative to that of boys. In the paper, I develop a simple human capital model where parents view children's education as a form of investment to illustrate these effects. I derive the equilibrium level of schooling (enrollment) and the cognitive skills of boys and girls and evaluate how enrollment and cognitive skills vary with income.

The empirical findings are broadly consistent with the model. I find negative deviations in rainfall from the long-term mean to have an immediate and negative effect on female enrollment in primary schools and the effect is strongest for older girls. Young girls are not affected by rainfall shocks. A decrease in rainfall (relative to average local rainfall) by 15 percent results in 118 fewer female students in grade 7, which corresponds to a decrease of 5 percentage points in female enrollment. Conversely, I do not find any relationship between rainfall shocks and male enrollment. A decrease in rainfall by 15 percent only decreases male enrollment in Primary 7 by 1 percentage point and this effect is insignificant at standard levels. I also find that when schooling is free of charge and both marginal boys and girls are enrolled, a negative income shock has an adverse effect on the test scores of female students while I do not find any effect on boys' academic performance. The results imply that households respond to income shocks by varying the enrollment and resources provided to older girls, while boys are to a large extent sheltered. Moreover, the finding that older girls who have a comparative advantage in home production are affected by rainfall shocks, and not boys or younger girls, suggests that the driving mechanism is the differential benefit from child labor in home production. Specifically, while there is suggestive evidence that households in poor countries respond to transitory income shocks by increasing child labor, I find that it is primarily older girls' labor that is used as a buffer.[7]

In the robustness section, I use a household dataset to examine whether there is aggregation bias in the district level data. The analysis indicates that findings from the district level data are consistent with findings at the household level. Older girls of primary school age living in households that experience negative deviations in rainfall from the long-term mean are less likely to attend school compared to girls living in households with average rainfall. The matching result from the household data suggests that the findings from the district level analysis are not subject to aggregation bias.

I also exploit a natural policy experiment – the removal of school fees in primary education – to estimate the effects of a reduction in the (formal) cost of schooling on the enrollment and academic performance of boys versus that of girls. While suggestive, the evidence suggests that the removal of school fees has a large and positive effect on the enrollment of both boys and girls, although it is stronger for girls. Moreover, after the abolishment of user fees in primary schools, a negative income shock has an even larger negative effect on female enrollment, while boys still remain unaffected.

The remainder of the paper is organized as follows. Section 2 presents the conceptual framework and Section 3 describes the data. The

[3] See Asiimwe and Mpuga (2007) for more information on the implications of rainfall shocks for household income in Uganda.

[4] Other studies that have used rainfall as an instrument for income in developing countries are i.e. Miguel, 2005; Miguel et al., 2004; Paxson, 1992; Rose, 1999.

[5] Examples of previous studies are Alderman et al., 1996; Behrman and Knowles, 1999; Jacoby and Skoufias, 1997; Qian, 2008.

[6] Townsend, 1994.

[7] I.e. Beegle et al. (2003) use data from Tanzania and find that households respond to transitory income shocks by increasing child labor.

identification strategy is discussed in Section 4. Section 5 presents the results for enrollment and test scores and Section 6 includes a robustness analysis. Additional results are presented in Section 7 and Section 8 concludes the paper.

2. Conceptual framework

This section presents a simple human capital model with intra-household choices on the quality and quantity of children's education. I use this framework to show how children's abilities in domestic work, learning efficiency, expected returns to education as well as parental income affect the household's investment in children's primary education.

When modeling intra-household decisions, the bargaining model is preferred, since it accommodates the conflict of preferences and asymmetric power relations within the household and further, women's relative well-being depends on the relative bargaining power of the spouses. However, this choice of model might not be obvious in a Sub-Saharan African setting where women are responsible for most of the agricultural production and domestic chores but have little to no decision-making power within the household and where the male head of household (i.e. husband, father or brother) determines the decisions.[8] The perception that women are less economically valuable is intensified by institutional constraints and a legal setting that limits women's rights to own property and access productive resources. In this type of setting (i.e. a Ugandan setting) where cultural, social, and legal institutions basically give all bargaining power to the man, one could simply look at the outcomes from a male head of household maximizing problem when modeling the household's decisions.

2.1. Basics

Assume that parents make decisions for their children. Each family i does, for simplicity, have two children — a boy (denoted with subscript b) and a girl (denoted with subscript g). There is a continuum of families, $i \in [0,1]$, that could potentially send their children to school. There are two periods. In period 1, the child works at home, goes to school, or both. In the second period, the child is an adult and works for a wage. In period 1, the parents derive a direct benefit from the child's domestic work, while in period 2 the parents benefit from transfers from their child (now an adult).

The parents' utility is

$$U_i = u\left(c_1^i\right) + \delta c_2^i\ , \tag{1}$$

where c_t^i is ith parents' consumption in period t, δ is a discount factor and u is a function, with $u' > 0$, $u'' < 0$, and $u''' \geq 0$.

Cognitive skills, a, are acquired according to

$$a_s^i = \alpha_s^i s_s^i\ , \tag{2}$$

where α_s^i is the learning efficiency of a child of sex s in family i (which depends on many factors, such as innate ability, child motivation, parental motivation etc.). For simplicity, I assume an equal learning efficiency between boys and girls in family i, $\alpha_b^i = \alpha_g^i = \alpha^i$, and that α^i is distributed according to $f(\alpha^i)$ over the unit interval. s_s^i is the fraction of time in period 1 spent in school by a child from family i of sex s, where $s_s^i \in [0,1]$. Alternatively, s could be interpreted as the division of resources (food) within the family, i.e. between children of different sexes.

Parents' consumption in each time period is given by

$$c_1^i = y_1 - pe_b^i - pe_g^i + \eta_b\left(1 - s_b^i\right) + \eta_g\left(1 - s_g^i\right), \tag{3}$$

and

$$c_2^i = y_2 + \gamma_b y_b^{ai} + \gamma_g y_g^{ai}\ , \tag{4}$$

where y_t is parental income (exogenous), p is the price of schooling for one child, e_s^i is a dummy variable taking the value of 1 if family i sends the child of sex s to school, $\eta_s(1 - s_s^i)$ is the income generated from home production by the child in period 1, y_s^{ai} is the child's income when working as an adult in period 2, and $\gamma_s y_s^{ai}$ is the share of the child's income transferred to his/her parents.

Eq. (5) completes the model, relating child cognitive skills to the child's income as an adult

$$y_s^{ai} = \omega_s a_s^i, \tag{5}$$

where ω_s is the return to education of a child of sex s. In this simple model, parents are liquidity constrained and cannot borrow or save. The only way of shifting income between periods is to alter the investment in children's education.[9]

In a Ugandan setting, it is reasonable to assume that girls' labor in home production is viewed as more valuable than that of boys. Poor households use their children's labor and in particular when parents are working (informally or formally) outside the household and girls often have to assist in the care of younger siblings and domestic chores such as food preparation, fetching water, collecting firewood, washing clothes, and taking care of the sick and old.[10] Although biological conditions might suggest sons to be stronger and more able to perform certain household tasks, daughters are often viewed as better substitutes for mothers and more productive than sons in taking care of siblings and other home production tasks such as cleaning and cooking. Normalizing η_b, I thus assume that $\eta_g > \eta_b \equiv 1$. Lacking pension programs and saving options, families in Uganda rely on their children as a source of income at old age. In a society of patrilocal exogamy, it is reasonable to assume that households perceive that a boy's contribution to the household as an adult will be larger than a girl's transfer, since the girl will marry and leave the natal household while the son will remain in the family, implying that $\gamma_b > \gamma_g$. Finally, while the limited evidence on the social returns to education suggests that $\omega_g \geq \omega_b$, it is plausible to assume that the expected private return (viewed by the parents or the male head of household) is $\omega_b > \omega_g$.[11]

To simplify the notation, without loss of generality, I assume that $p = 1$ and normalize θ_b to 1, where $\theta_s \equiv \delta\gamma_s\omega_s$. Given the above assumptions, this implies that $\theta_g < \theta_b \equiv 1$.

2.2. Private optimum and equilibrium outcomes

The optimal choice of children's education can be found by maximizing the parents' expected utility, Eq. (1), subject to the budget constraints (3)–(4). The first-order condition of household i for a child of sex s is

$$-u'(c_1)\eta_s + \alpha_s^i\theta_s^i \leq 0 \ \text{ for } \ s_s \in [0,1], \tag{6}$$

Thus, for a given ability of the child, α^i, parents will choose to invest in education up to the point where the marginal cost of more schooling, taking the form of reduced time for domestic production, is equal to the marginal gain, taking the form of increased transfer from a more educated and hence higher paid child as an adult. The properties from this simple model are summarized below.

Proposition 1. *a) It is always optimal for parents to invest in more (or at the minimum as much) education of the boy as compared to the girl, $s_b \geq s_g$; b) A girl will attend school iff the boy is sent to school full-time; c) a reduction in the boy's education is only optimal iff the girl works full-*

[8] African Development Fund (2005).

[9] Introducing savings and borrowing would reduce parents' incentives to invest in education but would not eliminate them. Specifically, if all investments are assumed to be risky, parents will diversify their investments along several different alternatives, including children's education (Glewwe, 2002).

[10] The World Bank, 1993.

[11] See Summers (1994) and the discussion in Section 6.

time in domestic work; d) If both $s_b > 0$ and $s_g > 0$, a reduction in parental income, y_1, will on the margin only reduce investment in the girl's education; e) a reduction in parental income, y_1, primarily affects the girl's cognitive skills, a_g.

Proof. See the Appendix A.

These results are intuitive. From the first-order condition, it follows that for a given α, the marginal cost of schooling is lower and the marginal return is higher for boys as compared to girls. Thus, parents have incentives to increase s_b as far as possible, i.e. $s_b = 1$, before investing in the girls' education. For the same reasons, a change in parental income, y, will affect s_g as long as $s_g > 0$ and thus, it will on the margin primarily affect girls' amount of schooling and academic performance.[12]

Given the assumption that parents differ in the innate ability of their children, α^i, we can derive the equilibrium number of boys and girls in school and their average cognitive skills ($\alpha^i s^i$) for a given (average) income y_1. Formally, the shares of boys (BiS) and girls (GiS) in school are

$$BiS = 1 - F(\alpha_1(y)) \tag{7}$$

and

$$GiS = 1 - F(\alpha_3(y)), \tag{8}$$

where $\alpha_1 = u'(y + \eta_g)$ and $\alpha_3 = u'\left(y + \eta_g \text{-} 2\right)^{\frac{\eta_g}{\theta_g}}$ are the threshold values of α for boys and girls, for a given y, and when $s_s = 0$. Note that $\alpha_1 < \alpha_3$.

Proposition 2. *If $f(\alpha)$ is symmetric and unimodal and at least half the population of girls is enrolled in school, then an income shock will have a larger effect on the enrollment of girls than on that of boys.*[13]

Proof. Differentiating BIS and GIS with respect to y gives:

$$\frac{dBiS}{dy} = -f(\alpha_1(y))\alpha_1'(y) < \frac{dGiS}{dy} = -f(\alpha_3(y))\alpha_3'(y).$$

If $f(\alpha)$ is symmetric and unimodal, then $f(\alpha)$ is maximized at $\alpha = \frac{1}{2}$ and $f'(\alpha) > 0$ for $\alpha < \frac{1}{2}$. Since $\alpha_1 < \alpha_3$, it follows that $f(\alpha_3(y)) > f(\alpha_1(y))$ for all $\alpha_3 \le \frac{1}{2}$. Moreover, $-\alpha'_1(y) = -u''\left(y + \eta_g\right) < -u''\left(y + \eta_g - 2\right)^{\frac{\eta_g}{\theta_g}} = -\alpha_3'(y)$ since $u''' > 0$, $\eta_g > 1$ and $\theta_g < 1$.[14]

Intuitively, if the marginal cost is higher and the marginal return is lower for girls' schooling, parents will, if possible, adjust the amount of schooling and the resources provided to girls in response to an income shock. These individual effects will also guide the aggregate outcome, provided that most households send both their children to school.

In the empirical part, I test the predictions in Proposition 2 by using rainfall shocks as a proxy for income shocks to empirically evaluate whether an income shock has a larger effect on girls' schooling than that for boys. The theoretical model is explored to say something about the mechanism underlying the results. The model depicts three possible mechanisms: child labor in home production, transfers from children (as adults), and returns to education. The last two channels, transfers from children as adults and returns to education, are independent of the child's age. However, the child labor mechanism which assumes that girls have a comparative advantage in home production should only be true for older girls and hence, it should not affect younger girls who are still too immature to take on responsibility at home. If the sex-specific results in the empirical part are also age-specific, so that only older girls are affected by rainfall shocks, this would suggest that the driving mechanism is the parents' value of older girls' labor in home production.

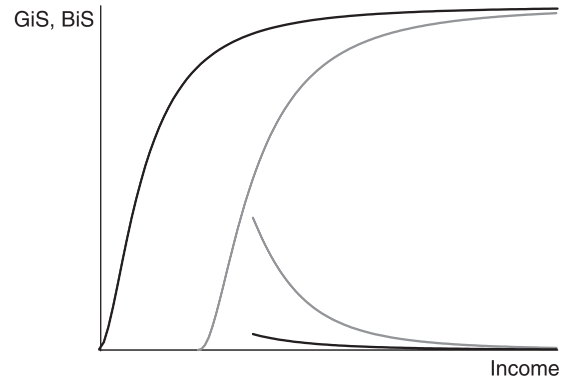

Fig. 1. Enrollment of girls (in gray) and boys (in black) conditional on income is illustrated by the concave curves and the derivatives of children's enrollment with respect to income shown by the convex curves.

Fig. 1 illustrates the property of Propositions 1 and 2 for the case where α^i is distributed according to a beta distribution over the interval [0,1].[15]

The two concave curves show the enrollment of girls (gray line) and boys (black line) and the convex lines in the lower part of the figure depict the derivatives of the enrollment of girls and boys with respect to y.

I can also do comparative statics on changes in the formal cost of schooling, p, as well as study whether the effect of income shocks on boys' and girls' enrollment is different in poor and wealthy districts. The results are rather complicated and discussed in Section 7 where I empirically estimate conditional effects.

The average ability of boys enrolled in school is

$$a_b = BiS^*(y)\int_{\alpha_2}^{1} \alpha f(\alpha|\alpha \ge \alpha_2)d\alpha + BiS^{**}(y)\int_{\alpha_1}^{\alpha_2} \alpha S_b(\alpha, y) f(\alpha|\alpha_1 < \alpha < \alpha_2)d\alpha \tag{9}$$

and for girls

$$a_g = GiS^*(y)\int_{\alpha_4}^{1} \alpha f(\alpha|\alpha \ge \alpha_4)d\alpha + GiS^{**}(y)\int_{\alpha_3}^{\alpha_4} \alpha S_g(\alpha, y) f(\alpha|\alpha_3 < \alpha < \alpha_4)d\alpha, \tag{10}$$

where $BiS^*(y)$ [$GiS^*(y)$] is the share of enrolled male [female] students that are in school full time ($s_s = 1$) and $BiS^{**}(y)$ [$GiS^{**}(y)$] is the share of enrolled male [female] students that are in school less than full time ($0 < s_s < 1$). The first part of Eq. (9) [Eq. (10)] is the average ability of boys [girls] who are sent full-time to school, weighted by the size of this group and the second part of the equation is the average ability of boys [girls] who are sent part-time to school weighted by the size of the group. The threshold values for α, for a given y and when $s_s = 1$, are $\alpha_2(y) = u'(y + \eta_g - 1)$ and $\alpha_4(y) = u'(y - 2)^{\frac{\eta_g}{\theta_g}}$. $s_s = S_s(\alpha, y)$ is the fraction of time spent in school for children not enrolled full time and s_s is a function of α and y and defined from the first-order condition, Eq. (6).

[12] Given the assumption that η_s and θ_s differ across boys and girls, assuming decreasing returns to education by using a concave cognitive ability function does not render the predictions in the model.

[13] The share of girls enrolled in primary school in Uganda during the period 1989–2002 was higher than 50% (The World Bank, 2002).

[14] Proposition 2 states the sufficient conditions for $dBiS/dy < dGiS/dy$. Assuming a specific distribution, it can be shown that the results typically hold more generally. For example, if $f(\alpha)$ is uniform, then $dBiS/dy < dGiS/dy$ independent of the population of girls enrolled in school, provided that some girls are indeed in schools. If $f(\alpha)$ is a beta distribution, the results will hold for $\alpha_3 > \frac{1}{2}$ (i.e. when less than half the population of girls is in school), provided that η_g is large.

[15] In Fig. 2, I assume $u(c) = ln(c)$, $\eta_g = 1.1$, and $\theta_g = 0.9$.

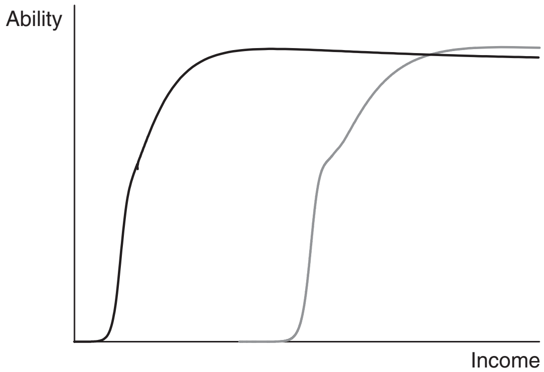

Fig. 2. Average ability of girls (gray line) and boys (black line) conditional on income.

Proposition 3. *a) The average innate ability of girls in school, α_g, will be higher than that of boys, α_b, although the girls' average cognitive skills, a_g, may not be higher. b) a negative income shock has two effects on average cognitive skills: (i) more marginal students drop out which will raise the average ability of the remaining students and (ii) less resources will be provided to the child, or less time will be spent in school, which will reduce the average ability. Which of the two effects that dominates is ambiguous.*

How does an income shock affect average ability? As is evident from Eqs. (9) and (10), even in this simple model, the total impact depends on a number of factors and it is difficult to derive closed form solutions. However, numerical simulations suggest that the resource effect dominates for low y, while the selection effect dominates for high y. Moreover, as long as a sufficient number of girls are in school, the effect is larger for girls than for boys.

Fig. 2 depicts one of these numerical simulations.[16] In this particular example, for sufficiently low income levels, there are no girls (gray line) enrolled in school. The average ability of boys (black line) is always higher than the average ability of girls up to some high level of income. For relevant parameter values (i.e. when more than 20% and less than 80% of the girls are enrolled in schools) $\frac{da_g}{dy} > \frac{da_b}{dy}$, an income shock will reduce the average ability of girls more than the average ability of boys.

In the empirical part, I am testing these predictions by using primary school test scores as a proxy for cognitive skills. I estimate the effect of an income shock (using rainfall shock as a proxy) on girls' and boys' performance on the standardized test in primary school. If the empirical results show a larger effect for girls than for boys, the theoretical model suggests that the resource effect dominates.

3. Data

The main analysis in this paper uses and merges data from three different administrative records. Enrollment data, by grade and gender, for the period 1979 to 2003 and for each of Uganda's districts, were collected from the Education Management Information System (EMIS) of the Uganda Ministry of Education. The underlying information is collected at the beginning of each year from all primary schools by the Ministry of Education and is then aggregated to the district level.

Test score data for the years 1989 to 2002 were collected from the Uganda National Examination Bureau. Obviously, a useful measure of students' academic performance must be comparable between all schools and districts. Primary Leaving Exams (PLE) fulfill this requirement as these standardized tests in Math, English, Science, and Social Studies are conducted yearly on all grade 7 students in Uganda.[17] Each of the subjects is graded from 0 to 8 and I use the sum of the four tests as the primary test score measure. There are pros and cons with using PLE as a measure of academic performance. One advantage is that almost all students in the last grade of primary school (grade 7) take the test. Passing the test is a requirement for acceptance into secondary school, so students have strong incentives to do their very best on the test. A possible concern is that average PLE scores across districts may not display a great deal of variation.

Rainfall data, used as a proxy for household income at the district level, have been collected from the Meteorological Department at the Ministry of Water, Land and Environment. The data include monthly records for each station from 1951 through 2003, although most stations had shorter time-series of data from 1975; some stations also lack data for some months per year. The rainfall measure was used to construct a basic weather variable representing total rainfall (in millimeters) by district d and year t, R_{dt}.[18] To construct a transitory rainfall variable, it should also be known how current rainfall deviates from its expected value, $\overline{R_{dt}}$. If rainfall were serially correlated across years, one would have to forecast the expected value of rainfall for each region in each year. However, rainfall does not appear to be serially correlated: I am unable to reject the hypothesis that rainfall follows a white-noise process.[19] Thus, I can set $\overline{R_{dt}} = \overline{R_d}$, historical rainfall over time in district d. The transitory rainfall variable is given by the deviation of yearly rainfall from the historical mean of rainfall for the district. Specifically, the variable is the natural log of yearly rainfall minus the natural log of long-term mean annual rainfall in the given district, i.e. $\ln R_{dt-1} - \ln \overline{R_d}$, where the long-term mean annual rainfall for a particular district is calculated over the years 1975–2003. The log rainfall variable simplifies the interpretation and should be interpreted (roughly speaking) as the percentage deviation from mean rainfall (e.g. a value of −0.05 means that rainfall was approximately 5 percent less than normal).[20]

A negative rainfall shock refers to rainfall less than the long-term mean for the district and a positive rainfall shock refers to rainfall above the mean in the district. According to the regressions on rainfall shocks and crop production in Table 2, a negative rainfall shock is also equivalent to a bad shock to crop production and a positive rainfall shock is equivalent to a beneficial shock to agricultural production. Besides this core variable, I also experiment with several other rainfall shock measures.[21] Fig. 3 depicts scatterplots of yearly rainfall in a subset of districts from different regions and rainfall zones in Uganda. The plots illustrate the large variation in yearly rainfall within districts and that rainfall data do not appear to be auto-correlated.

The schooling data used in the main analysis are the full census of all primary school children in Uganda aggregated at the district level. The benefit of using these data is that it is not subject to sampling error as a representative sample of micro data might be. By using the aggregate data for an extensive period of 24 years, I am able to establish the effect of a rainfall shock on the enrollment of a large number of cohorts of children in primary school in Uganda. Moreover, by focusing on aggregate sources of risk, such as district level rainfall shocks, I am studying risks that might be more critical to rural households since it is more difficult to insure away aggregate risk than idiosyncratic risks.[22] Findings from the district level analysis could be used by policy makers to understand larger scale effects of income shocks on educational achievements,

[16] I simulate ability using the same parameter values and distribution function as in Fig. 1.

[17] Note that the test score (unlike enrollment) data are available at the individual level.

[18] For further information on how the basic rainfall variable, R_{dt}, was constructed, see the Appendix A.

[19] When testing for serial correlation across years by district, I was unable to reject the hypothesis that rainfall follows a white-noise process in all but 10% of the districts. Excluding these districts in my sample does not quantitatively change the result. Moreover, when computing Newey–West robust standard errors, allowing for serial correlation with up to 10 lags, the result does not change.

[20] The results are robust to using the level of rainfall but I choose to use the log version because of the simpler interpretation. The log variable is also used in other papers exploring rainfall shocks (e.g. Maccini and Yang, 2009).

[21] See section 6 on robustness analysis.

[22] Townsend (1994).

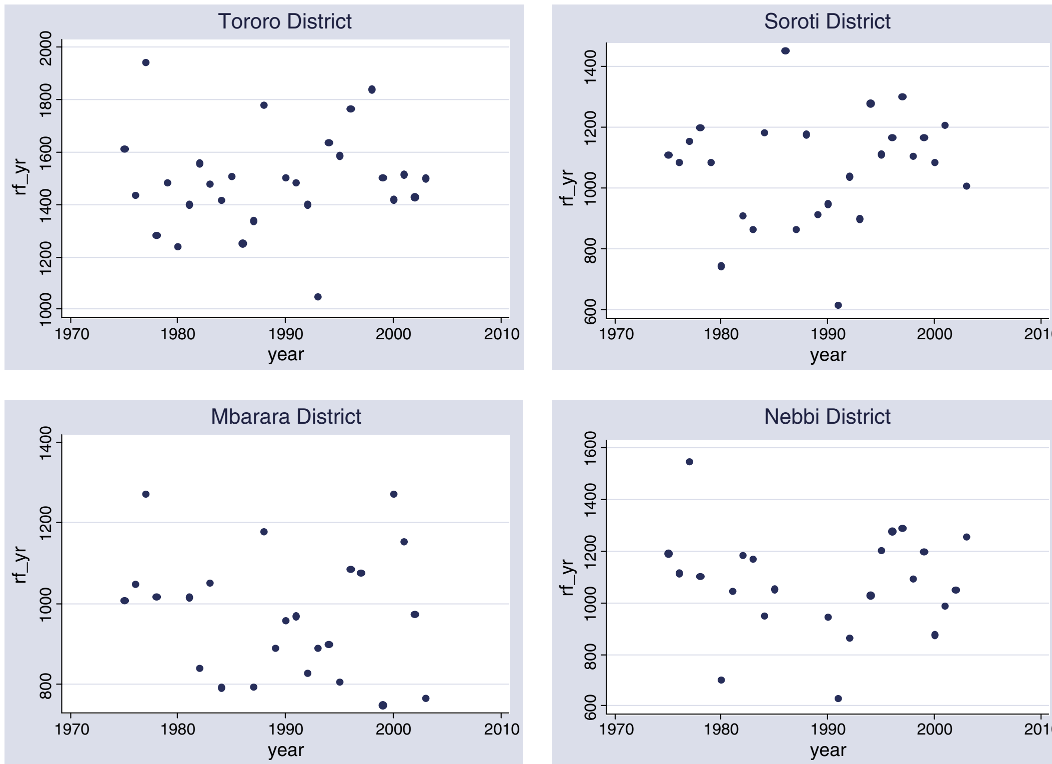

Fig. 3. Scatterplots of average yearly rainfall for some sample districts.

conditional on gender across districts. The downside of using district level data is the possible aggregation bias. Therefore, I have collected household data for the three years available in order to study if the district level results are consistent with findings from micro data. I use a pooled cross section of household data from the 1999/2000, 2002/03, and 2005/06 Uganda National Household Surveys. Random samples were drawn during each survey round and hence, pooling the resulting random samples provides an independent pooled cross section.[23] The benefit of using a pooled cross section is that it helps achieve more precise estimators and test statistics with more power. As a source of exogenous weather variation for the households, I use the TRMM Multi-Satellite Precipitation Analysis (TMPA) database of monthly rainfall estimates for the period 1998–2006.[24] The resolution of the TMPA precipitation data is at a 0.25 latitude and longitude degree interval. I match the GPS coordinates of the villages in the household survey with the 0.25 gridbox for rainfall to get a good estimate of village precipitation.[25] In total, the TMPA database has 690 gridpoints covering Uganda and I match these with the household dataset. The final sample includes 228 localities across all regions in Uganda for which I have matched household and rainfall data. In the household analysis, I use the same definition of a rainfall shock as in the district level analysis. I.e. the transitory rainfall variable is given by the natural log of yearly rainfall minus the natural log of mean annual rainfall in a given community.

3.1. Descriptive statistics

Summary statistics by district are presented in Table 1. Information on enrollment for the period 1979–2003 shows that, on average, more boys than girls are enrolled in primary school and the sex imbalance in enrollment increases by grade. In grade 1 there are, on average, 9% more boys than girls enrolled while in grade 7, the difference between male and female enrollment has increased to 32%. Summary statistics

Table 1
Summary statistics by district.

	Mean	Median	St.dev	Obs
Female students in P1	11,336	8839	9408	652
Female students in P2	7717	6076	6054	652
Female students in P3	7081	5487	5637	652
Female students in P4	6006	4565	5155	652
Female students in P5	4813	3535	4195	652
Female students in P6	3682	2749	3202	652
Female students in P7	2462	1896	2148	652
Male students in P1	12,541	9755	10,875	652
Male students in P2	8528	7029	6190	652
Male students in P3	7905	6712	5615	652
Male students in P4	6892	5492	5907	652
Male students in P5	5572	4526	4289	652
Male students in P6	4604	3843	3384	652
Male students in P7	3623	3113	2536	652
Deviation of log rainfall from mean	−0.006	−0.005	0.15	652
Yearly average rainfall (mm)	1291	1272	344	652
Aggregate test score	11.50	11	8.43	1,667,447
Aggregate test score female students	11.09	10	8.53	682,206
Aggregate test score male students	11.79	11	8.34	982,580
Total students taking the PLE per district and year	8072	7274	4748	1,667,447
Female students taking the PLE per district and year	3976	3216	2771	682,206
Male students taking the PLE per district and year	4454	3952	2323	982,580

Notes: Summary information for the period 1979–2003. Enrollment data are disaggregated by grade, district and year. Test score data are the average district score of the grade 7 students' Primary Leaving Exam (PLE) for the years 1989–2002. The rainfall measure is the natural log of rainfall minus the log of mean rainfall in the district for the period 1975–2003.

[23] Deaton (1997) and Wooldridge (2000).
[24] The TMPA data are publicly available on the web at http://disc2.nascom.nasa.gov/tovas/.
[25] As suggested by NASA's Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

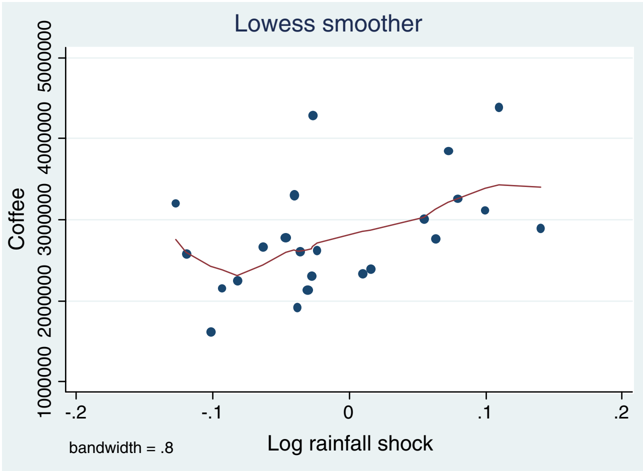

Fig. 4. The effect of rainfall shocks on coffee production at the national level.

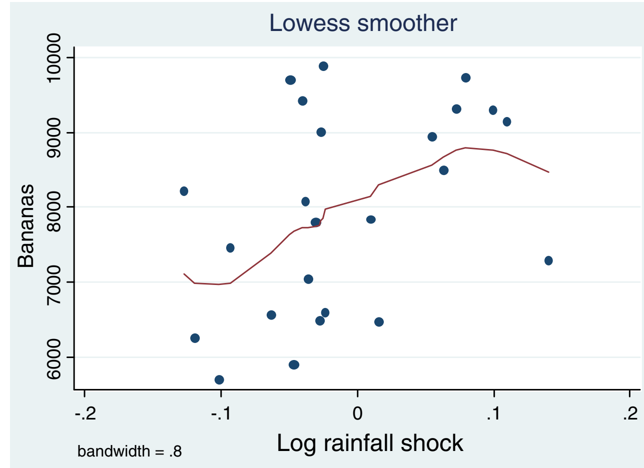

Fig. 5. The effect of rainfall shocks on banana production at the national level.

on rainfall for the period show that the average deviation between natural log of yearly rainfall and the natural log of historical rainfall is −0.006 log points and the standard deviation is 0.15 log points per year. On average, it rains 1291 mm per district and year. Descriptive statistics of Primary Leaving Exam (PLE) test scores for the years 1989–2002 indicate that the test score ranges from 0 to 32 and the average test score is 11.50. On average, female students score 0.70 points worse than male students on the test. The average number of students taking the test per district and year is 8072.

4. Identification strategy

The main problem in identifying the link between households' economic conditions and differential investments in children's education is that both may be partly related to omitted household characteristics. The failure to control for these household variables will bias the estimates since household income and girls' education might be jointly determined by this omitted variable. Omitted district-specific variables may also create some concern if, for example, households in some districts are more progressive than households in other districts (i.e. because of different cultural norms across ethnic groups), which could affect both girls' enrollment in primary school and household income.

I avoid these problems by looking at plausibly exogenous income shocks in Uganda across districts and over time. Rainfall shocks constitute a good proxy to household income shocks in Uganda for several reasons.[26] Rainfall is the most important dimension of weather variation in Uganda. Because of its equatorial location, temperature shows very little variation, both within years and across them. Uganda is an agricultural country and much of Uganda's agricultural production activities are rainfed, meaning that changes in weather conditions have important implications for the households' total agricultural production and wellbeing.[27] The mainly rainfed agricultural sector employs more than 80% of the workforce and 40% of Uganda's total output is obtained from rainfed agriculture. Hence, rainfall shocks constitute an important determinant of variation in household income. Figs. 4 and 5 show that deviations of rainfall from the historical mean are associated with deviations of crop yields in Uganda (years with rainfall higher than average have larger crop yields, and years with low rainfall have lower crop yields).

Table 2
Effect of rainfall shocks on crop yields and income at the national level.

Dependent variable	Coffee	Bananas	Peas	log (GDP)
Deviation of log rainfall from mean	4358**	7.89**	0.022**	5.65**
	(1796)	(3.38)	(0.01)	(2.77)
R^2	0.23	0.19	0.19	0.09
Observations	23	24	24	25

Notes: The dependent variable is average crop production and gross domestic product in the country. Coffee data are measured in 1000 tons and available for the years 1977–2002, data on bananas and peas are measured in 1000 tons and are available for the years 1978–2002, and gross domestic product data are measured in USD billion and are available for the years 1980–2005.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

In an ideal setting, I would use rainfall as an instrument for household income in a first-stage regression and income as a determinant of investment in education in a second-stage regression. Unfortunately, district-specific income data over time are not available and I will therefore study the reduced form relationship between rainfall shocks and investment in boys' and girls' education. However, national time-series data on crop production and national income are available and are used to show that there exists a relationship between rainfall shocks and agricultural production and income.[28] Table 2 and Figs. 4–6 depict the relationship between the rainfall shock measure at the national level $(\ln R_t - \ln \overline{R})$ and the yield of the main staple food (banana), the main export crop (coffee), and the Gross Domestic Product. For both the crops and the Gross Domestic Product, $(\ln R_t - \ln \overline{R})$ enters highly significant and with the predicted positive sign. This positive relationship between the rainfall measure, agricultural production, and income at the national level provides confidence for using rainfall shocks as a proxy for exogenous household income shocks at the district level.

A concern when looking at the reduced form relationship between rainfall shocks and educational outcomes is if rainfall affects education through some other channel other than income. However, it is unlikely that rainfall shocks would have a direct impact on differential investment in the education of boys and girls. One possible channel through which rainfall shocks could alter children's enrollment is through

[26] Note that Miguel et al. (2004) find a close relationship between rainfall and GDP at the cross-country level. Moreover, Levine and Yang (2006) find that deviations of rainfall from the district level mean are positively associated with agricultural output in Indonesia in the 1990s.

[27] Asiimwe and Mpuga (2007) show that rainfall shocks have important implications for household production and agricultural income for households in Uganda.

[28] National data on income are available for the years 1980–2005 and data on crop production are available for the period 1979–2003.

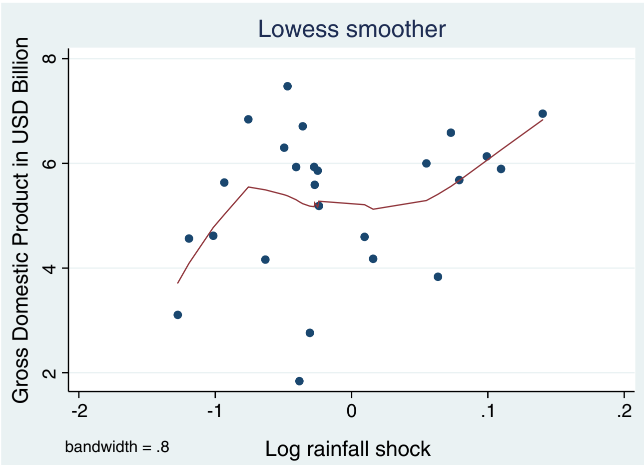

Fig. 6. Rainfall shocks and GDP at the national level.

damaged infrastructure. However, the difficulty in attending school due to damaged roads would affect boys' and girls' enrollment equally. Alternatively, if the unusual rainfall damaged the family's property and the boy needed to help in reconstruction, there would be a downward bias in the estimates. Another concern would be if unusual rainfall alters the disease environment, and these diseases differentially affect boys and girls. However, according to the Ministry of Health in Uganda, there is no prevalent disease appearing in unusual rainfalls that affects boys and girls differently.[29] Further supportive evidence that there are no diseases affecting boys and girls differently in periods of unusual rainfall is provided in Specification 3 in Table 9. Household data show no difference between boys' and girls' reported number of days sick in periods of rainfall shocks. Hence, this potential channel is also ruled out.

Another possible concern is the post-birth sex ratio that may vary across time and over space due to endogenous mechanisms like migration and mortality. Variation in the cross-section over time driven by differences across age groups would challenge the identification. Fig. 7 plots the fraction of males in Uganda by age and time using the 1969, 1991, and 2002 Population Census. The fraction of males in all age groups below 15–19 (i.e. primary school age children) is stable over time. Further, in the 2002 Population Census, the fraction of males among newborn children was 49.6% and there is no evidence of differences in population growth rates of boys and girls of primary school age between 1991 and 2002. Fig. 8 looks at each district separately and compares the sex ratio between children aged 0–4 in 1991 to the sex ratio of the same cohort 10 years later, i.e. children aged 10–14 in 2002. The pattern reveals a stable cohort sex composition across districts. In addition, the 2002 Population Census also collected detailed information on internal migration by gender and age groups. The sex ratio of internal migrants aged less than 15 years old was close to 50% and there was no difference in the sex ratio between migrants to rural and urban areas or between different regions in Uganda. The sex composition among school age children in Uganda, across time and over space, can be interpreted as exogenous.

In the empirical analysis, I first consider the relationship between income shocks and primary school enrollment. More specifically, I test for differences in female and male enrollments, Y_{dt}, in district d and year t in the wake of transitory rainfall shocks using the following reduced-form linear regression:

$$Y_{dt} = \alpha + \beta\left(\ln R_{dt\text{-}1} - \ln \overline{R_d}\right) + \delta_t + \mu_d + \varepsilon_{dt}. \tag{11}$$

The coefficient of interest is β, the impact of the natural log of deviation in rainfall in year $t-1$ from the log of historical mean in district d $(\ln R_{dt-1} - \ln \overline{R_d})$ on enrollment conditional on gender. δ_t is a linear time trend, μ_d is district fixed effects, and ε_{dt} is the error term. Due to serial and spatial correlation in error terms, standard errors allow for an arbitrary variance–covariance structure within districts (clustering by district). This regression is estimated separately for female students, male students, and total students.

The identifying assumption in Eq. (11) is that the natural log of deviation in rainfall from its long-term mean, $(\ln R_{dt-1} - \ln \overline{R_d})$, is uncorrelated with the error term ε_{dt}, i.e. there are no omitted variables correlated with $(\ln R_{dt-1} - \ln \overline{R_d})$. Clearly, the education levels of girls and boys are affected by other factors than income. Some of these factors may also be related to the pattern of rainfall over time. For example, districts with a high average rainfall may be populated by more households and these households may, on average, have different characteristics than households in districts with less average rainfall (e.g. higher income, more progressive view towards women). These characteristics may, in turn, influence the educational choice of boys and girls. Similarly, districts with a high variation in rainfall may be populated by more risk-averse people and they may also have different preferences for optimal levels of education for girls and boys. In the robustness section, I control for the mean and standard deviation of rainfall and this does not affect the estimates. Other potential time-invariant variables may also influence Y_{dt}. The inclusion of district fixed effects controls for these persistent fixed effects of rainfall in the localities where the children live. Any omitted variable with common effects on the enrollment of girls and boys living in the same districts should be absorbed by the district fixed effects.

In addition to studying the effect of income shocks on enrollment, I study the relationship between rainfall shocks and school performance measured by test scores. To this end, I use the following specification:

$$TS_{idt} = \alpha + \theta_1\left(\ln R_{dt-1} - \ln \overline{R_d}\right) + \delta_t + \mu_d + \varepsilon_{idt}, \tag{12}$$

where TS_{idt} is the test score of individual student i in district d and year t. $(\ln R_{dt-1} - \ln \overline{R_d})$ is the natural log of rainfall in year $t-1$ minus the natural log of the historical mean in the given district d. δ_t is a linear time trend, μ_d is district fixed effects, and ε_{idt} is the individual specific error component. The regression is run separately for girls and boys and the parameter θ_1 gives an estimate of the effect of a short-term rainfall shock on girls' and boys' academic performance.[30] Errors are assumed to be iid between districts but correlated within districts, i.e., standard errors are clustered by district.

5. Results

5.1. Enrollment

Fig. 9 depicts the correlation between district income and girls' and boys' enrollment in grade 7. Although the locally weighted regression does not provide any causal evidence, it reveals a relationship between income and enrollment similar to that predicted in the theoretical model. For low levels of income, very few girls attend school and there is a large gap between girls' and boys' enrollment. When a sufficiently large number of girls are enrolled in primary school, an income shock will have a larger effect on girls than on boys (indicated by the slope of the curves) up to some high level of income.[31]

[29] Personal communication with Dr. Solome K. Bakeera and Dr. Francis Runumi Mwesigye.

[30] I am not able to estimate the effect at the school level, since the collected raw data are only coded by district level.

[31] The inverted-U shaped relationship suggested by Fig. 9 is similar to the predictions in the theoretical model (Fig. 1). For low levels of income, mainly boys are enrolled in school and an income shock would primarily affect boys. For higher levels of income, marginal girls are also enrolled in school and the effect of an income shock would be larger for girls than for boys.

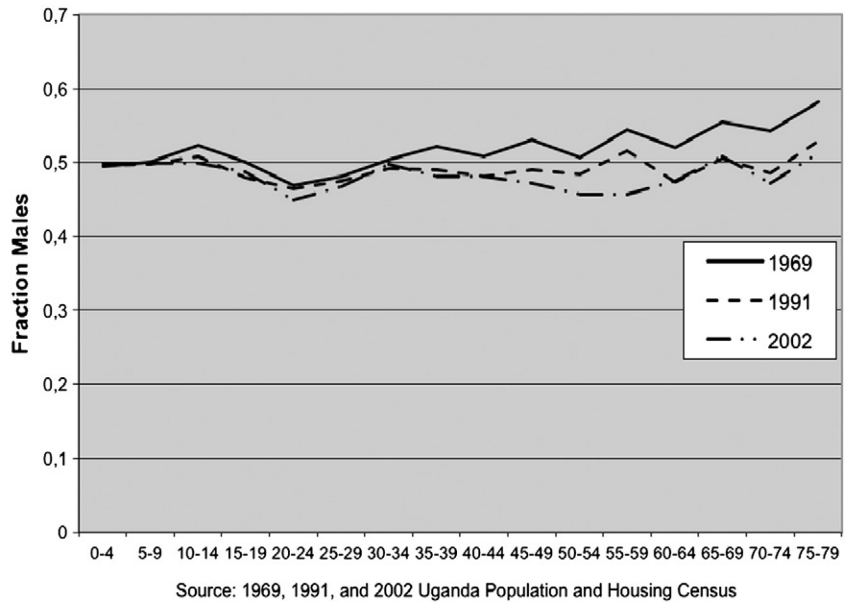

Fig. 7. Sex ratios by age and time period in Uganda.

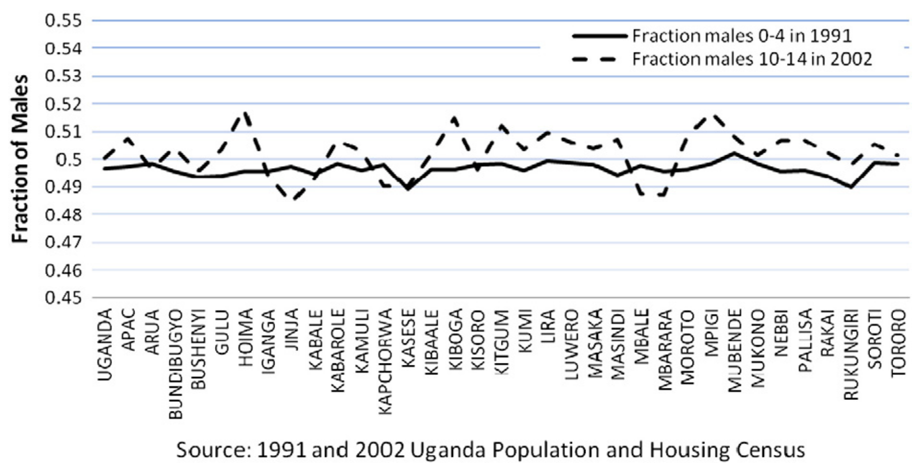

Fig. 8. Sex ratios for children 0–4 and 10–14 years old by district in Uganda.

The correlation depicted in Fig. 9 is characterized by omitted variables. To study the causal effect of household income shocks on investment in education conditional on gender, I use exogenous rainfall shocks as a proxy for household income. Estimates of the reduced-form equations are presented in Table 3. Regressions are run separately for each primary grade in order to study age-specific effects. For each grade, the coefficient on the rainfall variable is presented for girls (Panel A), boys (Panel B), and all students (Panel C). Panels D and E study whether lagged rainfall shocks matter. Standard errors are presented in parentheses and the sample size in brackets. Coefficients for the many fixed effects and time trends are not shown.

The result in Panel A of estimating Eq. (11) for the female sample shows that there is a positive relationship between the deviation of rainfall from the mean and girls' enrollment for all grades. This indicates that rainfall below the long-term mean decreases female enrollment in primary schools. Looking at the effect across grades, it is also apparent that rainfall shocks are especially important for older girls. The effects are highly significant for girls in grades 6 and 7. When discussing the magnitudes of the estimated effects, I focus on the impact of a 0.15 log point change – one standard deviation – in the rainfall variable (the

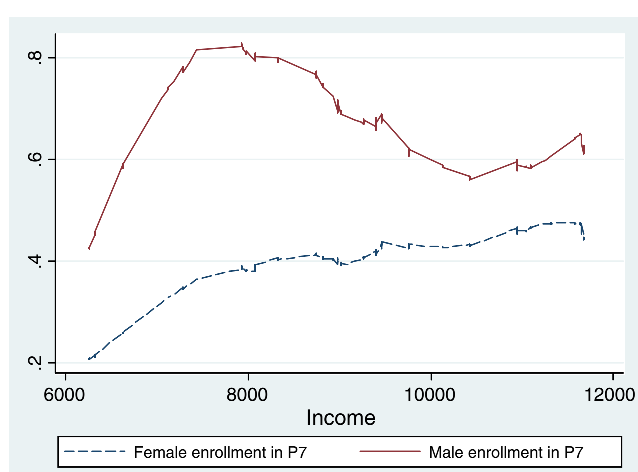

Fig. 9. Correlation between district income and enrollment in grade 7.

Table 3
Effect of rainfall shocks on enrollment.

Primary school grade	P1	P2	P3	P4	P5	P6	P7
Panel A: female enrollment							
Observations: 652							
Deviation of log rainfall from mean	1662	233	1339	576	885	961**	787***
	(2076)	(1147)	(945)	(947)	(617)	(456)	(292)
Panel B: male enrollment							
Observations: 652							
Deviation of log rainfall from mean	2355	634	1050	530	759	365	409
	(2195)	(1166)	(985)	(851)	(678)	(549)	(430)
Panel C: total enrollment							
Observations: 652							
Deviation of log rainfall from mean	3991	866	2389	1105	1644	1325	1197*
	(4163)	(2290)	(1920)	(1641)	(1245)	(982)	(699)
Panel D: female enrollment							
Observations: 599							
Deviation of log rainfall from mean	1953	235	1473*	656	1078*	1118***	852***
	(1961)	(1049)	(852)	(1020)	(588)	(426)	(274)
Deviation of log rainfall from mean ($t-1$)	−1210	77	−79	649	280	255	395
	(1838)	(1053)	(1068)	(1040)	(672)	(555)	(370)
Panel E: male enrollment							
Observations: 599							
Deviation of log rainfall from mean	2935	811	1180	902	991	603	618
	(1986)	(1060)	(878)	(759)	(611)	(475)	(377)
Deviation of log rainfall from mean ($t-1$)	839	183	157	−339	79	−31	54
	(2748)	(1067)	(1005)	(992)	(731)	(556)	(418)

Notes: The sample is all children enrolled in primary school between 1979 and 2003 inclusive. Each coefficient (standard error) is from a separate regression of the dependent variable on rainfall (deviation of log rainfall in district from log of 1975–2003 district mean rainfall). Panel A is regressions on girls separately; Panel B is regressions on boys; Panel C is regressions on all students; Panels D and E are regressions on girls and boys, respectively, when including lagged rainfall shocks. The standard errors are clustered by district. All regressions include district fixed effects and a linear time trend.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

deviation of log rainfall from the log of district mean rainfall). A 15%, or a one standard deviation, decrease in rainfall from the historical mean cuts enrollment by 118 girls in grade 7 in a typical district and the effect is precisely estimated at the 1% level. This is a substantial effect and corresponds to a 5 percentage point reduction of the 2462 girls enrolled in grade 7 in an average district.

The results for boys' enrollment are presented in Panel B. In stark contrast, there is no indication that the enrollment of boys is affected by deviation in rainfall from the mean. Almost all coefficients on the rainfall variable in all regressions are smaller in magnitude than the corresponding coefficients in the female regressions, and none are statistically significantly different from zero. According to the point estimates, a 15% (or one standard deviation) decrease in rainfall from the long-term mean leads to a reduction of 61 boys in grade 7 in the average district, which only corresponds to 1.6 percent of the average number of male students (3623). Moreover, the effect is statistically insignificant.

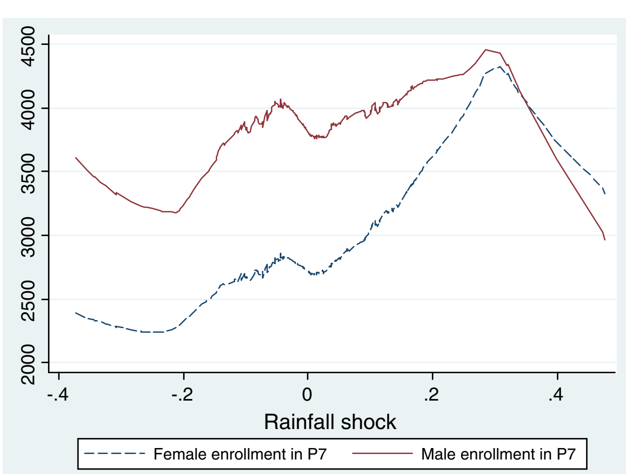

Fig. 10. Locally weighted regression on rainfall and female and male enrollment.

Fig. 10 depicts the results from a locally weighted non-parametric estimation on rainfall shocks and female and male enrollment in grade 7. This graph visually shows the positive relationship between rainfall shocks and enrollment and the differential effects by gender as indicated in Table 3. There are always more boys than girls enrolled. In periods of rainfall below the mean, the ratio of boys to girls in primary school increases since girls are dropping out of school. This effect becomes stronger for larger negative rainfall shocks. When districts experience rainfall above the mean, the ratio of boys to girls enrolled decreases and both more girls and boys enroll in primary school. For extreme positive rainfall shocks (floods), there is a decrease in the enrollment of both boys and girls. Fig. 10 (and similarly also Figs. 4 through 6) suggests that there are non-monotonicities and I test for this in Section 6, the robustness section, but I do not find any effects. These non-monotonicities are driven by outliers and a low number of observations at the end points. The enrollment results become even stronger if I run the regressions excluding these outliers.[32]

The plots of rainfall shocks on agricultural production in Figs. 4 and 5 joint with the results in Table 3 suggest an explanation for these results. Rainfall less than the long-term mean has a negative impact on agricultural output which leads to lower household incomes and worsened

[32] The outliers causing the non-linearity in Fig. 10 are less than 1% of the data points. For transparency, I run the regressions including them but if I exclude them the results become even stronger. I also think it is interesting to observe that very large positive rainfall shocks, floodings, have a negative impact on enrollment (result shown in Table 5).

conditions for girls to go to school. Furthermore, the theoretical model discussed in Section 2 highlights several mechanisms that could explain differential investment in children's education as a response to negative income shocks and in particular why mainly older girls would be affected. I will use these mechanisms to say something about what is driving the empirical results. The returns to education (ω_s in the model) and the share of the child's income (as an adult) transferred to his/her parents (γ_s in the model) are not likely to be a function of the child's age or the primary grade in which the child is enrolled. Hence, if the returns to education and transfers to parents as an adult were the driving channels, I should not find an age-specific effect. On the other hand, children's productivity in domestic work is increasing in age since older children are able to perform more tasks and take on more responsibility in home production than younger children, $\frac{\partial \eta_s}{\partial age} > 0$. The results in Table 3 depict that sex-specific effects are also age-specific since only older girls are affected by rainfall shocks. Thus, using the model to interpret the evidence in Table 3 suggests that the key mechanisms behind the differential treatment of girls versus boys are related to the fact that parents' values of child labor (for home production) differ across sexes. Teenage girls with a comparative advantage in home production (and who are better substitutes for mothers) are withdrawn from school in periods of negative rainfall shocks while there is no effect for very young girls or boys. That is, older girls must bear the bulk of the additional work required at home in bad times.

Negative income shocks have a negative effect on girls' schooling, while boys remain relatively unaffected, but do income shocks have a permanent or temporary effect on girls' enrollment? To study this, I examine how lagged rainfall shocks affect children's enrollment. Specifically, I include lagged rainfall shocks in regression Eq. (11) and the results are found in Panels D and E in Table 3. If income shocks have a temporary effect on enrollment, students withdrawn from primary school in the year of an income shock will return to school in the following year, which will be indicated by a small, if not nil, value on the lagged variables. The result depicted in Panel D suggests that lagged rainfall deviations from the mean do not affect girls' current enrollment. However, when including lagged rainfall, the results are more precisely estimated and the coefficient on girls in grade 5 is now also significant. Older girls, in grades 5 to 7, are affected by rainfall shocks while boys and younger girls are not affected by temporary income shocks.

To summarize, the results clearly indicate that households respond to income shocks by varying the quantity of girls' education, while boys are to a large extent sheltered. This finding is consistent with the simple theoretical model presented in Section 2. Older girls are withdrawn from primary school in periods when rainfall is deviating from the long-term mean, while boys and younger girls are not affected and the mechanism explaining this result, as outlined in the model, is girls' perceived higher productivity in performing domestic chores. When taking lagged rainfall shocks into account, the results become stronger and more precisely estimated.

5.2. Educational achievements

Table 4 shows the results of estimating regression (12) and indicates that transitory income shocks do not only affect investment in children's education, but also children's academic performance. The results show that rainfall shocks affect children's test scores differently depending on the cost of schooling and the type of children enrolled in school. I explore this by studying the differential effect of rainfall shocks on boys and girls before and after user fees were abolished in primary schools (the Universal Primary Education reform). The abolishment of user fees relaxes the financial constraint of poor households and affects the household's decision of whom to send to school. As illustrated in the simple model in Section 2, the type of students enrolled in school differs before and after the abolishment of user fees. Fig. 4 shows that in a district with both less poor and poor households prior to the abolishment of user fees, only the less poor group sent their girls to school while both the poor and the less poor group sent their boys to school. After the cost reduction, also the poor group starts sending its girls to school which leads to an increase in the enrollment of marginal girls.

Table 4
Effect of rainfall shocks on academic performance.

Specification	Dependent variable: test scores (1)	(2)
Panel A: female enrollment		
Deviation of log rainfall from mean	−0.42	−2.21*
	(0.79)	(1.11)
Universal primary education reform		2.14**
		(0.83)
Universal primary education reform x deviation of log rainfall from mean		2.89**
		(1.31)
R^2	0.15	0.15
Observations	682,206	682,206
Panel B: male enrollment		
Deviation of log rainfall from mean	−0.78	−2.12***
	(0.70)	(0.95)
Universal primary education reform		−1.04
		(0.72)
Universal primary education reform x deviation of log rainfall from mean		2.24
		(1.47)
R^2	0.12	0.12
Observations	982,580	982,580

Notes: The sample is grade 7 children participating in the Primary Leaving Exam between 1989 and 2002 inclusive. Specification: (2) Controlling for the universal primary education reform in primary schools. UPE is an indicator variable for the years when user fees were abolished in primary schools. Each coefficient (standard error) is from a separate regression of the dependent variable on rainfall (deviation of log rainfall in district from log of 1975–2003 district mean rainfall). Panel A is regressions on female students and Panel B is regressions on male students. All regressions include district fixed effects and a linear time trend. Standard errors are clustered by district.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

Specification 1 in Table 4 shows no impact of rainfall shocks on average test scores when merging the periods before and after the abolishment of user fees. However, Specification 2 shows that there is a differential effect on children's enrollment depending on whether there was a cost for attending school. As outlined in the theoretical model, prior to the Universal Primary Education reform, when there was a cost for attending school, boys and girls from the upper quintile attended primary school but also marginal boys were prioritized. Specification 2 in Table 4 shows that a negative income shock in this period affected both boys and girls. In periods of a negative rainfall shock and hence, low household income, parents withdraw the worst performing girls and the marginal boys and, as a consequence, there is an increase in the average test scores. After the abolishment of user fees, there was no cost of having children enrolled in school and hence, also marginal girls were enrolled in school. The results now indicate that a rainfall shock only affects girls. Girls score 2.89 points less on the standardized test in periods of negative income shocks and this effect is significant at the 5 percentage level. Rainfall shocks do not have a significant impact on boys' test scores after the removal of user fees. This result is consistent with the model in Section 2 when both marginal boys and girls are enrolled. The selection effect causes marginal girls to be withdrawn from school to work at home and only the girls in the upper quintile remain in school during negative rainfall shocks while boys belonging to both the upper and the lower quintile remain in school. At the same time, in periods with low rainfall, the upper quintile girls who remain in school are provided with less resources within the household or have to spend more time on domestic chores as compared to boys and this resource effect causes even these upper quintile girls to perform worse on the test as compared to boys. The effect of rainfall shocks on

Table 5
Effect of extreme rainfall shocks on enrollment.

Specification	Dependent variable: female enrollment in grade 7				
	(1)	(2)	(3)	(4)	(5)
Deviation of log rainfall from mean >2 st.dev from mean	1359* (745)				
Negative log rainfall deviation >2 st.dev from mean		−371* (207)			
Positive log rainfall deviation >2 st.dev from mean		828 (534)			
Extreme positive log rainfall dev >700 mm from mean			−2782*** (559)		
Deviation of log rainfall from mean				787*** (292)	787*** (292)
No. years with negative deviations >1.5 st.dev				−221** (84)	
No. years with positive deviations >1.5 st.dev				1455** (216)	
Log of yearly average rainfall					42.2 (435)
Log of standard deviation of rainfall					2805*** (379)
R^2	0.91	0.91	0.91	0.80	0.80
Observations	652	652	652	652	652

Notes: The dependent variable is girls enrolled in grade 7 between 1979 and 2003 inclusive and rainfall deviations are measured as the deviation of log rainfall in district from log of 1975–2003 district mean rainfall. Standard errors are clustered by district. All regressions include district fixed effects and a linear time trend.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

girls' academic performance suggests that the resource effect dominates when school is free of charge.[33]

The results show that when households have to pay for schooling, both boys and girls are affected by rainfall shocks and the selection effect dominates so that households withdraw the worst performing children (both boys and girls) when they face a negative income shock which makes average test scores in the districts increase. However, when schooling is free of charge, a transitory income shock only affects girls' school performance. Boys are left in school since they are prioritized within the family and there is no cost of having them in school, while the marginal girls are withdrawn and the girls who remain in school are provided with less resources and/or have to spend more time on household chores, which consequently has negative effects on girls' performance in school.

6. Robustness analysis

To study the robustness of the key explanatory variable, I experiment with other rainfall shock measures alongside the core variable that is the natural log of rainfall minus the natural log of mean annual rainfall in the given districts $\left(lnR_{dt-1} - ln\overline{R_d}\right)$. Other rainfall shock measures used are rainfall deviations of at least 2 standard deviations from the historical mean, a normalized dependent variable with the total number of children in the district, and the log of rainfall minus the natural log of mean annual rainfall in the district excluding the current year in order to eliminate biases in the measure of the deviation from the mean for those regions with shorter time-series. According to Table 5, the effect of rainfall shocks on female enrollment is maximized when using rainfall shocks larger than 2 standard deviations from the historical mean as the explanatory variable. According to specification (2), when looking at positive and negative rainfall shocks larger than 2 standard deviations, I find that negative income shocks drive the result.

I also study the individual effect of positive and negative rainfall shocks using the main rainfall shock variable. First, I determine the effect of positive and negative deviations of log rainfall from the log mean of rainfall on coffee production. The results in Table 6 show that negative rainfall shocks decrease the coffee production in the district. Then, I estimate the differential effect of positive and negative rainfall shocks on children's schooling conditional on gender using the most precisely estimated specification and the results are presented in Table 7. I find that negative rainfall shocks decrease the number of girls in primary grade 5 to 7 while it has not effect on younger girls. I also find that negative rainfall shocks affect boys in primary grade 7 and the effect is significant at the 10% level. After exploring several different rainfall shock variables, I conclude that the results in this paper are robust to alternative rainfall shock measures.

I also study how extreme positive rainfall shocks affect enrollment, i.e. whether large positive rainfall shocks have a negative effect on enrollment. According to specification (3) in Table 5, a large positive rainfall shock, i.e. a flood, decreases the enrollment of girls in school and the effect is significant at the 1% level. The result indicates that an extreme positive rainfall shock has a negative effect on enrollment and this

Table 6
Differential effects of rainfall shocks on crop yields.

Dependent variable	Coffee/hectare
Negative deviation of log rainfall from mean	−164* (92)
Positive deviation of log rainfall from mean	29 (116)
R^2	0.46
Observations	320

Notes: The dependent variable is the average coffee output per hectare by district for the period 1990–2002, measured in 1000 kg. The regression contains year and district fixed effects.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

[33] The explanation described in the paper relies on the theoretical predictions from the model that low quality girls are withdrawn from school in periods of negative income shocks which bias the estimates on the girl–boy gap towards zero. However, if we disregard the predictions from the model, the results that a negative rainfall shock decreases girls' test scores could also be because higher quality girls left school to conduct home production, in which case the estimated effects on the girl–boy gap in the test score are biased upwards.

Table 7
Differential effects of rainfall shocks on enrollment.

Primary school grade	P1	P2	P3	P4	P5	P6	P7
Panel A: female enrollment							
Observations: 596							
Negative deviation of log rainfall from mean	568	−286	−1086	−542	−1794**	−1311**	−722*
	(2299)	(1348)	(1331)	(2042)	(800)	(612)	(423)
Positive deviation of log rainfall from mean	5125	489	2489	862	607	1158	1216
	(3319)	(1685)	(1707)	(1363)	(1318)	(991)	(730)
Lagged negative deviation of log rainfall from mean	1904	462	486	−1431	−151	286	189
	(2885)	(1290)	(1202)	(2091)	(902)	(715)	(469)
Lagged positive deviation of log rainfall from mean	−397	503	313	−329	281	746	1011
	(3502)	(2265)	(2348)	(2076)	(1576)	(1231)	(896)
Panel B: male enrollment							
Observations: 596							
Negative deviation of log rainfall from mean	−1909	−787	−1354	−1832	−1394	−747	−948*
	(2821)	(1325)	(1202)	(1161)	(901)	(633)	(480)
Positive deviation of log rainfall from mean	4393	1103	1383	321	727	412	411
	(3328)	(1740)	(1557)	(1384)	(1260)	(986)	(839)
Lagged negative deviation of log rainfall from mean	−2691	732	43	741	168	372	612
	(5701)	(1360)	(1104)	(1112)	(932)	(665)	(587)
Lagged positive deviation of log rainfall from mean	−1133	978	217	−138	178	178	655
	(3857)	(2410)	(2117)	(1964)	(1580)	(1191)	(935)

Notes: The sample is children enrolled in primary school between 1979 and 2003 inclusive. Each coefficient (standard error) is from a separate regression of the dependent variable on rainfall (the deviation of log rainfall in district from log of 1975–2003 district mean rainfall). Panel A is regressions on female students and Panel B is regressions on male students. All regressions include district fixed effects and a linear time trend and standard errors are clustered by district.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

pattern is consistent with that found for crop production (see Figs. 4 and 5). I exclude extreme outliers and reestimate Eq. (11) and the effect of rainfall shocks on enrollment becomes stronger, which implies that large positive shocks are pushing the measured effect of rainfall shocks towards zero.[34] The results depicted in specification (4) in Table 5 indicate that girls living in districts with more years of negative and positive rainfall shocks are even more affected and the effect is fairly precisely estimated. Districts that have had more years of negative rainfall shocks have less girls enrolled in grade 7 as compared to districts with less years of negative shocks.

I am also running regression (11) with controls for the long-term mean and variance (results shown in Specification 5 in Table 5) and my findings are robust, which should clearly be the case since the mean and variance of rainfall are, by construction, orthogonal to the key explanatory variable. The variation in rainfall seems to have a larger impact on enrollment than the mean of rainfall.

Finally, Figs. 4 through 6 and 10 suggest a non-linear relationship between rainfall and output and rainfall and enrollment. However, as indicated earlier, the nonlinearities are driven by the low number of observations at the endpoints (less than 0.5% of the data points). As a robustness check, I also test for non-linearity in the effect and the results do not change. In my regressions, I am assuming that rainfall shocks have a linear effect on enrollment and by testing for the polynomial effect, I investigate whether rainfall shocks have a diminishing or increasing marginal effect on enrollment. I find that the results are robust to the inclusion of polynomial variables and the coefficient on polynomial rainfall shocks is insignificant.[35] Accordingly, I can exclude the possibility that rainfall shocks have a non-linear effect on the enrollment of girls and boys.

6.1. Results on enrollment using household data

The schooling data used in the main analysis are aggregate district level data. The advantage of using the full census of school age children in Uganda is that it is not subject to sampling error as a representative sample of micro data might be. The full census of primary school age children for 24 years allows me to establish the effect of rainfall shocks across a large number of cohorts for all districts in Uganda which confirms the external validity of the study. In order to show that the results from the district level data are consistent with the micro level findings, I use household data to estimate the effect of rainfall shocks on the enrollment of children of primary school age. I use a pooled cross section of households from the 1999/2000, 2002/2003, and 2005/2006 national household surveys in Uganda. I match the GPS position of the village where the household is located with rainfall data for that community.

Table 8 provides summary statistics for the household data. The sample includes roughly 30,000 rural households with children of primary school age. A community is the 0.25 latitude and longitude degree area and it is the level at which I have corresponding rainfall data. There are, on average, 228 communities in the dataset and each community contains on average 7.8 villages and 421 households. The fraction of female children is 50% and the average age of primary school age children is 10.2 years. 84% of the children are enrolled in primary school. Summary statistics of rainfall show that the average deviation of log rainfall from the log mean of rainfall in the community is 0.024 log points and

Table 8
Descriptive statistics of household data.

	Mean	St.dev	Obs
Number of communities	228	12.7	30,489
Number of villages per community	7.8	6.9	30,489
Number of households per community	421	354	30,489
Fraction of females	0.50	0.50	30,489
Age	10.2	2.9	30,489
Primary school enrollment	0.84	0.37	30,489
Females	0.83	0.37	30,489
Males	0.84	0.37	25,057
Household size	8.1	3.47	30,489
Deviation of log rainfall from mean	0.024	0.20	30,489
Yearly average rainfall (mm)	1229	381	30,489

Notes: Household data are collected from Uganda National Household Survey 1999, 2002, and 2005. The rainfall measure is the natural log of rainfall minus the log of mean rainfall in the community for the period 1996–2006. A community is defined by 0.25 × 0.25 degree gridpoints and contains all villages within this area.

[34] The results are available upon request.
[35] The results are available upon request.

Table 9
Effect of rainfall shocks on schooling and health using household data.

Dependent variable	Attendance	Attendance	Days sick
Specification	(1)	(2)	(3)
Age	0.013***	0.014***	−0.043***
	(0.002)	(0.002)	(0.014)
Female	0.027	0.027	−0.021
	(0.021)	(0.021)	(0.21)
Female x age	−0.003	−0.003	0.006
	(0.002)	(0.002)	(0.19)
Deviation of log rainfall from mean	0.179**	0.189**	−0.60
	(0.08)	(0.07)	(0.73)
Age x deviation of log rainfall from mean	−0.019**	−0.019**	0.068
	(0.007)	(0.007)	(0.064)
Female x deviation of log rainfall from mean	−0.11	−0.14	0.49
	(0.096)	(0.09)	(0.97)
Female x age x deviation of log rainfall from mean	0.015*	0.017**	−0.034
	(0.009)	(0.008)	(0.092)
Year fixed effect	Y	Y	Y
Community fixed effect	Y	Y	Y
Controls	N	Y	Y
R^2	0.06	0.07	0.03
Observations	31,759	30,489	30,266

Notes: Data from the 1999/2000, 2002/03, and 2005/06 Uganda National Household Survey and the sample includes households with primary school age children. Specification: (1) and (2) the dependent variable is primary school attendance; (3) the number of days lost due to illness. The control variables in specifications (2) and (3) are agricultural household, household size, mother working, and a poverty measure. All regressions include community fixed effects and a linear time trend. Standard errors are clustered at the village level.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

the standard deviation of rainfall per year is 0.19 log points. On average, it rains 1229 mm per community and year.

Table 9 shows the results from estimating the following regression:

$$
\begin{aligned}
Y_{ijct} = \partial + \alpha F_{ijct} + \theta age_{ijct} + \beta(\ln R_{ct-1} - \ln\overline{R_c}) + \eta\left[F_{ijct} * age_{ijct}\right] \\
+\zeta\left[age_{ijct} * (\ln R_{ct-1} - \ln\overline{R_c})\right] + \lambda\left[F_{ijct} * (\ln R_{ct-1} - \ln\overline{R_c})\right] \\
+\gamma\left[F_{ijct} * age_{ijct} * (\ln R_{ct-1} - \ln\overline{R_c})\right] + X_{ijct} + \delta_t + \mu_c + \varepsilon_{ijct}
\end{aligned}
\tag{13}
$$

where Y_{icdt} is school enrollment of child i, in household j, community c, and year t. F_{ijct} is a dummy variable indicating the child's gender and age_{ijct} is the age of the child. The rainfall shock variable is the natural log of deviation in rainfall from the log of historical mean in community c and year $t-1$ $(\ln R_{ct-1} - \ln\overline{R_c})$. δ_t is a linear time trend, μ_c is community fixed effects, and ε_{ijct} is the error term. X_{ijct} is a set of control variables used in Specifications 2 and 3. γ is the coefficient of interest which indicates whether older girls are more affected by rainfall shocks.

The result without control variables in Specification 1 in Table 10 shows a positive relationship between the deviation in log rainfall from the log mean of rainfall and children's school enrollment. The general positive relationship between deviations in rainfall from the mean decreases with the age for boys, but for girls I find matching results as in the aggregate district level data — older girls are more affected by rainfall shocks. The effect is significant at the 10% level. When including control variables in the regressions, the result becomes more precisely estimated and the effect on older girls is now significant at the 5% level. A decrease (increase) in rainfall of one standard deviation from the long-term mean in the community decreases (increases) the enrollment of teenage girls by 4.4 percent, which should be compared to the 5 percent effect found using district level data. Hence, the result using micro level data matches the result from the district level analysis. Therefore, the aggregation bias in the main analysis does not seem to be a major issue in this study.

7. Additional results

Uganda experienced a large primary education sector reform in the mid 1990s when the Government outlawed school fees in primary education, the so-called universal primary education reform (UPE). Under certain conditions (discussed below), I can use this policy experiment to estimate the effects of a reduction in the (formal) cost of schooling on boys' versus girls' amount of schooling and to assess the extent to which the price elasticity of education varies with changes in the economic conditions.

The UPE reform was implemented country-wide in 1997. Prior to this, all primary schools in Uganda charged user fees. To identify the effects of income shocks after the reform on boys' versus girls' schooling, I estimate the following regression

$$
\begin{aligned}
Y_{dt} = \partial + \gamma_1 UPE + \gamma_2\left(R_{dt-1} - \overline{R_d}\right) \\
+ \gamma_3\left[UPE * \left(R_{dt-1} - \overline{R_d}\right)\right] + \delta_t + \mu_d + \varepsilon_{idt},
\end{aligned}
\tag{14}
$$

where Y_{dt} is the measure of enrollment in district d and year t, δ_t is a linear time trend, μ_d is district fixed effects, and γ_1 is the direct effect of lower school fees on education. Note that γ_1 is purely identified from the time-series variation. Clearly, this is not an uncontroversial identification strategy. There are other changes that could have occurred during the same time period as the UPE reform that could affect Y_{dt} and thus, cause biased estimates. Note, however, that no other (major) policy reform which could be affecting primary school enrollment was introduced at the same time. Still, the evidence should be viewed as suggested.

Table 10 reports results from the regression determining the effect of the UPE reform on enrollment. According to Panels A and B, the abolishment of user fees had a large and positive effect on both female and male enrollment. It increased female enrollment by 1.4 to 2 standard deviations in all grades and similarly for males, the UPE reform resulted in a large increase in boys' enrollment and the effects are significant at standard levels. Relative to the average enrollment of boys and girls in the districts, girls' enrollment increased more relative to that of boys for all grades, as compared to the years prior to the UPE reform and the effect is largest in lower grades. Table 10 also shows whether deviations in log rainfall from the log mean of rainfall had different effects on girls' and boys' schooling before and after the abolishment of user fees. According to the results, a negative rainfall shock had a negative effect on female students in grade 7 after the UPE reform, while boys' enrollment was still not affected by a rainfall shock after the abolishment of user fees.

What can account for these results? Let us now return to the model formulated in Section 2. As illustrated in Fig. 11, removing school-fees would lead to an inward shift of the *BiS* and *GiS* curves.[36] In the figure, the solid lines represent the enrollment of boys (black line) and girls (gray line) before user fees were abolished and the dashed lines represent the enrollment after the cost reduction.

If all households had the same income (and at least 50% of the girls were enrolled in school), the model suggests that the abolishment of user fees would result in an increase in the enrollment of marginal students (both boys and girls), but also that the effects of an income shock would fall. However, I find the effect of a negative income shock on girls' enrollment to be larger after the UPE reform. In reality, of course, not all households within a district have the same income. To illustrate this, consider the case with two population groups in each district; poor and less poor people. Prior to the abolishment of user fees, only the less poor group sent their girls to school while both the poor and the less poor group sent their boys to school. This is illustrated in Fig. 12, where the vertical line to the right corresponds to the less poor group and the vertical line to the left illustrates the poor group.

[36] I simulate ability using the same parameter values and distribution function as in Fig. 1.

Table 10
Additional results of rainfall shocks on enrollment.

Primary school grade	P1	P2	P3	P4	P5	P6	P7
Panel A: female enrollment							
Observations: 652							
Deviation of log rainfall from mean	1253	44	752	−16.9	284	394	246
	(2018)	(955)	(758)	(997)	(549)	(399)	(269)
Universal primary education (UPE) reform	15,598***	11,496***	11,590***	9845***	7205***	5237***	3048***
	(1715)	(1197)	(1127)	(988)	(807)	(658)	(471)
UPE x deviation of log rainfall from mean	1256	579	1805	1823	1850	1741	1662*
	(2944)	(2114)	(2048)	(1858)	(1583)	(1333)	(942)
Panel B: male enrollment							
Observations: 652							
Deviation of log rainfall from mean	2540	206	640	4.86	331	−178	−0.19
	(2432)	(1071)	(858)	(896)	(689)	(545)	(416)
Universal primary education (UPE) reform	15,645***	11,286***	11,272***	9954***	7521***	5677***	3228***
	(1790)	(1292)	(1174)	(1216)	(838)	(715)	(520)
UPE x deviation of Log rainfall from mean	−570	1316	1261	1615	1313	1668	1259
	(3091)	(2074)	(1872)	(1819)	(1635)	(1300)	(1045)
Panel C: female enrollment							
Observations: 646							
Deviation of log rainfall from mean	1752	310	1976**	867	1704***	1729***	1345***
	(1710)	(1025)	(924)	(1376)	(608)	(501)	(343)
Poor district	908	−2086***	−1525**	−1431**	−529	−67	472
	(1213)	(725)	(708)	(695)	(638)	(481)	(338)
Poor district x deviation of log rainfall from mean	−224	−199	−1503	−694	−1932*	−1820**	−1317**
	(2957)	(1752)	(1568)	(1724)	(1062)	(760)	(506)
Panel D: male enrollment							
Observations: 646							
Deviation of log rainfall from mean	2887	1147	1989**	886	1581**	1299**	1227***
	(1775)	(1020)	(913)	(1403)	(698)	(547)	(443)
Poor district	1169	−1423*	−1237*	−1299**	−499	−337	−215
	(1265)	(710)	(663)	(643)	(542)	(484)	(372)
Poor district x deviation of log rainfall from mean	−1268	−1223	−2220	−842	−1945	−2211**	−1929**
	(3190)	(1818)	(1634)	(2323)	(1197)	(947)	(754)

Notes: The sample is children enrolled in primary school between 1979 and 2003 inclusive. Specification: Panels A and B control for the universal primary education reform in primary schools. UPE is an indicator variable for the years when user fees were abolished in primary schools; Panels C and D control for poor districts where poor district is an indicator variable for whether the district has an income less than the median income in the country. Each coefficient (standard error) is from a separate regression of the dependent variable on rainfall (deviation of log rainfall in district from the log of 1975–2003 district mean rainfall). Standard errors are clustered by district. All regressions include district fixed effects and a linear time trend.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

As a response to the cost reduction, the poor group starts sending its girls to school which increases the enrollment of marginal girls. Hence, the effect of an income shock on enrollment after the UPE reform will be larger for girls than for boys, since marginal girls who did not attend school before the UPE reform are now also affected.

More insight into how rainfall shocks affect households' differential investment in children's education is obtained by examining the impact in districts with different income levels. I am using a specification similar to (14), replacing the UPE dummy with a dummy for rich and poor districts when studying whether investment in children's education conditional on gender differs between districts with different income levels.[37] Panels C and D in Table 10 show these result for female and male students. The results indicate that fewer girls are enrolled in the poorer districts as compared to the more wealthy districts and the effect is most prevalent for girls in the lower grades. The result further indicates that the effect of a negative rainfall shock is lower for girls in poorer districts as compared to girls in districts with an average income above the median. To exemplify, consider Fig. 1 with poor and wealthy districts and two population groups in each district: poor and less poor people. In the wealthier districts, parents from both the poor and the less poor group send their girls and boys to school, while in the poorer districts only parents in the less poor group send their girls to school. Hence, the effect of an income shock on girls' enrollment will be larger in the wealthier districts as compared to the effect in the poorer districts, since also the marginal girls are affected. The results for boys are presented in Panel D. When taking into account the district income, there is also an effect of rainfall shocks on boys' enrollment. Boys in the poorer districts are less affected by a rainfall shock than boys in richer districts. This would be consistent with the hypothesis that in

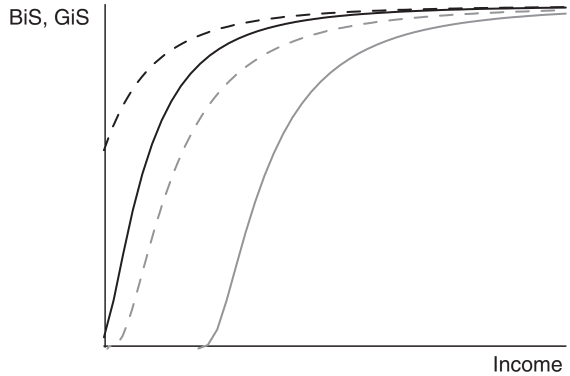

Fig. 11. Enrollment of girls (gray lines) and boys (black lines) after the abolishment of school fees.

[37] I do not have access to income data for all districts and therefore, the sample is slightly smaller in the regressions with income data. Moreover, the income data are only available for one year (1995) and the results should therefore be viewed as suggestive since the distribution of incomes across districts could have changed during the period.

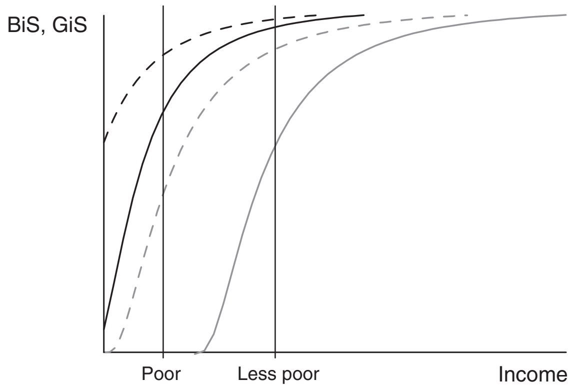

Fig. 12. Enrollment of girls (gray lines) and boys (black lines) after the abolishment of school fees, with two population groups in the district.

more wealthy districts, more marginal boys are sent to school and hence, these boys are sensitive to changes in the economic conditions.

Table 11 studies the effect of rainfall shocks on students' academic performance conditional on the income level in the district before and after user fees were abolished. As pointed out earlier, the results should be viewed as suggestive since income data are only available for one year (1995) and the distribution of incomes across districts could have changed during the period. I find that students in poorer districts (income below the median in the country) before the abolishment of user fees performed better on the PLE test as compared to students in wealthier districts which suggests that mainly high performing children were sent to school in the poorer areas when there was a cost associated with enrollment. After the UPE reform, there is no significant difference in test scores between poor and richer districts indicating that the reform led to a more equalized distribution of students across poor and rich districts. After the abolishment of user fees, the effect of a negative income shock is largest in the wealthier districts, indicating that more marginal students, who are sensitive to income shocks, were enrolled in wealthier districts. I also find that when dividing the districts into poorer and richer, both boys' and girls' test scores in wealthier districts are affected by negative rainfall shocks after the UPE reform. These results, although suggestive, are coherent with both the results on enrollment conditional on district income and the theoretical predictions suggesting that in more wealthy districts, more marginal boys are sent to school and if there is not a sufficiently large number of girls enrolled, these marginal boys are also affected by economic changes.

Table 11
Additional result of rainfall shocks on academic performance.

	Dependent variable: test scores
Panel A: female enrollment	
Deviation of log rainfall from mean	−2.69***,*
	(1.06)
Universal primary education reform	−0.62
	(0.86)
Universal primary education reform ∗ deviation of log rainfall from mean	3.37**
	(1.32)
Poor district	9.18***
	(0.93)
Poor district ∗ universal primary education reform	0.34
	(0.69)
Poor district ∗ deviation of log rainfall from the mean	1.76
	(1.35)
Poor district ∗ universal primary education reform ∗ deviation of log rainfall from the mean	−1.15
	(2.79)
R^2	0.15
Observations	678,454
Panel B: male enrollment	
Deviation of log rainfall from mean	−2.27***
	(0.79)
Universal primary education reform	−0.81
	(0.85)
Universal primary education reform ∗ deviation of log rainfall from mean	2.60**
	(1.12)
Poor district	5.43***
	(0.79)
Poor district ∗ universal primary education reform	0.76
	(0.63)
Poor district ∗ deviation of log rainfall from the mean	0.82
	(1.19)
Poor district ∗ universal primary education reform ∗ deviation of log rainfall from the mean	−0.36
	(2.64)
R^2	0.12
Observations	976,730

Notes: The sample is grade 7 children participating in the Primary Leaving Exam between 1989 and 2002 inclusive. Specification: controlling for the universal primary education reform in primary schools and poor district. UPE is an indicator variable for the years when user fees were abolished in primary schools and poor district is an indicator variable for whether the district has an income less than the median income in the country. Each coefficient (standard error) is from a separate regression of the dependent variable on rainfall (deviation of log rainfall in district from log of 1975–2003 district mean rainfall). Panel A is regressions on female students and Panel B is regressions on male students. All regressions include district fixed effects and a linear time trend. Standard errors are clustered by district.
*** Denotes significance at 1% level.
** Denotes significance at 5% level.
* Denotes significance at 10% level.

8. Conclusion

In many developing countries, boys are more likely to complete primary school than girls. Economists have long argued that boys' and girls' differential educational outcomes can be explained by the underlying economic conditions. Methodologically, it is challenging to establish a link between households' economic conditions and investments in children's education, since households' economic conditions and schooling may be associated with omitted variables. This paper develops and implements a strategy capturing the causal effect of changes in households' economic conditions on the differential investment in children's primary education. In particular, it uses the exogenous variation in rainfall across districts in Uganda to estimate the causal effects of household income shocks on children's enrollment and academic performance conditional on gender. This is the first paper to study how changes in household income affect children's test scores and whether there are any differences depending on gender. I show that negative deviations in log rainfall from the log mean of rainfall have negative, and highly significant, effects on female enrollment in primary schools and the effects are particularly strong for older girls. A 15% decrease in rainfall from its historical mean cuts female enrollment in grade 7 by 5 percentage points. Boys, on the other hand, are not affected by rainfall shocks and neither are younger girls. Additionally, I find that when school is free of charge and both marginal boys and girls are enrolled, a negative income shock has an adverse effect on the test scores of female students. A decrease in rainfall from its historical mean decreases the academic performance of girls to a larger extent than that of boys. The findings in this paper indicate that an exogenous transitory income shock to the household has a different effect, not only on investments in girls' and boys' education, but also on girls' and boys' academic performance. The results imply that households respond to income shocks by varying the enrollment and resources provided to girls, while boys are to a large extent sheltered — a finding consistent with a model where parents' values of child labor (in home production) differ across sexes. Older girls with a comparative advantage in home production are withdrawn from school in periods of negative rainfall shocks while there is no effect on the enrollment of younger girls and boys.

I also find that the universal primary education reform introduced in 1997, which abolished user fees in all primary schools, had a large and positive effect on the enrollment of both boys and girls, although the effect is stronger for girls. After the abolishment of user fees in primary schools, a negative income shock had a negative effect on the enrollment of female students, while the enrollment of boys was still not affected.

The two main implications of the simple model and the empirical results in this paper are (i) income is a key determinant of educational choices, in particular for girls; and (ii) households appear to use girls for consumption smoothing in periods of negative income shocks, i.e. girls are perceived as a buffer and used as an insurance (for domestic work and reduced consumption) in periods of transitory shocks. Considering that gender equality has been identified as one of the most important goals of the donor community, e.g. the Millennium Development Goal, my paper shows that policies boosting income and increasing access to insurance against aggregate risk and saving options for households are likely to affect the speed of reaching the gender equality goal. Moreover, my findings also provide additional arguments for interventions, such as weather insurance, that shield children and especially girls from consequences of temporary environmental and economic shocks.

Appendix A

A.1. Proof of Proposition 1

The optimal intra-household choices for a girl's and a boy's education are determined by the following first-order conditions:

$$-u'[c_1] + \alpha_b^i \leq 0 \quad \text{for } s_b \in [0, 1] \tag{15}$$

$$-u'[c_1]\eta_g + \alpha_g^i \theta_g^i \leq 0 \quad \text{for } s_g \in [0, 1]. \tag{16}$$

Given assumption $\theta_g^i < \theta_b^i = 1$, $\eta_g > \eta_b = 1$ and $\alpha_g^i = \alpha_b^i = \alpha^i$, it follows that $\alpha_g^i \theta_g^i < \alpha_b^i$ and $u'(c_1)\eta_g > u'(c_1)$; the marginal cost for the girl's schooling is higher and the marginal gain is lower as compared to the boy's schooling. When y is sufficiently high, $y_1 = u_c'^{-1}[\alpha_b^i] - \eta_g$, so that it is not optimal to choose $s_b^i = s_g^i = 0$, then $s_b^i > 0$. The household will increase the boy's schooling until $s_b^i = 1$ when $y_2 = u_c'^{-1}[\alpha_b^i] + 1 - \eta_g$. Thereafter, Eq. (15) does not hold and the only way of transferring funds to the next period is by sending the girl to school which happens at $y_3 = u_c'^{-1}\left[\frac{\alpha_g^i \theta_g^i}{\eta_g}\right] + 2 - \eta_g$. The girl's schooling increases until $s_g^i = 1$ when $y_4 = u_c'^{-1}\left[\frac{\alpha_g^i \theta_g^i}{\eta_g}\right] + 2$.

If $s_g^i > 0$ and $s_b^i > 0, y_5 = u_c'^{-1}\left[\frac{\alpha_g^i \theta_g^i}{\eta_g}\right] + 1 - \eta_g$, a reduction in y will force the parents to decrease s_g^i on the margin. Parents will decrease s_g^i until $s_g^i = 0$ and only then will they start to decrease s_b^i. The girl's ability, $a_g^i = \alpha_g^i s_g^i$, is primarily affected by a reduction in y since parents choose to decrease s_g^i when y decreases.

A.2. Construction of rainfall measure

Uganda has 16 demarcated rainfall zones and 31 rainfall stations. Thus, each rainfall zone does, on average, have 2 rainfall stations. The rainfall stations collect monthly precipitation in millimeters which is compiled by the Department of Meteorology. Most stations have collected monthly precipitation from the mid 1970s until today, although some stations have monthly precipitation from 1950 until today. Each rainfall zone on average contains three districts and most districts are located within the boundaries of one rainfall zone, but some few districts are divided into two rainfall zones.

The Department of Meteorology provided me with a detailed map of Uganda showing the demarcated rainfall zones as well as the location of the 31 rainfall stations. I also received data on monthly rainfall for all rainfall stations in Uganda for all years. I merged the map of demarcated rainfall zones with a map of the district boundaries in Uganda to determine which districts are located in which rainfall zones and which rainfall stations are located in which districts.

Three distinct rules have been used when determining the rainfall of each district. The rules are set so as to use all possible rainfall stations in the most accurate way and generate as much variation across districts as possible. First, districts that are located within one rainfall zone and have a rainfall station located in the district are assigned to have a yearly rainfall according to this specific station. Approximately 70% of the districts in Uganda fall into this category. Second, for districts that are located within one rainfall zone but do not have their own rainfall station, I have used average yearly rainfall in the rainfall zone. Third, for districts located in two rainfall zones, I have calculated average yearly rainfall in the two zones and use this as a measure of yearly rainfall in the district. This rule also applies if the district has its own rainfall station but more than one fifth of the district is located in another rainfall zone.

For each district, I constructed a rainfall measure of total rainfall (in millimeter) by district d and year t, R_{dt}, by summing the monthly precipitation in the district over the 12-month period. Some stations lack data for some months per year and if more than one month is missing for a specific year, this is indicated by a missing value for that specific year.

References

African Development Fund, 2005. Uganda — multi-sector country gender profile. Agriculture and Rural Development North East and South Region (ONAR).

Alderman, Harold, Behrman, Jere R., Ross, David R., Sabot, Richard, 1996. Decomposing the gender gap in cognitive skills in a poor rural economy. J. Hum. Resour. 31 (1), 229–254.

Asiimwe, John Bosco, Mpuga, Paul, 2007. Implications of rainfall shocks for household income and consumption in Uganda. African Economic Research Consortium (AERC) Reserch Paper. 168.

Becker, Gary, 1981. A Treatise on the Family. Harvard University Press, Cambridge, Massachusetts.

Beegle, Kathleen, Dehejia, Rajeev H., Gatti, Robert, 2003. Child labor, income shocks, and access to credit. World Bank Policy Research Working Paper, p. 3075.

Behrman, Jere R., 1988. Intrahousehold allocation of nutrients in rural India: are boys favored? Do parents exhibit inequality aversion? Oxf. Econ. Pap. 40 (1), 32–54.

Behrman, Jere R., Knowles, J.C., 1999. Household income and child schooling in Vietnam. World Bank Econ. Rev. 13 (2), 211–256.

Deaton, Angus, 1997. The analysis of household surveys: a microeconometric approach to development policy. Johns Hopkins University Press, Baltimore.

Duflo, Esther, 2012. Women's empowerment and economic development. Journal of Economic Literature 50 (4), 1051–1079.

Foster, Andrew, 1995. Prices, credit constraints and child growth in low-income rural areas. Econ. J. 105 (430), 551–570.

Glewwe, Paul, 2002. Schools and skills in developing countries: education policies and socioeconomic outcomes. J. Econ. Lit. 40 (2), 436–482.

Jacoby, Hanan G., Skoufias, Emmanuel, 1997. Risk, financial markets, and human capital in a developing country. Rev. Econ. Stud. 64 (3), 311–335.

Levine, David I., and Dean Yang. 2006. "A Note on the Impact of Local Rainfall on Rice Output in Indonesian Districts." Mimeo, University of California, Berkeley and University of Michigan.

Maccini, Sharon, Yang, Dean, 2009. Under the weather: health, schooling, and economic consequences of early-life rainfall. Am. Econ. Rev. 99 (3), 1006–1026.

Miguel, Edward, 2005. Poverty and witch killing. Rev. Econ. Stud. 72 (4), 1153–1172.

Miguel, Edward, Satyanath, Shanker, Sergenti, Ernest, 2004. Economic shocks and civil conflict: an instrumental variables approach. J. Polit. Econ. 112 (4), 725–753.

Paxson, Christina H., 1992. Using weather variability to estimate the response of savings to transitory income in Thailand. Am. Econ. Rev. 82 (1), 15–33.

Qian, Nancy, 2008. Missing women and the price of tea in China: the effect of sex-specific income on sex imbalance. Q. J. Econ. 123 (3), 1251–1285.

Rose, Elaine, 1999. Consumption smoothing and excess female mortality in rural India. Rev. Econ. Stat. 81 (1), 41–49.

Rosenzweig, Mark R., Schultz, T. Paul, 1982. Market opportunities, genetic endowments, and intrafamily resource distribution: child survival in rural India. Am. Econ. Rev. 72 (4), 803–815.

Summers, Lawrence, 1994. Investing in all the people: educating women in developing countries. The World Bank Policy Research Working Paper Series number 905.

The World Bank, 1993. Uganda growing out of poverty. A World Bank Country Study.

The World Bank, 2002. Achieving universal primary education in Uganda: the 'big bang' approach. Human Development Network.

Townsend, Robert M., 1994. Risk and insurance in village India. Econometrica 62 (3), 495–526.

Wooldridge, Jeffrey M., 2000. Introductory Econometrics: A Modern Approach. South-Western College Publishing, Thomson Learning, Crawfoldsville, USA.