NBER WORKING PAPER SERIES

START WHAT YOU FINISH! *EX ANTE* RISK AND SCHOOLING INVESTMENTS IN THE PRESENCE OF DYNAMIC COMPLEMENTARITIES

Andrew D. Foster
Esther Gehrke

Working Paper 24041
http://www.nber.org/papers/w24041

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

The authors would like to thank the editor, James Heckman, three anonymous referees, Jishnu Das, Francisco Ferreira, Michael Grimm, Hanan Jacoby, Stephan Klasen, Holger Strulik, Sebastian Vollmer, and numerous conference participants for many helpful comments and suggestions. The REDS data used in this paper were collected by the National Council for Applied Economic Research (NCAER), India, in collaboration with Brown University, Yale University, and the World Bank. The SEPRI data were collected by the Institute for Rural Management, Anand (IRMA), India, in collaboration with the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) and the World Bank. Financial support by the Population Studies and Training Center at Brown University as well as the German Federal Ministry of Economic Cooperation and Development is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Start What You Finish! Ex Ante Risk and Schooling Investments in the Presence of Dynamic Complementarities
Andrew D. Foster and Esther Gehrke
NBER Working Paper No. 24041
November 2017, Revised July 2020
JEL No. I25,O13,O15

ABSTRACT

We study the relationship between risk and schooling investment in a low income setting, with a particular focus on possible *ex ante* effects. We first present a model that shows that such effects can arise if the human capital production function exhibits dynamic complementarity and parental preferences for human capital are not too concave. We then estimate the key parameters of the model using multiple rounds of panel data from rural India that contain, in each round, three seasons of time allocation for each sampled child. These estimates suggest an elasticity of schooling investments with respect to risk of -0.09 in this context. We then use cross-round differences in village-level irrigation interacted with rainfall variability to estimate the relationship between income risk and school time. Using this variation, we find an estimated elasticity of study time with respect to risk between -0.05 and -0.04. Finally, we simulate the effects of an implicit social insurance program, modeled after the National Rural Employment Guarantee Scheme (NREGS). Our results suggest that the risk-reducing effect of the NREGS may offset adverse effects on child education that were evident during the NREGS phase-in due to rising wages.

Andrew D. Foster
Department of Economics and Community Health
Brown University
64 Waterman Street
Providence, RI 02912
and NBER
afoster@brown.edu

Esther Gehrke
University of Goettingen
Department of Agricultural Economics
and Rural Development
Platz der Goettingen Sieben 5
37073 Goettingen, Germany
esther.gehrke@uni-goettingen.de

1 Introduction

By the standards of most forms of household investment that have been studied in rural areas of low-income countries, the education of a household's children has a very long time horizon. Successful completion of a particular credential involves a series of day to day decisions that accumulate and interact with each other over multiple years. Given the nature of this process and the uncertainties of rural life, one might reasonably expect forward looking households to forego such a long term project in favor of activities with a less risky and immediate return. This paper sets out to explore the role of uncertainties about future schooling on investments in current schooling that can be attributed to the presence of dynamic complementarities in the production of human capital.

Dynamic complementarity arises when school investments for a given child in different periods are complements in the production of human capital (Cunha and Heckman, 2007). It implies that early period investments in education have to be followed up by investments in later periods in order to make that early stage investment productive. If future school investment is not known, because parents might have to reduce schooling investments as response to economic shocks, households may be reluctant to invest in schooling in early periods, as they are uncertain about the ability to capture the returns to this investment at later stages.[1] Under such circumstances, the cost of risk in terms of human capital accumulation could be considerably higher than the sum of the direct effects of shocks.

It is by now well established that economic shocks affect child schooling, either because children have to drop-out of school and work (Jacoby and Skoufias, 1997; Beegle et al., 2006; Duryea et al., 2007; Shah and Steinberg, 2017), or because households lack the financial resources to send children to school (see e.g. Jensen, 2000; Skoufias and Parker, 2006; Gubert and Robilliard, 2008; Bjorkman-Nyqvist, 2013). While this relationship is well established in the developing world and often attributed to the existence of credit market imperfections, it also plays a role in high income countries, particularly for the most disadvantaged. Page et al. (2009), as well as Coelli (2011) find that parental income shocks have severe repercussions for the schooling achievements of disadvantaged children in the United States and Canada, respectively. The existence of dynamic complementarities could also explain why such shocks, while transient in nature, seem to have adverse consequences for achieved human capital in the long-run (Shah and Steinberg, 2017).

In this paper, we focus on the question of whether there is an *ex ante* response to risk in the form of study time (*i.e.* the time per day that a child spends studying in school or at home) that can be attributed to the presence of dynamic complementarities in the production of human capital. While there exists previous work that explores the relationship between risk and schooling

[1] For this mechanism to work, parents need to understand the existence of dynamic complementarities. While we are not aware of any evidence from developing countries, recent work from the US seeks to elicit maternal expectations about the technology of skill production and finds evidence that socioeconomically disadvantaged African-American women do indeed expect positive complementarities between current investments and skills in the production of future skills (Cunha et al., 2013). Perhaps more importantly, our paper identifies the degree of complementarity from seasonal variation in observed study time within the school year. As such it characterizes what parents believe and thus what determines their investment choices.

investments (Fitzsimons, 2007; Kazianga, 2012; Colmer, 2019a), we are not aware of any work that explores this question through the lens of dynamic complementarities.

In order to address the question outlined above, we first develop a two-period model of child school investment and human capital accumulation that highlights the effect of uncertainty regarding future household income on a child's study time when schooling investments in different periods are perfect complements in the production of human capital and utility is linear in the final stock of human capital. We then explore how the *ex ante* effects of risk will vary according to the elasticity of substitution in production and the curvature of preferences for human capital. We show that the *ex ante* effect of risk on school investments could be positive or negative. We also show that the effect is negative as long as the demand for human capital exhibits dynamic complementarity, that is if households invests more in schooling the higher the stock of human capital of a given child.

A reduction in study time as a response to an economic shock might arise through any number of channels. Previous work has stressed the importance of a child-labor market effect in which children drop out of school and start working to cover an income shortfall. School fees may also lead to drop out if parents no longer feel able to pay. A potential limitation of these mechanisms is that they may not be applicable to settings in which children are rarely employed as workers outside the family and school fees are zero. Yet analogous effects could be generated by the substitution of time between parents and children in home production (household chores or own agricultural production), if adult household members increase their labor supply in response to adverse economic shocks (as has been shown by Kochar, 1999; Rose, 2001; Jayachandran, 2006). Given that home production in India – the context of our study – takes up a substantial share in the time allocation of adults, particularly of women, any increase in parental work time is likely to increase the need for children to work at home (Skoufias, 1993; Ilahi, 2000; Shah and Steinberg, 2017).[2]

While we treat schooling investment as market good for reasons of tractability in model, we also establish that analogous risk results could be obtained with any mechanism that relates income shocks to study time. Specifically, we show that our basic specification can be mapped into a model where the primary alternative for children is time spent in home production. We also explore the implications of allowing for precautionary savings in our model, and show that this extension would yield a negative effect of risk *ex ante* and (largely) positive effect *ex post* to the shock.

We then discuss the data. As our primary data set we use the three rounds of the Rural Economic and Demographic Survey (REDS). These data are unique in that they include time allocation of children and of mothers for three seasons during the year in each round. The REDS data are representative of rural India in 1967 and span multiple decades, allowing us to use the long-term variance of rainfall interacted with irrigation as a source of variation in risk within villages over time.

[2]The India Time Use survey from 1998/99 revealed that women spend on average 18.7 hours per week on labor market activities (incl. agricultural production) and 34.6 hours per week on household-related activities. These activities are classified as extended SNA activities in the Time Use Survey and include household maintenance, as well as care for children, sick and elderly. Men, in contrast, spend 42.0 hours on labor market activities, and 3.7 hours on household-related activities. In this paper we focus on home production (own agricultural production and household chores), because these are the main activities that we observe children to be carrying out next to studying.

Our empirical strategy proceeds in two steps. First, we structurally estimate the key parameters of a three-period version of our model using indirect inference (Gourieroux et al., 1993). The parameters of the auxiliary model are: the coefficients from a quadratic regression of third-period study time on first and second period study time and the interaction thereof, the coefficients from a quadratic regression of second-period study time on first period study time, and the first two moments of study time in each season. We show that the quadratic regressions are quite robust to different specification choices including specifications that control for village-by-round of interview fixed effects, and, using the panel households in our data, for household fixed effects and household-specific time trends. The simulated data seem to match the parameters from the auxiliary model well. Our estimated structural parameters point to substantial dynamic complementarity in the demand for human capital, and imply that an increase in income risk should reduce schooling investments due to the presence of dynamic complementarities in production. Finally, we use the estimated model parameters to predict the elasticity of schooling investments with respect to income variance, and find that an increase in variance by 100% would reduce study time by roughly 8.9%.

We then test the predictions of our model at the estimated parameter values by directly incorporating village heterogeneity in income risk. For this purpose we combine the REDS data with monthly precipitation data (ERA5 reanalysis data) obtained from the European Center for Medium-Range Weather Forecasts (ECMWF). We start by replicating results from previous studies on the role of rainfall shocks on agricultural incomes, consumption, and - given the focus of this article - on the time use of children. We then explicitly test the empirical fit of different functional forms in the relationship between precipitation and consumption, and show that – at least in our data – consumption can be best explained by log rainfall interacted with the village level average of area under irrigation.

We use this specification to predict village-level risk. Specifically, we estimate the relationship between household consumption per capita, rainfall, village-level agricultural area under irrigation, and the interaction of these two variables. Using these estimates, we predict consumption outcomes at the village level for each observed rainfall outcome, given the current share of area that is irrigated. We then use the historical rainfall distribution to obtain the probability distribution of rainfall outcomes, and then calculate the probability distribution of predicted consumption for each village and round. Risk, finally, is defined as the variability in the distribution of predicted consumption. This concept of risk exploits the fact that labor markets in rural India are largely dominated by agriculture over the time period we study. Hence, rainfall shocks not only affect farmers' consumption through on-farm production, they also affect consumption levels of landless households who mostly engage in casual agricultural employment.

We explore variation in this variable within villages over time to estimate the effect of risk on study time, and find that risk significantly reduces the probability that children attend school. These results are extremely robust to various specifications, including specifications that control for household wealth, three lags of rainfall, and state-specific shocks.

An important challenge for this work is that, in the presence of dynamic complementarities, past variation in study time will generally lead to lower levels of human capital than would be the case with lower variation but the same mean. This could lead to a cross-sectional negative relationships been school attendance and risk even in the absence of forward looking behavior. But the forward looking behavior, under certain circumstances, will importantly exacerbate any such effects, and thus it is important to distinguish the two effects. We do this by conditioning on the recent history of rainfall shocks. We also interact the history of the rainfall shocks with a dummy that equals one if the child was of school-age in that particular year, thereby allowing the effects of shocks to be more pronounced for school-age children. Our results are robust to these controls, suggesting that we are indeed observing a forward looking risk effect. Our findings imply that a 100% increase in our risk variable would reduce the probability that a child attends school by 4-5%, which is only slightly smaller than the value estimated from the structural model.

We also perform a battery of robustness checks. We show that our results are not driven by underlying differences in household or village characteristics, in the availability of schools, or in the history of weather shocks, nor driven by differential time trends. Finally we explore the potential for alternative explanations being the driver of our results. An important challenge for our analysis is that precautionary savings might generate another kind of *ex ante* relationship between risk and school investment. We explore this avenue, but cannot find any evidence that our results are driven by a savings motive. Our results also do not seem to be driven by differences in the returns to human capital, nor by differential fertility or gender-imbalances.

In order to assess the scope for public policy, we simulate the effects of an income-smoothing policy, modeled after the National Rural Employment Guarantee Scheme (NREGS) in India, on human capital investments. We estimate the extent to which NREGS reduces variability in consumption and use this reduction in risk to simulate the program's effect on child school attendance. We find that a similar program, that held the level of wages fixed, would increase school attendance by 1 percentage point.

The wage effects of such programs should not be ignored, however. Shah and Steinberg (2019) and Li and Sekhri (2020) find negative effects of the NREGS on schooling using the rolled phase-in of NREGS as a source of variation, and argue that the NREGS increased the opportunity costs of time of adolescents and therefore led them to drop out of school at younger ages. Arguably a risk mitigation effect could not be fully internalized by households at the early stages of the program when implementation was spotty and long-term viability was unclear. It is unclear how the direct effect of rising wages and the indirect effect of less variable incomes balance out in the longer term. However, our results suggest the negative wage consequences for schooling of the NREGS might be mitigated to the extent that the program is recognized as a reliable source of support during periods of adverse shocks.

By combining the structural estimation of key model parameters with a more reduced-form analysis of the effect of risk on child schooling, this paper is placed at the intersection of two important strands in the economics literature. We contribute to an exiting literature that explores the determinants of human capital investments in low and middle income countries (see *e.g.* Foster and Rosenzweig, 1996; Glewwe and Jacoby, 2004; Jensen, 2010, 2012; Oster and Steinberg, 2013; Atkin, 2016; Shah and Steinberg, 2017). Most closely related to our work, are the papers by Fitzsimons (2007); Kazianga (2012) and Colmer (2019a). These papers explore the implications of weather variability on schooling in Indonesia, Burkina Faso and Ethiopia, respectively, and focus on precautionary savings motives in explaining the negative effect of risk on schooling (Fitzsimons, 2007; Kazianga, 2012), and the desire to diversify out of agriculture in explaining the positive effect of risk on schooling (Colmer, 2019a). Our paper, in contrast, highlights the existence of dynamic complementarities in the production of human capital, and explores household investment decisions in response to these. There is other work that focuses on how income realizations affect schooling outcomes in a setting where school investments at different points in time are complementary (in particular Jacoby and Skoufias, 1997; Shah and Steinberg, 2017). However, we are not aware of other work that considers the question from an *ex ante* perspective.

We also contribute to an emerging literature that seeks to structurally estimate the parameters of the human capital production function, and the elasticity of inter-temporal substitution in human capital investments specifically (Cunha et al., 2010; Attanasio et al., 2019, 2020; Agostinelli and Wiswall, 2020). Most of this work uses detailed data on skills and investments to estimate a fully dynamic model over the life-time of the child, and focuses on time-intervals of several years. With one exception (*i.e.* Agostinelli and Wiswall, 2020), these papers largely find an elasticity of inter-temporal substitution between early and late investments that is close to one.[3] Our paper concentrates on the substitutability in schooling investments within the school year, and finds an elasticity of substitution that is substantially lower (0.210). Perhaps more importantly, in contrast to previous works, our structural parameters reveal what parents believe to be the true production function, which is arguably the relevant object if one wishes to understand the parental response to risk.[4]

The remainder of the paper proceeds as follows. Section 2 discusses the theoretical model. Section 3 estimates the structural parameters of the model and generates predictions about the effect of risk in our context, and then empirically estimates the effect of risk in India. Section 4 simulates the effect of the NREGS on the outcome of interest, and Section 5 concludes.

2 A Model of Schooling Investment with Dynamic Complementarity in Human Capital Production and Risk

To understand the concept of *ex ante* risk for schooling it is helpful to consider a simple-two period model. In this model child investment is treated as a market good and human capital accumulates

[3]Castro and Villacorta (2019) focus on heterogeneity in the return to schooling and find evidence that inputs and skills are complements at low levels of skills and the reverse, *i.e.* that they are dynamic substitutes, at high levels of initial skill.

[4]Of course, if parental beliefs are incorrect and one wishes to estimate the consequences of risk for achieved human capital it is also necessary to know the true production function.

over both periods. The model emphasizes the dynamic complementarity in the production function of child human capital as in Jacoby and Skoufias (1997); Cunha and Heckman (2007); Cunha et al. (2010). An analogous model in which the only cost of schooling is the opportunity cost of foregone time in home production is presented in appendix A.1. In that model, the parent responds to a negative income shocks by supplying more labor in the labor market *ex post* (as in Kochar, 1999; Rose, 2001; Jayachandran, 2006), while children are required to spend more time in home production (but not in the labor market).[5] The two models are isomorphic as long as the wage effects of shocks are not too large, which is why we work with the simpler model here.

2.1 General Setup

Our initial two period model has an objective function

$$\max \mathrm{L}(s_1, \tilde{s_2}) = ln(y_1 - s_1) + \mathbb{E}_1[ln(y_2 + \tilde{\theta_2} - \tilde{s_2}) + \zeta v(H_2)], \tag{1}$$

where y_i denotes non-stochastic income and s_i denotes school investment in period i. θ_i is an income shock realized in period i, and can can take on two values $[-r, +r]$ with equal probability. The tildes represent variables that are determined at the start of period 2. We abstract from savings for now. H_2 denotes the stock of human capital at end of period 2, ζ describes the utility return to child human capital, and the function $v(H)$ captures curvature in the demand for human capital.

The human capital stock in period 2 increases in school time in periods 1 and 2 according to

$$H_2 = (\nu_1 s_1^{\rho} + \nu_2 \tilde{s_2}^{\rho})^{\frac{1}{\rho}}. \tag{2}$$

for $\nu \in [0, 1]$ and $\rho \in (-\infty, 1]$. The ν_i are share parameters. The degree of dynamic complementarity is entirely governed by ρ, with $\sigma = 1/(1-\rho)$ being the elasticity of substitution between school time in periods 1 and 2.

We first consider the extreme case in which $\rho \to -\infty$ and $v(H) = H$. We will subsequently numerically illustrate the effects for intermediate values of ρ and alternative specifications of $v(H)$. We show numerical solutions for an extended model that incorporates precautionary savings in appendix A.2.[6]

2.2 Special Case: Perfect Complements and Linear Utility

When first and second period school times are perfect complements, second period schooling is never chosen to be higher than first-period investment. We can distinguish two cases. In the first

[5] Note that, in practice, a negative income shock will be correlated with a negative wage shock. In order to see a negative effect of rainfall on parental labor supply, it must be the case that the income effect dominates the wage effect. We show that this is the case in Section 3.2.

[6] Allowing for savings essentially yields a negative predicted effect of risk throughout, which is consistent with the notion of deterring investments in the presence of risk. Yet, in a model with savings, the expected value of second-period (*ex post*) investment actually increases with risk, which is not what we find empirically. We also find no evidence that risk increases or decreases savings across years in our context. We return to this point in Section 3.3.

case, the realized income in the second period is sufficiently high that the household chooses second-period schooling to be equal to first-period schooling but would choose higher investment in period 2 if period 1 investment were exogenously higher. We call this the binding case. In the second case, second-period income is sufficiently low that $s_2 < s_1$. In this non-binding case second-period schooling would not be affected by an exogenous increase in first period investment. In this sense we have the binding case if $y_2 + \theta_2 > s_1 + \frac{1}{\zeta}$, in which case $s_2 = s_1$. Otherwise, $s_2 = y_2 + \theta_2 - \frac{1}{\zeta}$. Note that in the first (binding) case, child schooling does not depend on the income shock, but in the second (non-binding) case, schooling increases with the income shock (θ_2).

By substituting optimal s_2 into Ł and taking expectations over the assumed distribution of shocks, we can then write the maximand as a function of child school time in the first period only. This expression is straightforward in the case that the variance of the shock is small enough so the binding case is achieved – that is as long as things cannot be not so bad in the second period that the household chooses to provide less schooling in period 2 than was chosen in the first period. In particular, assuming we have an interior solution, we can differentiate with respect to s_1. While the resulting expression cannot be solved in closed form, we can write a series approximation (to second order) in the size of the shock for small values of r:

$$s_1 = y1 - \frac{2}{\zeta} - \frac{\zeta}{4}r^2 + O\left(r^3\right) \tag{3}$$

where we assume that $y_2 = y_1$ to keep the solution simple. This result shows clearly that first period schooling is decreasing in the variance of the shock in the binding case.

On the other hand, if r is larger, such that $s_2 < s_1$ in the bad state of the world, s_1 can be approximated by

$$s_1 = y_1 - \frac{3}{\zeta} + \frac{1}{3}r - \frac{2\zeta}{27}r^2 + O\left(r^3\right). \tag{4}$$

This expression suggests that, at least for moderate values of r, first-period schooling could rise with r.[7] In the bad state, when $s_2 < s_1$, human capital depends only on second period investment given the assumption of perfect complements. This result implies that the *ex post* return to first-period schooling is zero and does not decline further with increasing variance. On the other hand, in the good state, human capital depends on both first and second period schooling. As the variance increases, the utility gain of additional schooling in the second period under the good state increases, as does the *ex post* return to first-period investment. Thus, in expectation over both states, the expected return to first-period investment rises in r.[8]

But interpreting these expressions is challenging because (a) they are approximations in r

[7] However, note that in the bad state, $s_2 = y_1 - r - 1/\zeta$ such that it declines faster in r than s_1 (and s_2) increases in r in the good state. Therefore, the expected value of second-period investment and thus expected human capital is always declining in the variance the shock.

[8] This result has an analogy with respect to borrowing in the case of investment risk: increasing return variance can increase the willingness to take a loan at a given interest rate because it raises the return in the good state and does not decrease it in the bad state, because defaulting is the only option in the bad state throughout.

for small r and (b) which equation applies depends itself on the magnitude of r. A numerical simulation helps to clarify these issues. In figure 1, we graph s_1, s_2 by state, and $\mathbb{E}[H_2]$ by the standard deviation of the shock ($\sigma_\theta = r$), assuming parameter values of $y_1 = 10$, $y_2 = 10$, and $\zeta = 1$.

For low values of r, second period investment equals first period investment, which then also equals total human capital, regardless of the state of the world, and all three measures decline in risk. At $r \approx 1.6$ the measures divide. Now the household chooses $s_2 = s_1$ in the good state of the world, but in the bad state s_2 is lower than s_1. Note that initial period investment is now rising in r over this range. However, second period schooling falls with risk (and faster than s_1 rises), such that the expected value of human capital declines with risk throughout. We can think of this as a combination of an *ex ante* and *ex post* response in the presence of dynamic complementarity. Clearly, for given first period investment, a given mean preserving increase in the gap between school time in the good and bad states results in lower human capital accumulation as would be the case in an *ex post* model. But the fact that first-period schooling is lower throughout than in the no-risk case, and thus caps good-state investment, reflects the *ex ante* response.

Note that this pattern is the opposite of what would happen under a precautionary savings motive. In a typical precautionary savings model, period one investment is a substitute for income in the second period, and thus one avoids the downside risk by reducing consumption in the first period as risk increases. For human capital with a complementary production function, first period investment is a complement with second period investment. Thus one invests less in the asset (human capital) when faced with greater risk.

2.3 The Effect of Risk for General Values of ρ and ϕ

While the case of perfect complements nicely illustrates the *ex ante* effect of risk, it imposes strong assumptions of the production function of human capital. We therefore present numerical solutions for more general values of ρ and curvature in the demand for child human capital.

Before we do this, we introduce a new concept, namely dynamic complementarity in the demand for human capital. The demand for child human capital exhibits dynamic complementarity if $d^2v(H_2(s_1, s_2))/ds_1 ds_2 > 0$, so that an increase in first period schooling leads to an increase in the marginal utility of second period schooling.

In particular, let the utility function for child human capital exhibit constant relative risk aversion (CRRA):

$$v(H) = \frac{1}{1-\phi}(H)^{1-\phi}, \tag{5}$$

with $\phi \geq 0$.[9] With parental utility as described by eq. (5) and a child human capital production function as described by eq. (2), the demand for child human capital will exhibit dynamic complementarity if and only if $1 - \phi - \rho > 0$. To see this, substitute eq. (2) into (5), and consider the

[9] At $\phi = 1$, $v(H) = \ln(H)$.

cross-derivative with respect to s_1 and s_2:

$$\frac{d^2v(H_2(s_1,s_2))}{ds_1ds_2} = (1-\phi-\rho)H_2^{(1-\phi-2\rho)}(s_1s_2)^{\rho-1}v_1v_2. \tag{6}$$

For any $v_1>0$, $v_2>0$, $s_1>0$ and $s_2>0$, this expression is strictly positive as long as $1-\phi-\rho>0$.

A few numerical examples illustrate the relationship between $d^2v(H_2(s_1,s_2))/ds_1ds_2$, ϕ and ρ. For $\phi=0$, i.e. linear utility with respect to child human capital, there will always be dynamic complementarity in demand, except in the extreme case where the human capital production function exhibits perfect substitutability ($\rho=1$). On the other hand for logarithmic utility ($\phi=1$), school investments at different points in time will be dynamic complements if and only if the production function exhibits dynamic complementarity (*i.e.* $\rho<0$). Finally, with $\phi=2$ school investments at different points in time will be dynamic complements for any $\rho<-1$.

Equipped with this insight, we numerically solve for optimal s_1 at three representative values of ϕ and at various values of ρ. In particular, we calculate outcomes for three functional forms of $v(H_2)$: $v(H_2)=H_2$, $v(H_2)=\ln(H_2)$ and $v(H_2)=-1/H_2$, which correspond to the three values of ϕ that were discussed above (i.e. $\phi=0$, $\phi=1$, $\phi=2$).[10]

Figure 2 shows the derivative of s_1 with respect to r for these three functional forms and $-3.5\le\rho<0.8$. Since we are plotting against $\ln(\sigma)$, the y-axis captures the Cobb-Douglas case. As $\ln(\sigma)\to 1.5$, the production function approaches a simple additive function in school times, *i.e.* the case of perfect substitutes. And as $\ln(\sigma)\to -1.5$, the human capital production function exhibits increasing dynamic complementarity.

As can be seen, the numerical solutions suggest that the sign of the effect of an increase in r on first-period human capital investment maps 1:1 with the presence of dynamic complementarity in demand for human capital as described above. That is, with linear $v(H_2)$ the predicted effect of risk remains negative throughout, and approaches zero only as the human capital production function exhibits perfect substitutability ($\rho=1$). With concave $v(H_2)$, this is no longer the case. Second period risk will then affect the expected marginal utility of the human capital stock, and households choose to invest more in first-period schooling in order to avoid very low stocks of human capital for their children. If $v(H_2)=\ln(H_2)$, for example, first period schooling will increase in risk for all $\rho>0$. If $v(H_2)=-1/H_2$, the predicted effect of risk on human capital investments turns positive at values of $\rho>-1$.[11]

In contrast, child human capital declines for almost all specifications in these simulations. Figure 3 presents the effect of an increase in r on the expected value of human capital for different values of ϕ and ρ. Not surprisingly the largest effects are observed for linear preferences and highly complementary inputs in the production function and the smallest effects (that turn weakly

[10] Here we set $y_1=y_2=10$, $\zeta=1$, $\nu_1=\nu_2=1/2$ and $r=1$.

[11] Note, that the effect of an increase in risk is not constant in r. Importantly, with linear v, and values of $r\ge\frac{y_1}{2}$, the effect of risk on first-period human capital investment turns positive at very low values of ρ, which is analogous to the case of perfect complements described above. The effect of risk on first-period school maintains its sign for values of $\phi>0$. Numeric approximations for $y_1=y_2=10$, $\zeta=1$, $\nu_1=\nu_2=1/2$ and $r=5$ are depicted in figure C.1 (figures C.1 to C.7 are available in appendix C).

positive) arise with a strong degree of risk aversion and very substitutable production. This latter result arises because, with substitutability, human capital becomes a mechanism for precautionary savings. In any case, there are moderate negative effects over much of the range. For example, when the elasticity of substitution in human capital production is one, the effect of 0.1 increase in r on $\mathbb{E}[H]$ is -0.009 when $\phi = 1$ and -0.026 when $\phi = 0$.

Whether variance increases or decreases first-period investment, at least in a two-period model, thus depends on the presence of dynamic complementarity in the demand for human capital. Dynamic complementarity in the demand for human capital is essentially a composite of the dynamic complementarity in human capital production and the curvature in parental demand for child human capital. While there is emerging consensus about the complementarity in the production, this does not necessarily map perfectly with what parents believe to be the true parameter. Furthermore, the shape of human capital preferences is unclear, and – to the best of our knowledge – has not been assessed empirically so far. Within the school year, which is how we approach the problem in this paper, one might anticipate that parents have a strong desire for their children to accumulate enough human capital to pass on to the next grade but care less about marginal returns beyond that. This perspective would then yield quite concave preferences. Likewise, one could expect complementarity in the production of human capital to be more pronounced within the school year than it would be across school years. Yet, deriving clear predictions on the effect of risk on human capital investment remains impossible without an intuition about the values of ϕ and ρ. Below we propose a novel approach for estimating these values.

3 Empirical Strategy and Results

To empirically estimate the effect of risk on investments in schooling we use the Rural Economic and Demographic Survey (REDS) data, which were collected in rural India in 1981-82, 1998-99 and 2007-08, and combine it with high-resolution precipitation data (Copernicus Climate Change Service, 2017). The REDS collects detailed information on time use for all women in the household and their children. This information refers the hours per day allocated to an extensive list of activities on a typical day, and is collected for three seasons of the year, which are also marked by very different levels of agricultural activity. The reference months are: October/November (season one), February (season two), and April/May (season three). We provide a detailed description of the data used in the empirical analysis in appendix B.[12] Our main analysis proceed in two steps. First, we structurally estimate the key model parameters in the REDS data using indirect inference. This allows us to generate predictions about the effect of risk on schooling investments in the context we study. Second, we test the predictions empirically by exploiting the variation across survey rounds in village-level irrigation, and income risk.

[12]Summary statistics at the household level and the individual level are reported in Tables D.1 and D.2 (Tables D.1 - D.29 are available in appendix D.

3.1 Estimating Key Model Parameters

We now consider how to best use the time allocation data available in the REDS survey to structurally estimate the parameters ϕ and ρ. One approach would be to set up a fully dynamic model over the lifetime of the child. But while we do have panel data, the fact that rounds are collected at intervals of multiple years means that we really only have a single year of each child's school-age life.[13] Moreover, the only measures of human capital available in the data are years of schooling completed at the beginning of the reference period, which does count repeated grades and years in (voluntary) pre-school. Even if we did have information on whether a child progressed to the next grade that is likely to be an extremely imperfect measure of human capital acquisition over the year. We also don't have data on income or consumption at the level of the season. The special value of these data is the time allocation data within a year, so we design our approach around the relationship between these measures.

3.1.1 Time Use across Seasons in the REDS Data

Figure 4 presents a non-parametric smoothed version of season three school time as function of s_1 and s_2 data. The two horizontal axes are school time in the first two seasons and the the vertical axis is school time in the third season. As might be expected, the surface slopes upward toward the point where s_1 and s_2 are highest. This pattern may reflect complementarity in human capital production but it also may simply indicate that those households with higher demand for human capital are likely to have higher school time in all three periods. More interesting is the fact that along a cross-section where $s_1 + s_2$ is a constant, s_3 rises when s_1 and s_2 are more different.

This pattern turns out to reflect a robust empirical relationship between the three time allocation measures. We focus, in particular, on a quadratic approximation to this surface of the form

$$s_{3i} = \alpha_0 + \alpha_1 s_{1i} + \alpha_2 s_{2i} + \alpha_3 s_{1i}^2 + \alpha_4 s_{2i}^2 + \alpha_5 s_{1i} s_{2i} + \alpha_6' x_i + \epsilon_{3i} \tag{7}$$

where x_i denotes a vector of individual, household, and village characteristics.

Table 1 presents estimates for various specifications. We note that the coefficient vectors are very stable across various specifications.[14] In particular we see that third-period schooling is increasing and convex in the schooling of each previous period but that the coefficient on the interaction between s_1 and s_2 is negative. It is tempting to interpret the positive linear coefficients as evidence of complementarity but given that the schooling times are jointly determined it is easy to imagine that this is generated by heterogeneity in parental return to human capital (ζ) across households and children. It is perhaps less clear what combination of forces would generate the negative

[13] It is also not possible to track individuals across survey rounds in the REDS data. While we can follow a subset of households over time (and control for household fixed effects in part of the analysis), we cannot know the final human capital stock of children interviewed in previous rounds.

[14] These results are also unaffected by controlling for household fixed effects, or household-specific time trends, in the subset of panel households, see table D.3. Because season three could overlap with summer vacations in southern states of India, but not in the northern states, we also test if our results are robust to restricting our data to the northern states. We show in table D.4 that this is indeed the case.

interaction effect, which was also evident in the descriptive graph.

3.1.2 Three-period Model of Schooling Investments

Consider then an augmented version of eq. (1) that covers three periods

$$\begin{aligned}\max \mathrm{L}(s_1, \tilde{s_2}, \tilde{s_3}, a_1, \tilde{a_2}) = & ln(y_1 - s_1 - a_1) \\ & + \mathbb{E}_1[ln(\tilde{y_2} - \tilde{s_2} + a_1 - \tilde{a_2}) + ln(\tilde{y_3} - \tilde{s_3} + \tilde{a_2}) + \zeta v(H_3)]\end{aligned} \tag{8}$$

where the a_i denote savings in period i.[15] Human capital production is given by:

$$H_3 = (\nu_1 s_1^{\rho} + \nu_2 \tilde{s_2}^{\rho} + +\nu_3 \tilde{s_3}^{\rho})^{\frac{1}{\rho}}. \tag{9}$$

with $\nu_i \in [0, 1]$ and $\sum_1 \nu_i = 1$ for $i \in \{1, 2, 3\}$. $v(H)$ is defined by eq. (5). In the structural analysis we will endogenize savings but constrain it to be non-negative.

Some intuition for the forces at play in this model may be garnered by a simple comparative static exercise with savings set to zero. Consider fixing $s_2 + s_1$ but allowing the difference between the two values to vary (as discussed in the context of figure 4). In the case of perfect substitutes ($\rho = 1$), the difference should not matter for the choice of s_3 because the stock of human capital at the end of period 2 will be invariant to the difference. As the two initial times become more complementary the difference will matter net of the sum – the larger the difference for given sum the lower will be the human capital at the end of the second period. But how much second period human capital affects third period investment will depend not only on the elasticity of substitution in production but also on ϕ. If ϕ is large, such that preferences for human capital are strongly concave, high second period human capital could result in lower third period investment, and vice versa.

To formalize this idea we can reparametrize $st_{12} = s_1 + s_2$ and $sd_{12} = s_1 - s_2$. An optimal choice of s_3 implies that

$$\frac{d\mathrm{L}(st_{12}, sd_{12}, s_3)}{ds_3} = 0 \tag{10}$$

We then implicitly differentiate the first order condition to get

$$\frac{ds_3^*(st_{12}, sd_{12})}{dsd_{12}} = -\frac{\partial^2 \mathrm{L}^*(st_{12}, sd_{12}, s_3)}{\partial s_3 \partial sd_{12}} / \frac{\partial^2 \mathrm{L}^*(st_{12}, sd_{13}, s_3)}{\partial s_3^2} \tag{11}$$

and analogously for $\frac{ds_3^*(st_{12}, sd_{12})}{dst_{12}}$. Because the denominator of (11) must be negative for an interior solution, the sign of this implicit derivative will be determined by the sign of the numerators. These

[15] Note that we let the ys be stochastic rather than adding a season-specific shock (θ) in this specification. Of course both approaches yield identical results.

are:

$$\text{sign}(\frac{\partial^2 \mathbb{L}^*(st_{12}, sd_{12}, s_3)}{\partial s_3 \partial st_{12}}) = \text{sign}(1 - \phi - \rho) \tag{12}$$

and

$$\text{sign}(\frac{\partial^2 \mathbb{L}^*(st_{12}, sd_{12}, s_3)}{\partial s_3 \partial sd_{12}}) = \text{sign}(1 - \phi - \rho)\,\text{sign}(s_1^{\rho-1} - s_2^{\rho-1}). \tag{13}$$

The first expression suggests as one might expect that as long there is complementarity in demand an increase in the total $s_1 + s_2$ should increase third period schooling time. But if this is the case then one should also see that an increase in the absolute difference $|s_1 - s_2|$ will decrease third period school time as long as the school time in the first and second period school time are not perfect substitutes.[16]

But this is not what is observed in figure 4. One does observe that, along a cross-section in which $s_1 - s_2$ is fixed, second period schooling rises with the sum of school times in the first two periods. The effect of moving along a cross-section in which $s_1 + s_2$ is fixed is u-shaped. In order to reconcile the data on the model it is necessary to think more carefully about the processes that generate s_1 and s_2 in the first place. Or, from another perspective, we need to think about possible biases that might arise in trying to apply a structural interpretation to the coefficients in eq. (7). A likely source of misspecification, as noted, is heterogeneity parental return to human capital (ζ), which might generate a very specific pattern of how study time reacts to shocks for different groups of households and children.

3.1.3 Indirect Inference

The specific structure of the REDS data makes it attractive to estimate our model using an indirect inference procedure (Gourieroux et al., 1993) rather than trying to fit the data to the model directly. We use the relationship between study times in the three periods as primary source of identification.

We construct the underlying data generating process as follows. First, we assume that income in each period takes on two values with equal probabilities y_L and y_H, with realizations of income being independently drawn across time. We also incorporate a buffer stock model in which savings are nonnegative and must be fully consumed by the end of the year. This basic structure fits with the idea of a rural village with underdeveloped financial markets in which most savings are in the form of stored agricultural products and there is little systematic savings from year to year.[17] Finally, to account for heterogeneity we assume that the parameter ζ takes on two values ζ_L and ζ_H with equal probability.

[16] Note that the sign is on $(s_1^{\rho-1} - s_2^{\rho-1})$ is negative as long as $s_1 > s_2$ and $\rho < 1$, while the difference (sd_{12}) is positive, and vice versa, such that s_3 is always declining in the absolute difference $|s_1 - s_2|$. The full derivatives are shown in appendix A.3.

[17] Note that because we will estimate the values of income in the two states we implicitly allow for other forms of state-contingent transfers if present. E.g., for a fully insured household one should estimate that there is no difference between high and low incomes.

The simulated data will take on sixteen values: eight combinations of realized states and two types, while the actual data have an effectively continuous distribution. However, the simulated data can be used to estimate the same auxiliary model described by eq. (7) as well as a comparable equation that captures the relationship between first and second period schooling:

$$s_{2i} = \gamma_0 + \gamma_1 s_{1i} + \gamma_2 s_{1i}^2 + \gamma_3' x_i + \epsilon_{2i}. \tag{14}$$

While estimating equations (7) and (14) in the actual data, we also control for variables such as age and village characteristics that are relevant to the actual data but are not explicitly modeled in the data generating process, which focuses on the allocation process within a year for a child of given age.

Our structural estimation proceeds as follows. We start with a vector of parameters $\Psi = [\zeta_L, \zeta_H, \phi, \sigma, \delta, y_L, y_H]$ where we translated ρ into the elasticity of substitution $\sigma = 1/(1-\rho)$ for ease of interpretation, and δ captures, in a simple way, the relative impact of study time on human capital at different points of the year so that

$$\nu_i = \frac{\delta^{i-1}}{(1+\delta+\delta^2)}, \tag{15}$$

for $i \in \{1, 2, 3\}$. For each type of household and set of realization of y_i in periods 1, 2, and 3, we then numerically solve for the optimal values of the state contingent s_i and a_i given the parameter vector Ψ. We then regress the resulting observed s_3^* on a quadratic function of s_1^* and s_2^* using the fact that different states (combinations of household type and income realizations) are, by assumption, realized with equal probability. We similarly regress s_2^* on s_1^*.

We then compute the squared difference between the vector of estimated coefficients and the first two-moments of the s_i from the simulated data $B(\Psi)$ and their corresponding results from the actual data, $\hat{B}$, using the inverse of the variance-covariance matrix V of the estimated $\hat{B}$ to weight the squared differences

$$(B(\Psi) - \hat{B})' V^{-1} (B(\Psi) - \hat{B}). \tag{16}$$

We then iterate over Ψ to minimize the resulting quadratic form.[18]

Of particular interest is the extent to which the model captures the basic structure of the data on school time over the relevant ranges. Table 2 presents the targeted parameters and the simulated parameters of the auxiliary model. The coefficients are appropriately signed and of the correct magnitude and within about 30% of the correct value. Also, the mean school times are a bit high the standard deviations are about half of what they should be.

More important than the specific coefficients is the ability of the model to capture the overall pattern of the relationship between the three schooling measures. Figure 5 provides a picture of

[18] We use a global optimization method (Storn and Price, 1997). In this iteration, ϕ and σ are constrained to be nonnegative, the ζs and ys are nonnegative, and $\delta \in [0, 2]$. Maple code available on request.

the quadratic surfaces for the regressions based on the actual (lighter) and simulated data (darker). The horizontal axes are s_1 and s_2 and the vertical axis is s_3. The basic shapes of the curve are similar, with the simulated data lying above the actual data for low values of s_1 and high values of s_2 but roughly coinciding elsewhere.

Table 3 presents the estimated parameter vector Ψ with standard errors along with the sensitivity matrix Λ (Andrews et al., 2017) in the form of semi-elasticities, that describes how the matched moments (α and γ estimates) map into the parameters of interest and thus illustrate the primary sources of identification for each parameter of the model. Standard errors are calculated based on the variance-covariance matrix of the estimates from the actual data.[19]

The estimates of the structural parameters (presented in Panel A) are relatively precise in percentage terms. The elasticity of substitution (σ) is 0.210, which corresponds to a value of $\rho = -3.76$. There is thus a strong degree of complementarity in the production function. But preferences over human capital are also quite concave. The important difference $1 - \phi - \rho = 3.218$ indicating that demand is overall complementary. We also see substantial heterogeneity in demand as the gap in the preference parameter ζ is 8.470, a difference of 96% over the lower value. There is also considerable uncertainty in income with the high income level being being 70% above the low income level. Finally, the δ of 0.919 indicates that the ν_i are close to equal. Study time in the three periods contributes almost equally to human capital.

Panel B of table 3 describes the sensitivity matrix Λ as noted above. The pattern is not straightforward to interpret, suggesting a complex interrelationship between the coefficients from the auxiliary model and the structural parameter estimates. There is little sense that one or more parameters is particularly "identified by" any one of the coefficients. It is notable that all the coefficient values (αs and γs) lead to increases parental return to human capital estimates (both ζ_L and $\Delta\zeta = \zeta_H - \zeta_L$). Households with higher human capital returns should have uniformly higher study time than those with lower returns, thus leading to a more positive relationship between the study time measures in the population as the ζs becomes more heterogeneous. The effects are quite large. A one standard deviation (.048) increase in α_1, with other parameters fixed, results in a 98.0% increase in the estimate $\Delta\zeta$. Conversely, increases in the coefficient estimates lead to decreases in the estimates of the elasticity of substitution. This result reflects the fact that higher complementarity induces higher correlations in school-time across periods. The negative semi-elasticity of -2.993 for α_5 and σ indicates that a one standard deviation increase (.005) in the negative season 1 $\times$ season 2 interaction would yield a 1.5% decrease in the elasticity of substitution. We can also use the sensitivity matrix to assess by how much our structural parameters would change if we used the estimates in Table D.3 (panel households) in the auxiliary model. Using these alternative moments would reduce the estimate of ϕ by 19%, and increase the estimate of σ by about 6%, such that $1 - \phi - \rho = 3.249$, which is very close to the original estimate.

We can now turn to the central question of interest, whether, given the estimated parameters

[19] If $S = dB(\Phi)/d\Phi'$ and because our weighting matrix is V^{-1} sensitivity $\Lambda = (S'V^{-1}S)^{-1}S'V^{-1}$ and the parameter variance-covariance matrix is $\Sigma = \Lambda V \Lambda'$.

of the model, an expansion in the variance of the effective income distribution would result in lower investment in each period as well as in a lower stock of child human capital. We do this by numerically estimating the derivative of each measure of s_i by state and of H_i to variance, by comparing baseline estimates to estimates from a mean preserving perturbation of the estimated parameters for y_L and y_H. The estimated levels of investment and elasticities with respect to variance for the lower and higher human-capital return households and for the population on average are presented in table 4. Note that because our model allows for buffer stocks savings, some of the estimated effect may incorporate a precautionary savings effect. To distinguish the dynamic complementarity effect from the precautionary effects we thus compute these elasticities with the model as specified and then with the same model parameters but constraining savings to zero. Panel A presents the results that incorporate a potential savings effect, as our model allows for buffer stock savings. In panel B, we present the corresponding results using the same parameter estimates but with savings constrained to be zero.

We find that – for the estimated model parameters – the effect of an increase in variance is to lower human capital investment and this effect is proportionately larger in the households with higher returns to human capital. For example, doubling the variance of income for given mean results in a 7.6% decline in study time in season 1 on average over both types of households. The figure in Panel A is 8.4% for high return households and 6.7% for low return households. The negative effects are observed for study time in each period and this same change leads to an 8.4% percent decline in human capital by the end of the year on average. Note that the elasticities vary over season when we allow for savings, which is essentially driven by the fact that households who experience positive income shocks in periods one and two save part of their income. The period 2 elasticity thus combines a dynamic complementarity effect with a precautionary savings effect. In Panel B, when savings is restricted to zero, The resulting elasticities are very similar on average over all seasons, with substantially lower variation across season. The elasticities of total human capital are almost identical to the ones in Panel A, *i.e.* doubling the variance of income would lower investments by 8.9% on average over all seasons, and total human capital by 8.7%.

3.2 Estimating the Effect of Risk on Schooling Investments

With these estimates in hand, we can now turn to estimating the effect of income risk on schooling in our data. Our concept of risk exploits two key facts from rural India. First, agriculture is the dominant source of income for the majority of households over the time period of our study, either through own agricultural production or through casual agricultural employment. Second, agriculture is inherently risky. A large body of literature has documented that agricultural yields, wages and employment levels in rural India are strongly influenced by rainfall conditions (see e.g. Jayachandran, 2006; Shah and Steinberg, 2017; Kaur, 2019). High rainfall leads to good harvests, high demand for labor, and high wages. In contrast, low rainfall levels lead to poor harvests, and low demand for agricultural labor.[20] Second, the use of irrigation in agricultural production expanded rapidly in rural India over the time-period covered by our data, reducing the exposure to weather shocks in some places but not in others (Duflo and Pande, 2007; Mondal et al., 2015; Ambika et al., 2016).[21]

3.2.1 Rainfall Shocks in Rural India

Before discussing the construction of our risk variable, we first show that rainfall shocks matter economically in our context. In particular, we test if rainfall shocks affect agricultural income per capita (in logs), consumption expenditures per capita (in logs), and – given the focus of this article – the time allocation of children per day to home production (own agricultural production and household chores) and to studying (in school or at home).[22] Similarly to previous studies that look at the role of rainfall shocks in determining economic outcomes in India, we focus on variation in relative rainfall (see e.g. Jayachandran, 2006; Shah and Steinberg, 2017; Kaur, 2019). Given the structure of our data, we construct historical rainfall distributions for each village and estimate if deviations from village average rainfall affect the outcomes of interest. Exploring the exogenous nature of rainfall shocks, we estimate

$$y_{ijst} = \beta_0 + \beta_1'\theta_{jst} + \beta_2 p_{jst} + \Sigma_{n=1}^{3}\beta_{2+n}\theta_{jst-n} + \gamma_j + \psi_{st} + \epsilon_{ijst}, \tag{17}$$

where y_{ijst} is any of the outcome variables discussed above, observed in year i, village j, state s and year t. θ_{jst} is a vector of dummies that indicate in which quintile of the historical rainfall distribution (of each village) the current realization of rainfall is located. We also control for the share of village agricultural area that is irrigated p_{jst}, for rainfall (in log mms) in the last 3 years $\Sigma_{n=1}^{3}\theta_{jst-n}$, for village fixed effects γ_j, and for state-by-round of interview fixed effects ψ_{st}.[23] We report standard errors that are corrected to allow for clustering at the level of the grid cell by round of interview, as well as Conley (1999) standard errors that allow for arbitrary spatial auto-correlation (300km cut-off, Bartlett weights).[24]

We present the fitted values obtained from estimating eq. (17) in figure 6. In line with previous work, our findings suggest that low rainfall, measured as rainfall outcome that falls below the 20th percentile of the historical rainfall distribution in a particular village, severely reduces incomes from agricultural production and consumption expenditures, while rainfall above the 80th

[20] More recent work also highlights the role of temperature on agricultural output in India: directly, because it affects the pace of crop growth, and indirectly, because temperature influences the rate at which water evaporates from the ground, and thus soil moisture levels (Mondal et al., 2015; Colmer, 2019b).

[21] Irrigation water in India mainly stems from two sources: dams (surface water irrigation) and bore-wells (groundwater irrigation).

[22] Agricultural income is the total value of agricultural production corrected for (paid-out) operating expenses (but not for the value of family labor or the rental value of owned machinery).

[23] When the dependent variable is child time use, we additionally control for age-by-gender and season-by-round of interview fixed effects.

[24] We use the spatial-HAC standard error correction implemented by Hsiang (2010) and Colella et al. (2019). The Conley standard errors are largely identical to standard errors clustered at the grid cell by round, c.f. table D.6.

percentile (60th percentile) of the historical rainfall distribution seems to increase agricultural output (consumption), see panels A and B. We also find that low rainfall is associated with more time spent in home production (i.e. own agricultural production or household chores), and less time spent studying (in school or at home). Conversely, high rainfall (i.e rainfall above the 80th percentile of the rainfall distribution) reduces the time spend in home production, and increases the time spent studying (see panels C and D of figure 6). These effects are economically meaningful: Moving from the lowest rainfall quintile to the highest rainfall quintile increases the time spent in school on a typical day by a little over 40 minutes on average. These results differ from previous work: Shah and Steinberg (2017) find that positive rainfall shocks are associated with lower school attendance among children in rural India. Given that the rainfall-school effect reflects the sum of opposite-signed income and substitution effects it is perhaps not surprising that the overall effect will vary given different sampling schemes, time periods, and measures of educational investment. In any case, what matters for our purpose is that rainfall affects the choice of study time, not the sign of this effect.

3.2.2 Consumption Variability as Measure of Risk

After having established that rainfall shocks matter economically in our sample, we turn to developing a measure of risk. In order to do this, we need to move away from a relative measure of rainfall to an absolute one. In other words, we need an estimate of the absolute effect of rainfall (conditional on village characteristics) on the outcomes of interest. Given the growing importance of irrigation in our context, we allow the absolute effect of rainfall to vary with the availability of irrigation. We hypothesize that a higher the share of irrigated land in a village will mediate the effect of rainfall shocks on household-level outcomes. This could happen because households are more likely to have irrigation on their own land in villages with a high share of area under irrigation or because (wages and employment in) casual agricultural employment will be less dependent on good rainfall realizations (as only a small share of agricultural production in the village would be solely rain-fed).

We estimate

$$y_{ijst} = \beta_0 + \beta_1 \theta_{jst} + \beta_2 p_{jst} + \beta_3 \theta_{jst} p_{jst} + \Sigma_{n=1}^{3} \beta_{3+n} \theta_{jst-n} + \gamma_j + \psi_{st} + \epsilon_{ijst}, \tag{18}$$

where θ_{jst} now is the total precipitation in the current agricultural year (in log mms).[25] All remaining controls are defined as in eq. (17). y_{ijst} is the outcome of interest. As above, we estimate the effect of rainfall on agricultural income per capita (in logs), and consumption expenditures per capita (in logs). We also estimate the effect of rainfall on maternal time allocation to household chores and to the labor market, and on the time children allocate to own agricultural production, household chores, studying, and to leisure. In the time-use specifications, we disaggregate the

[25] We show in tables D.7 and D.8 that specifications that use rainfall in levels or in squares do not fit the data better than those that use the log of rainfall.

sample by gender (child-level), and control for age and for season-by-round (of interview) fixed effects. Standard errors are again corrected to allow for arbitrary spatial auto-correlation.

Identification in this exercise relies on the exogenous nature of rainfall, and – because we are interested in the coefficients on irrigation and on its interaction with rainfall – on the assumption that irrigation at the village level is uncorrelated with the error term. The village fixed effect removes any time invariant heterogeneity, but cannot control for unobserved time-varying shocks that might affect the outcomes of interest and irrigation availability simultaneously. The identifying assumption is violated – for instance – if unobserved shocks or underlying trends affect the availability of irrigation at the village level as well as the covariance between consumption outcomes and precipitation in this village. Under which conditions would these violations generate a negative bias on the coefficient on the interaction term? If villages that are wealthier initially experience a faster expansion in irrigation, and also have better means of self-insuring against shocks, then this could bias our estimates. Similarly, unobserved shocks that improve self-insurance, and irrigation availability, would generate a similar negative bias. In both cases this would not necessarily mean that our risk variable does not capture risk, since improved self-insurance still generates a consumption portfolio that is less exposed to shocks. More worrisome are violations of the identifying assumption that might have direct or indirect effects on investments in child human capital through other mechanisms than risk. An example could be that places that have experienced adverse climatic conditions had to diversify out of agriculture. The few farmers that remain are profitable only because they irrigate their land. These places would probably also display a lower correlation between rainfall and consumption, but not because agricultural production became less risky. And places with a more diversified economic structure probably also have higher returns to human capital, which would incentivize parents to invest more in the human capital of their children. We return to this point in the robustness checks.

Estimates of eq. 18 are reported in tables 5 and 6. In table 5, we present results with and without controlling for village fixed effects to explore the robustness of our estimates. Controlling for village fixed effects seems to attenuate the coefficients somewhat, but the interpretation remains largely identical. Analogously to figure 6, we first explore the relationship between rainfall, irrigation and agricultural income (cols. (1) and (2)). Our results suggest that higher rainfall increases agricultural income; but that rainfall affects these outcomes to a smaller extent in villages with high irrigation levels. Columns (3) and (4) explore the relationship between rainfall, irrigation and consumption expenditures. The effects are somewhat smaller, but more precisely estimated, and suggest a similar relationship: higher rainfall increases consumption, but less so in villages with a higher share of agricultural area that is irrigated. Consistent with the notion of an *ex-post* labor supply response by the parent, log rainfall is negatively (positively) associated with the time mothers allocate to the labor market (household chores). Again these associations are weakened by the availability of irrigation.[26] Note, that we can only observe the time-allocation of the mother in our data, not of the father. We can expect her response to affect children's time allocation as long

[26]The coefficients are statistically significant only for time allocated to the labor market, not for household chores.

as the mother's response to shocks is not perfectly undone by the father's response (i.e. the mother increases her working hours but the father reduces working hours by exactly the same amount, and vice versa).

The effects of rainfall on child time use are reported in table 6. For brevity, we only report estimates that control for village fixed effects.[27] Again, rainfall is negatively associated with children's time spent in own agricultural production and on household chores, and positively associated with time spent in school or studying at home. As with mothers, the availability of irrigation at the village level seems to weaken these associations. These findings are consistent with a model in which parental and child time are substitutes in home production.[28] While responses to shocks seem to dominate in the farm-sector for boys, they are more pronounced in household chores for girls (at least in absolute terms). The schooling effects, then, are roughly similar for boys and girls (if anything slightly more pronounced for girls). Finally, we explore if rainfall shocks affect the leisure time of children. We find weak evidence that boys' leisure increases with rainfall, but no evidence of an association between leisure and rainfall for girls. This suggests that boys can reduce (increase) leisure time to some extent when faced with negative (positive) shocks, while girls cannot.[29]

With the estimates of column (3) of table 5, we then simulate the variability of consumption for each household in our sample at a particular point in time.[30] First, using the annual rainfall data from 1979 to 2019, we calculate the probabilities of all rainfall outcomes per grid cell.[31] Second, we predict consumption expenditures per capita (in logs) for each rainfall outcome observed in our data given the availability of irrigation in the village in a particular round of interview. Third, we construct the probability distribution of consumption outcomes for each household in a particular village and round of interview by multiplying the probability of each of these rainfall outcomes with the predicted log consumption at that rainfall level (given irrigation). Finally, we calculate the interquartile range as well as the standard deviation of a household's predicted log consumption per capita as measures of risk.

Figure 7 illustrates our calculations. Panel A depicts the historical rainfall distribution of all REDS villages, grouped by quartiles of the share of agricultural area that was irrigated in 1982. We can see small differences in the distribution of rainfall, but no clear link between the distribution of rainfall and the availability of irrigation. Panels B to D, then, plot the distribution of predicted consumption per capita (in logs) in the years 1981-82, 1998-99 and 2007-08, respectively. Again the villages are grouped into quartiles of irrigation availability in 1981-82. As expected,

[27]The corresponding results without village fixed effects are largely identical, as reported in table D.9.

[28]We cannot observe paternal time allocation in our data. Given that the effect of maternal time allocated to household chores is not statistically significant, while we still see strong responses of children, it seems likely that fathers also reduce the time they spend on chores and on-farm.

[29]Note that leisure time includes sleep time.

[30]This approach is inspired by Dercon and Christiaensen (2011). We also use the estimates in column (4) to calculate the risk variable, and repeat the main regression as robustness check. The results are largely identical, as reported in table D.10.

[31]The probabilities are obtained by dividing the sample rainfall distribution in 0.025 intervals of annual log rainfall. We then calculate the historical probability of village-level rainfall to fall in each of these intervals.

the villages with the highest levels of irrigation in 1981-82 show the least variability in predicted consumption in that same year. In 1998/99 this general patterns seems to remain, although the initial differences in terms of variance are somewhat attenuated. In 2007/08 then, the picture looks substantially different with villages of the lower 3 quartiles (of initial irrigation) displaying roughly similar variance in predicted consumption, while the villages in the highest quartile of initial irrigation seem to have developed in an entirely different direction, displaying substantially lower variability in consumption.

3.2.3 The Effect of Risk on Schooling

In a multi-period model with savings, the reduced form school investment can be described by

$$s_t = s_t(\sigma_\theta, A_t, H_t, \theta_t) \tag{19}$$

where σ_θ is the variance of the shock, A_t is the stock of assets, H_t the stock of human capital accumulated at the beginning of t, and θ_t is the current realization of the shock. A detailed derivation of this decision rule is outlined in appendix A.4.

In order to quantify the effect of risk on child schooling, we estimate a linear approximation to the decision rule described by in eq. (19). In particular, we estimate

$$s_{ijst} = \beta_0 + \beta_1 R_{jst} + \beta_2' A_{ijst} + \beta_3 \theta_{jst} + \Sigma_{n=1}^{3} \beta_{3+n} \theta_{jst-n} + X_{ijst} + \gamma_j + \psi_{st} + \xi_t + \epsilon_{ijst}, \tag{20}$$

where the dependent variable s_{ijst} is the study time of child i, living in village j, and state s in year t.[32] We focus on extensive margin and intensive margin responses by using school attendance (*i.e.* a dummy that equals one if study time is non-zero) and hours spent studying (in school and at home) on a typical day of each season as dependent variables. We are particularly interested in estimating β_1, *i.e.* the effect of village-level risk R_{jst} on schooling.

So that we can isolate an *ex-ante* risk effect from the *ex-post* effect of accumulated shocks, our specification needs to control for household wealth A_{ijst}, for the current realization of the shock θ_{jst}, and for the accumulated stock of human capital H_{ijst}.[33] We use household income per capita (in logs) during the reference period, the value of all assets owned at the beginning of the reference period (again in per capita terms and in logs), the area owned (per capita, in log acres), and household size, as approximations for household wealth. We use annual precipitation (in log mms) as measure of the current shock. In the absence of information about past school attendance or attainment, we control for rainfall realizations in the three years prior to the survey $\Sigma_{n=1}^{3} \theta_{jst-n}$,

[32] Note the slight abuse of notation t here. We observe three seasons of study time for each child, and also control for season-by-round of interview fixed effects ξ_t. Yet, risk, rainfall and state-by-round (of interview) fixed effects are measures that remain constant within the year, and do not vary per child.

[33] Suppose that human capital stocks are lower in riskier areas due to more regularly occurring shocks. Given dynamic complementarity, the stock of human capital determines the marginal return to current school investments; and therefore, current school attendance should be lower if there was more variation in school attendance previously. Similarly, households exposed to regularly occurring shocks are likely poorer, less well educated, and consequently less able (or willing) to invest in the education of their children.

as well as for the standard deviation of annual rainfall over these three years, as approximation for H_{ijst}. We also interact the rainfall realizations with a dummy that takes the value one if the child was of school age (*i.e.* 6 and above) in that particular year in order to allow for past rainfall realizations to matter more for older children, *i.e.* we add $\Sigma_{n=1}^{3}[\theta_{jst-n} \times 1(SchoolAge_{ijst-n})]$ to eq. (20).

We also control for X_{ijst}, a vector of age-by-gender fixed effects, that captures for example any differences in the implicit price of schooling by age and gender (see appendix A.1). γ_j is a vector of village fixed effects, ψ_{st} a vector of state-by-round (of interview) fixed effects, and ξ_t a vector of season-by-round (of interview) fixed effects. This specification allows us to compare children of the same age and gender from the same village, who are surveyed in different rounds. State-by-round of interview fixed effects control for aggregate shocks at the state level, and season-by-round effects control for temporal variation in time allocation. The identifying assumption is that – conditional on the control variables and the fixed effects – no other (unobserved) characteristics that determine hours in school are correlated with our regressor of interest, *i.e.* risk. This assumption is quite strong and we discuss several concerns about identification (and how we address these) in the following section.

The term ϵ_{ijst} reflects unobserved time-specific shocks to study time. We cluster our standard errors at the level of the village by round of interview in all specifications (the level of variation of our main regressor of interest). However, because our measure of consumption risk is not directly observed but predicted for each village and round, these standard errors might not be correct. We therefore additionally report p-values obtained from percentile-t cluster bootstrap (null-imposed, 999 replications) in which we re-sample at the village level. The risk variable is re-calculated for every bootstrap sample, and the distribution of t-statistics is then obtained by regressing the outcome of interest on a newly predicted risk variable in every bootstrap sample.[34] Note that this is a highly conservative approach, as entire villages, rather than individual observations, are being re-sampled to calculate risk, and to estimate the effect of risk.

OLS estimates of extensive margin responses are reported in table 7. We add control variables step-wise to gauge the sensitivity of our estimates to the choice of specification. We have seasonal data for 22,766 children in the age group 6 to 15. We use two variables as proxies for risk: the interquartile range (panel A), and the standard deviation of predicted log consumption (panel B). Column (1) controls for village and for season-by-round of interview fixed effects, column (2) adds rainfall controls to account for the effect of current and past shocks. In column (3), we additionally interact each lag with a dummy that is unity if the child was of school age in that particular year. This allows for the effect of shock to differ by age. Column (4), then, adds wealth controls, and column (5) adds state-by-round (of interview) fixed effects.

We find that risk considerably reduces study time at the extensive margin. The point estimates are largely insensitive to the inclusion of any of the controls. According to the most conservative

[34] The percentile-t cluster bootstrap has a better asymptotic refinement than the simple pairs cluster bootstrap, and seems more appropriate in a sample with a limited number of clusters (i.e. 242). For more details see Cameron and Miller (2015).

estimates (column 5), a one standard deviation increase in the interquartile range of predicted log consumption (0.030) reduces the probability that children attend school by 2.7 percentage points. Likewise, an increase in the standard deviation of predicted log consumption by one standard deviation (0.023) reduces the probability that children are in school by 2.4 percentage points. The effect is statistically significant throughout: the percentile-t bootstrap p-values range from 0.028 to 0.079. These point estimates suggest that the elasticity of school attendance with respect to risk ranges between -0.047 and -0.041, which is fairly close to the predicted elasticity from the structural estimation.

We explore the effect of risk on the study time (in school or at home) on a typical day in table D.11. To isolate extensive margin responses from intensive margin responses, we estimate eq. (20) in OLS for the full sample of children, and for the sub-sample of children who have positive hours in school. Our results suggest that changes in school time are driven by extensive margin responses, and not by intensive margin responses. The estimated elasticity of study time with respect to risk at the extensive margin ranges between -0.050 and -0.042. These findings imply that children are less likely to attend school due to risk, while there seems to be little scope for adjusting the time in school at the intensive margin. This finding could be suggestive of an (unmodeled) fixed cost of school attendance, such as travel or material costs, that causes households to withdraw a child from school entirely when faced with higher risk. It could also be a sign for limited flexibility on the side of the teacher, that forces children to withdraw from school entirely rather than reducing hours.

We split the sample by gender and by age group to explore some heterogeneity in effect sizes in table D.12. We find some differences in effect size by gender, with effects being more pronounced for girls than for boys (and being statistically significant for girls only). This is in line with our earlier finding that the effect of rainfall on school time is slightly stronger for girls than for boys, which suggests that girls are exposed to higher risk of dropping out. It could also be interpreted as evidence for differential parental preferences towards boys' schooling.[35] We find hardly any differences in effect size by age group. If anything the effect of risk might be slightly more pronounced in older age groups, but again the difference in effect size is very small. We also interact the risk variable with age dummies to explore more nuanced changes by age in figure C.6. Again, the results suggest that the effect of risk is more pronounced in older age groups, however, we find the strongest effects at age 10, thus around the transition to middle school.

3.2.4 Robustness Checks

The identification strategy of his paper largely explores differential changes over time in the availability of irrigation across villages as approximation for changes in risk.[36] There are three major

[35] An earlier draft of this paper explored only the 2007-08 round of data and found effects of risk for girls but not not for boys. The fact that we now find risk effects for boys and girls (when the data are pooled) might be due to gains in precision (the sample size increased almost three-fold), or due to the fact that households were poorer in 1981-82 and 1998-99, which made it more difficult to protect boys from the adverse effects of risk.

[36] Remember that the construction of the risk variable from the availability of irrigation and the historical rainfall distribution adds precision to our estimates (we specifically allow an increase in irrigation to have a differential

reasons why this strategy might be inappropriate. First, village level irrigation (or any other variable that correlates with irrigation) might have an effect on schooling decisions that does not work through the risk portfolio of a household. For example, one would expect wealthier villages to exhibit more irrigation and, consequently, lower consumption variability, while also investing more in education. Similarly, differences in village characteristics that co-evolve with the availability of irrigation might affect the supply side of education. Second, in in a similar vein, we might not be adequately distinguishing the effect of *ex-ante* risk from the accumulated effects of past shocks (*ex-post* effect), even after controlling for lagged rainfall and household wealth. Third, any endogeneity concerns about the expansion of irrigation will necessarily also concern our risk variable.

We conduct several robustness checks to address these concerns. The first set of robustness checks essentially seeks to rule out the possibility that we are not capturing risk but differences in wealth, or in other household or village-level characteristics, or the cumulative effect of past shocks. We begin by exploring household characteristics, and whether these are systematically associated with our measure of risk. In particular, we focus on household income, consumption expenditures, asset ownership (as measure of household wealth), savings accumulated during the reference period, whether the household is involved in agricultural production, area owned, and maternal labor supply. As reported in table 8, we find some evidence that households exposed to more risk may be slightly poorer. While the point estimates on income, wealth and land ownership are not statistically significant, we do find significant negative effects of risk on consumption expenditures per capita. This finding could be consistent with households saving a higher share of their income if exposed to more risk, but we do not find this to be the case, at least not for households that report positive savings over the reference period. It seems more likely that these households are indeed somewhat poorer, and that consumption expenditures are just more precisely measured than incomes.[37] Finally, households exposed to more risk are no more or less likely to be involved in agricultural production. And we do not find any evidence that mothers in our sample are working more or less hours in the labor market due to risk.

We then explore whether school infrastructure, and village characteristics more generally, are systematically co-evolving with our measure of risk. We focus on educational infrastructure because we have to worry that any schooling response of children is driven by supply side constraints rather than demand side factors. We find no evidence that risk is associated with a lower number of primary, middle, lower secondary or higher secondary schools in a village. We also test if our results are affected by controlling for the presence of primary, middle or secondary schools, and find no change in the magnitude or statistical significance of the effect of risk (*c.f.* table D.13). We take this as evidence that our results are not driven by the differential supply of education infrastructure. We then address the concern raised in Section 3.2, namely that villages with a higher share of irrigated

effect on schooling outcomes conditional on the amount of variance in each village's rainfall distribution), but does not introduce any additional source of variation (since we use the same rainfall distribution for all three rounds of interviews).

[37]Note that we control for these measures in order to rule out the possibility that differences in wealth are driving our results.

land might be villages that have have diversified out of agriculture (leaving rain-fed land largely uncultivated), and thus are less exposed to rainfall shocks. These would also be places that invest more in education, if the returns to education are higher in the non-agricultural sector than in the agricultural sector (as is typically assumed). In table D.14, we show that the economic structure in places with more or less risk is not systematically different: neither the share of households that are involved in agriculture, nor the share of income that is generated in agriculture are affected by our risk variable. We also explore if these villages might differ in the presence of factories, or in the number of villagers employed in factories, and find no systematic differences. Finally, we explore if villages with higher risk experience differential access to cooperatives or banks, and again find no evidence that this could be explaining our results.

One caveat with the exercise above is that all these variables might be measured with considerable error, such that we are biased against finding any significant effects. Since we are using a regressor that is calculated for the entire village we are somewhat limited in the amount of tests that we can do. We can, however, explore if any variation within the village (in terms of risk exposure) is pointing in a similar direction, after accounting for village-by-round of interview fixed effects. In order to do this, we first restrict the sample to households that were observed and could be matched in at least two rounds. We then augment equation (18) by an interaction term of rainfall with the log of cash-equivalent savings (i.e. savings that can relatively easily be sold in the case of a shock, such as jewelery or cattle). We these estimates, we predict consumption risk at the household level, and re-estimate table 7 with the new risk variable. After controlling for household fixed effects, for household-specific time-trends, and for village-by-round fixed effects, we still find a large negative effect of risk on school attendance, that is statistically significant at the 0.1% level.[38]

Taken together, these results are consistent with the notion that we are indeed capturing *ex-ante* risk, rather than the cumulative effect of past shocks or some other underlying household or village characteristics that correlate with our risk variable.

The second set of robustness checks seeks to address potential sources of endogeneity in the expansion of irrigation. Endogeneity could arise from differences in initial levels of risk that might put villages on very different paths of agricultural production, wealth accumulation, and investments in education. To address the concern that underlying trends are driving the contemporaneous effect of risk on schooling, we conduct robustness checks in which we control for (non-linear) time trends that vary with baseline levels of village average irrigation, education of the household head, income, wealth, and land ownership, and our results are robust in four out of five specifications (*c.f.* table D.19). We also slice our data to perform placebo tests. In particular, we regress current schooling

[38]Results are reported in tables D.15, D.16 and D.17. While the effect of risk on education is clearly negative, and the effect is robust to controlling for the standard controls, as well as for initial conditions multiplied by time fixed effects, we cannot rule out that this micro-level risk variable also captures other differences between households. Note, that the same risk variable seems to be negatively associated with land ownership, wealth, consumption expenditures and with the probability of being engaged in agricultural production. Yet, it is reassuring to see that we find similar associations within villages as we find across villages. We also exploit the panel data to control for household fixed effects in the original specification, *i.e.* with village-level risk, and find that our results are robust to using this subset of the data and to controlling for household fixed effects (see table D.18).

(in rounds 1981-82 and 1998-99) on future risk in table D.20, and find that these are unrelated.

Endogeneity could also arise from unobserved shocks that are not captured by rainfall nor by state-by-round (of interview) fixed effects, and that simultaneously affect irrigation availability (and therewith risk) at the village level and investments in education at the child level. Lagged negative shocks, for example, could slow down investments in irrigation, and – through wealth effects – also negatively affect investments in education. While we control for contemporaneous and lagged rainfall in all our specifications, we might not be capturing the full effect of these shocks. We therefore also explore if our results are robust to flexibly controlling for lagged (3 years) rainfall and temperature. As we show in table D.21, this is indeed the case.[39] It remains to be said, however, that we cannot fully rule out the possibility that other unobserved shocks simultaneously affect irrigation and schooling outcomes in ways that could bias our results.

3.3 Alternative Explanations

The results presented so far are consistent with the existence of dynamic complementarities in the human capital production function. However, other mechanisms could also result in a negative association between risk and human capital investments, leading us to erroneously attribute this effect to dynamic complementarities. We analyze each of these explanations in detail to show that they are unlikely to produce the results we are finding.

The first alternative explanation to the negative effect of risk on child schooling could be a precautionary savings motive. Study time might be lower in high-risk villages, because parents are less willing to spend their limited resources on education if concerned about next seasons income and rather save more, or because parents are working more in order to accumulate savings, and therefore spend less time on home production, which has to be taken on by the children of the family. *Ceteris paribus*, this would increase the time children spend on household chores or agricultural production, and reduce study time.[40] However, the available empirical evidence does not suggest that this is what is driving our results. As discussed previously, we do not find that households save more year-to-year nor that women work more in areas with higher risk.[41] While year-to-year saving does not necessarily capture the mechanism we outlined above, the simulations with savings showed clearly that the negative effect of risk would vary substantially over the year (in particular *ex ante* and *ex post* to shocks) in a precautionary savings world, while it is mostly constant over time if driven by dynamic complementarities. Empirically, we find the effect of risk

[39] We cannot use more than three lags since the weather data we are using only goes back to 1979. We also show that neither lagged, nor current or future rainfall is systematically related to our measure of consumption risk (*c.f.* table D.22).

[40] That households tend to save more in the presence of uninsured risk is by now well established. However, it is unclear to which extent this affects labor supply decisions. Typically, the literature has looked at foregone current consumption for the sake of higher (and more secure) future consumption (see e.g. Zeldes, 1989; Deaton, 1991). On the other side, one might think that households would tend to work harder in order to accumulate higher savings. A positive labor supply response to risk in the context of India was found by Rose (2001).

[41] Table 8 shows the effect of risk on savings and on maternal time allocation. Unfortunately, the same data is not available for fathers, because the time-use questions are included in the demographic questionnaire (that is administered to women exclusively).

to be surprisingly constant over the three seasons of the year, with the most pronounced negative effects *de facto* arising in season one (October/November), thus *ex post* to the main shock (*i.e.* monsoon), see table D.23. This result is consistent with the dynamic complementarity mechanism, yet less so with a savings motive.

And finally, liquidity (or savings) concerns should generally be less of concern in contexts like India, where schooling is generally for free. Obviously, one-time expenditures for school material or uniforms could be just as prohibitively expensive, but this seems not to be driving our results. In table D.24, we allow the effect of risk on school attendance to vary by quartiles of consumption, income, and wealth, and contrast the results with the average expenditures on education (per child), and average income (per capita) in these subsamples. This exercise reveals that the effect of risk is generally most pronounced in the poorest quartile. Yet, we also find effects of similar magnitudes in the second and third quartiles of the consumption, income or wealth distributions, and in these groups the share of income that is spent on education is substantially lower than in the poorest quartile.

The second alternative explanation is that risk could negatively affect returns to human capital, either through the economic structure or due to a reduced propensity to adopt technological innovations, causing parents to invest less in the education of their children. We suggest two strategies to address this concern. First, we test directly whether the wage return to education varies with risk. Since we only observe wage data in the 2007-08 round of data, we restrict this exercise to an analysis of cross-sectional associations. We regress the log wages (averaged over all observed activities for the individual) on risk interacted with the level of education, and on a number of individual-level characteristics in table D.25. We do not find any evidence that risk in fact reduces the returns to education either for men or for women.[42] To the contrary, if anything evidence suggests that returns to education are positively associated with risk, which suggests that diversification out of agriculture is higher in more risk prone areas or that risk prone areas rely on production technologies that require higher levels of human capital. Second, for girls at least, who are likely to leave the village following marriage, we can examine if the negative effect of risk is driven by the level of risk in the village of birth or by the level of risk in the relevant marriage market. The concern would be that risk is likely correlated across space, such that we would be confounding the level of risk in the the village of origin with the level of risk in the marriage market. We explore this idea by adapting a strategy from Foster and Rosenzweig (2001). In particular, that paper uses a radius of 67km to define the marriage market (the 90th percentile of the marriage distance). It then distinguishes the returns to human capital for boys (that in the village) and the returns to human capital for girls (among sample villages in the broader marriage market). In this case, we control for the average risk among sample villages within 67km of each village of interest. We exclude the village of interest from this mean and drop those villages in which there are no other sample villages in the marriage market. As reported in table D.26, the coefficient on marriage-market risk

[42]Of course, this approach does not account for other possible consequences of risk such as the returns to entrepreneurship or the returns to migration. But it is at least consistent with the idea that our effect of interest is not driven by the return to schooling.

is close to zero and its inclusion does not substantially affect the point estimate on the variable of interest. Thus, as posited in our model, the risk estimates appear to reflect the consequences of risk during childhood rather than adulthood.

A third alternative explanation could be differences in fertility – or (gender specific) child survival rates – that are due to risk. If households exposed to higher risk have more children, they might consequently invest less in each of them. Likewise, a higher gender-imbalance might increase (or reduce) investments in girls (and boys) due to selection effects or due to changes in marriage-market returns. Yet, as reported in table D.27, we find no evidence that these mechanisms are empirically meaningful in our data.

4 Simulating the Effect of the NREGS on Consumption Variability and Schooling Investments

Given the magnitude of the effects of risk observed above, it seems worthwhile to explore potential policy tools to mediate these. Obviously, any policy that helps farmers insure against agricultural production risk, could be a viable option. But as Mobarak and Rosenzweig (2013) pointed out, providing insurance to farmers might actually increase overall risk in village economies as farmers become more risk taking in their production decisions. This would then be particularly harmful for the poorest households with no own land and no access to agricultural insurance. An alternative could be an workfare program, such as the Indian National Rural Employment Guarantee Scheme, that was introduced in rural India in 2006.[43]

A workfare program like the NREGS can affect schooling decisions via two mechanisms: first, the program could provide employment at higher wages than casual agricultural wages, which might affect total labor supply and the amount of labor supplied to the private sector (with potential general equilibrium effects on wages, as documented by Imbert and Papp (2015)). Second, such a program could reduce consumption risk in rural areas, by providing employment at a fixed wage independently of rainfall shocks. Assuming income effects of rainfall shocks dominate in the time allocation of adults, the fact that wages do not fall during periods of need is likely to reduce the magnitude of the labor supply response of households relative to the case in which there are negative general equilibrium effects on wages. This effect should reduce the extent of the response among children *ex post* to shocks. This lower response, in turn, increases the return to first-period school time.

To understand the effectiveness of a workfare program in mediating the negative effects on income risk on school attendance, we explore the follow-up data to the REDS collected between 2014 and 2016, and simulate the effect of the NREGS.[44] This simulation is expected to shed light on the importance of the second mechanism and is intended to complement existing empirical evidence

[43]The National Rural Employment Guarantee Act is India's flagship anti-poverty program; it entitles every household in rural India to a maximum of 100 days of employment per year at state minimum wages.

[44]Summary statistics of the SEPRI data are shown in table D.28. More details about these data are available in appendix B.3.

on the short-term impact of the NREGS on school enrollment, which estimates that the NREGS decreased enrollment by 1-3.5 percentage points (Shah and Steinberg, 2019).[45]

Using the data from 1982 to 2016, we estimate the extent to which the presence of the NREGS, or more specifically, the amount of employment generated by the NREGS per year, mediates the effect of rainfall on household consumption. Formally, we re-estimate eq. (18), but now add employment per capita (in person-days) generated in a given village by the NREGS and its interaction with rainfall to the estimation. We find that a one unit increase in employment per capita, reduces the effect of rainfall on consumption by 4.2 percentage points (see table D.29).

We use these results to predict the standard deviation of predicted log consumption at different levels of NREGS employment per capita. At the mean of irrigation and the minimum of NREGS employment (i.e. zero employment), an increase in NREGS employment by 2.24 days per capita would reduce the standard deviation of log consumption by 34%.[46] A reduction of risk by this magnitude would increase the probability of attending school by 1.0 percentage points, according to estimates presented in table 7, column (4).

Our simulation thus suggests that an employment guarantee such as the NREGS could have positive effects on child school attendance by reducing consumption risk. However, as documented by previous literature the NREGS had substantial wage effects, which seem to explain the negative effect on school enrollment documented by Shah and Steinberg (2019); Li and Sekhri (2020). If the wage effect outweighs the risk reducing effect, or in other words, if the risk reducing effect is not high enough to offset the negative effect of increasing wages, then an employment guarantee such as the NREGS decreases school attendance not only in the first years of program implementation, but permanently.

5 Conclusions

This paper analyzes the joint effect of risk and dynamic complementarity on investments in human capital, and concentrates on the study time of children in rural India. Our findings highlight that risk can reduce schooling investments even in the absence of a labor market for children and in contexts of no school fees, because shocks affect the opportunity costs of time of children if they participate in home production.

The model of human capital investment developed in this paper shows that risk reduces schooling as long as there is dynamic complementarity in the demand for child human capital, which is a

[45]We expect that the risk mitigating effect of the NREGS is internalized by households only after a certain period of time. Therefore this effect cannot be captured by impact evaluations that explore the sequenced phase-in of the NREGS and which can thus only provide information about the immediate effects of the program. This simulation is intended to provide insights about potential long-run effects associated with public works programs such as the NREGS.

[46]We use 2.24 days of employment to make our results comparable to previous work. Imbert and Papp (2015) and Shah and Steinberg (2019) use NSS data collected in 2004-05 (pre) and 2007-08 (post). In 2007-08, the NREGS had been implemented in 330 (out of 625) rural districts of India, covering a population of 642 Million (projections for March 2007). To get the population estimates, we apply the India-wide projected population growth rate to the Census 2001 population count of these districts. That year, the total amount of employment-days created within the NREGS was 1,435.9 Million (Ministry of Rural Development, 2012).

function of the curvature in the parental demand for human capital and dynamic complementarity in the production of child human capital.

Using indirect inference, we then estimate the structural parameters of the model exploiting – in the auxiliary model – a surprisingly robust relationship in study times across seasons observed in the data. Our results suggest that households perceive a strong degree of dynamic complementarity in child schooling, and substantial curvature in the demand for human capital in the form of grade completion. Importantly, demand for human capital is estimated to be complementary in this context, which implies a negative risk prediction. We test the model's predictions by exploring village-level variation over time in the use of irrigation in agriculture, and find strong evidence for a negative effect of risk on the probability that children attend school.

These findings contribute to a better understanding of the consequences of dynamic complementarity in the production of human capital. Dynamic complementarity not only calls for early lifetime interventions in order to remedy the consequences of shocks, and to reduce inequality of opportunities. We also need to think of better ways to support low-income households, both in developed as well as in developing countries, in dealing with various sources of risk.

Arguably, a public works program could be a policy that supports households in contexts of high exposure to climate risk, because it off-sets the negative effects of rainfall shocks on wages and incomes. We simulate the effect of the National Rural Employment Guarantee Scheme on the child schooling. We estimate the extent to which NREGS reduces variability in consumption and use this reduction to simulate the program's effect on child study time. We find that the program would increase school attendance by 1 percentage point, if it were holding the level of wages fixed. These results suggest that income transfers, either in the form of cash transfers or public works programs could increase investments in human capital by reducing risk.

However such programs need to be carefully designed. As shown in the case of the NREGS, the fact that public works wages were above market wages led to declines in the enrollment rates of adolescents. While we cannot assess if the risk effect could have later mitigated the negative wage effect on human capital investments that was evident during the program's phase-in, our results do support the idea that policies that support households in smoothing incomes without raising wages, will also increase schooling investments. Our results also speak to the design of such programs: welfare effects can be maximized by making access to work transparent and reliable, and by differentially targeting such programs to villages that are more regularly affected by adverse shocks.

References

Agostinelli, F. and M. Wiswall (2020). Estimating the Technology of Children's Skill Formation. Working Paper 22442, National Bureau of Economic Research.

Ambika, A. K., B. Wardlow, and V. Mishra (2016). Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015. *Scientific data 3*, 160118.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017, 06). Measuring the Sensitivity of Parameter Estimates to Estimation Moments. *The Quarterly Journal of Economics 132*(4), 1553–1592.

Atkin, D. (2016). Endogenous Skill Acquisition and Export Manufacturing in Mexico. *American Economic Review 106*(8), 2046–85.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020, January). Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., C. Meghir, and E. Nix (2019). Human capital development and parental investment in India. *Review of Economic Studies (accepted).*

Beegle, K., R. H. Dehejia, and R. Gatti (2006). Child labor and agricultural shocks. *Journal of Development Economics 81*(1), 80–96.

Bjorkman-Nyqvist, M. (2013). Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics 105*, 237 – 253.

Cameron, C. A. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Castro, J. F. and L. Villacorta (2019). *Too Hard, Too Easy, or Just Right: Dynamic Complementarity and Substitutability in the Same Input of Cognitive Skill.* unpublished manuscript.

Coelli, M. B. (2011). Parental job loss and the education enrollment of youth. *Labour Economics 18*(1), 25 – 35.

Colella, F., R. Lalive, S. O. Sakalli, and M. Thoenig (2019). *Inference with arbitrary clustering.* IZA Discussion Paper.

Colmer, J. (2019a). *Rainfall variability, child labor and human capital accumulation in rural Ethiopia.* unpublished manuscript.

Colmer, J. (2019b). *Temperature, labor reallocation and industrial production: Evidence from India.* unpublished manuscript.

Conley, T. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Copernicus Climate Change Service (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.* Copernicus Climate Change Service Climate Data Store (CDS).

Cunha, F., I. Elo, and J. Culhane (2013). *Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation.* NBER Working Paper No. 19144.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Deaton, A. (1991). Saving and liquidity constraints. *Econometrica 59*(5), 1221–1248.

Dercon, S. and L. Christiaensen (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics 96*(2), 159–173.

Directorate of Economics and Statistics (2019). *Agricultural Statistics at a Glance 2018*. Ministry of Agriculture & Farmers Welfare. Government of India.

Duflo, E. and R. Pande (2007). Dams. *The Quarterly Journal of Economics 122*(2), 601–646.

Duryea, S., D. Lam, and D. Levison (2007). Effects of economic shocks on children's employment and schooling in Brazil. *Journal of Development Economics 84*(1), 188–214.

Fitzsimons, E. (2007). The Effects of Risk on Education in Indonesia. *Economic Development and Cultural Change 56*(1), 1–25.

Foster, A. and M. Rosenzweig (2001). Missing women, the marriage market, and economic growth. Manuscript, Brown University.

Foster, A. D. and M. R. Rosenzweig (1996). Technical change and human-capital returns and investments: Evidence from the green revolution. *American Economic Review 86*(4), 931–953.

Glewwe, P. and H. G. Jacoby (2004). Economic growth and the demand for education: is there a wealth effect? *Journal of Development Economics 74*(1), 33 – 51.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect Inference. *Journal of Applied Econometrics 8*, S85–S118.

Gubert, F. and A.-S. Robilliard (2008). Risk and Schooling Decisions in Rural Madagascar: A Panel Data-Analysis. *Journal of African Economies 17*(2), 207–238.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences 107*(35), 15367–15372.

Ilahi, N. (2000). *The intra-household allocation of time and tasks: what have we learnt from the empirical literature?* World Bank, Development Research Group/Poverty Reduction and Economic Management Network.

Imbert, C. and J. Papp (2015). Labor Market Effects of Social Programs: Evidence from India's Employment Guarantee. *American Economic Journal: Applied Economics 7*(2), 233–63.

Jacoby, H. G. and E. Skoufias (1997). Risk, financial markets, and human capital in a developing country. *Review of Economic Studies 64*(3), 311–335.

Jayachandran, S. (2006). Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. *Journal of Political Economy 114*(3), 538 – 575.

Jensen, R. (2000). Agricultural volatility and investments in children. *American Economic Review 90*(2), 399–404.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics 125*(2), 515–548.

Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India. *Quarterly Journal of Economics 127*(2), 753–792.

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review 109*(10), 3585–3616.

Kazianga, H. (2012). Income Risk and Household Schooling Decisions in Burkina Faso. *World Development 40*(8), 1647 – 1662.

Kochar, A. (1999). Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural India. *Review of Economics and Statistics 81*(1), 50–61.

Li, T. and S. Sekhri (2020). The Spillovers of Employment-based Safety Net Program on Child Labor and Education. *The World Bank Economic Review (forthcoming)*.

Ministry of Rural Development (2012). *MGNREGA Sameeksha: An Anthology of Research Studies on the Mahatma Gandhi National Rural Employment Guarantee Act, 2005. 2006-2012.* New Delhi: Orient BlackSwan.

Mobarak, A. M. and M. R. Rosenzweig (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review 103*(3), 375–80.

Mondal, P., M. Jain, R. S. DeFries, G. L. Galford, and C. Small (2015). Sensitivity of crop cover to climate variability: Insights from two Indian agro-ecoregions. *Journal of Environmental Management 148*, 21 – 30. Land Cover/Land Use Change (LC/LUC) and Environmental Impacts in South Asia.

Oster, E. and B. M. Steinberg (2013). Do it service centers promote school enrollment? evidence from india. *Journal of Development Economics 104*(Supplement C), 123 – 135.

Page, M., A. H. Stevens, and J. Lindo (2009). *Parental Income Shocks and Outcomes of Disadvantaged Youth in the United States*, pp. 213–235. University of Chicago Press.

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics 64*(2), 371–388.

Shah, M. and B. M. Steinberg (2017). Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital. *Journal of Political Economy 125*(2), 527–561.

Shah, M. and B. M. Steinberg (2019). Workfare and Human Capital Investment: Evidence from India. *Journal of Human Resources (published ahead of print)*.

Skoufias, E. (1993). Labor market opportunities and intrafamily time allocation in rural households in South Asia. *Journal of Development Economics 40*(2), 277–310.

Skoufias, E. and S. W. Parker (2006). Job loss and family adjustments in work and schooling during the Mexican peso crisis. *Journal of Population Economics 19*(1), 163–181.

Storn, R. and K. Price (1997). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization 11*, 341–359.

Zeldes, S. P. (1989). Optimal consumption with stochastic income: Deviations from certainty equivalence. *Quarterly Journal of Economics 104*(2), 275–298.

Figures

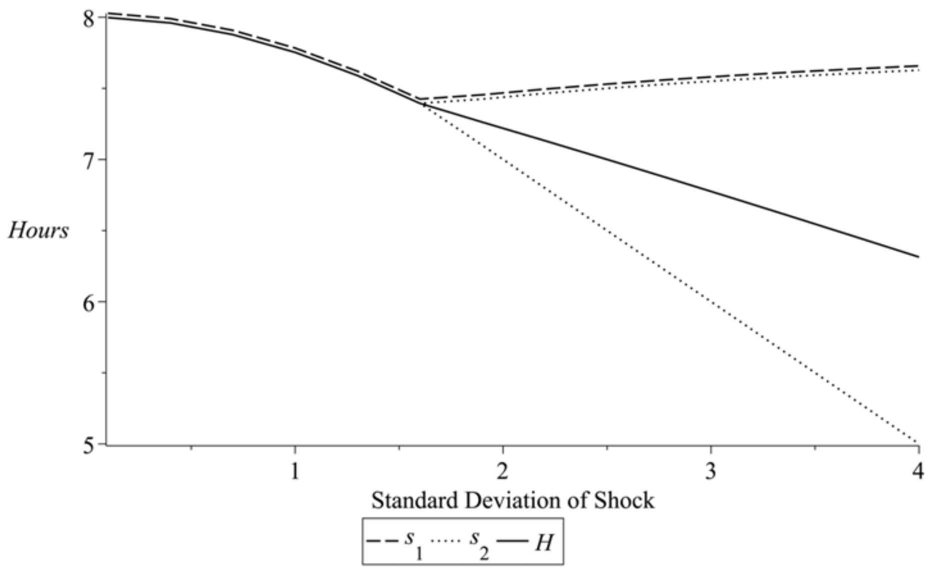

Figure 1: Risk, Study Time and Expected Human Capital

Notes: This figure presents numerical solutions to the optimal choice of s_1, s_2 by state and $\mathbb{E}[H_2]$ by the standard deviation of the shock ($\sigma_\theta = r$), for a two-period model as described by eq. (1) with perfect complements and linear $v(H)$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, and $\zeta = 1$.

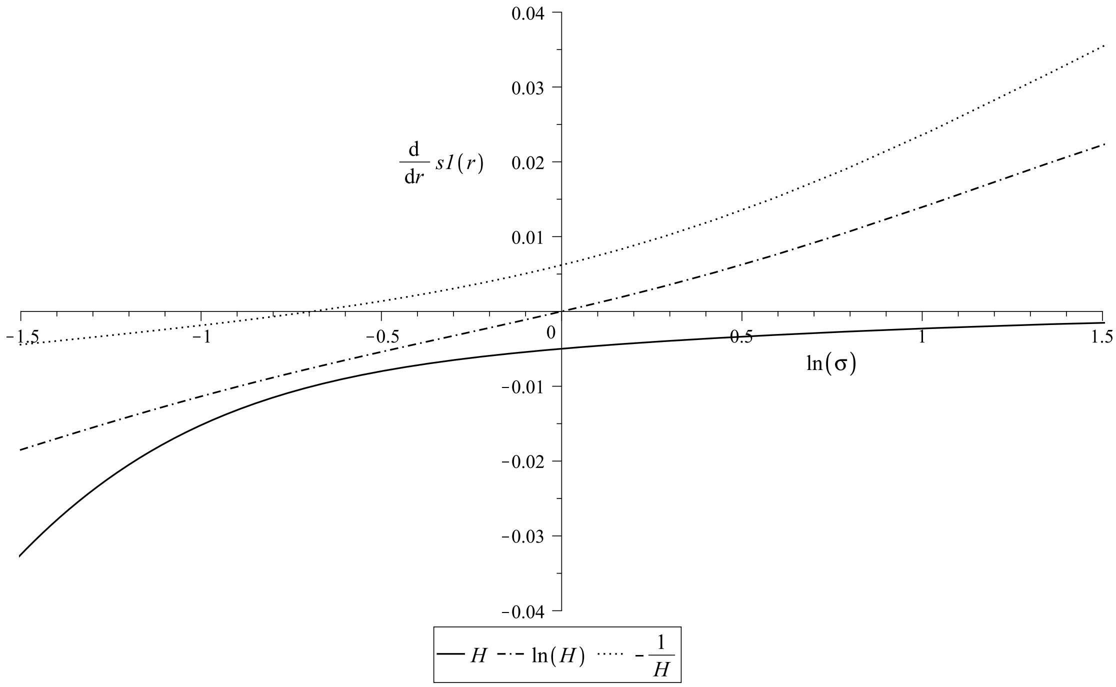

Figure 2: Numeric Approximation of the Effect of Risk on First Period Schooling

Notes: This figure presents numerical solutions to the derivative of s_1 wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the *ex ante* effect of risk, for a two-period model as described by eq. (1) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 1$ and $\zeta = 1$.

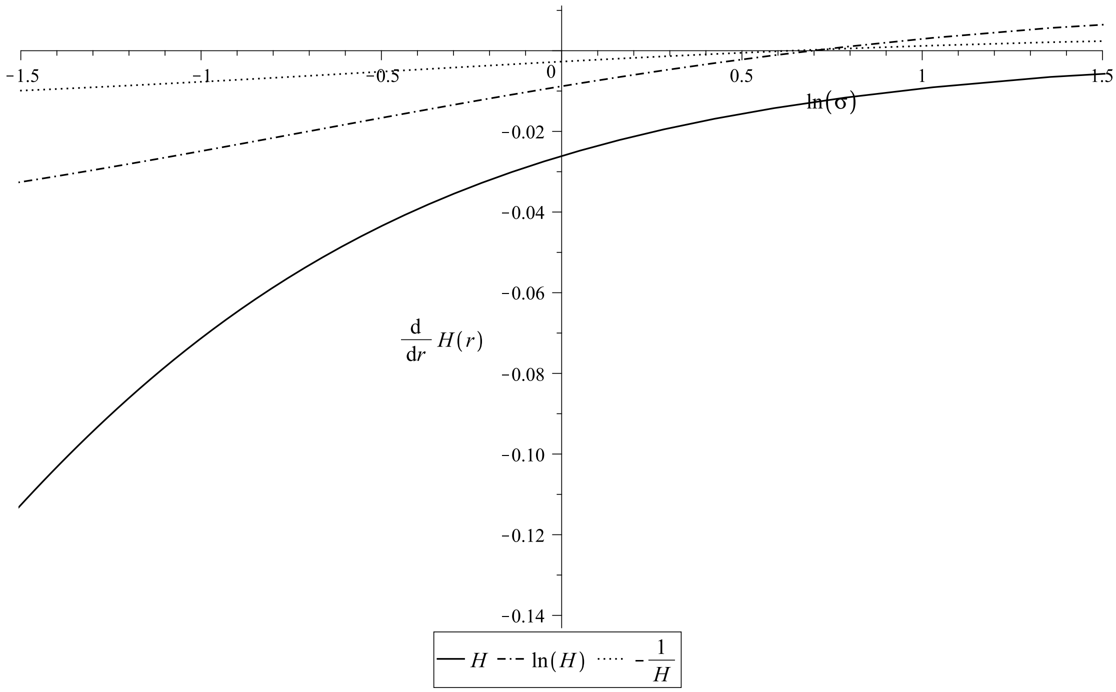

Figure 3: Numeric Approximation of the Effect of Risk on the Expected Value of Human Capital
Notes: This figure presents numerical solutions to the derivative of $\mathbb{E}(H_2)$ wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the combined *ex ante* and *ex post* effect of risk, for a two-period model as described by eq. (1) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 1$ and $\zeta = 1$.

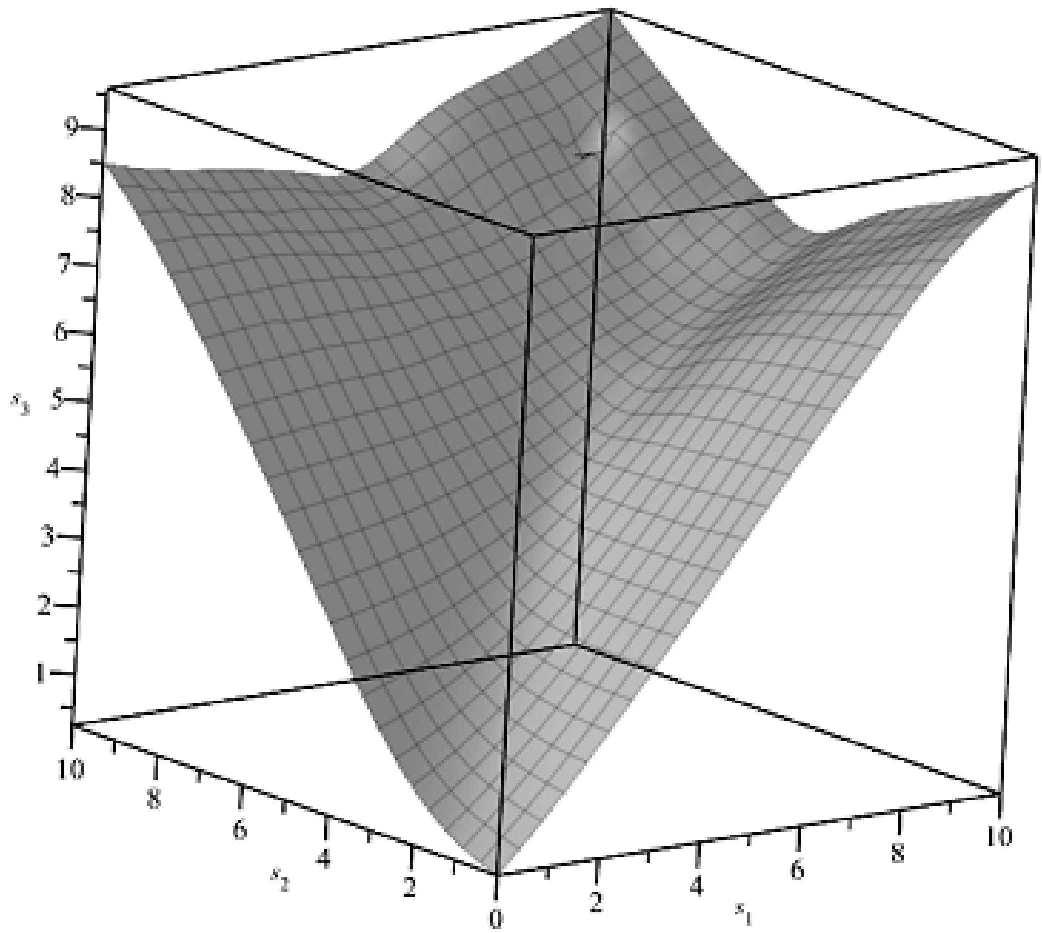

Figure 4: Lowess-smoothed Study Time by Season
Notes: This figure plots the fitted values of a lowess-regression (bandwidth = 0.3, degree = 1) of s_3 on s_1 and s_2.

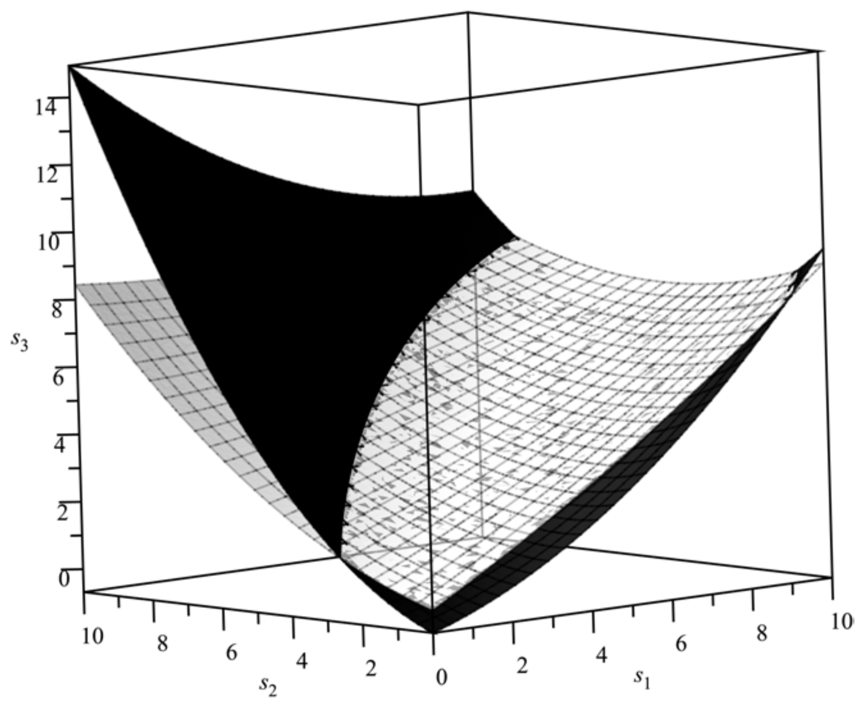

Figure 5: Study Time by Season: Model and Data Estimates
Notes: This figure plots the estimated quadratic relationship between s_3 and s_1 and s_2 from the simulated data (black) and actual data (grey). The parameter estimates ($\alpha_1 - \alpha_5$) are given by table 2.

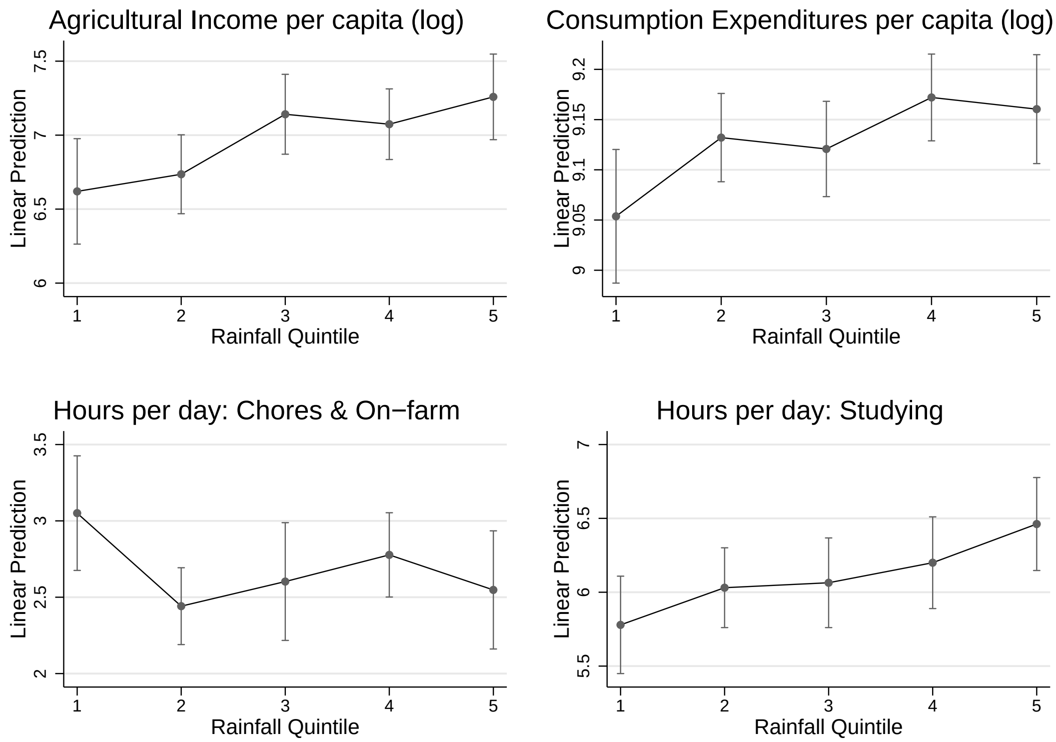

Figure 6: Effect of Rainfall Shocks

Notes: The figures plot linear predictions and 95% confidence intervals from regressions of each outcome variable on dummies for each quintile of the rainfall distribution. Each quintile dummy equals 1 if total annual rainfall in the current agricultural year fell within the given quintile of the village's usual rainfall distribution and equals 0 otherwise. Each regression controls for state-by-round of interview fixed effects, for the share of area that is irrigated within a village, and for lagged rainfall (in log mms) in the past 3 years (2 years in Panel A). Panels C and D additionally control for age-by-gender and season-by-round of interview fixed effects. Standard errors are corrected to allow for clustering at the level of the grid cell by round of interview.

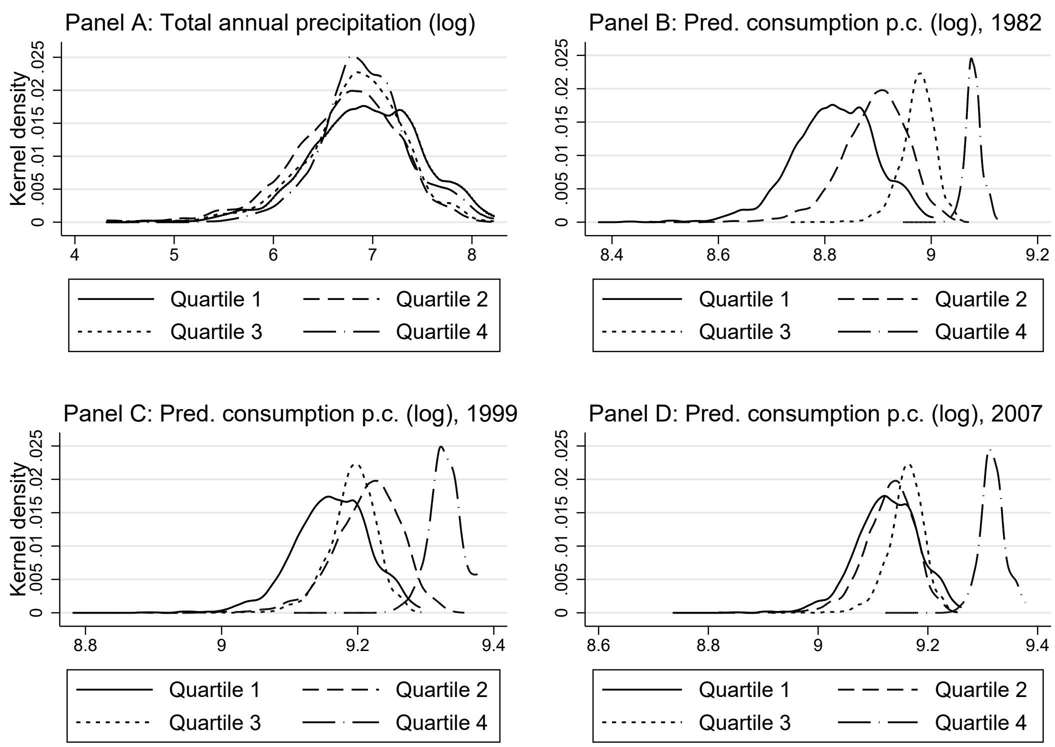

Figure 7: Distribution of Rainfall and Predicted Consumption by Quartiles of Baseline Irrigation

Notes: This figure plots the kernel densities of total annual rainfall (log), and predicted consumption expenditures per capita (log) in REDS villages in the rounds 1981-82, 1998-99 and 2007-08. In all four graphs, the REDS villages are split into quartiles of irrigation share in 1981-82.

Tables

Table 1: Study Time across Seasons

	(1)	(2)	(3)	(4)	(5)
Study time, season 1	0.662***	0.665***	0.665***	0.678***	0.687***
	(0.044)	(0.045)	(0.046)	(0.048)	(0.048)
Study time, season 1 (square)	0.019***	0.018***	0.018***	0.019***	0.018***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Study time, season 2	0.362***	0.360***	0.359***	0.346***	0.338***
	(0.049)	(0.050)	(0.050)	(0.052)	(0.054)
Study time, season 2 (square)	0.049***	0.049***	0.049***	0.050***	0.051***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Study time, season 1 × season 2	-0.079***	-0.079***	-0.079***	-0.081***	-0.082***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Rainfall controls	N	Y	Y	Y	Y
Income, wealth, area, hh size, risk	N	N	Y	Y	Y
State-by-round FE	N	N	N	Y	Y
Village-by-round FE	N	N	N	N	Y
Observations	22766	22766	22766	22766	22766
Adjusted R^2	0.811	0.818	0.819	0.823	0.834

Notes: The dependent variable is study time in season 3. Each regression controls for village, round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms). Income, wealth and area are measured in per capita terms (and are in logs). Risk is the interquartile range of predicted log consumption. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Match of the Auxiliary Model

Auxiliary model parameter	Symbol	Simulated	Target	Difference
Effect of s_1 on s_3	α_1	0.410	0.687	0.277
Effect of s_2 on s_3	α_2	0.430	0.338	-0.092
Effect of $(s_1)^2$ on s_3	α_3	0.057	0.018	-0.039
Effect of $(s_2)^2$ on s_3	α_4	0.113	0.051	-0.062
Effect of $(s_1 \times s_2)$ on s_3	α_5	-0.150	-0.082	0.068
Effect of s_1 on s_2	γ_1	0.826	1.040	0.214
Effect of $(s_1)^2$ on s_2	γ_2	-0.017	-0.018	-0.001
Mean of s_1	μ_{s_1}	8.210	6.426	-1.784
Standard deviation of s_1	σ_{s_1}	1.512	3.069	1.557
Mean of s_2	μ_{s_2}	8.023	6.384	-1.639
Standard deviation of s_2	σ_{s_2}	1.275	3.056	1.781
Mean of s_3	μ_{s_3}	8.607	6.146	-2.461
Standard deviation of s_3	σ_{s_3}	1.529	3.082	1.552

Notes: Target parameter estimates α_1 to α_5 are given by table 1, col. (5), and γ_1 and γ_2 by table D.4, col (5).

Table 3: Estimated Structural Parameters and Sensitivity Matrix

Symbol	ζ_L	$\Delta\zeta$	ϕ	σ	δ	y_L	Δy
Panel A: Structural Parameters							
Estimate	8.841	8.470	1.544	0.210	0.919	11.390	7.992
	(1.231)	(1.334)	(0.080)	(0.005)	(0.007)	(0.224)	(0.517)
Panel B: Sensitivity Matrix							
α_1	16.301	20.416	6.317	-0.960	0.110	0.632	3.828
α_2	10.773	14.146	4.456	-0.388	0.223	0.307	3.138
α_3	152.336	187.936	57.870	-9.033	1.071	5.555	31.814
α_4	86.139	118.588	31.657	-0.423	1.964	-3.075	11.145
α_5	119.852	154.184	44.523	-2.993	1.669	0.975	19.462
γ_1	4.812	7.431	2.055	-0.327	0.048	0.223	1.317
γ_2	39.266	63.995	17.956	-2.176	0.389	2.487	14.390
μ_{s_1}	1.590	2.216	0.712	0.074	0.406	0.167	0.637
σ_{s_1}	0.227	1.878	0.003	0.536	0.303	-0.264	-0.295
μ_{s_2}	1.297	1.186	0.111	-0.361	-0.205	-0.179	-0.722
σ_{s_2}	2.775	1.511	0.925	-0.891	-0.062	0.305	0.453
μ_{s_3}	-2.336	-2.822	-0.699	0.201	-0.190	0.084	0.106
σ_{s_3}	-4.547	-5.091	-1.320	0.534	-0.262	0.085	0.278

Notes: Parameters of the three-period model $\{\zeta_L, \zeta_H, \phi, \sigma, \delta, y_L, y_H\}$ are estimated by indirect inference. $\Delta\zeta = \zeta_H - \zeta_L$, and $\Delta y = y_H - y_L$. Standard errors are in parentheses. The parameters of the auxiliary model are $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \gamma_1, \gamma_2, \mu_{s_1}, \sigma_{s_1}, \mu_{s_2}, \sigma_{s_2}, \mu_{s_2}, \sigma_{s_3}\}$. The Sensitivity Matrix is expressed in semi-elasticities.

Table 4: Predicted Study Times, Human Capital and Elasticities with respect to Risk

	Predicted levels				Elasticities wrt Risk			
	s_1	s_2	s_3	H_3	s_1	s_2	s_3	H_3
Panel A: Main model estimates								
Low return household (ζ_L)	7.382	7.249	7.739	7.367	-0.067	-0.104	-0.045	-0.072
High return household (ζ_H)	9.288	8.996	9.679	9.254	-0.084	-0.166	-0.111	-0.095
Average household	8.335	8.123	8.709	8.310	-0.076	-0.138	-0.081	-0.084
Panel B: Model estimates with savings = 0								
Low return household (ζ_L)	7.194	7.274	7.382	7.199	-0.080	-0.088	-0.087	-0.083
High return household (ζ_H)	9.109	9.192	9.312	9.111	-0.086	-0.095	-0.096	-0.090
Average household	8.152	8.233	8.347	8.155	-0.084	-0.092	-0.092	-0.087

Notes: Levels of s_1, s_2, s_3 and H_3 are obtained from numerically solving the the three-period model described by eq. (8) and parameter values shown in table 3, and averaging over 8 possible combinations of y_H and y_L for each type. The levels shown for the average household are obtained from averaging over the high and the low type. Elasticities with respect to risk are obtained from numerically estimating the effect of a mean-preserving perturbation in the incomes on schooling investments. Estimates of panel B use the same parameters as panel A, but now solve a model in which savings are constrained to be zero.

Table 5: Effect of Rainfall on Economic Outcomes

	Household				Maternal time			
	Agr. income		Cons.expend.		Labor market		Household chores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Precipitation, current year (log mms)	1.015**	0.866**	0.229***	0.138**	-1.964***	-0.876*	1.007**	0.636
	(0.453)	(0.408)	(0.076)	(0.061)	(0.565)	(0.520)	(0.510)	(0.460)
Irrigated area (share of village agr. area)	6.721**	3.482	1.608***	0.907***	-1.933	-4.526*	0.757	2.451
	(2.917)	(2.408)	(0.396)	(0.338)	(2.547)	(2.656)	(2.669)	(2.204)
Precipitation × Irrigated area	-0.924**	-0.490	-0.210***	-0.116**	0.282	0.670*	-0.121	-0.308
	(0.420)	(0.349)	(0.057)	(0.047)	(0.368)	(0.381)	(0.389)	(0.316)
State-by-round FE	Y	Y	Y	Y	Y	Y	Y	Y
Village FE	N	Y	N	Y	N	Y	N	Y
Dependent variable mean	$ln(8741)$	$ln(8741)$	$ln(9219)$	$ln(9219)$	3.05	3.05	9.06	9.06
Observations	15849	15849	20445	20445	33052	33052	33052	33052
Adjusted R^2	0.062	0.144	0.266	0.334	0.244	0.310	0.227	0.308

Notes: Agricultural income and consumption expenditures are in log INR per capita. Female time allocation is in hours per day for three seasons. Each regression controls for state-by-round of interview fixed effects, and for lagged rainfall (in logs) in the last 3 years (2 years in cols. (1) and (2)). Columns (5) to (8) also control for season-by-round, and age fixed effects. Columns (2), (4), (6) and (8) additionally control for village fixed effects. Conley (1999) standard errors in parentheses (cut-off 300km, Bartlett weights). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Effect of of Rainfall on Child Time Use

Hours per day:	On-farm		Chores		Studying		Leisure	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Precipitation, current year (log mms)	-0.822**	-0.446*	-0.167	-0.532	1.155***	1.516***	-0.092	-0.522
	(0.344)	(0.254)	(0.573)	(0.643)	(0.381)	(0.457)	(0.495)	(0.507)
Irrigated area (share of village agr. area)	-4.670***	-3.363***	-5.260**	-8.686***	5.835**	13.088***	5.031**	1.079
	(1.701)	(1.300)	(2.325)	(2.778)	(2.538)	(2.998)	(2.329)	(2.642)
Precipitation × Irrigated area	0.664***	0.476***	0.694**	1.198***	-0.761**	-1.814***	-0.738**	-0.159
	(0.240)	(0.182)	(0.334)	(0.398)	(0.359)	(0.428)	(0.334)	(0.372)
Observations	37002	31296	37002	31296	37002	31296	37002	31296
Adjusted R^2	0.180	0.156	0.310	0.402	0.218	0.288	0.498	0.485

Notes: Time allocation is reported in hours per day over three seasons. Each regression controls for village, state-by-round of interview, season-by-round, and age fixed effects, and for lagged rainfall (in logs) in the last 3 years. Conley (1999) standard errors in parentheses (cut-off 300km, Bartlett weights).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Effect of Risk on School Attendance

	(1)	(2)	(3)	(4)	(5)
Panel A:					
IQR of predicted log consumption	-1.000**	-1.027**	-1.021**	-0.971**	-0.892*
	(0.297)	(0.299)	(0.299)	(0.279)	(0.302)
	[0.038]	[0.038]	[0.036]	[0.028]	[0.056]
Treatment Mean	0.035	0.035	0.035	0.035	0.035
Treatment Std. Dev.	0.030	0.030	0.030	0.030	0.030
Adjusted R^2	0.207	0.208	0.209	0.228	0.240
Panel B:					
SD of predicted log consumption	-1.156*	-1.169*	-1.161*	-1.099**	-1.009*
	(0.379)	(0.388)	(0.388)	(0.358)	(0.383)
	[0.052]	[0.051]	[0.053]	[0.044]	[0.079]
Treatment Mean	0.027	0.027	0.027	0.027	0.027
Treatment Std. Dev.	0.023	0.023	0.023	0.023	0.023
Adjusted R^2	0.207	0.208	0.208	0.228	0.239
Rainfall controls	N	Y	Y	Y	Y
Rainfall by schoolage	N	N	Y	Y	Y
Income, wealth, area, hh size	N	N	N	Y	Y
State-by-round FE	N	N	N	N	Y
Dep. Var. mean	0.762	0.762	0.762	0.762	0.762
Observations	68298	68298	68298	68298	68298

Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Rainfall by schoolage additionally interacts each rainfall lag with a dummy that equals 1 of the child was 6 years or older in that year. Income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table 8: Effect of Risk on Economic Outcomes

	Income	Wealth	Land	Cons. exp	Curr. savings	Agr. prod. (=1)	Mat. labor supply
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A:							
IQR of predicted log consumption	-1.781	-1.853	1.411	-1.513**	-3.244	0.337	-3.427
	(1.129)	(0.845)	(1.438)	(0.429)	(1.673)	(0.286)	(2.703)
	[0.344]	[0.142]	[0.463]	[0.023]	[0.207]	[0.407]	[0.334]
Treatment Mean	0.035	0.035	0.035	0.035	0.035	0.035	0.035
Treatment Std. Dev.	0.030	0.030	0.030	0.030	0.030	0.030	0.030
Adjusted R^2	0.134	0.280	0.209	0.334	0.275	0.171	0.305
PANEL B:							
SD of predicted log consumption	-2.036	-2.325	2.313	-1.985**	-3.634	0.624	-3.168
	(1.346)	(1.094)	(1.763)	(0.563)	(2.162)	(0.391)	(3.508)
	[0.345]	[0.170]	[0.338]	[0.031]	[0.261]	[0.290]	[0.507]
Treatment Mean	0.027	0.027	0.027	0.027	0.027	0.027	0.027
Treatment Std. Dev.	0.023	0.023	0.023	0.023	0.023	0.023	0.023
Adjusted R^2	0.134	0.280	0.209	0.334	0.275	0.171	0.305
Dependent variable mean	ln(15696)	ln(29990)	ln(0.778)	ln(10909)	ln(6796)	0.789	3.044
Observations	20445	20445	20445	20445	11496	20445	33486

Notes: Income, land owned, wealth (value of all assets owned), consumptions expenditures and savings (if > 0) are expressed in per capita terms and logs. Maternal labor supply is in hours per day for three seasons. Each regression controls for village, state-by-round of interview fixed effects, and for current and lagged rainfall in past 3 years (1 years in cols. (1) and (6), in log mms), and for the standard deviation of rainfall over the last three years. Col. (6) additionally controls for age and for season-by-round of interview fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile Bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

A Mathematical Appendix

A.1 Incorporating child home production

A model that fits our data better links risk and schooling via the time children allocate to home production. While the share of time that children in our sample allocate to the labor market is declining strongly over time, we find that boys and girls allocate time to household chores and to own agricultural production throughout. Importantly, this time-allocation seems to be responsive to rainfall shocks (with children spending more time on these activities when faced with a negative income shock).

To generate this pattern in the absence of a child labor market, we incorporate a home produced good in the model, as well an *ex post* labor supply response to income shocks on the part of the parent. The idea is that parents supply more labor to the labor market in the face of an adverse shock to cope with the shock, as found by Kochar (1999); Rose (2001); Jayachandran (2006) and – at least for mothers – also in our data.[47] As long as there is substitution between the parent's and the child's time in home production, the opportunity cost of time of the child must rise during bad periods even though the child is unlikely to be active in the labor market (Skoufias, 1993; Ilahi, 2000).

The single-period utility function u over the market-produced good c_t^m and the home-produced good c_t^h is assumed to be log Cobb-Douglas with share parameter α:

$$U_t = u(c_t^m, c_t^h) = \alpha \ln c_t^m + (1-\alpha) \ln c_t^h. \tag{A.1}$$

Market goods are purchased using cash income that is earned by the parent with wage w_t and a time allocation of $T^a - h_t^a$, where T^a is her time endowment and h_t^a is the time spent by the parent in home production,

$$c_t^m = I_t + \theta_t + w_t(T^a - h_t^a). \tag{A.2}$$

I_t is the non-stochastic part of income (for example non-labor income), and θ_t is a shock realized at time t. Consumption of the home-produced good is linear in the allocation of time by the parent h_t^a and the child h_t^c,

$$c_t^h = h_t^a + \tau h_t^c, \tag{A.3}$$

where τ captures the relative productivity of the child in home production. School time, then, is the difference between the child's time endowment T^c and the amount of time spent in home production:

$$h_t^c = T^c - s_t \tag{A.4}$$

Parental time at home in each period is:

$$h_t^a = (1-\alpha)T^a - \alpha\tau h_1^c + \frac{(1-\alpha)}{w_t}(I_t + \theta_t). \tag{A.5}$$

The result is decreasing in the child's home time and the wage, and increasing in the income endowment and shock. With a more favorable shock or a lower wage, the parent spends less time

[47] As described in the empirical section, we do not observe paternal time use.

in the labor market and more time at home. If we then define

$$y_t = T^a + \tau T^c + \frac{(I_t + \theta_t)}{w_t}, \tag{A.6}$$

we can rewrite eq. (A.1) to get a period utility that corresponds to that used in our initial model plus a constant, in which the price of schooling now varies with τ

$$U_t = (1-\alpha)\ln(1-\alpha) + \alpha\ln(w_t) + \alpha\ln(\alpha) + \ln(y_t - \tau s_t) \tag{A.7}$$

In the theoretical section, we normalize τ to 1 without loss of generality. In the empirical analysis with we control for a vector of age-by-gender fixed effects, to capture any differences in the price of schooling that can be attributed to τ.

A.2 Incorporating savings

If we extend the simple two-period model to allow for savings or borrowings in period one against period two, the objective function becomes

$$\max \text{Ł}(s_1, \tilde{s_2}) = ln(y_1 - s_1 - a_1) + \mathbb{E}_1[ln(y_2 + \tilde{\theta_2} + a_1 - \tilde{s_2}) + \zeta v(H_2)], \tag{A.8}$$

where a_1 is the amount saved or borrowed in period one, and we abstract from the interest rate (*i.e* we assume it equals 1). Note that this formulation does not allow the household to borrow beyond the year. We again solve this expression numerically for various ρ and three functional forms of $v(H)$. As before, we use parameter values of $y_1 = y_2 = 10, r = 1, \nu_1 = \nu_2 = 1/2, \zeta = 1$.

We now find a negative effect of risk on first-period schooling over all values of ϕ and ρ as depicted in Figure C.2. This can be explained by a additional income-smoothing mechanism, that encourages households to save in the first period, in order to smooth consumption in the second period. Yet, this effect reverses in the second period, where the effect of risk on the expected value of second period schooling is positive except for very small ρ (see Figure C.3). The net effect on total human capital, finally, is exactly identical to the case without savings (see Figure C.4).

This exercise highlights the importance of distinguishing a negative effect of risk on first-period schooling that is due to income smoothing from a negative effect of risk that is due to dynamic complementarity. Yet, it also identifies a potential strategy to distinguish both mechanisms: While the effect of risk on schooling is positive in expectation *ex post* to shocks in a situation where the *ex ante* negative effect is driven by income smoothing concerns, it continues to be negative throughout for very concave preferences over human capital and no savings (see figure C.5). For log-utility and for linear preferences, the predicted effect of risk on the expected value of s_2 turns positive for moderate complementarity in production, but remains close to zero throughout.

A.3 The choice of season 3 schooling as function of previous investments

In a three-period model with buffer-stock savings, and $\nu_i = 1/3$ for simplicity, the household maximizes

$$\begin{aligned}\max \text{Ł}(s_1, \tilde{s_2}, \tilde{s_3}, a_1, \tilde{a_2}) =& ln(y_1 - s_1 - a_1) \\ &+ \mathbb{E}_1[ln(\tilde{y_2} - \tilde{s_2} + a_1 - \tilde{a_2}) + ln(\tilde{y_3} - \tilde{s_3} + \tilde{a_2}) \\ &+ \zeta\frac{1}{1-\phi}(\frac{1}{3}s_1^{\rho} + \frac{1}{3}\tilde{s_2}^{\rho} + +\frac{1}{3}\tilde{s_3}^{\rho})^{\frac{1-\phi}{\rho}}]\end{aligned} \tag{A.9}$$

We reparametrize $st_{12} = s_1 + s_2$ and $sd_{12} = s_1 - s_2$ and substitute into eq. (A.9). An optimal choice of s_3 implies that

$$\frac{d\mathrm{L}(st_{12}, sd_{12}, s_3)}{ds_3} = 0 \tag{A.10}$$

We then implicitly differentiate the first order condition to get

$$\frac{ds_3^*(st_{12}, sd_{12})}{dsd_{12}} = -\frac{\partial^2 \mathrm{L}^*(st_{12}, sd_{12}, s_3)}{\partial s_3 \partial sd_{12}} / \frac{\partial^2 \mathrm{L}^*(st_{12}, sd_{13}, s_3)}{\partial s_3^2} \tag{A.11}$$

and analogously for $\frac{ds_3^*(st_{12}, sd_{12})}{dst_{12}}$. Because the denominator of (A.11) must be negative for an interior solution, the sign of this implicit derivative will be determined by the sign of the numerators. The numerators are:

$$\frac{\partial^2 \mathrm{L}^*(st_{12}, sd_{12}, s_3)}{\partial s_3 \partial st_{12}} = \frac{(1-\phi-\rho)\zeta(s_1^{\rho-1} + s_2^{\rho-1})s_3^{\rho-1} H_3^{1-\phi}}{2s_1^{2\rho} + 2s_2^{2\rho} + 2s_3^{2\rho} + (4s_2^\rho + 4s_3^\rho)s_1^\rho + 4s_2^\rho s_3^\rho} \tag{A.12}$$

and

$$\frac{\partial^2 \mathrm{L}^*(st_{12}, sd_{12}, s_3)}{\partial s_3 \partial sd_{12}} = \frac{(1-\phi-\rho)\zeta(s_1^{\rho-1} - s_2^{\rho-1})s_3^{\rho-1} H_3^{1-\phi}}{2s_1^{2\rho} + 2s_2^{2\rho} + 2s_3^{2\rho} + (4s_2^\rho + 4s_3^\rho)s_1^\rho + 4s_2^\rho s_3^\rho} \tag{A.13}$$

A.4 A multi-period formulation

A more general multi-period formulation can be written as

$$\begin{aligned} \max V_0 = & u(c_1) + \mathbb{E}_1 \left[\sum_{t=2}^{T} \beta^{t-1} u(c_t) + \zeta v(H_T) + f(A_T) | \sigma_\theta \right] \\ & s.t. \\ & \quad c_t \leq y_t + \theta_t + A_t - a_t - s_t \end{aligned} \tag{A.14}$$

where θ_t denotes a mean-zero village shock with standard deviation σ_θ. a_t is net saving, and A_t the stock of assets at the beginning of t. The stock is updated in each period by $A_{t+1} = A_t + ba_t$, with b being the interest rate. The human capital production function is

$$H_{t+1} = (H_t^\rho + v_t s_t^\rho)^{1/\rho}. \tag{A.15}$$

Substituting the budget constraint into the utility function, and using $u(c) = \ln(c)$ as before, we can write this maximization problem recursively as:

$$V_1(H_1, A_1) = \ln(y_1 + \theta_1 + A_1 - b_1 - s_1) + \beta \, \mathbb{E}_1 \left[V(H_2, A_2) | \sigma_\theta \right]. \tag{A.16}$$

The first order condition with respect to s_1 is

$$\frac{-1}{y_1 + \theta_1 + A_1 - a_1 - s_1} + \beta \, \mathbb{E}_1 \left[\frac{\partial V(H_2, A_2, \sigma_\theta)}{\partial H_2} \frac{\partial H_2}{\partial s_1} \right] = 0. \tag{A.17}$$

Based on eq. (A.17), we can write an unconditional decision rule for s_t as

$$s_t = s_t(\sigma_\theta, A_t, H_t, \theta_t). \tag{A.18}$$

B Data Appendix

B.1 REDS Data

Our primary data set is the Rural Economic and Demographic Survey (REDS). The REDS is the follow-up survey of the Additional Rural Incomes Survey (ARIS), which was carried out between 1969 and 1971. The ARIS sample was designed to represent the rural population of India across 17 major states, and covers 4,527 households in 259 villages. The REDS expanded the thematic coverage of the ARIS, and in addition to revisiting the original ARIS households, increased the sample size by randomly sampling additional households from the same villages. The REDS data were collected in 1981-82, 1998-99, and 2007-08.[48] We use data from all three REDS rounds, but restrict the sample to the 242 villages from which data were collected consistently.[49] Because the sample size increased substantially over time, we re-weight all observations to give equal weight to each of the three survey rounds.

The REDS is particularly suited for the analysis proposed here, as it collects detailed information on time use for all women in the household and their children (linked through a mother identifier). This information refers the hours per day allocated to an extensive list of activities on a typical day, and is collected for three seasons of the year, which are also marked by very different levels of agricultural activity. The reference months are: October/November (season one), February (season two), and April/May (season three).[50] Importantly, this data explicitly differentiates between time in school, time allocated to household chores, to farm work, to the labor market and to leisure.

We restrict the sample to households with complete information on village irrigation, and on household consumption consumption and income. This sample consists of 4,704 households in 1981-82, 7,148 households in 1998-999 and 8,593 households in 2007-08. Table D.1 reports summary statistics for these households. In order to use the available information efficiently, we use this sample to estimate the effect of shocks on household level outcomes, and to calculate the risk variable.

During the time period we study, agriculture is the dominant source of income for the majority of households in rural India, either through own agricultural production or through casual agricultural employment. In the REDS sample, the share of households involved in agricultural production decreased from 86% in 1982 to 71% in 2007-08. Yet, even in 2007-08, total household income still consists to 55% of income from casual employment in agriculture and from own agricultural production (see Table D.1).[51]

Over this time period, the share of the net area cultivated that is under irrigation has increased from 51% to 60% in our sample (see Table D.1). This is higher than the national average, but follows similar trends: According to the Directorate of Economics and Statistics, the share of the net area sown that is under irrigation at the national level increased from 28% in 1980-81 to 45% in 2007-08 (Directorate of Economics and Statistics, 2019).

[48] A reduced version of the questionnaire was also administered between 2014 and 2016. This survey did not collect information on time allocation, but can be used to simulate the effects of NREGS in Section 5.

[49] Due to ongoing conflict, no data were collected in Jammu & Kashmir in the 1998-99 and 2007-08 round of interviews (11 villages), and in Assam in the 1981-81 and 2007-08 round of interviews (8 villages). The 242 villages for which data were collected in all three rounds are distributed across 15 states (17 states since the foundation of Chhattisgarh and Jharkhand in 2000).

[50] As such, season one corresponds to the harvest period of rice (and planting period of wheat). Season two is the slack season, and season three the planting period of rice (and harvest period of wheat). In 2007-08, the questionnaire did not make reference to precise months.

[51] The agricultural year in India typically lasts from July to June, and the main harvest is strongly influenced by the onset of the monsoon and precipitation levels during the monsoon. The monsoon onset varies across places in India, but is usually between May and July, and its end is mostly around September or October.

In our main analysis, then, we focus on all children aged 6 to 15 with complete information on time allocation across all three seasons. We focus on the age group 6 to 15 because most children still live at home at this age, and because the questionnaire only collects this information for women in the household and her own (co-resident) children (not her sons or daughters in-law) in 1981-82 and in 1998-99. By restricting ourselves to younger ages, we seek to minimize any selection bias that could arise if children with low education were more likely to leave the house at younger ages (*e.g.* in order to get married). Because not all households have children in this age group, the child sample comprises 5,581 children from 2,388 households in 1981-82, 8,424 children from 3,818 households in 1998-99, and 8,761 children from 4,472 households in 2007-08. The unit of observation in the main estimation is the child by agricultural season and round of interview, resulting in a final sample size of 68,298 observations. Child-by-season level summary statistics are reported in Table D.2. Note that is is not possible to track individuals across time in the REDS data. While we can follow a subset of households over time (and control for household fixed effects in part of the analysis), we cannot control for individual fixed effects.

Achieving universal education has been the declared goal of Indian governments since independence. The Right to Education Act of 2002 declares free and compulsory education a fundamental right of children aged 6 to 14. Since then, substantial improvements have been made in the enrollment rates of boys and girls and in closing the gender gap in primary school enrollment, which are also reflected in our data. In the REDS sample, the share of boys aged 6 to 15 that are attending school increased from 68% in 1981-81 to 93% in 2007-08. Likewise, the share of girls that are attending school increased from 53% to 91% in the same age group.[52]

The time use data presented in Table D.2 shows that boys and girls spend most of the day in school but that other activities are important, too. Boys in our sample spend on average 5.5 (7.1) hours per day in school or studying in 1981-82 (2007-08). They also spend an average of 2.9 (1.0) hours per day on household chores, and 1.6 (0.1) hours per day working on-farm or on the labor market. Girls spend slightly less time in school (4.2 hours per day in 1981-82 and 6.9 hours in 2007-08), and more time on household chores (4.8 hours per day in 1981-82 and 1.3 hours in 2007-08). As found in previous work, agricultural work and wage work are a little less relevant for girls: they spend about 1.1 (0.1) hours per day these activities in 1981-82 (2007-08). In terms of leisure, finally, we do not observe substantial differences between boys and girls, but an increase in time allocated to leisure over time.[53] In figure C.7, we graph the average time children spend on different activities on a typical day by age and gender.

B.2 Weather Data

To obtain precise rainfall data, we merge our data with ERA5 (reanalysis) monthly averaged precipitation data (Copernicus Climate Change Service, 2017). ERA5 is the fifth generation reanalysis for the global climate and weather by the European Center for Medium-Range Weather Forecasts (ECMWF). The data is available since 1979, and in grid cells of 0.25x0.25 degrees (approximately 31x31 km). We calculate total annual rainfall in Millimeters, and merge the data point that is closest to the village centroid.

[52]Note that these figures are likely smaller than actual enrollment rates, because this variable captures whether a child is actually attending school in a given time period. Also, the reference period of season three (April/May) could overlap with summer vacations in the southern states of India.

[53]Note, that the questionnaire changed in how it elicited leisure over survey rounds. Leisure includes sleep time in all cases, but in the rounds 1981-82 and 1998-99, the questionnaire only accounts for hours between 04:00 and 24:00, whereas the 2007-08 questionnaire accounts for 24 hours. Thus by construction, leisure should be around 4 hours higher in 2007-08.

Some villages fall in the same cell, resulting in 195 unique data points per round of interview. We merge the precipitation data by the agricultural year to explain agricultural production and income, i.e. we use May 1998 to April 1999 rainfall to explain agricultural income, area cultivated and yields of the agricultural year 1998-99, and so forth. Note, that the consumption module and the time use module in the REDS do not refer to the last agricultural year, but to the last 12 months prior to the interview (except in 1998-99, where the consumption module also refers to the AY 1998-99). We therefore merge precipitation data of the current year, i.e. May 1982 to April 1983, for interviews conducted in the fall of 1982, and so forth. Apart from generating the best fit, this seems sensible because the bulk of household consumption consists of food products, and food consumption is typically extrapolated from last month's consumption to the entire year.

B.3 SEPRI Data

The Socio-Economic Profiles of Rural Households in India (SEPRI) data of 2014 and 2016 is a follow-up survey to the REDS. It that was collected in 13 states of India: Andhra Pradesh, Bihar, Chhattisgarh, Gujarat, Haryana, Jharkhand, Madhya Pradesh, Maharashtra, Orissa, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal.

It samples the entire population of the REDS survey villages, but applies a questionnaire that is considerably shorter than the REDS. We drop Gujarat from the analysis because information about the NREGS was not collected for all villages in that state. The rainfall data covers the agricultural year 2012/13 for interviews conducted in 2014, and the agricultural year 2014/15 for interviews conducted in 2016. We also use a one year lag of employment generation within the NREGS (captured in the village questionnaire) to minimize concerns about reverse causality.

C Supplementary Figures

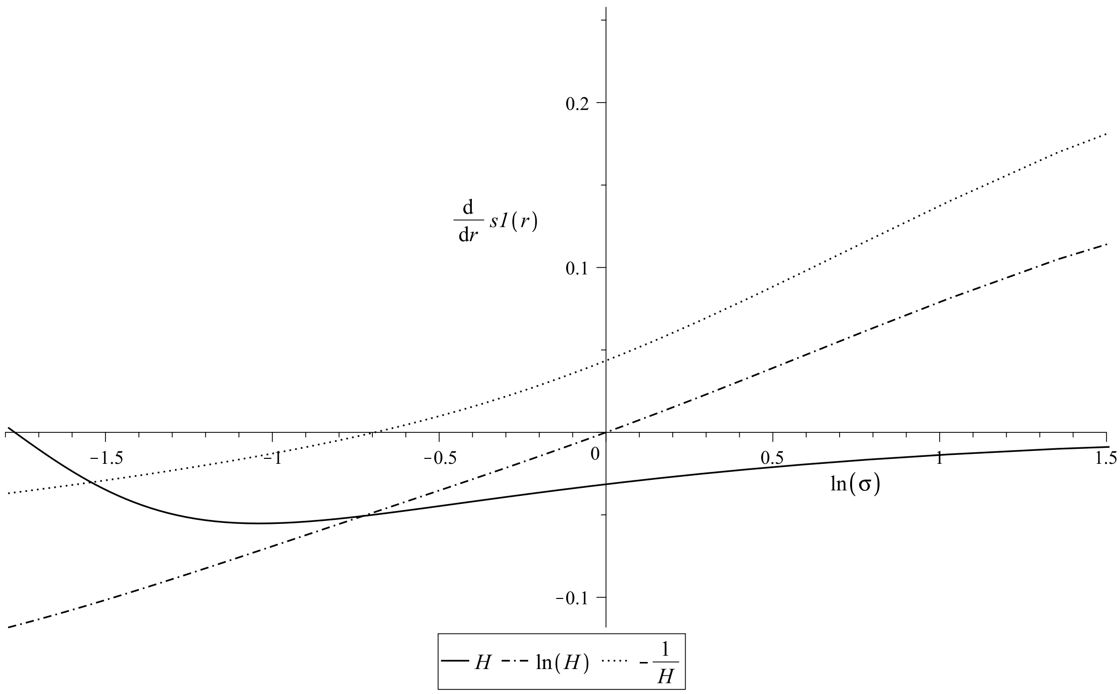

Figure C.1: Numeric Approximations of the Effect of Risk on First Period Schooling with High Risk

Notes: This figure presents numerical solutions to the derivative of s_1 wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the *ex ante* effect of risk, for a two-period model as described by eq. (1) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 5$ and $\zeta = 1$.

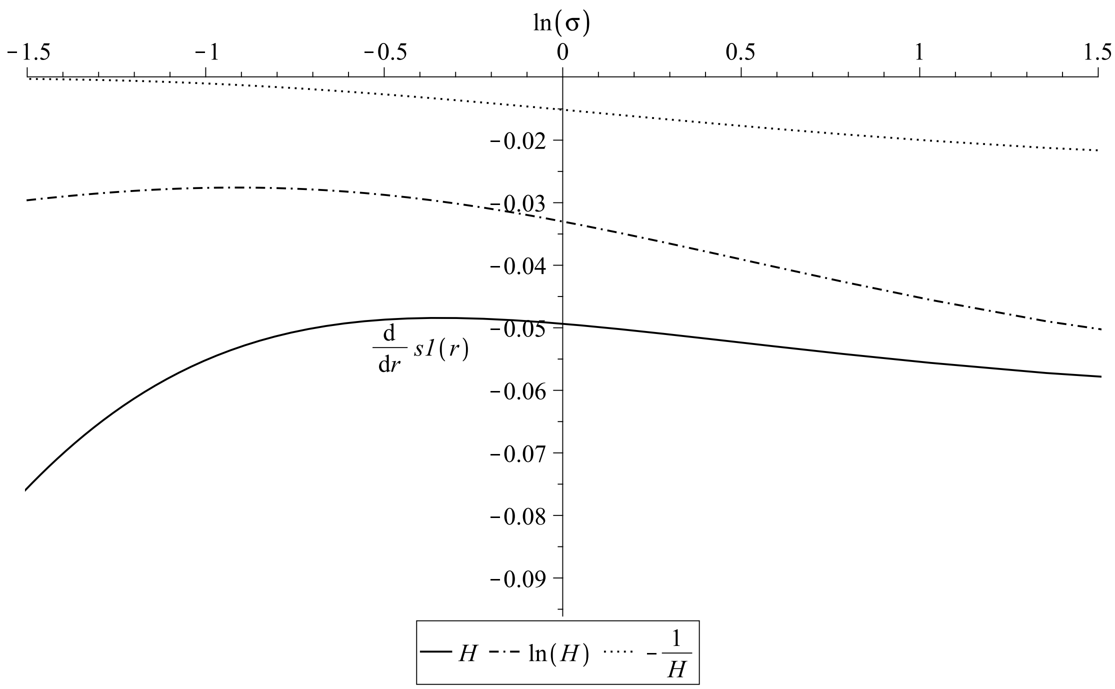

Figure C.2: Numeric Approximations of the Effect of Risk on First Period Schooling with Savings

Notes: This figure presents numerical solutions to the derivative of s_1 wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the *ex ante* effect of risk, for a two-period model that allows for precautionary savings as described by eq. (A.8) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 1$ and $\zeta = 1$.

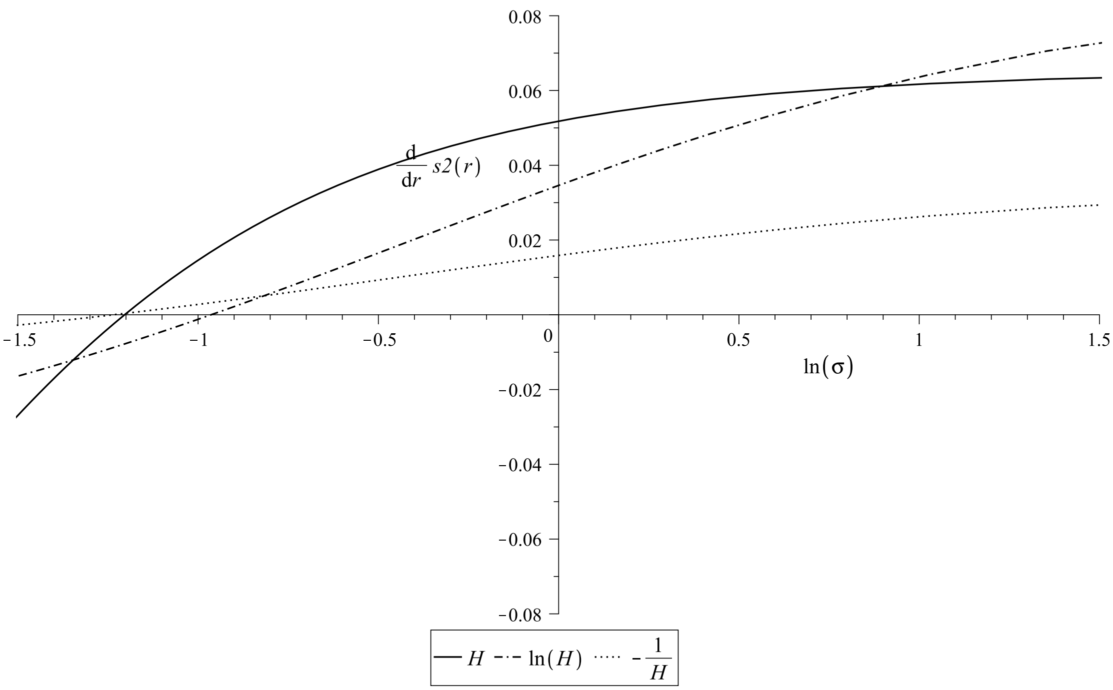

Figure C.3: Numeric Approximations of the Effect of Risk on Expected Second Period Schooling with Savings

Notes: This figure presents numerical solutions to the derivative of $\mathbb{E}[s_2]$ wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the *ex post* effect of risk, for a two-period model with precautionary savings as described by eq. (A.8) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 1$ and $\zeta = 1$.

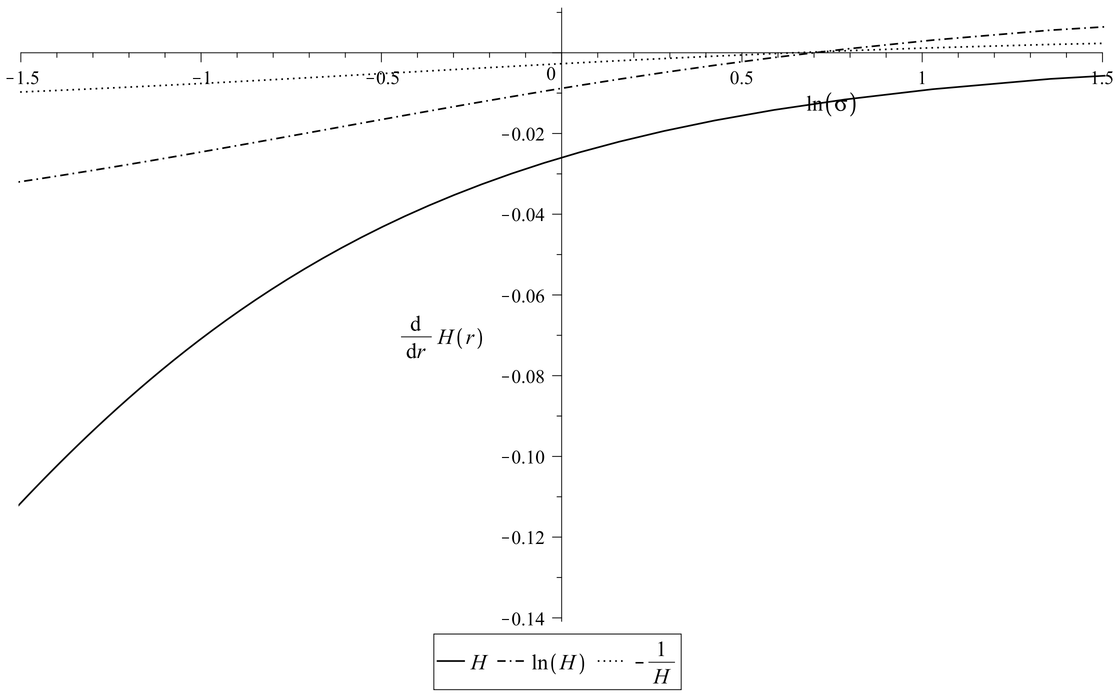

Figure C.4: Numeric Approximations of the Effect of Risk on Final Human Capital with Savings

Notes: This figure presents numerical solutions to the derivative of $\mathbb{E}[H_2]$ wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the combined *ex ante* and *ex post* effect of risk, for a two-period model with precautionary savings as described by eq. (A.8) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 1$ and $\zeta = 1$.

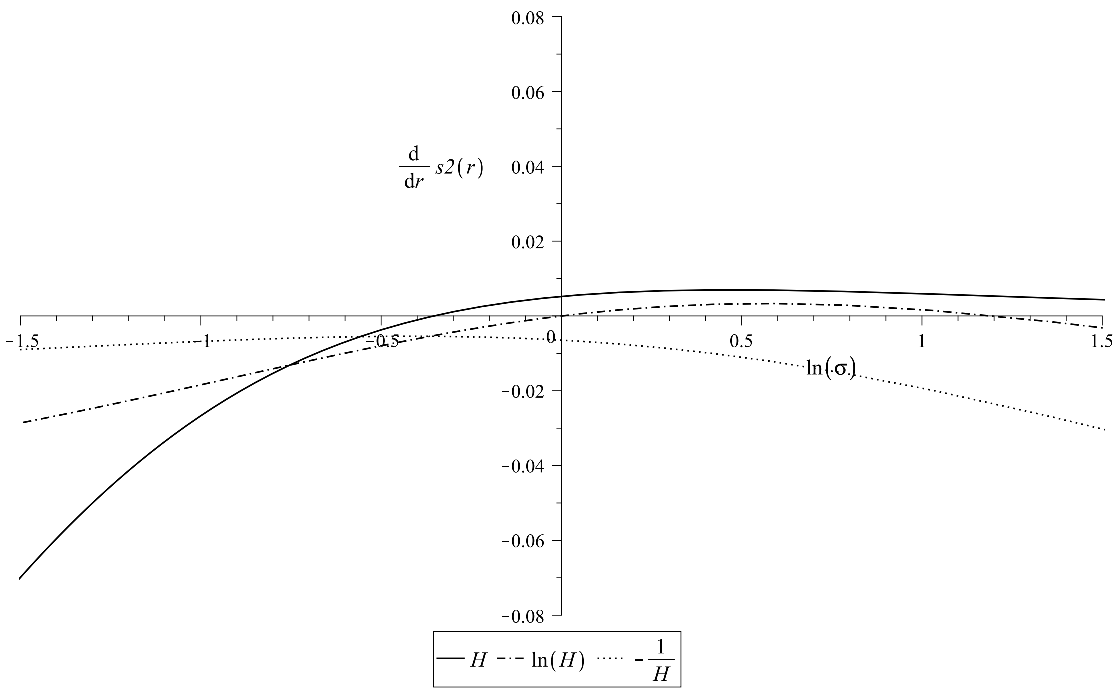

Figure C.5: Numeric Approximations of the Effect of Risk on Expected Second Pperiod Schooling – No Savings

Notes: This figure presents numerical solutions to the derivative of $\mathbb{E}[s_2]$ wrt to the standard deviation of the shock ($\sigma_\theta = r$), i.e. the *ex post* effect of risk, for a two-period model as described by eq. (1) and for various values of $\sigma = 1/(1-\rho)$ and three specifications of $v(H)$: $v(H) = H$, $v(H) = \ln(H)$ and $v(H) = -1/H$. Assumed parameter values are $y_1 = 10$, $y_2 = 10$, $r = 1$ and $\zeta = 1$.

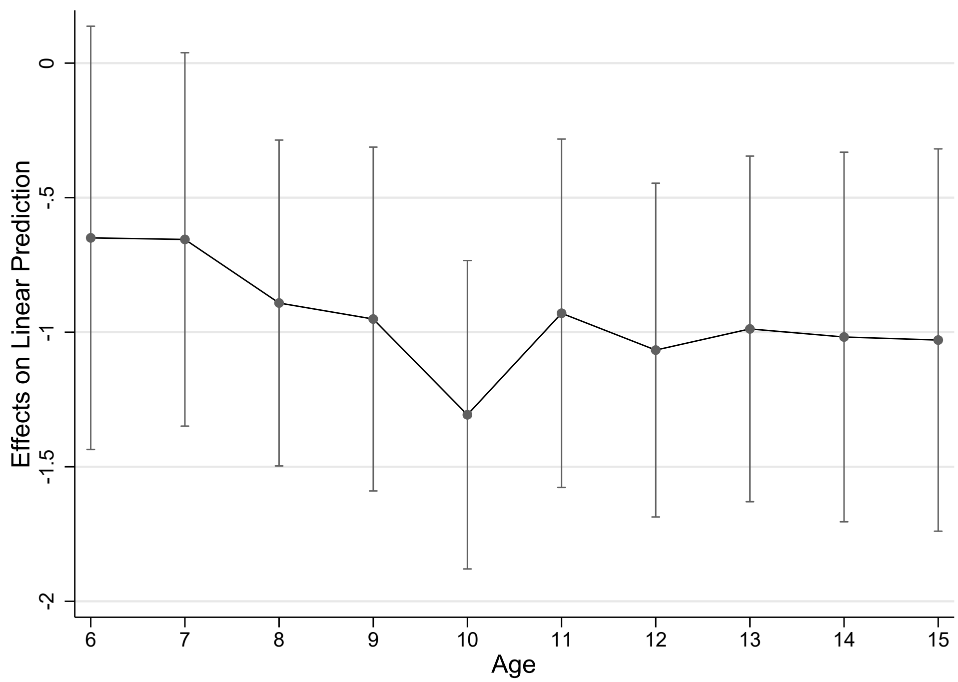

Figure C.6: Effect of Risk on School Attendance by Age

Notes: This figure plots the age-specific coefficients of risk and 90% confidence intervals from a regression of school attendance on risk interacted with dummies for each age. The regression controls for village, season-by-round of interview and gender fixed effects, for current and for lagged rainfall in the past 3 years (in log mms), and the standard deviation of rainfall over the last 3 years. The regression also controls for income, wealth and area (all measured in per capita terms, in logs) and for household size. Standard errors are corrected to allow for clustering at the level of the grid cell by round of interview.

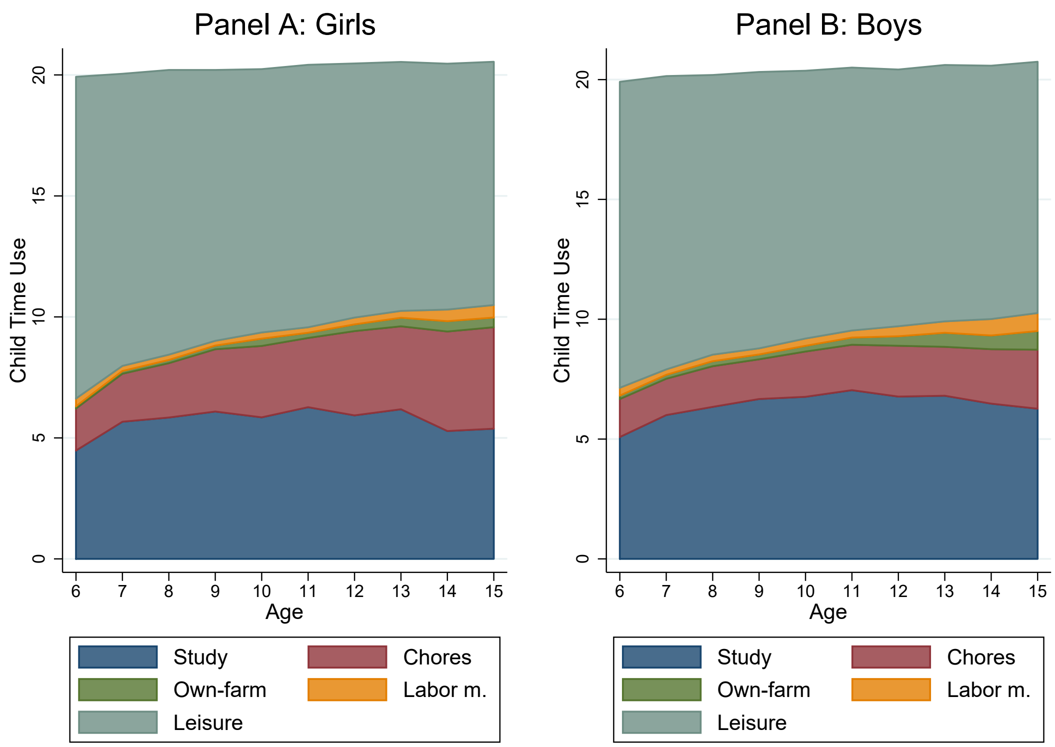

Figure C.7: Time-Use of Children by Gender and Age

Notes: This figure shows the average reported time use of children in hours per day by age and gender.

D Supplementary Tables

Table D.1: Household Characteristics

	1981-82		1998-99		2007-08	
	Mean	SD	Mean	SD	Mean	SD
Household size	6.86	3.56	6.06	3.40	5.17	2.60
Hh owns any agr. land	0.78	0.42	0.75	0.43	0.63	0.48
Area owned (acres)	6.37	9.36	4.05	7.42	2.73	5.44
Area owned per capita	1.01	1.67	0.74	1.36	0.59	1.32
Household cultivated any land	0.86	0.35	0.80	0.40	0.71	0.45
Income from wages	14709.3	25610.0	32335.0	51654.5	25587.9	40605.1
Income from agr. wage work	4371.2	9473.7	6424.2	14302.8	5942.1	13012.9
Annual income, agriculture and livestock	40506.3	57921.5	43282.0	85627.1	47886.4	90568.9
Income from self-employment	7107.5	33741.6	7073.5	35060.2	16210.4	78835.0
Non-labor income	2737.5	21902.7	3546.0	19269.8	8395.6	33309.7
Annual income, total	65060.6	74125.4	86236.5	103161.0	98080.3	126554.7
Annual income per capita	10043.8	11327.7	16026.8	21766.6	21016.4	29588.7
Total consumption expenditures	56239.7	39150.9	66026.0	54179.0	54077.3	38633.6
Consumption expenditures per capita	9092.6	6320.7	11983.7	8538.1	11650.5	8849.8
Household wealth	165323.5	340774.7	175813.7	377987.9	139373.7	211952.3
Household wealth per capita	26977.3	57602.9	32451.1	68614.3	30542.1	49236.9
Irrigated area (share of village agr. area)	0.51	0.39	0.58	0.36	0.60	0.34
Precipitation, current year	864.5	406.8	1087.1	506.8	1212.1	516.3
Precipitation, one year lag	1012.0	502.0	1182.3	521.7	1148.1	660.8
Precipitation, two years lag	1080.7	560.5	1113.9	500.7	1111.8	508.5
Precipitation, three years lag	867.7	482.0	1074.9	553.6	1053.0	494.5
Rainfall Quintile	1.96	1.15	3.32	1.21	3.80	1.27
IQR of predicted log consumption	0.039	0.032	0.033	0.028	0.032	0.027
SD of predicted log consumption	0.030	0.025	0.026	0.022	0.025	0.022
Observations	4704		7148		8593	

Notes: All nominal values are reported in constant 2010 INR. Prices are deflated using the All India Consumer Price Index for Agricultural Labourers in the month of July.

Table D.2: Individual Characteristics

	1981-82		1999-99		2007-08	
	Mean	SD	Mean	SD	Mean	SD
PANEL A: Boys						
Age	10.4	2.79	10.2	2.81	10.9	2.76
Attending school	0.68	0.47	0.81	0.39	0.93	0.25
Hours per day: wage work	0.78	2.26	0.27	1.15	0.10	0.90
Hours per day: on-farm agriculture	0.85	2.56	0.13	0.83	0.037	0.42
Hours per day: household chores	2.91	3.56	1.69	2.37	1.04	1.56
Hours per day: studying	5.45	4.22	6.88	3.70	7.08	2.62
Hours per day: leisure	8.87	3.70	11.0	3.57	14.3	1.40
Observations	9048		13440		14514	
PANEL B: Girls						
Age	10.2	2.78	10.2	2.79	10.8	2.77
Attending school	0.53	0.50	0.74	0.44	0.91	0.29
Hours per day: wage work	0.53	1.82	0.26	1.10	0.058	0.65
Hours per day: on-farm agriculture	0.57	2.02	0.11	0.73	0.011	0.21
Hours per day: household chores	4.75	4.39	2.48	2.95	1.33	1.84
Hours per day: studying	4.19	4.31	6.23	4.01	6.89	2.81
Hours per day: leisure	8.76	3.85	10.9	3.82	14.2	1.48
Observations	7695		11832		11769	
PANEL C: Mothers						
Age of mother	34.9	6.90	34.3	6.65	36.0	7.27
Number of children born alive to mother	4.88	2.21	3.53	1.67	3.35	1.67
Hours per day: wage work	1.54	2.98	1.54	2.81	1.29	2.85
Hours per day: on-farm agriculture	2.39	3.55	1.43	2.32	0.88	1.96
Hours per day: household chores	9.38	3.78	9.71	2.93	7.92	2.83
Observations	7806		12504		13176	

Notes: The unit of observation is the individual by season.

Table D.3: Study Time Across Seasons (Panel households only)

	(1)	(2)	(3)	(4)	(5)
Study time, season 1	0.603***	0.603***	0.603***	0.558***	0.558***
	(0.048)	(0.048)	(0.048)	(0.065)	(0.066)
Study time, season 1 (square)	0.019***	0.019***	0.019***	0.018***	0.016**
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)
Study time, season 2	0.438***	0.440***	0.441***	0.534***	0.546***
	(0.049)	(0.050)	(0.049)	(0.067)	(0.069)
Study time, season 2 (square)	0.041***	0.041***	0.041***	0.030***	0.028***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)
Study time, season 1 × season 2	-0.075***	-0.075***	-0.075***	-0.068***	-0.066***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)
Rainfall controls	N	Y	Y	Y	Y
Income, wealth, area, hh size, risk	N	N	Y	Y	Y
Household-specific time trend	N	N	N	Y	Y
Village-by-round FE	N	N	N	N	Y
Observations	18403	18403	18403	18403	18402
Adjusted R^2	0.860	0.860	0.860	0.895	0.893

The dependent variable is study time in season 3. Each regression controls for household, state-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms). Income, wealth and area are measured in per capita terms (and are in logs). Risk is the interquartile range of predicted log consumption. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4: School time across seasons (Northern States only)

	(1)	(2)	(3)	(4)	(5)
Study time, season 1	0.719***	0.720***	0.720***	0.742***	0.755***
	(0.052)	(0.054)	(0.054)	(0.057)	(0.057)
Study time, season 1 (square)	0.014***	0.014***	0.014***	0.014***	0.013***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Study time, season 2	0.307***	0.311***	0.308***	0.277***	0.264***
	(0.062)	(0.063)	(0.063)	(0.066)	(0.068)
Study time, season 2 (square)	0.054***	0.053***	0.053***	0.057***	0.058***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Study time, season 1 $\times$ season 2	-0.081***	-0.081***	-0.081***	-0.084***	-0.084***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Rainfall controls	N	Y	Y	Y	Y
Income, wealth, area, hh size, risk	N	N	Y	Y	Y
State-by-round FE	N	N	N	Y	Y
Village-by-round FE	N	N	N	N	Y
Observations	13981	13981	13981	13981	13981
Adjusted R^2	0.809	0.810	0.810	0.815	0.825

The dependent variable is study time in season 3. Sample is restricted to northern states of India, *i.e.* we drop Karnataka, Kerala, Tamil Nadu, Andhra Pradesh, and Maharashtra. Each regression controls for village, round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms). Income, wealth and area are measured in per capita terms (and are in logs). Risk is the interquartile range of predicted log consumption. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: Study Time Across Seasons 1 & 2

	(1)	(2)	(3)	(4)	(5)
Study time, season 1	0.982***	0.992***	0.990***	1.017***	1.040***
	(0.026)	(0.026)	(0.026)	(0.024)	(0.026)
Study time, season 1 (square)	-0.010***	-0.012***	-0.012***	-0.015***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Rainfall controls	N	Y	Y	Y	Y
Income, wealth, area, hh size, risk	N	N	Y	Y	Y
State-by-round FE	N	N	N	Y	Y
Village-by-round FE	N	N	N	N	Y
Observations	22766	22766	22766	22766	22766
Adjusted R^2	0.816	0.822	0.822	0.829	0.834

Notes: The dependent variable is study time in season 2. Each regression controls for village, round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms). Income, wealth and area are measured in per capita terms (and are in logs). Risk is the interquartile range of predicted log consumption. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: Effect of Rainfall Shocks on Consumption and Time Use

	Household		Child time	
	Agr.inc.	Cons.exp.	Home prod.	Studying
	(1)	(2)	(3)	(4)
Rainfall Quintile=1	-0.521**	-0.067	0.448	-0.285
	(0.232)	(0.045)	(0.283)	(0.242)
	[0.233]	[0.048]	[0.303]	[0.250]
Rainfall Quintile=2	-0.405**	0.011	-0.161	-0.033
	(0.196)	(0.033)	(0.229)	(0.199)
	[0.205]	[0.033]	[0.221]	[0.203]
Rainfall Quintile=4	-0.067	0.051	0.175	0.136
	(0.182)	(0.031)	(0.248)	(0.217)
	[0.185]	[0.033]	[0.239]	[0.242]
Rainfall Quintile=5	0.117	0.040	-0.055	0.398*
	(0.203)	(0.038)	(0.292)	(0.225)
	[0.233]	[0.040]	[0.298]	[0.229]
Dependent variable mean	ln(8741)	ln(9219)	2.69	6.10
Observations	15849	20445	68298	68298
Adjusted R^2	0.062	0.264	0.371	0.177

Notes: Agricultural income and consumption expenditures are in log INR per capita. Child time allocation is in hours per day for three seasons. Each regression controls for state-by-round of interview fixed effects, for irrigated area (share of village area), and for lagged rainfall (in logs) in the last 3 years (2 years in col. (1)). Columns (3) and (4) also control for season-by-round, and age-by-gender fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the grid cell by round of interview. Conley (1999) standard errors (300km cut-off, Bartlett weights) in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.7: Effect of Rainfall on Consumption Expenditures per Capita (log)

	(1)	(2)	(3)	(4)	(5)
Precipitation, current year (1,000 mms)	0.105*	0.115			
	(0.053)	(0.122)			
	[0.056]	[0.132]			
Irrigated area (share of village agr. area)	0.172***	0.172***	0.170***	1.608***	0.907***
	(0.034)	(0.034)	(0.034)	(0.394)	(0.352)
	[0.035]	[0.035]	[0.035]	[0.396]	[0.338]
Precipitation, current year (square)		-0.004			
		(0.033)			
		[0.033]			
Precipitation, current year (log mms)			0.119*	0.229***	0.138**
			(0.063)	(0.071)	(0.061)
			[0.071]	[0.076]	[0.061]
Precipitation × Irrigated area				-0.210***	-0.116**
				(0.057)	(0.050)
				[0.057]	[0.047]
Dependent variable mean	$ln(8741)$	$ln(8741)$	$ln(8741)$	$ln(8741)$	$ln(8741)$
Observations	20445	20445	20445	20445	20445
Adjusted R^2	0.263	0.263	0.263	0.266	0.334

Notes: Consumption expenditures are in log INR per capita. Each regression controls for state-by-round of interview fixed effects, and for lagged rainfall (in logs) in the last 3 years. Column (5) also controls for village fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the grid cell by round of interview. Conley (1999) standard errors in brackets (cut-off 300km, Bartlett weights). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, according to Conley standard errors.

Table D.8: Effect of Rainfall on Study Time

	(1)	(2)	(3)	(4)	(5)
Precipitation, current year (1,000 mms)	0.616**	1.764***			
	(0.325)	(0.690)			
	[0.299]	[0.650]			
Irrigated area (share of village agr. area)	0.854***	0.844***	0.850***	7.856***	9.432***
	(0.203)	(0.202)	(0.201)	(2.745)	(2.277)
	[0.204]	[0.204]	[0.203]	[2.794]	[2.376]
Precipitation, current year (square)		-0.425**			
		(0.206)			
		[0.190]			
Precipitation, current year (log mms)			0.955***	1.525***	1.334***
			(0.357)	(0.434)	(0.375)
			[0.337]	[0.416]	[0.379]
Precipitation × Irrigated area				-1.028**	-1.286***
				(0.393)	(0.324)
				[0.403]	[0.336]
Dependent variable mean	6.10	6.10	6.10	6.10	6.10
Observations	68298	68298	68298	68298	68298
Adjusted R^2	0.176	0.177	0.177	0.179	0.245

Notes: Time spent studying is in hours per day for three seasons. Each regression controls for state-by-round of interview fixed effects, for age-by-gender fixed effects, for season-by-round fixed effects, and for lagged rainfall (in logs) in the last 3 years. Column (5) also controls for village fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the grid cell by round of interview. Conley (1999) standard errors in brackets (cut-off 300km, Bartlett weights). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, according to Conley standard errors.

Table D.9: Effect of Rainfall on Child Time Use – No village FE

Hours per day:	On-farm		Chores		Studying		Leisure	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Precipitation, current year (log mms)s	-0.796	-0.478	-0.168	-0.678	1.186***	1.872***	0.276	-0.203
	(0.504)	(0.299)	(0.520)	(0.485)	(0.433)	(0.478)	(0.603)	(0.503)
Irrigated area (share of village agr. area)	-0.968	-0.543	-4.383*	-7.303***	6.106**	9.517***	2.406	0.975
	(1.890)	(1.200)	(2.627)	(2.781)	(2.563)	(3.482)	(2.468)	(2.473)
Precipitation × Irrigated area	0.124	0.072	0.553	0.972**	-0.771**	-1.274**	-0.357	-0.128
	(0.266)	(0.168)	(0.381)	(0.404)	(0.369)	(0.506)	(0.359)	(0.357)
Dep. Var. Mean	0.35	0.24	1.92	2.94	6.44	5.70	11.27	11.11
Observations	37002	31296	37002	31296	37002	31296	37002	31296
Adjusted R^2	0.112	0.096	0.252	0.356	0.150	0.209	0.466	0.456

Notes: Time allocation is reported in hours per day over three seasons. Each regression controls for state-by-round of interview, season-by-round, and age fixed effects, and for lagged rainfall (in logs) in the last 3 years. Conley (1999) standard errors in parentheses (cut-off 300km, Bartlett weights). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.10: Effect of Risk on School Attendance (with village FE in risk calculation)

	(1)	(2)	(3)	(4)	(5)
Panel A:					
IQR of predicted log consumption	-2.086	-2.420	-2.404	-2.358	-1.806
	(0.651)	(0.685)	(0.684)	(0.627)	(0.642)
	[0.143]	[0.138]	[0.141]	[0.126]	[0.187]
Treatment Mean	0.025	0.025	0.025	0.025	0.025
Treatment Std. Dev.	0.017	0.017	0.017	0.017	0.017
Adjusted R^2	0.207	0.209	0.209	0.228	0.240
Panel B:					
SD of predicted log consumption	-2.447	-2.835	-2.814	-2.798	-2.111
	(0.836)	(0.901)	(0.900)	(0.818)	(0.816)
	[0.186]	[0.195]	[0.199]	[0.182]	[0.219]
Treatment Mean	0.019	0.019	0.019	0.019	0.019
Treatment Std. Dev.	0.013	0.013	0.013	0.013	0.013
Adjusted R^2	0.207	0.208	0.209	0.228	0.239
Rainfall controls	N	Y	Y	Y	Y
Rainfall by schoolage	N	N	Y	Y	Y
Income, wealth, area, hh size	N	N	N	Y	Y
State-by-round FE	N	N	N	N	Y
Dep. var. mean	0.762	0.762	0.762	0.762	0.762
Observations	68298	68298	68298	68298	68298

Notes: In this table, risk is calculate with the estimates of column (4), table 5. Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Rainfall by schoolage additionally interacts each rainfall lag with a dummy that equals 1 of the child was 6 years or older in that year. Income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.11: Effect of Risk on Study Time

	All children					School children				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IQR of predicted log consumption	-8.498**	-8.706**	-8.646**	-8.109**	-7.373**	-1.198	-1.123	-1.109	-1.057	-1.228
	(2.799)	(2.803)	(2.805)	(2.772)	(2.382)	(2.744)	(2.651)	(2.654)	(2.662)	(2.118)
	[0.048]	[0.037]	[0.037]	[0.047]	[0.033]	[0.749]	[0.761]	[0.763]	[0.781]	[0.659]
Adjusted R^2	0.221	0.223	0.223	0.242	0.264	0.219	0.221	0.221	0.222	0.279
SD of predicted log consumption	-9.755*	-9.833*	-9.749*	-9.077*	-8.475*	-1.279	-1.167	-1.147	-1.048	-1.412
	(3.489)	(3.511)	(3.514)	(3.484)	(3.027)	(3.566)	(3.388)	(3.391)	(3.411)	(2.723)
	[0.065]	[0.073]	[0.073]	[0.089]	[0.054]	[0.805]	[0.825]	[0.829]	[0.836]	[0.717]
Adjusted R^2	0.221	0.222	0.223	0.242	0.264	0.219	0.221	0.221	0.222	0.279
Rainfall controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
Rainfall by schoolage	N	N	Y	Y	Y	N	N	Y	Y	Y
Income, wealth, area, hh size	N	N	N	Y	Y	N	N	N	Y	Y
State-by-round FE	N	N	N	N	Y	N	N	N	N	Y
Dep. var. mean	6.10	6.10	6.10	6.10	6.10	8.01	8.01	8.01	8.01	8.01
Observations	68298	68298	68298	68298	68298	54092	54092	54092	54092	54092

Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current rainfall and lagged rainfall in the past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Rainfall by schoolage additionally interacts each lag with a dummy that equals 1 of the child was 6 years or older in that year. Income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.12: Effects of Risk on School Attendance by Gender and Agegroup

	Boys		Girls		Age 6 -10		Age 11 -15	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQR of predicted log consumption	-0.768	-0.699	-1.283**	-1.161**	-0.928*	-0.798	-1.024**	-1.007*
	(0.350)	(0.370)	(0.313)	(0.333)	(0.320)	(0.355)	(0.315)	(0.327)
	[0.141]	[0.205]	[0.017]	[0.039]	[0.06]	[0.131]	[0.028]	[0.051]
0.141 Adjusted R^2	0.189	0.201	0.281	0.294	0.236	0.247	0.247	0.261
SD of predicted log consumption	-0.875	-0.791	-1.422**	-1.287*	-0.981	-0.813	-1.234*	-1.244*
	(0.448)	(0.472)	(0.400)	(0.419)	(0.410)	(0.452)	(0.406)	(0.413)
	[0.185]	[0.242]	[0.034]	[0.053]	[0.106]	[0.229]	[0.058]	[0.051]
Adjusted R^2	0.189	0.201	0.281	0.294	0.236	0.247	0.247	0.261
Rainfall controls	Y	Y	Y	Y	Y	Y	Y	Y
Rainfall by schoolage	Y	Y	Y	Y	Y	Y	Y	Y
Income, wealth, area, hh size	Y	Y	Y	Y	Y	Y	Y	Y
State-by-round FE	N	Y	N	Y	N	Y	N	Y
Dep. var. mean	0.802	0.802	0.715	0.715	0.763	0.763	0.761	0.761
Observations	37002	37002	31296	31296	35625	35625	32673	32673

Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current rainfall and lagged rainfall in the past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Rainfall by schoolage additionally interacts each lag with a dummy that equals 1 of the child was 6 years or older in that year. Income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.13: Effect of Risk on School Infrastructure

TREATMENT VARIABLE:	IQR of exp. cons		SD of exp. cons	
	(1)	(2)	(3)	(4)
Panel A: Village-level outcomes				
Primary school in village (=1)	-0.847	0.318	-1.178	0.286
	(0.972)	(1.012)	(1.134)	(1.209)
	[0.429]	[0.760]	[0.363]	[0.813]
Middle school in village (=1)	-0.864	-0.950	-0.854	-0.871
	(1.125)	(1.151)	(1.399)	(1.433)
	[0.461]	[0.460]	[0.547]	[0.567]
Lower secondary school in village (=1)	-1.095	-0.893	-1.509	-1.261
	(0.920)	(0.954)	(1.127)	(1.163)
	[0.281]	[0.379]	[0.234]	[0.310]
Higher secondary school in village (=1)	-0.036	0.169	-0.204	0.107
	(0.729)	(0.740)	(0.911)	(0.913)
	[0.951]	[0.814]	[0.826]	[0.905]
Village and round FE	Y	Y	Y	Y
State-by-round FE	N	Y	N	Y
Observations	710	710	710	710
Panel B: Child-level outcomes				
School attendance (=1)	-1.096**	-1.035**	-1.270**	-1.203**
	(0.284)	(0.297)	(0.368)	(0.378)
	[0.023]	[0.037]	[0.034]	[0.042]
Village and round FE	Y	Y	Y	Y
Village & hh controls	Y	Y	Y	Y
School controls	Y	Y	Y	Y
State-by-round FE	N	Y	N	Y
Observations	67482	67482	67482	67482

Notes: Each coefficient relates to an individual regression. The unit of observation is the village-by-round (Panel A), and the child (Panel B). Each regression controls for village and round of interview fixed effects. Cols. (2) and (4) add state-by-round of interview fixed effects. Child-level regressions additionally control for season-by-round of interview, and age-by-gender fixed effects, and for current rainfall and for lagged rainfall (in log mms) in the past 3 years. Each lag is additionally interacted with a dummy that equals 1 of the child was 6 years or older in that year. Child regressions also control for income, wealth and area (all measured in per capita terms, in logs), for household size, and for the presence of all four school types in the village (primary, middle, lower secondary and higher secondary). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile Bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.14: Effect of Risk on Village Characteristics

TREATMENT VARIABLE:	IQR of pred. log cons		SD of pred. log cons	
	(1)	(2)	(3)	(4)
Agricultural income (share of village income)	0.138	0.313	0.319	0.581
	(0.418)	(0.397)	(0.589)	(0.539)
	[0.757]	[0.469]	[0.636]	[0.360]
Share of households involved in agr.	-0.469	-0.575	-0.419	-0.575
	(0.550)	(0.452)	(0.733)	(0.606)
	[0.447]	[0.241]	[0.605]	[0.354]
Any factory in village	0.252	-0.216	0.254	-0.450
	(1.081)	(1.085)	(1.413)	(1.380)
	[0.817]	[0.830]	[0.850]	[0.735]
No. of villagers employed in factories	116.826	142.625	121.169	149.594
	(88.854)	(91.913)	(125.223)	(127.540)
	[0.258]	[0.211]	[0.441]	[0.338]
Any cooperative in village	-0.190	0.953	-0.384	1.236
	(1.149)	(1.094)	(1.461)	(1.373)
	[0.885]	[0.379]	[0.799]	[0.372]
Any bank in village	-0.680	-0.870	-1.331	-1.379
	(1.501)	(1.417)	(1.887)	(1.777)
	[0.664]	[0.535]	[0.563]	[0.515]
Village and round FE	Y	Y	Y	Y
State-by-round FE	N	Y	N	Y
Observations	713	713	713	713

Notes: The unit of observation is the village-by-round. Each coefficient relates to an individual regression. Each regression controls for village and round of interview fixed effects. Cols. (2) and (4) add state-by-round of interview fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile Bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.15: Effect of Rainfall on Outcomes of Interest – Cash Savings

	Household	Maternal time	
	Cons. exp.	Household chores	Labor market
	(1)	(2)	(3)
Precipitation, current year (log mms)	0.488***	1.554*	-0.806
	(0.124)	(0.833)	(0.861)
Cash savings (per capita, log)	0.433***	0.979*	-0.053
	(0.093)	(0.563)	(0.633)
Precipitation × Cash savings	-0.042***	-0.120	-0.010
	(0.013)	(0.080)	(0.090)
Irrigated area (share of village agr. area)	0.472	0.914	-5.105*
	(0.306)	(2.628)	(3.063)
Precipitation × Irrigated area	-0.056	-0.076	0.773*
	(0.043)	(0.379)	(0.441)
Observations	15693	26736	26736
Adjusted R^2	0.583	0.500	0.572

Notes: Consumption expenditures are in log INR per capita. Female time allocation is in hours per day for three seasons. Each regression controls for household fixed effects, for state-by-round of interview fixed effects, and for lagged rainfall (in logs) in the last 3 years. Columns (2) and (3) also control for season-by-round, and age fixed effects. Conley (1999) standard errors in parentheses (cut-off 300km, Bartlett weights). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.16: Effect of Risk (calculated at household-level) on School Attendance

	(1)	(2)	(3)	(4)
Panel A:				
IQR of predicted log consumption	-5.423***	-5.374***	-4.324***	-4.499***
	(1.035)	(1.039)	(1.126)	(1.142)
	[0.001]	[0.000]	[0.001]	[0.000]
Adjusted R^2	0.573	0.573	0.573	0.571
Panel B:				
SD of predicted log consumption	-7.518***	-7.452***	-6.172***	-6.395***
	(1.273)	(1.278)	(1.408)	(1.420)
	[0.000]	[0.000]	[0.001]	[0.001]
Adjusted R^2	0.573	0.573	0.573	0.571
Rainfall controls	Y	Y	Y	Y
Rainfall by schoolage	N	Y	Y	Y
Income, wealth, area, hh size	N	N	Y	Y
Initial savings-by-round	N	N	N	Y
Dep. var. mean	0.762	0.762	0.762	0.762
Observations	55452	55452	55452	48543

Notes: Each regression controls for household, village-by-round of interview, season-by-round, and age-by-gender fixed effects, and for household-specific linear time trends. Rainfall by schoolage interacts each rainfall lag with a dummy that equals 1 of the child was 6 years or older in that year. Income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the household by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.17: Effect of Risk (calculated at household-level) on Economic Outcomes

	Income	Land	Wealth	Cons. exp	Agr. prod. (=1)	Mat. labor supply
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A:						
IQR of predicted log consumption	-7.557	-36.384*	-44.422**	-9.272*	-6.620**	21.583
	(2.778)	(4.647)	(1.819)	(0.829)	(0.823)	(12.141)
	[0.427]	[0.061]	[0.017]	[0.062]	[0.038]	[0.144]
Observations	13701	13701	13701	13701	13701	23280
Adjusted R^2	0.320	0.654	0.729	0.716	0.424	0.732
PANEL B:						
SD of predicted log consumption	-9.931	-45.121*	-56.867**	-11.721	-8.410*	23.289
	(3.475)	(5.863)	(2.324)	(1.099)	(1.072)	(15.079)
	[0.394]	[0.077]	[0.020]	[0.114]	[0.050]	[0.173]
Observations	13701	13701	13701	13701	13701	23280
Adjusted R^2	0.320	0.653	0.730	0.716	0.424	0.732

Notes: Income, land owned, wealth (value of all assets owned), and consumptions expenditures are in per capita terms and logs. Maternal labor supply is in hours per day for three seasons. Each regression controls for household and village-by-round of interview fixed effects, and for household-specific linear time trends. Col. (6) additionally controls for age and for season-by-round of interview fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the household by round of interview. Percentile Bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.18: Effect of Risk on School Attendance - Panel Households

	(1)	(2)	(3)	(4)	(5)
Panel A:					
IQR of predicted log consumption	-1.069**	-1.140**	-1.124**	-1.102**	-1.018*
	(0.285)	(0.275)	(0.275)	(0.270)	(0.311)
	[0.029]	[0.022]	[0.023]	[0.022]	[0.063]
Adjusted R^2	0.434	0.435	0.435	0.437	0.443
Panel B:					
SD of predicted log consumption	-1.289**	-1.375**	-1.354**	-1.323**	-1.226*
	(0.367)	(0.355)	(0.356)	(0.351)	(0.396)
	[0.045]	[0.035]	[0.039]	[0.033]	[0.078]
Adjusted R^2	0.434	0.435	0.435	0.436	0.443
Rainfall controls	N	Y	Y	Y	Y
Rainfall by schoolage	N	N	Y	Y	Y
Income, wealth, area, hh size	N	N	N	Y	Y
State-by-round FE	N	N	N	N	Y
Dep. Var. mean	0.762	0.762	0.762	0.762	0.762
Observations	56919	56919	56919	56919	56919

Notes: Each regression controls for household, season-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current and lagged rainfall in past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Rainfall by schoolage additionally interacts each rainfall lag with a dummy that equals 1 of the child was 6 years or older in that year. Income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.19: Robustness of Effects to Differential Time Trends

	(1)	(2)	(3)	(4)	(5)
Panel A:					
IQR of predicted log consumption	-0.619	-0.735*	-0.933**	-0.837**	-0.894*
	(0.319)	(0.282)	(0.278)	(0.282)	(0.292)
	[0.148]	[0.068]	[0.030]	[0.044]	[0.050]
Observations	66705	68298	68298	68298	68298
Adjusted R^2	0.208	0.218	0.209	0.210	0.208
Panel B:					
SD of predicted log consumption	-0.642	-0.892*	-1.080*	-0.958*	-1.054*
	(0.401)	(0.350)	(0.357)	(0.359)	(0.366)
	[0.234]	[0.079]	[0.053]	[0.063]	[0.057]
Observations	66705	68298	68298	68298	68298
Adjusted R^2	0.208	0.218	0.209	0.210	0.208
Irrigation (village mean, 1982) × round FE	Y	N	N	N	N
Hh head literate (village mean, 1982) × round FE	N	Y	N	N	N
Log hh income p.c. (village mean, 1982) × round FE	N	N	Y	N	N
Log hh wealth p.c. (village mean, 1982) × round FE	N	N	N	Y	N
Log hh area p.c. (village mean, 1982) × round FE	N	N	N	N	Y

Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the grid-cell by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.20: Effect of Risk on School Attendance - Placebo Tests

	(1)	(2)	(3)	(4)
IQR of predicted log consumption, current	-0.886			
	(0.406)			
	[0.160]			
IQR of predicted log consumption, lead		0.426		
		(0.413)		
		[0.453]		
SD of predicted log consumption, current			-1.016	
			(0.488)	
			[0.206]	
SD of predicted log consumption, lead				0.518
				(0.568)
				[0.538]
Observations	41886	41886	41886	41886
Adjusted R^2	0.240	0.240	0.240	0.240

Sample: REDS 1981-82 and 1998-99. Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects, and for current rainfall and lagged rainfall in the past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Each lag of rainfall is additionally interacted with a dummy that equals 1 of the child was 6 years or older in that year. Additional controls are income, wealth and area are measured in per capita terms (and in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.21: Effect of Risk on School Attendance – Weather Controls

	(1)	(2)	(3)	(4)	(5)
Panel A:					
IQR of predicted log consumption	-0.976***	-0.823***	-0.962***	-0.949***	-1.009***
	(0.279)	(0.286)	(0.287)	(0.296)	(0.291)
	[0.027]	[0.056]	[0.041]	[0.045]	[0.065]
Observations	68298	68298	68298	68298	68298
Adjusted R^2	0.226	0.228	0.230	0.232	0.233
Panel B:					
SD of predicted log consumption	-1.106***	-0.918**	-1.125***	-1.090***	-1.164***
	(0.358)	(0.359)	(0.363)	(0.374)	(0.371)
	[0.045]	[0.089]	[0.055]	[0.060]	[0.066]
Observations	68298	68298	68298	68298	68298
Adjusted R^2	0.226	0.228	0.230	0.232	0.233
Rainfall controls	Y	Y	Y	Y	Y
Rainfall controls (square)	N	Y	Y	Y	Y
Temperature controls	N	N	Y	Y	Y
Temperature controls (square)	N	N	N	Y	Y
Rainfall × Temperature controls	N	N	N	N	Y

Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects. Rainfall controls are current rainfall and lagged rainfall (in log mms) in the past 3 years. Temperature controls are current temperature and lagged temperature (degree celcius) in the past 3 years. Each regression also controls for income, wealth and area (all measured in per capita terms, in logs) and for household size. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile Bootstrap-t p-values (null-imposed, 1,000 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.22: Effect of Risk on Rainfall Outcomes

TREATMENT VARIABLE:	IQR of pred. cons		SD of pred. cons	
	(1)	(2)	(3)	(4)
Dep. var.: Annual Rainfall, log (t+3)	0.574	-0.400	0.698	-0.597
	(0.757)	(0.556)	(0.979)	(0.696)
	[0.491]	[0.472]	[0.520]	[0.413]
Dep var.: Annual Rainfall, log (t+2)	0.550	0.011	0.877	-0.073
	(0.714)	(0.514)	(0.896)	(0.666)
	[0.452]	[0.990]	[0.373]	[0.915]
Dep var.: Annual Rainfall, log (t+1)	0.893	1.011	1.432	1.452
	(0.661)	(0.531)	(0.878)	(0.721)
	[0.217]	[0.120]	[0.173]	[0.115]
Dep var.: Annual Rainfall, log (t)	0.904	-0.061	1.135	-0.121
	(0.585)	(0.483)	(0.729)	(0.599)
	[0.138]	[0.899]	[0.228]	[0.838]
Dep var.: Annual Rainfall, log (t-1)	0.022	-0.355	0.143	-0.498
	(0.621)	(0.529)	(0.759)	(0.663)
	[0.965]	[0.535]	[0.867]	[0.500]
Dep var.: Annual Rainfall, log (t-2)	0.462	-0.692	0.754	-0.898
	(0.725)	(0.487)	(0.980)	(0.602)
	[0.538]	[0.187]	[0.516]	[0.159]
Dep var.: Annual Rainfall, log (t-3)	0.282	0.141	0.585	0.189
	(1.071)	(0.510)	(1.523)	(0.696)
	[0.769]	[0.781]	[0.715]	[0.777]
Village and round FE	Y	Y	Y	Y
State-by-round FE	N	Y	N	Y
Observations	718	718	718	718

Notes: The unit of observation is the village-by-round. Each coefficient relates to an individual regression. Each regression controls for village and round of interview fixed effects. Cols. (2) and (4) add state-by-round of interview fixed effects. Standard errors (in parentheses) are corrected to allow for clustering by grid cell by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.23: Effect of Risk on School Attendance by Season

	Season 1 (Oct./Nov.)		Season 2 (February)		Season 3 (April/May)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
IQR of predicted log consumption	-1.063**	-0.953*	-0.882**	-0.836*	-0.968**	-0.886*
	(0.289)	(0.309)	(0.295)	(0.313)	(0.297)	(0.314)
	[0.020]	[0.054]	[0.047]	[0.077]	[0.042]	[0.068]
Adjusted R^2	0.220	0.232	0.230	0.244	0.224	0.236
Panel B:						
SD of predicted log consumption	-1.193**	-1.083*	-0.986*	-0.938	-1.118**	-1.005*
	(0.374)	(0.393)	(0.380)	(0.402)	(0.379)	(0.397)
	[0.047]	[0.067]	[0.070]	[0.122]	[0.047]	[0.076]
Adjusted R^2	0.220	0.232	0.230	0.244	0.223	0.236
Rainfall by schoolage	Y	Y	Y	Y	Y	Y
Income, wealth, area, hh size	Y	Y	Y	Y	Y	Y
State-by-round FE	N	Y	N	Y	N	Y
Dep. Var. mean	0.769	0.769	0.762	0.762	0.755	0.755
Observations	22766	22766	22766	22766	22766	22766

Notes: Each regression controls for village, season-by-round of interview, and age-by-gender fixed effects, for current and lagged rainfall in past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Rainfall controls are additionally interacted with a dummy that equals 1 of the child was 6 years or older in that year. Each regression also controls for income, wealth and area, measured in per capita terms (and in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.24: Effect of Risk on School Attendance by Quartiles of Consumption, Income and Wealth

Quartiles of:	Cons.exp. (1)	Income (2)	Wealth (3)
PANEL A:			
IQR of predicted log consumption× 1st Quartile	-1.025*	-1.001*	-1.325**
	(0.388)	(0.355)	(0.409)
	[0.057]	[0.054]	[0.025]
IQR of predicted log consumption× 2nd Quartile	-0.871*	-1.204**	-1.075*
	(0.349)	(0.359)	(0.366)
	[0.085]	[0.026]	[0.064]
IQR of predicted log consumption× 3rd Quartile	-0.895*	-0.769	-0.980*
	(0.348)	(0.342)	(0.345)
	[0.075]	[0.129]	[0.060]
IQR of predicted log consumption× 4tht Quartile	-0.673	-0.784	-0.439
	(0.398)	(0.382)	(0.365)
	[0.215]	[0.156]	[0.380]
Adjusted R^2	0.241	0.239	0.239
PANEL B:			
SD of predicted log consumption× 1st Quartile	-1.163*	-1.107*	-1.382**
	(0.502)	(0.457)	(0.529)
	[0.093]	[0.078]	[0.048]
SD of predicted log consumption× 2nd Quartile	-0.832	-1.276*	-1.194*
	(0.442)	(0.462)	(0.474)
	[0.151]	[0.054]	[0.090]
SD of predicted log consumption× 3rd Quartile	-1.092*	-0.948	-1.180*
	(0.441)	(0.439)	(0.430)
	[0.08]	[0.130]	[0.062]
SD of predicted log consumption× 4th Quartile	-0.832	-0.928	-0.566
	(0.498)	(0.475)	(0.453)
	[0.210]	[0.154]	[0.347]
Adjusted R^2	0.240	0.239	0.238
Educ. exp. (per child)/ income p.c. (1st Quartile)	15.18%	31.45%	13.62%
Educ. exp. (per child)/ income p.c. (2nd Quartile)	9.93%	12.96%	11.38%
Educ. exp. (per child)/ income p.c. (3rd Quartile)	9.45%	10.08%	8.92%
Educ. exp. (per child)/ income p.c. (4th Quartile)	6.73%	5.03%	6.66%
Observations	68298	68298	68298

Notes: Each regression controls for village, for state-by-round of interview, for season-by-round of interview, and age-by-gender fixed effects, and for current and lagged rainfall in past 3 years (in log mms), as well as for income, wealth and area are measured in per capita terms (and are in logs). Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.25: Effect of Risk on Returns to Education for Men and Women

	Wages (men)		Wages (women)	
	(1)	(2)	(3)	(4)
PANEL A:				
IQR of predicted log consumption	0.270	0.383	-0.104	2.626
	(0.785)	(0.914)	(1.037)	(2.002)
Highest degree: Illiterate	-0.229***	-0.215***	-0.094*	-0.006
	(0.022)	(0.028)	(0.046)	(0.078)
Highest degree: Primary	-0.105***	-0.090**	-0.038	0.015
	(0.021)	(0.028)	(0.047)	(0.080)
Highest degree: Higher secondary	0.278***	0.242***	0.176*	0.156
	(0.035)	(0.047)	(0.082)	(0.123)
Highest degree: University	0.792***	0.759***	1.093***	1.151***
	(0.063)	(0.071)	(0.142)	(0.221)
Illiterate × IQR		-0.442		-3.329+
		(0.762)		(1.919)
Primary × IQR		-0.511		-2.233
		(0.694)		(2.076)
Higher secondary × IQR		1.304		0.555
		(1.374)		(3.923)
University × IQR		1.199		-2.233
		(2.198)		(7.623)
R^2	0.403	0.404	0.455	0.457
PANEL B:				
SD of predicted log consumption	0.520	0.814	-0.161	4.022
	(0.995)	(1.192)	(1.382)	(2.802)
Highest degree: Illiterate	-0.230***	-0.211***	-0.094*	0.009
	(0.021)	(0.028)	(0.046)	(0.081)
Highest degree: Primary	-0.105***	-0.085**	-0.038	0.020
	(0.021)	(0.028)	(0.047)	(0.082)
Highest degree: Higher secondary	0.277***	0.236***	0.176*	0.158
	(0.035)	(0.048)	(0.082)	(0.122)
Highest degree: University	0.791***	0.757***	1.093***	1.243***
	(0.063)	(0.072)	(0.142)	(0.231)
Illiterate × SD		-0.788		-5.032+
		(0.978)		(2.634)
Primary × SD		-0.896		-3.204
		(0.928)		(2.817)
Higher secondary × SD		1.905		0.526
		(1.891)		(5.213)
University × SD		1.594		-7.418
		(2.947)		(10.318)
Observations	16547	16547	5180	5180
R^2	0.403	0.404	0.455	0.458

Notes: Sample consists of men and women aged 15 to 64, with non-missing wages (2007-08 round of interviews only). Dep. var.: Log daily wage (averaged over all activities). Omitted category: Lower secondary. Each regression controls for state and season fixed effects, for age fixed effects, marital status, and for current rainfall and lagged rainfall in the past 3 years (in log mms), and the standard deviation of rainfall over the last three years. Standard errors (in parentheses) are corrected to allow for clustering at the village by round of interview. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.26: Effect of Risk on School Attendance of Girls – Marriage Market Risk

	(1)	(2)	(3)	(4)
PANEL A:				
IQR of predicted log consumption	-1.237**	-1.138**	-1.092*	-1.061*
	(0.315)	(0.315)	(0.335)	(0.333)
	[0.020]	[0.023]	[0.069]	[0.064]
Average IQR in marriage market		-0.625		-0.652
		(0.389)		(0.403)
Adjusted R^2	0.281	0.279	0.295	0.292
PANEL B:				
SD of predicted log consumption	-1.367**	-1.246**	-1.206*	-1.161*
	(0.401)	(0.403)	(0.421)	(0.420)
	[0.044]	[0.052]	[0.092]	[0.094]
Average SD in marriage market		-0.720		-0.757
		(0.497)		(0.511)
Adjusted R^2	0.278	0.278	0.292	0.292
State-by-round FE	N	N	Y	Y
Observations	30909	30909	30909	30909

Notes: Each regression controls for village, for season-by-round of interview, and age-by-gender fixed effects, and for current rainfall and lagged rainfall (in log mms) in the past 3 years. Each lag is additionally interacted with a dummy that equals 1 of the child was 6 years or older in that year. Each regression also controls for income, wealth and area (all measured in per capita terms, in logs), and for household size. Cols. (3) and (4) add state-by-round of interview fixed effects. Villages belonging to the same marriage market are defined as villages located in a 67 km radius around village of interest. When calculating average risk in these villages, we exclude the village of interest itself. Sample: Girls aged 6 to 15 living in a village that has at least one more REDS village in a 67 km radius. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile Bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.27: Effect of Risk on Fertility and Gender-Composition

	No. of children born		Boys per girl	
	(1)	(2)	(3)	(4)
PANEL A:				
IQR of predicted log consumption	0.006	1.915	-0.283	-1.939
	(1.339)	(1.315)	(1.646)	(1.680)
	[0.999]	[0.317]	[0.849]	[0.243]
Adjusted R^2	0.429	0.446	0.055	0.093
PANEL B:				
SD of predicted log consumption	0.707	2.939	-0.256	-2.320
	(1.683)	(1.660)	(2.138)	(2.183)
	[0.782]	[0.264]	[0.894]	[0.288]
Adjusted R^2	0.429	0.446	0.055	0.093
State-by-round FE	N	Y	N	Y
Observations	11125	11125	710	710

Notes: The unit of observation is the mother in cols. (1) and (2), and the village-by-round in cols. (3) and (4). Each regression controls for village and round of interview fixed effects, and for current rainfall and lagged rainfall (in log mms) in the past 3 years, and the standard deviation of rainfall over the past 3 years. Cols. (1) and (2) also control for age (of the mother) fixed effects, and cols. (2) and (4) add state-by-round of interview fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the village by round of interview. Percentile bootstrap-t p-values (null-imposed, 999 replications) are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ according to bootstrap p-values.

Table D.28: Summary Statistics (SEPRI)

	Mean	SD
Consumption expenditures p.c. (log)	9.34	(0.58)
Precipitation, current year	6.90	(0.35)
Irrigated area (share of village agr. area)	0.79	(0.30)
Employment generated per capita, NREGS (lag)	0.74	(1.46)
SD of predicted log consumption	0.012	(0.014)
SD of predicted log consumption (Av. NREGS employment = 0)	0.014	(0.0035)
SD of predicted log consumption (Av. NREGS employment = 2.24)	0.0088	(0.0023)
Observations	77,938	

Notes: Consumption expenditures are in constant July 2010 INR.

Table D.29: Effect of Rainfall on Consumption Expenditures per Capita (1982-2016)

	(1)	(2)	(3)	(4)
Precipitation, current year (log mms)	0.190***	0.194***	0.135**	0.135**
	(0.063)	(0.063)	(0.061)	(0.061)
	[0.067]	[0.066]	[0.059]	[0.059]
Irrigated area (share of village agr. area)	1.371***	1.361***	1.007***	0.993***
	(0.380)	(0.380)	(0.353)	(0.349)
	[0.395]	[0.394]	[0.328]	[0.325]
Precipitation × Irrigated area	-0.174***	-0.172***	-0.129**	-0.127**
	(0.055)	(0.055)	(0.050)	(0.049)
	[0.058]	[0.058]	[0.046]	[0.046]
Employment generated per capita, NREGS (lag)		0.307**		0.266*
		(0.149)		(0.149)
		[0.017]		[0.020]
Precipitation × Employment generated, NREGS		-0.042**		-0.037*
		(0.021)		(0.021)
		[0.124]		[0.143]
Village FE	N	N	Y	Y
Observations	98383	98383	98383	98383
Adjusted R^2	0.290	0.290	0.344	0.344

Notes: Each regression controls for state-by-round of interview, and for lagged rainfall (in logs) in the last 3 years. Columns (3) and (4) also control for village fixed effects. Standard errors (in parentheses) are corrected to allow for clustering at the level of the grid cell by round of interview. Conley (1999) standard errors in brackets (cut-off 300km, Bartlett weights). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, according to Conley standard errors.