Three essays on the production and investment decisions of households living in rural India

Esther Gehrke

Wirtschaftswissenschaftliche Fakultät

Universität Passau

Dissertation zur Erlangung des akademischen Grades des Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

Juni 2016

Erstgutachter: Prof. Dr. Michael Grimm, Universität Passau

Zweitgutachter: Prof. Dr. Stefan Bauernschuster, Universität Passau

Drittgutachter: Prof. Dr. Andreas König, Universität Passau

Tag der Disputation: 21.10.2016

To Benjamin

Abstract

In order to end poverty by 2030, the declared goal of the United Nations, a better understanding is needed which policies help poor households to escape poverty and how to end its inter-generational transmission.

Since the Millennium Declaration in September 2000, and the adoption of the Millennium Development Goals (MDGs), the delivery of basic social services, such as education, health, water supply and sanitation, has become the central focus of international development assistance. However, the provision of basic social services is not necessarily sufficient to lead to an accumulation of human and productive capital, which would allow households to escape poverty and interrupt its inter-generational transmission.

To understand why people are poor, we need to understand what productive decisions poor households take, and to identify what constraints households face in their attempt to accumulate human, as well as, productive capital. A better understanding of such constraints could guide policies that have a long-term impact on poverty reduction and on development.

A number of factor could explain why poor households operate at unprofitable levels and why they are constrained in their investment decisions. Empirical evidence points to different explanations: cost of learning and access to information, insufficient education, risk, credit constraints, non-convex production technologies, and behavioral patterns that are inconsistent with standard neoclassical models. Currently, one of the major challenges in formulating policies that foster productive investments among the poor, seems to be to disentangle the effects of scale, credit constraints, and the lack of insurance mechanisms.

This thesis seeks to shed further light on the relative role of these three constraints. In the context of rural India, it analyzes what production and

investment decisions households take and how important risk and credit constraints as well as scale effects are in these decisions. Finally, it evaluates potential policy tools that could support households in overcoming these constraints. Today, 33% of the world's poor live in India, the vast majority of them (80.5%) in rural areas. The economic structure of rural India is still dominated by agricultural production, and consequently, this thesis concentrates on agricultural production decisions and employment in agriculture.

In particular, this thesis addresses three questions in three individual papers: First, are farm households constrained in their crop choices by agricultural production risk and to which extent can India's public works program support households in overcoming this constraint? Second, how profitable is cattle farming in rural India at different levels of investment and which barriers do households face in reaching optimal investment levels? And third, can risk in agricultural wages explain limited investment in girls' education in the presence of intra-household substitution in household chores?

The first paper focuses on crop choice of farm households. It reassesses the stylized fact that households have to trade-off between returns and risk in their crop choice in the context of Andhra Pradesh, a state in the south of India. It then explores the effect of India's flagship anti-poverty program, the National Rural Employment Guarantee Scheme (NREGS) on households' crop choice using a representative panel data set. The NREGS guarantees each household living in rural India up to a hundred days of employment per year, at state minimum wages. The paper shows theoretically, and empirically, that the introduction of the NREGS reduces households' uncertainty about future income streams because it provides reliable employment opportunities in rural areas independently of weather shocks and crop failure. With access to the NREGS, households can compensate income losses emanating from shocks to agricultural production. Households with access to the NREGS can therefore shift their production towards riskier but also more profitable crops. These shifts in agricultural production have the potential to considerably raise the incomes of smallholder farmers.

The paper concludes that employment guarantees can, similarly to crop insurance, help households in managing agricultural productions risks. It also argues that accounting for the effects of the NREGS on crop choice and profits from agricultural production affects the cost-benefit analysis of such a program considerably.

The second paper concentrates on the profitability of farming cattle in Andhra Pradesh. The paper also uses a representative panel dataset, and examines average and marginal returns to cattle at different levels of cattle investment. It finds average returns in the order of -8% at the mean of cattle value. These returns vary across the cattle value distribution between negative 53% (in the lowest quintile) and positive 2% (in the highest). While marginal returns are positive on average, they also vary considerably with cattle value and breed. The paper shows that average and marginal returns are considerably higher for modern variety cows, i.e. European breeds and their crossbreeds, than for traditional varieties of cows or for buffaloes. It also shows that cattle farming becomes most profitable at minimum herd sizes of five animals, due to decreasing average labor costs with increasing herd sizes.

The results of this paper suggest that cattle farming is associated with sizable non-convexities in the production technology and that substantial economies of scale, as well as high upfront expenses of acquiring and feeding high-productivity animals, might trap poorer households in low-productivity asset levels. The fact that wealthier households and households with lower costs to access veterinary services are more likely to overcome these barriers, supports this idea.

The second paper concludes that cattle farming might well generate positive returns for households in rural India, but that most households seem to operate at unprofitable levels. This could also explain the apparent paradox between widespread support of cattle farming through agricultural policy interventions and negative returns to cattle, as stressed in recent works. It argues that policy interventions that target productive assets will only be beneficial if transfers are high enough to allow households to overcome these entry barriers.

The third paper concentrates on the effect of risk on the productive decisions of households, and analyzes the effect of wage risk in agricultural employment on women's labor supply and time allocated to home production. It seeks to understand the extent to which risk raises labor supply of women to levels that can become harmful for other members of the household. The hypothesis is that in the presence of intra-household substitution effects – for instance in the performance of household chores – increased female labor supply might have negative effects on the time allocation of girls. If women have less time available for home production and childcare, and such activities can only be foregone at high cost, they might be forced to take older girls out of school or to cut down on the time these girls study at home in order for them to fill in for these tasks.

The paper uses cross-sectional data on the time allocation of different household members and predicts wage risk at the village level as a function of the historical rainfall distribution and a village's share of land that is under irrigation. The results show that wage risk affects the time allocation of women, increasing their labor supply and reducing the time they allocate to home production. Wage risk also increases the time girls spend on household chores and reduces their time in school. Because the observed effect of wage risk on girls' time allocated to household chores corresponds very closely to the effect observed for women, it seems plausible to attribute it to intra-household substitution effects. The observed effect of risk on girls' school time, however, is greater than the observed effect of risk on the home-production time of girls. This can be due to two reasons: First, in the presence of intra-household substitution effects, shocks in wages will not only increase female labor supply but also girls' time on household chores. And the model predicts that risk-averse households invest less in education when future school time becomes uncertain, because future school time affects the returns to current schooling. Second, if school attendance is indivisible, then girls might be forced to drop out of school temporarily or even permanently.

The paper then simulates the effect of the NREGS on the time-allocation decisions of working women and school-age girls. The results suggest that the NREGS could increase the time working women spend on household duties, because it reduces uncertainty regarding future earnings, and alleviates the need to accumulate savings. Thereby, the NREGS would reduce the pressure on girls to perform household tasks and allow them to increase the time they spend in school or studying by 6 minutes daily.

Wit these findings, this thesis contributes to a better understanding of the choices poor households in rural India face in their day-to-day decision making, and offers insights into what policies could support households in escaping poverty, and interrupt its inter-generational transmission.

Acknowledgements

I would like to thank Michael Grimm for his dedication, guidance and support. This thesis would not have been possible without him.

My thanks also go to Tilman Altenburg and Markus Loewe, who supported me in all the phases of the PhD, and whose comments and suggestions I cherish. I would also like to thank Andrew Foster for stimulating discussions and his guidance during and after my time at Brown University, and Stefan Bauernschuster and Andreas König for serving on my dissertation committee.

This thesis benefited from the helpful comments of five anonymous referees as well as of Denis Cogneau, Jishnu Das, Marcel Fafchamps, Francisco Ferreira, Clement Imbert, Hanan Jacoby, Clement Jubard, Eeshani Kandpal, Stephan Klasen, Jann Lay, Abhiroop Mukhopadhyay, Berk Ozler, Simon Quinn, Elisabeth Sadoulet, and Dominique Van de Walle.

Special thanks go to my fellow PhD students at the German Development Institute, University of Passau and Brown University, Anna Luisa, Armin, Nicole, Renate, Rita, Samuel and Sebastian for their help and for their views and comments.

I would also like to acknowledge the numerous comments and suggestions from conference and seminar participants in Bari (EQINEC), Berlin (PEGNet), Bonn (IZA / World Bank), Copenhagen (PEGNet and NCDE), Göttingen (Courant), Helsinki (NCDE), Kiel (AEL), Lusaka (PEGNet), Munich (AEL), Passau (AEL), Providence (Brown University), Washington, D.C. (World Bank) and Zürich (AEL PhD seminar).

The data used in the first two papers come from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India (Andhra

Pradesh), Peru and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Department for International Development (DFID), with co-funding from 2010 to 2014 by the Netherlands Ministry of Foreign Affairs, and from 2014 to 2015 by Irish Aid. The REDS data used in the third 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.

Finally, I would like to thank Jochen for his patience and moral support, and my family for always encouraging and supporting me.

Contents

List of Figures

List of Tables

Glossary

ARIS Additional Rural Incomes Survey

ASER Annual Status of Education Report

CES Constant Elasticity of Substitution

CSAE Centre for the Study of African Economies

DAC Development Assistance Committee

GDP Gross Domestic Product

GOI Government of India

HYV High Yielding Variety

MDG Millennium Development Goal

MSE Micro and Small Enterprise

NBER National Bureau of Economic Research

NGO Non-Governmental Organization

NREGA National Rural Employment Guarantee Act

NREGS National Rural Employment Guarantee Scheme

OECD Organization for Economic Co-operation and Development

ODA Official Development Assistance

OLS Ordinary Least Squares

GLOSSARY

PPP Purchasing Power Parity

RCT Randomized Control Trial

REDS Rural Economic and Demographic Survey

SEPRI Socio-Economic Profiles of Rural Households in India

YLS Young Lives Survey

1

Introduction

In order to end poverty by 2030, the declared goal of the United Nations, a better understanding is needed which policies help poor households to escape poverty and how to end its inter-generational transmission.

It is largely recognized by now, that there is no simple answer to what drives growth and development (Cohen and Easterly, 2009). With the demise of one-size-fits-all approaches such as the Washington Consensus, micro-development studies have gained increasing attention (Ravallion, 2009; Rodrik, 2009). While the research of what macroeconomic policies could drive growth and poverty reduction is still prominent, the focus of analysis has largely shifted to individual and household level decision-making. With the boom in randomized evaluations, we have learned much about effective methods of delivering public goods, such as education or vaccines (Banerjee, 2009).

At the same time political priorities also shifted. With the Millennium Declaration in September 2000, and the adoption of the Millennium Development Goals (MDGs), the delivery of basic social services has become the main focus of international development assistance (Temple, 2010). In the last few years, the bulk of official development assistance (ODA) has been channeled towards social services and infrastructure, including education, health, water supply and sanitation, and reproductive health.[1]

[1] Social services and infrastructure accounted for 38% of total ODA in 2014. Data extracted on 07 June 2016 from OECD.Stat. The Development Assistance Committee (DAC) of the OECD only collects aid data from OECD member states. This figure thus does not cover new donors, such as China, nor private development assistance channeled through NGOs and foundations.

However, poverty reduction policies are only effective and sustainable to the extent that they address the root causes of poverty (Rodrik and Rosenzweig, 2010). And the provision of basic social services is not necessarily sufficient to lead to an accumulation of human and productive capital, which would allow households to escape poverty and interrupt its inter-generational transmission.

By far most of the poor are engaged in self-entrepreneurial activities (both on the farm and in non-farm enterprises). According to Banerjee and Duflo (2011), 44% of the extreme poor in urban areas operate their own business. Even in rural areas, 24% of the poor operate a non-agricultural business, in addition to the 64% of the poor who are self-employed in agriculture.[2] These activities are largely unprofitable, and only in very few cases ever turn into growing businesses (Banerjee and Duflo, 2011). A better understanding of the why these household enterprises are unprofitable, of the options households have and the constraints they face in their production and investment decisions, should inform policies that have a long-term impact on poverty reduction and on development.

In line with this idea, this thesis goes beyond the viewpoint of households as mere beneficiaries of policies and investigates their production and investment decisions. In particular, it seeks to understand to which extent risk constrains households in their agricultural production choices, as well as in investing in the education of their children. It also analyzes the profitability of farming cattle at different investment levels and seeks to understand which constraints households face in the accumulation of cattle.

This thesis therewith contributes to a better understanding of the choices poor households face in their day-to-day decision making, and offers insights into what policies could support households in escaping poverty, and interrupt its inter-generational transmission.

1.1 Literature review: Constraints faced by households in their production and investment decisions

In their seminal book, Singh et al. (1986) show that consumption and production decisions in farm households are not independent of each other. In farm households,

[2]According tho their 18-country dataset. The figures above refer to individuals living on less than 1$ per day in PPP.

some inputs are purchased and others supplied by the household, similarly some outputs are retained for self-consumption while others are sold. Profits made on the farm affect household income, labor supply and consumption. At the same time, shocks to agricultural production also affect household consumption.

The non-separability of consumption and production decisions is not only relevant from a theoretical perspective, it crucially determines the effect of public policies on these households. For example, policies to increase consumption of food staples, i.e. price policies, might have unintended consequences on household production decisions; rural policies that target farm households will have spillover effects on landless households, etc. (Singh et al., 1986).

In the presence of market imperfections, the interrelations between consumption and production decisions become more pronounced (Sadoulet and De Janvry, 1995). Constraints faced on the farm, such as lack of credit or riskiness in returns, affect the household's decision to consume or to save. Similarly, constraints faced in the household, such as the necessity to hold cash in absence of formal protection against health shocks, prevent households from investing in their farms or firms.

The observation that production and consumption decisions are interlinked is commonly made for farm households, but applies as well to households operating informal enterprises and engaged in non-farm self-entrepreneurship in developing countries (Krishna, 1964). The following literature review therefore picks up examples from both strands of literature in order to highlight existing research gaps.

It is by now well established that farm households and owners of micro and small enterprises (MSEs) are subject to a number of constraints which prevent them from adopting profit-maximizing technologies (Duflo et al., 2008; Feder et al., 1985; Foster and Rosenzweig, 2010b; Suri, 2011). This manifests in delayed technology adoption, low investment in fixed capital, a preference for conservative crop choices and, more generally, a lack of innovative capacity. And has potentially severe and long-lasting effects on income and well-being in developing countries as a large share of their populations rely on self-entrepreneurial activities (including farming) as a major source of income.

Empirical evidence points to different explanations for the low propensity to innovate: cost of learning and access to information, insufficient education, risk, credit constraints, non-convex production technologies, and behavioral patterns that are inconsistent with standard neoclassical models. These factors are explained in the following using prominent examples from the literature. Rather than being exhaustive, the following literature review intends to highlight seminal work as well as the current state of research and ongoing debates.

1.1.1 Learning and access to information

If new technologies with higher returns are not adopted, one reason might be limited knowledge about the profitability of this technology or the lack of knowledge on how to apply it. Duflo et al. (2008) offer subsidized fertilizer over different periods of the agricultural cycle, and find that farmers applying fertilizer in the first round of their experiment are also more likely to apply it in the second round.

Own experience and experience from others can be substitutable in a number of cases. Conley and Udry (2010), for example, find that Ghanaian farmers are more likely to apply fertilizer if their neighbors successfully adopted fertilizer. However, this seems to be possible only for technologies that are largely non-specific. Munshi (2004) explores regional differences in soil suitability for rice and wheat in his analysis of the adoption of high-yielding varieties (HYV) during the Indian green revolution. He argues that social learning cannot happen if a crop's production technology is sensitive to individual and farm characteristics, such as rice.

If a technology is farmer specific, learning also involves acquiring information about how best to apply this new technology. Foster and Rosenzweig (1995) argue that the optimal amount of fertilizer application is farmer specific, as it varies with soil characteristics and climate. Therefore, farmers need to experiment with the technology on their own land in order to learn about its optimal application.

Learning processes arguably discourage technology adoption, because the cost of learning may be greater than the benefit of the new technology (Besley and Case, 1993). Consequently the probability of adoption would also be greater for larger farmers. Such scale effects can drive a wedge in the profitability of farming and or operating firms between large operations and small operations, and further exacerbate existing inequalities. When learning from others is effective, it can also lead to externalities and free-riding behavior (Bandiera and Rasul, 2006), further reducing the speed of technology diffusion.

1.1.2 Education

Education and levels of human capital may also determine the amount of innovative capacity; the argument is that more educated farmers or workers are better able to "decode" new technologies or new information in general (Foster and Rosenzweig, 2010b).

In his paper, Welch (1970) provides evidence on the importance of education in technology adoption in the United States. He finds that the relative earnings of more educated U.S. farmers vis-à-vis less educated farmers increase with the amount of research and development related to farming in a particular area. Similarly, Bartel and Lichtenberg (1987) assess the demand for educated workers across industries and find that industries that use newer technologies have a higher demand for an educated workforce.

Foster and Rosenzweig (1996) examine the question in a developing country context. Using household panel data, they find that the technological change associated with the green revolution increased returns to schooling. This also led to higher educational investments among farm households.

1.1.3 Risk

In contexts in which insurance markets are absent, risk aversion of farmers or entrepreneurs can prevent the adoption of or investment in new technologies. Because losses made in the farm or enterprises directly affect household consumption, households cannot afford to take great risks in their production decisions.

Morduch (1990) and Rosenzweig and Binswanger (1993) are among the first authors to provide evidence that uninsured risk prevents farmers from planting risky crops, and from holding profitable asset portfolios, respectively. Using household panel data from India, Morduch (1990) estimates that poorer households exposed to higher risk plant less risky crops. With the same data, Rosenzweig and Binswanger (1993) show that the composition of household assets and production factors influences the variability in profits from agricultural production. Furthermore, the authors estimate that the coefficient of variation in the monsoon onset is negatively associated with the variability in profits. Since more variable profits are also higher at the mean, the authors conclude that risk prevents households from taking profitable production decisions. Using panel data from Ethiopia, Dercon and Christiaensen (2011) assess the importance of uninsured risk by constructing an indicator of household risk exposure that combines a household's probability of facing a rainfall shock with its ability to cope with such a shock. They thereby circumvent the attribution problem of using only wealth as a proxy for a household's capacity to smooth consumption. The authors show that Ethiopian households with lower expected consumption outcomes - due to high risk exposure and low savings - are less likely to invest in fertilizer.

In the last few years, authors have used randomized variation in the availability of insurance mechanisms to estimate the importance of risk. These articles find that crop insurance is critical in stimulating fertilizer application (Karlan et al., 2014) and risky crop choice (Cole et al., 2013) and risk taking in agricultural production more generally (Mobarak and Rosenzweig, 2013).

In the context of MSEs, Grimm et al. (2012), Bianchi and Bobba (2013) and Dodlova et al. (2015) argue that uninsured risk is at least partly responsible for low investment rates. Bianchi and Bobba (2013) explore differences in the number of years, targeted households expect to have access to Mexico's cash transfer program, Progresa, to assess the relative importance of liquidity constraints versus uninsured risk in the decision to become an entrepreneur. They show that expected future transfers are more important determinants of occupational choices, than those currently received. Based on backward-looking data, Grimm et al. (2012) argue that firms facing higher self-reported business risk and higher sales variability have lower capital accumulation. Using similar proxies and firm-level panel data, Dodlova et al. (2015) argue that risk slows down the accumulation of capital in Peruvian MSEs.

1.1.4 Credit

Upfront expenses associated with any investment or change in technology may prevent adoption because expenses have to be made prior to the realization of profits from this technology. If households face credit constraints, for instance because they need to provide collateral, then wealth and the probability to adopt a new technology will be closely correlated.

In line with this argument, Bhalla (1979) finds that 48% of small-scale farmers report the lack of access to credit as reason for not adopting fertilizer as opposed to only 6% of large-scale farmers during the Indian green revolution. Rosenzweig and Wolpin (1993), and Fafchamps and Pender (1997) estimate structural models of investment decisions in rural India using household panel data. Both argue that the lack of credit is an important explanation for foregone investments.

More recently, Gine and Klonner (2008) have analyzed the adoption of boats built with plastic reinforced fiber among fishermen in Tamil Nadu, India. They find that households with higher wealth invested earlier in the new technology than households with lower wealth. Because most non-adopters cite the lack of financing as most important reason for not adopting, the authors conclude that the lack of credit is the main explanation for delayed adoption among poorer households.

In the context of MSEs, De Mel et al. (2008) use experimental data, and find that access to cash and in-kind grants stimulates investment and raises firm profits in Sri Lanka. Banerjee and Duflo (2014) explore policy changes in the eligibility for directed credit to estimate whether MSEs are credit constrained in India. Since improved access to credit leads to an acceleration in the rate of growth in the sales and profits of these firms, the authors conclude that micro entrepreneurs are indeed severely credit-constrained.

1.1.5 Scale effects

Scale effects might not only be due to learning costs. Some technologies might be profitable only when used on a larger scale. Indivisible investments, such as borewells for irrigation, or tractors, are not only costly; in the latter case they are also much less effective on very small plots.

In agricultural research, studying the relationship between farm size and productivity has a long tradition. Early studies by Deolalikar (1981) and Rao and Chotigeat (1981) suggest an inverse relation between farm-size and productivity. This stylized fact has been largely attributed to supervision costs (Feder, 1985), missing labor markets (Skoufias, 1994), credit constraints (Eswaran and Kotwal, 1986) and uninsured risk (Barrett, 1996) in the theoretical literature.

Contrasting evidence of the role of returns to scale in investments in agriculture was produced by Fafchamps and Pender (1997). The authors estimate a structural model of savings accumulation in the presence of risk aversion and shocks, and argue that households face substantial difficulties in accumulating sufficient savings to self-finance the investment in a well, which the authors identify as indivisible but highly profitable asset.

More recently, entry barriers and non-divisibilities have been studied more in the context of micro and small enterprises in the manufacturing or service sector. Using observational data, McKenzie and Woodruff (2006) find very high returns to capital at low levels of investment and argue that entry barriers do not play much of a role in the context of micro enterprises in Mexico. In a different article, they use randomized access to cash and in-kind grants, and again find very high marginal returns to capital at very low levels of investment. They conclude that entry barriers do not seem to be an issue for MSEs in Mexico (McKenzie and Woodruff, 2008). Based on findings from a randomized control trial (RCT) in Sri Lanka, De Mel et al. (2008) argue that indivisibilities in the capital stock are unlikely to explain why small firms do not reinvest profits despite very high returns to capital. In contrast, Grimm et al. (2011) find barriers to entry in most activities of micro and small enterprises in West Africa.

Banerjee and Duflo (2011) suspect that non-linearities in returns might explain why so many firms operate at small-scale despite high marginal returns to capital in these businesses. Based on a review of different studies, the authors suggest that entrepreneurs might face two different production technologies, with the more profitable one having very high entry barriers, such that it is beyond the reach of most MSEs.

1.1.6 Insights from behavioral economics

Recently, researchers have begun to test insight from behavioral economics in developing country contexts. The idea is that if individual behavior is at odds with standard economic models, then this could explain why some economic policies are not being effective.

In their experiment in Kenya, Duflo et al. (2011) find that farmers prefer purchasing subsidized fertilizer at the time of harvest rather than at the time of actual need in the following season. They argue that farmers seem to value the possibility to purchase fertilizer early on as a commitment device. The authors conclude that at least some farmers are present biased, and procrastinate in the sense that they keep postponing the purchase of fertilizer until later, because they underestimate the probability of being impatient (i.e. unwilling to decide on which fertilizer to use and to purchase it in the store) in the future. Some farmers therefore end up never using fertilizer, even though they originally intended to do so. The authors conclude that offering a commitment device such as early discounts could therefore be much more cost effective than heavy subsidies. Similarly, Fafchamps et al. (2014) show in their RCT that shopkeepers in Ghana are able to increase profits if supported by in-kind grants as opposed to cash grants. Since cash grants were mostly used to finance household needs, the authors conclude that entrepreneurs might face difficulties in translating intentions into action.

Behavioral economics also suggests that learning is more efficient in social networks: Vasilaky and Leonard (2013) find that learning about cotton farming practices in Uganda is most effective when farmers are assigned a partner with which they communicate throughout the season about growing practices than if each farmer just receives a standard training module. Likewise, BenYishay and Mobarak (2014) find in their RCT that farmers are more likely to adopt new technologies if they were trained by peer farmers than if they were trained by government extension agents.

Finally, Kremer et al. (2013) show that Kenyan shopkeepers choose not to invest despite high expected returns on that investment because they are small-stakes risk averse. This is at odds with expected utility theory, which argues that for very small stakes agents should be approximately risk neutral.

1.2 Contribution of this thesis

As we have seen, a number of factors can explain why households are not able to adopt optimal production and investment decisions. On the one hand, research on learning and on education has led to clear cut policy formulations. On the other hand, it is still difficult to assess the extent to which insights from behavioral economics are substitutes or complementary to policy formulations derived from more conventional economic models. Currently, one of the major challenges in formulating adequate policies, that foster productive investments among the poor, seems to be to disentangle the effects of scale, credit constraints, and the lack of insurance mechanisms (Foster and Rosenzweig, 2010b).

This thesis seeks to shed further light on the relative role of these three constraints in the context of rural India. Today, 33% of the world's poor live in India, the vast majority of them (80.5%) in rural areas. The economic structure of rural India is still dominated by agricultural production: the agricultural sector employed 59% of all male workers and 75% of all female workers in 2011/12 (National Sample Survey Office, 2014). Even nation-wide, agricultural employment still dominates: in 2010, 51% of the labor force were employed in agriculture. Meanwhile, the agricultural sector is much less profitable than other sectors, contributing only 18% to the gross domestic product (GDP) of India in the same year (World Bank, 2016).

Consequently, this thesis focuses on households engaged in agricultural production or employed in agriculture, and analyzes what production and investment decisions these households take. It shows that households face different constraints in their decisions, which prevent them from taking optimal production and investment decisions. Finally, this thesis evaluates potential policy tools that could support households in overcoming these constraints.

Specifically, it addresses three main questions in three papers: First, are farm households constrained in their crop choices by agricultural production risk and to which extent can India's public works program support households in overcoming this constraint? Second, how profitable is cattle farming in rural India at different levels of investment and which barriers do households face in reaching optimal investment levels? And third, can risk in agricultural wages explain limited investment in girls' education in the presence of intra-household substitution in household chores?

The first paper focuses on crop choice of farm households. It reassesses the stylized fact that households have to trade-off between returns and risk in their crop choice in the context of Andhra Pradesh, a state in the south of India. It then explores the effect of India's flagship anti-poverty program, the National Rural Employment Guarantee Scheme (NREGS), on households' crop choice using a representative panel data set. The NREGS guarantees each household living in rural India up to a hundred days of employment per year, at state minimum wages. The paper shows theoretically, and empirically, that the introduction of the NREGS reduces households' uncertainty about future income streams because it provides reliable employment opportunities in rural areas independently of weather shocks and crop failure. With access to the NREGS, households can compensate income losses emanating from shocks to agricultural production. Households with access to the NREGS can therefore shift their production towards riskier but also more profitable crops. These shifts in agricultural production have the potential to considerably raise the incomes of smallholder farmers.

Linking the employment guarantee to risk considerations is the key innovation of this paper. Therewith, it provides empirical evidence that employment guarantees can, similarly to crop insurance, help households in managing agricultural productions risks.

It also shows that accounting for the effects of the NREGS on crop choice and profits from agriculture affects the cost-benefit analysis of such a program considerably. This insight contributes to the ongoing debate on the effectiveness of the NREGS in reducing poverty.

Similarly to the first, the second paper analyzes the profitability of farming decisions. It concentrates on the profitability of farming cattle in Andhra Pradesh at different levels of investment in cattle, and then investigates potential constraints to investment. While the first paper shows that production risk is a major factor in explaining non-adoption of profitable crops, the second paper identifies non-convexities in the production technology as important constraint to investment in cattle. To which extent this constraint is aggravated by limited access to credit or the lack of insurance cannot conclusively be answered.

The paper also uses a representative panel dataset, and examines average and marginal returns to cattle at different levels of cattle investment. It finds average returns in the order of -8% at the mean of cattle value. These returns vary across the cattle value distribution between negative 53% (in the lowest quintile) and positive 2% (in the highest). While marginal returns are positive on average, they also vary considerably with cattle value and breed. The paper shows that average and marginal returns are considerably higher for modern variety cows, i.e. European breeds and their crossbreeds, than for traditional varieties of cows or for buffaloes. It also shows that cattle farming becomes most profitable at minimum herd sizes of five animals, due to decreasing average labor costs with increasing herd sizes.

These results suggest that cattle farming is associated with sizable non-convexities in the production technology and that substantial economies of scale, as well as high upfront expenses of acquiring and feeding high-productivity animals, might trap poorer households in low-productivity asset levels. The paper then analyzes which households are more likely to overcome these entry barriers, and finds that wealthier households and households with lower costs to access veterinary services are more likely to operate at profitable levels. What cannot be assessed with the data is whether insurance mechanisms or improved access to credit would support households in overcoming these entry barriers.

The second paper concludes that cattle farming might well generate positive returns for households in rural India, but that most households seem to operate at unprofitable levels. This could also explain the apparent paradox between widespread support of cattle farming through agricultural policy interventions and negative returns to cattle, as stressed in recent works. It argues that policy interventions that target productive assets will only be beneficial if transfers are high enough to allow households to overcome these entry barriers.

The third paper again concentrates on the effect of risk on the productive decisions of households. Similarly to the second paper, it focuses on investment decisions; only now the focus is on investment in human capital instead of productive capital.

The paper analyzes the effect of wage risk on women's labor supply and time allocated to home production. It shows that women increase their labor supply in the presence of wage risk in order to accumulate savings. Furthermore, it seeks to understand the extent to which risk raises labor supply of women to levels that can become harmful for other members of the household. The hypothesis is that in the presence of intra-household substitution effects – for instance in the performance of household chores – increased female labor supply might have negative effects on the time allocation of girls. If women have less time available for home production and childcare, and such activities can only be foregone at high cost, they might be forced to take older girls out of school or to cut down on the time these girls study at home in order for them to fill in for these tasks.

The paper uses cross-sectional data on the time allocation of different household members and predicts wage risk at the village level as a function of the historical rainfall distribution and a village's share of land that is under irrigation. The results show that wage risk affects the time allocation of women, increasing their labor supply and reducing the time they allocate to home production. Wage risk also increases the time girls spend on household chores and reduces their time in school. Because the observed effect of wage risk on girls' time allocated to household chores corresponds very closely to the effect observed for women, it seems plausible to attribute it to intra-household substitution effects. The observed effect of risk on girls' school time, however, is greater than the observed effect of risk on the home-production time of girls. This can be due to two reasons: First, in the presence of intra-household substitution effects, shocks in wages will not only increase female labor supply but also girls' time on household chores. And the model predicts that risk-averse households invest less in education when future school time becomes uncertain, because future school time affects the returns to current schooling. Second, if school attendance is indivisible, then girls might be forced to drop out of school temporarily or even permanently.

Similarly to the first paper, the third assesses the effect of the NREGS in this context. But in the absence of adequate data to estimate the effect of the employment guarantee, it simulates its effect on the time-allocation decisions of working women and school-age girls. The results suggest that the NREGS could increase the time working women spend on household duties, because it reduces uncertainty regarding future earnings, and alleviates the need to accumulate savings. Thereby, the NREGS would reduce the pressure on girls to perform household tasks and allow them to increase the time they spend in school or studying by 6 minutes daily.

Each of three papers takes a different angle and focuses on different aspects. They are complementary in the sense that they show that rural households face a multitude of decisions and engage in different activities. The thesis shows that policy needs to deal with this complexity adequately to avoid unintended consequences. It also shows that policy impacts can go well beyond intended immediate effects, and that such unintended effects are sometimes substantial. Consequently, a cost-benefit analysis of policies is incomplete if unintended effects (good and bad ones) are not accounted for.

This thesis contributes to an understanding of the importance of risk constraints and scale effects in household decision making. This thesis also shows that adequate policies can help households overcoming important constraints in their production and investment decisions and could therefore have a strong effect of poverty reduction, and on development more generally.

Uninsured risk seems to affect household decision making processes in very different ways. As we have seen, well designed public works programs can support households in managing these risks, and can therewith *inter alia* enhance the profitability of agricultural production and increase investments in education. The extent to which public work programs also enhance other forms of productive investments remains to be analyzed.

This thesis also shows that the direct transfer of assets to the poor should remain an important tool in development policy. But, as the second paper shows, such a transfer can only contribute to raising the incomes of the poor if assets transfers are large enough for households to operate at profitable levels. The extent to which these entry-barriers are context specific is not possible to assess given the small number of profitability analyses in the literature.

2

An employment guarantee as risk insurance? Assessing the effects of the NREGS on agricultural production decisions

2.1 Introduction

Previous research suggests that farmers in developing countries are constrained in their production and investment decisions. Evidence of delayed technology adoption, low investment in fixed capital, a preference for conservative crop choices and, more generally, a lack of innovative capacity is by now well established (Duflo et al., 2008; Foster and Rosenzweig, 2010b; Suri, 2011). This has potentially severe and long-lasting effects on income and well-being in developing countries as a large share of their populations still rely on agricultural production as a major source of income.

Empirical evidence suggests that uninsured risk prevents farmers from adopting new technologies. A number of studies have used randomized variation in the availability of index-based agricultural insurance to estimate the importance of uninsured risk in production decisions. These studies show that crop insurance is critical in stimulating fertilizer application (Karlan et al., 2014), risky crop choice (Cole et al., 2013) and risk taking in agriculture more generally (Mobarak and Rosenzweig, 2013). However, trust-related considerations and basis risk continue to limit the uptake of agricultural micro-insurance in many developing countries (Carter et al., 2014; Cole et al., 2013). Given these limitations, it seems worthwhile to explore other policy options that could help farmers to cope with shocks and manage risks.

This paper aims at contributing to the empirical evidence on the importance of risk management in households' production decisions. But instead of exploring variance in the availability of insurance, as do the studies cited above, it examines variation in the access to an alternative mechanism that could improve a household's risk management: an employment guarantee. The main argument is that public works programs or employment guarantees could help households to cope with income shocks by providing additional employment opportunities. This idea is not new; the potential of public works schemes in helping households to smooth income in the case of shocks has been highlighted *inter alia* by Barrett et al. (2005) and Binswanger-Mkhize (2012). However, to the best of my knowledge, no empirical evidence on the insurance effect of an employment guarantee on households' production decisions has been provided so far.

In this paper I present evidence that the introduction of the National Rural Employment Guarantee Scheme reduces households' uncertainty about future income streams and enables them to produce a higher share of high-risk, high-profit crops. The National Rural Employment Guarantee Act (NREGA) was passed in India in September 2005; the implementation thereof began in 2006. The NREGA entitles every rural household to up to a 100 days of work per year at the state minimum wage. In the financial year 2010/11 the NREGS provided work to close to 55 million rural households, generating a total of 2.5 billion person-days of employment (Ministry of Rural Development, Government of India, 2012).

For the empirical analysis I use the Young Lives data; a household panel that is representative of the state of Andhra Pradesh in southern India. The quality of implementation of the NREGS has been shown to vary immensely across India (Dutta et al., 2012). In most states the provision of work under NREGS is far too unpredictable to completely offset the effects of a shock. Under such circumstances, the NREGS would not affect households' risk expectations. Andhra Pradesh, however, is one of the states with the highest number of days of employment generated per rural household. I find that the provision of work in Andhra Pradesh does effectively respond to changes in household demand and thus supports households in managing agricultural production risks.

The estimation strategy builds on the sequenced introduction of the NREGS at the district level, and explores the fact that the scheme was introduced in four out of the six survey districts in 2006 and in the remaining two districts in 2008 and 2009. Because this approach relies heavily on the parallel trends assumption, I perform a number of robustness checks. The use of alternative treatment variables (e.g. block-level spending and employment days generated under the NREGS, as well as households' registration with NREGS) does not change the results. Several additional robustness checks rule out the possibility that the observed effect is due to alternative mechanisms.

The results of this paper suggest that employment guarantees can trigger important gains in agricultural productivity in the medium term. These gains go far beyond the direct income effect that the provision of employment in agricultural lean seasons has on the wellbeing of rural households. By providing households with the right to work, such programs have an insurance effect, which triggers additional increases in productivity and, in turn, in households' incomes. This is a very important lesson for other countries with planned or ongoing public works programs.

The remainder of this paper proceeds as follows: Section 2.2 introduces a theoretical framework for analyzing the effects of an employment guarantee on crop choice. Section 2.3 presents the data and summary statistics. Section 2.4 outlines the estimation strategy. Section 2.5 presents the empirical results, and Section 2.6 concludes.

2.2 Risk management and households' crop choices: A theoretical framework

Providing additional employment opportunities to a total of 55 million households has brought about considerable changes in the social and economic realities in India. The NREGS affects households in rural areas through various channels. The most obvious and so far most intensely researched effect is the increase in available income and wealth of those households participating in the program. This wealth effect is most pronounced for households with surplus labor - namely households whose labor supply exceeds the labor demand of their farm firm - and in regions where regular labor markets fail to absorb this excess. The increase in income resulting from NREGS participation has been shown to increase consumption levels (Jha et al., 2012) and to reduce poverty (Klonner and Oldiges, 2014).[1]

Another effect, which is much less well understood, is the insurance effect. It is particularly relevant for households that are highly exposed to covariate shocks such as droughts, floods or large-scale crop diseases. In rural areas of India wages were shown to fall with covariate shocks (Jayachandran, 2006). Such wage fluctuations severely limit households' possibilities to cope with shocks through the labor market. By giving households the right to work and making employment opportunities available independently of shocks, the NREGS greatly influences households' ability to smooth income in the case of a shock. The expectation of having access to the NREGS, households could take more risk in their production decisions, and reach higher expected incomes. If a shock then occurs, households can cope with it by working for the NREGS.[2]

Finally, the NREGS is expected to affect wage levels through general equilibrium effects in the village economy. The NREGS was shown to raise wage levels in the private sector because wages under the NREGS are in many cases higher than the wages paid for casual work and households consequently shift their labor supply from the private sector towards the public works program (Berg et al., 2012; Imbert and Papp, 2015). Increases in wages could also affect production levels or crop choice in agriculture because they raise production costs, particularly for large-scale farmers.[3]

In this paper, I focus specifically on the insurance effect, and how it affects the allocation of inputs to risky crops in a household's farm.[4] The following theoretical model of household decision-making under uncertainty shows more systematically how the introduction of NREGS can affect crop choice via the insurance effect. The model primarily builds on Dercon and Christiaensen (2011). Taking into account the ideas outlined by Fafchamps (1993) and Van Den Berg (2002), I particularly explore how the sequencing of input allocation, shock realization and harvesting influences production

[1] Increases in disposable income and wealth might also positively influence the capacity to take risks and investment behavior. This effect is different from the insurance effect, which is the main focus of this paper. I discuss how I attempt to isolate the insurance effect in Section 2.4.

[2] Without the shock, it is unlikely that all of these households would participate in the NREGS, because their shadow wages probably exceed the wage rate paid in the scheme.

[3] Bhargava (2014) for example shows, that the NREGS induces farmers to shift their production technology towards labor-saving equipment. I show in Section 2.5.4 that the results of this paper are not driven by differences in the labor intensity of crops.

[4] I focus on input allocation because of data constraints: information on land allocation was not consistently collected.

decisions. The possibility to smooth consumption over time is therein constrained by two main factors: the lack of adequate risk management strategies and limited access to credit. Crop choice is first modeled in a world without risk but with imperfect credit markets and then extended to a world with uncertainty. This allows for the isolation of the effects of uncertainty and risk aversion on production decisions. Finally, I show how the introduction of the NREGS can affect input allocation decisions in both scenarios.

2.2.1 General setup

Assume that a household engaging in agricultural production has the choice between two agricultural products Q^d and Q^s. Given that both products are well known to the household and have been produced in the region for some time, we can abstract from learning and other sunk costs. These products are produced with two different types of production functions: one is deterministic and the other stochastic.[5] It is also assumed that the risky crop is more productive on average. Both products can be sold at local markets at the same price p.

Agricultural production takes place over two periods, the planting and the harvesting seasons. The total yield of both products Q depends on land a, labor l_1 and input k allocation in period one:[6]

$$Q^d = f^d(k^d, l_1^d, a^d) \tag{2.1}$$

$$Q^s = \epsilon f^s(k^s, l_1^s, a^s) \qquad E[\epsilon] = 1. \tag{2.2}$$

Inputs k are defined as a bundle of variable inputs such as seeds, fertilizer and pesticides. The total yield of the risky product additionally depends on the realization of a multiplicative, random, serially uncorrelated shock ϵ at the end of the first period. The expected value of this shock is 1; thus in expectation, the yield of the risky crop is just $f^s(a^s, l_1^s, k^s)$. Total yield has to be harvested in the second period, and labor required for harvesting l_2 is a linear function of realized yields, e.g. $l_2 = \alpha(Q^d + Q^s)$, where α

[5]The assumption, that one production function is deterministic and the other stochastic is rather extreme. Instead, one would expect both production functions to depend on the realization of random shocks, although to a different extent. However, this simplification is without major impact on the results obtained here.

[6]I have abstracted from fixed capital because the marginal effect of productive capital was found to be close to zero.

is a parameter indicating how much labor is needed for harvesting given any realized yield.[7]

I assume that the household maximizes utility from consumption C in both the planting and the harvesting periods. The utility function is additive over both periods and future utility is discounted by the factor δ. The utility function satisfies the usual properties: it is twice differentiable and increases in C but at decreasing rates, $\partial U/\partial C > 0$ and $\partial^2 U/\partial C^2 < 0$. This also implies that the household is risk averse. I abstract from leisure in this model because it does not change the choice under uncertainty.[8] The household generates income from wage employment on local labor markets and from agricultural production. Building on the full-income approach, the household maximization problem can be described as follows:

$$
\begin{aligned}
max \qquad V =& U_1(C_1) + \delta U_2(C_2) \\
& s.t. \\
& \quad C_1 \leq w_1(T_1 - l_1^d - l_1^s) - g(k^d + k^s) + B \\
& \quad C_2 \leq p(Q^d + Q^s) + w_2(T_2 - l_2) - (1+r)B \\
& \quad B \leq B^m \\
& \quad a^d + a^s \leq 1.
\end{aligned}
\tag{2.3}
$$

Total time endowment is represented by T_1 and T_2. In both periods total time can be allocated between working in the labor market and working in own fields. In the first period, the household obtains income from wage work at level w_1 and from borrowing B. Inputs for agricultural production can be purchased at price g. In the second period, the household obtains income from the sale its own agricultural production $p(Q^d + Q^s)$ and wage work at level w_2. Note here that the household has to allocate labor to harvesting in order to generate income from its agricultural production. Because it seems plausible that the household always prioritizes its own harvest over wage employment, I assume that the household deems the cost of harvesting to equate to reservation wages rather

[7]Because labor allocation is linear in realized yields, it is profitable to harvest either the entire crop or nothing at all (depending on wage levels and output prices).

[8]By dropping leisure, I ignore possible income effects of increases in wage levels on a household's time allocation between labor and leisure. But since my main interest lies in crop choice rather than in production levels, ignoring leisure is not of major concern. Similar approaches can be found in Rosenzweig and Binswanger (1993), Fafchamps and Pender (1997) and Dercon and Christiansen (2011).

than market wages. It is therefore useful to replace the wage cost of harvesting $w_2 l_2$ in the budget constraint with $\alpha w_2^r(Q^d + Q^s)$, where w_2^r is the reservation wage and $\alpha(Q^d + Q^s)$ is the effort necessary for harvesting expressed in units of realized yield.

Incurred debts have to be repaid in the second period at an interest rate of r. Input credits are relatively common in rural Andhra Pradesh, although it seems that the amount of credit conceded is limited by a household's wealth. In the sample around 18% of the households that applied for credit reported not receiving the total amount of credit they applied for. Therefore, B^m describes the maximum amount a household can borrow for productive purposes. In contrast to input credit, consumption credit is much more difficult to obtain and highly expensive. Because households are expected to opt for that source of credit only under extreme circumstances, this model does not allow for any borrowing beyond the harvesting period.

In this setting local labor markets are assumed to function with the option to hire labor in as well as out. In fact, most households in the sample report a range of income sources - of which casual labor features prominently. However, harvest stage wages are assumed to be stochastic and to covary with covariant shocks, such as rainfall shortages. This was shown in the case of rural India by Jayachandran (2006). For most households, this means that they can only form expectations about harvest stage wages and face a double risk from rainfall fluctuations: First, their own harvest is likely to fail if there is a rain shortage. Second, they cannot find work at adequate wage levels in local labor markets.

Finally, $a^d + a^s = 1$ describes the restrictions on allocable land. I assume that there are no functioning land markets and that owned land is used for own agricultural production or left fallow. This is obviously a simplifying assumption that does not hold everywhere in India. Nonetheless, observed levels of land renting are relatively low in rural Andhra Pradesh and land sales are virtually absent.[9]

The model described so far deviates from standard neoclassical models in that credit and land markets are assumed to be dysfunctional. Given these constraints, households' production and consumption decisions are not separable even in the absence of risk.

[9] Part of this is due to a very restrictive legal environment that discourages land owners from renting out their land even if it is otherwise left fallow. Also, land prices are very high, which combined with low levels of credit availability makes land acquisition impossible for the majority of households. Those who could afford this rather seek to diversify out of agriculture and move to urban areas.

2.2.2 Deterministic case

First, consider a scenario without uncertainty. In such a world each household maximizes utility by maximizing profits from agricultural production plus income from wage employment.[10]. Because both production functions are deterministic in this scenario, optimal land, input and labor allocations are achieved when their marginal products equal respective prices. Solving the household maximization problem leads to the following decision rule for the allocation of variable inputs to each of the crops:[11]

$$\frac{\partial f^{d,s}}{\partial k^{d,s}} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial U_2}{\partial C_2}}. \tag{2.4}$$

In the absence of risk, the decision rule is equal for both crops, and optimal allocation implies that the marginal product of inputs in d is equal to the marginal product of inputs in s. Because realized yield is harvested in the second period, input allocation does not only depend on input and output prices but also on reservation wages in the harvest season and on the intertemporal marginal rate of substitution in consumption. If credit constraints bind, input allocation to both crops is lower, and the household allocates more time to the labor market.[12]

2.2.3 Introducing uncertainty

When introducing uncertainty, the household has to form expectations about the realized yield of the risky crop Q^s, the wage levels in the harvest period w_2, and the level of consumption that can be achieved in the second period C_2. The decision rules for input allocation under uncertainty change to

$$\frac{\partial f^{d}}{\partial k^{d}} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial EU_2}{\partial C_2}} \tag{2.5}$$

[10] Identical results would be obtained if the household were risk neutral

[11] As mentioned earlier, the main focus of this paper is on input allocation, but similar results can be obtained for the allocation of labor and land to each of the crops. A detailed derivation of all decision rules can be found in the Appendix, Section 2.A.

[12] C.f. Section 2.A of the Appendix for a derivation of this result.

for the deterministic crop, and to

$$\frac{\partial f^s}{\partial k^s} = \frac{g}{(p-\alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial EU_2}{\partial C_2}} - \frac{cov(\frac{\partial U_2}{\partial C_2}, \epsilon)}{(p-\alpha w_2^r)\delta \frac{\partial EU_2}{\partial C_2}} \tag{2.6}$$

for the stochastic crop. Equation (2.5) looks similar to equation (2.4), except that the household now maximizes expected utility of consumption in the harvest period. For any expected consumption level C_2, expected utility $EU_2(C_2)$ is lower than the utility of the expected value $U_2(E(C_2))$, and marginal expected utility is higher than the marginal utility of the expected value. Thus, under uncertainty, the right-hand side term is lower than in the deterministic case, implying that the household allocates more inputs to the safe crop than it would in the absence of risk. This reflects the greater weight households put on future consumption relative to current consumption when facing uncertainty. Equation (2.6) shows the effect of uncertainty on input allocation to the risky crop. Here the decision rule changes considerably and the overall effect is less clear. Again, marginal expected utility is higher than marginal utility, thus implying higher input allocation to the risky crop also. However, the covariance between marginal utility of consumption and the random shock ϵ is strictly negative.[13] This term increases the value of the right-hand side of equation (2.6), which means that input allocation to the risky crop is lower under uncertainty. Which of the two effects is stronger depends on the degree of risk aversion of the household, expected consumption levels C_2 and the amount of covariance between marginal utility and the random shock. Since the covariance is greater with lower wages in period two and with a higher interest rate r, the net effect of uncertainty on input allocation can be expected to be negative in this context. Irrespective of total levels of input allocation, it can be clearly seen that under uncertainty, input allocation shifts towards the safe crop k^d relative to the risky crop k^s. Thus under uncertainty, the share of risky crops in a household's portfolio is always lower than in the deterministic scenario.

[13]In a bad state of the world ($\epsilon = 0$) consumption in the second period is lower and marginal utility higher than in a good state of the world. Conversely, a high ϵ leads to higher consumption in period 2 and to lower marginal utility of consumption.

2.2.4 The insurance effect of an Employment Guarantee

The insurance effect of an employment guarantee, such as the National Rural Employment Guarantee Scheme (NREGS), on households' allocation rules is best represented by an increase in expected harvest stage wages.[14] For households with a labor surplus, other farms offer the best possibility of finding employment during harvest periods; in the case of major weather shocks, they have to expect to not find any employment at all (Jayachandran, 2006; Kaur, 2014). Because the NREGS provides reliable income opportunities throughout the year, households can expect to find employment in the harvest period even in bad years. In other words, the NREGS increases wage levels in bad years and therewith reduces the covariance between harvest stage wage levels and covariant shocks. The comparative statics in this section show that the introduction of NREGS affects optimal input allocation under certainty differently than under uncertainty.

Without uncertainty, an increase in average harvest period wages w_2 affects optimal input allocation by increasing consumption levels that can be realized in the second period (c.f. eq. 2.4). Households that hire labor out (i.e. those whose land is too small to produce at higher levels) increase consumption. One can thus see a decrease in input allocation for net lenders of labor because of increases in C_2, which reduces $\partial U_2/\partial C_2$ and increases the second part of the right-hand side of equation (2.4). The effect of increased wages on agricultural production levels (through consumption) can be understood as a substitution effect. Because working outside the farm becomes more profitable for households with little cultivated land, the allocation of inputs to those lands should decrease from very high levels to more efficient ones.

An entirely different effect can be observed if uncertainty reduces input allocation to risky crops as given by equation (2.6). If harvest stage wages increase, we can observe the same effects on marginal utility of consumption as in the deterministic case. Under uncertainty, however, the negative covariance term reduces input allocation to the risky crop, and this effect is now partially offset by the introduction of an employment guarantee. As possibilities to generate market income improve, the effect of shocks on harvest period consumption decreases. Because the household knows that it can earn additional income in instances of negative production shocks by spending more time

[14]Of course, in a scenario without uncertainty, expected wage levels need to be replaced by average wage levels.

working for the NREGS, it can afford to take a greater amount of risk in his agricultural production. The more the covariance term on the right-hand side of equation (2.6) approaches zero, the more the ratio of inputs allocated to the risky crop (versus the safe crop) approaches the deterministic scenario. This means that even if total input (or similarly labor) allocation is reduced due to the employment guarantee, the share of total inputs allocated to each of the crops approaches the ratio of the deterministic scenario. Interestingly, this effect holds independently of whether credit constraints reduce total input allocation or not.[15]

2.3 Data

When estimating the insurance effect of the NREGS, one must take into account considerable variation in the quality of implementation of the program across states (Dutta et al., 2012). The section above highlighted the importance of households' expectations about future income streams. Therefore it seems plausible to observe insurance effects only in states in which the demand for employment has been sufficiently met, already in the early years of program implementation.

Given these considerations, the model specified above is tested using the Young Lives Survey (YLS) data for Andhra Pradesh. Andhra Pradesh is particularly suited to studying the question of interest because it is one of the best performing states in India in terms of the number of workdays generated per household and meeting the demand for work (Dutta et al., 2012). Regarding outreach, only Chhattisgarh, West Bengal, Madya Pradesh and Rajastan reached higher proportions of rural households in the financial year 2009/10.[16]

The YLS data set covers 3,019 households living in six different districts, 17 subdistricts (blocks) and 87 villages. The selection process of districts for the YLS ensured that all three geographical regions were surveyed, as too were the poor and non-poor

[15] C.f. Section 2.A of the Appendix for a detailed derivation of this result.

[16] At the same time, Andhra Pradesh has been a forerunner in terms of innovative approaches to the implementation of the NREGS. First, it has a lot of experience with performing social audits to increase accountability within the scheme. Second, it was one of the first states to cooperate with IT enterprises to strengthen the efficiency of administrative processes. To increase transparency, entries on muster rolls and the number of workdays generated per job card holder, inter alia, are publicly accessible. Nonetheless, the program continues to be implemented in a top-down manner in Andhra Pradesh. Usually, work is not generated upon demand, rather work applications are only accepted if there is work available.

districts of each region, such that the YLS is broadly representative of the population of Andhra Pradesh (Galab et al., 2011).[17] Three rounds of interviews have been conducted so far (2002, 2007 and 2009/10). Panel attrition is relatively low: 2,910 households were revisited in 2009/10, giving an attrition rate of 3.6% (Galab et al., 2011). For reasons of comparability, only the second (2007) and third (2009/10) rounds are considered in the current analysis. Furthermore, the analysis is restricted to households with non-zero agricultural production in 2007 and 2009/10. This data is complemented by secondary data for the calculation of the dependent variable as well as for a number of controls.

Table 2.1 reports baseline summary statistics for the main variables and controls used in the paper. I split the sample in treatment and control group. Treatment indicates that a household has access to the NREGS at the district level at the beginning of the agricultural cycle. The period of reference for the 2007 round of interviews is the agricultural year 2005/06 (May 2005 to April 2006). Given that the introduction of the NREGS started in April 2006, no household had access to the NREGS in the baseline reporting period. The period of reference for the 2009/10 interviews is the agricultural year 2008/09. By that time, NREGS works had started in the districts Anantapur, Cuddapah, Karimnagar and Mahaboobnagar, the treatment districts. In Srikakulam and West Godavari, the control districts, the introduction of the NREGS was in August 2007 and in March 2008 respectively. Since activities started only very slowly in Srikakulam, I use this district as control district despite the introduction of the NREGS in mid 2007.[18]

For the calculation of the dependent variable - a risk index of each households' crop portfolio - data on input allocation to each crop from the questionnaire is combined with District-level crop production statistics. The time series of crop production statistics are used to calculate the coefficient of variation of each crop's yield.[19] With this information, a risk index R_i of each household's crop portfolio is constructed given the reported allocation of inputs to each of the crops.[20] The risk index for household i

[17] This is in reference to the State of Andhra Pradesh in 2013, prior to its division into the states of Andhra Pradesh and Telangana.

[18] For a detailed discussion of data sources and the construction of variables refer to the Appendix, Section 2.B.

[19] Crop risk can also stem from variability in prices, not only in yield. However, given the practice of setting and regularly adjusting Minimum Support Prices in India, it is impossible to compute price risk based on time-series of prices.

[20] Allocation of inputs refers to the share in total variable inputs such as seeds, fertilizer and pesticides that is allocated to each crop in a household's portfolio. This is the only information collected in the

Table 2.1: Baseline characteristics

	Treatment		Control		
	Mean	SD	Mean	SD	p-value
Household characteristics					
Male household head	0.96	0.20	0.97	0.18	0.41
Age of household head	41.93	12.13	41.01	11.83	0.24
Household head is literate	0.32	0.47	0.25	0.43	0.01
Household size	6.10	2.62	5.61	2.07	0.00
Wealth index	0.39	0.13	0.38	0.20	0.77
Annual income, off-farm activities	24.70	24.82	19.81	26.13	0.00
Hh benefits from credit/training program	0.62	0.49	0.58	0.49	0.18
Any serious debts	0.63	0.48	0.47	0.50	0.00
Able to raise 1000 rupees in one week	0.61	0.49	0.33	0.47	0.00
Farm characteristics					
Value of agr. production	28.49	45.76	24.38	124.96	0.56
Value of variable inputs	14.51	21.34	14.42	69.52	0.98
Area cultivated (acres)	4.14	4.57	2.73	5.47	0.00
Risk index of crop portfolio	0.36	0.12	0.26	0.08	0.00
Labor intensity of crop portfolio	0.27	0.07	0.28	0.08	0.01
Cost intensity of crop portfolio	21137	7411	25873	10455	0.00
Herfindahl index of crop portfolio	0.76	0.25	0.80	0.23	0.00
Number of crops	2.04	1.03	2.07	1.33	0.68
Irrigated area (% of total)	0.18	0.32	0.14	0.30	0.06
Fertilizer (dummy)	0.98	0.15	0.87	0.34	0.00
HYV seeds (dummy)	0.77	0.42	0.63	0.48	0.00
Participated in labor sharing (dummy)	0.75	0.43	0.78	0.41	0.22
Time in crop production (hours per year)	2085	2280	1365	1310	0.00
Shocks					
Rainfall (deviation)	0.33	0.28	-0.06	0.16	0.00
Rainfall (deviation, lag)	-0.39	0.10	-0.12	0.10	0.00
Self-reported shock	0.81	0.39	0.51	0.50	0.00
NREGS participation					
Household registered with NREGS	0.66	0.47	0.00	0.00	0.00
Household generated income from NREGS	0.54	0.50	0.00	0.00	0.00
Income, NREGS	1.24	2.39	0.00	0.00	0.00
Observations	750		338		

Notes: All values in constant INR 1,000 (July 2006). One US$ is equivalent to 46.38 INR (July 2006). Variable definitions and sources are described in the Appendix, Section B.

given input allocation k to crop m is defined as $R_i = \sum r_m k_m / \sum k_n$, where r_m is the coefficient of variation of the yield of crop m.[21] Note here, that r_m is only available for a subset of all crops n (26 out of 42), such that $m \subseteq n$. Still, $\sum k_m$ represents roughly 90% of the total allocation of inputs in the sample. To reduce potential bias, I drop all observations from the sample which have no crop in their portfolio for which risk information is available, e.g. $\sum k_m = 0$ or $R_i = 0$, in one or both of the survey rounds.[22] As can be seen in Table 2.1, The risk index at baseline is higher in the treatment group (0.36) than in the control group (0.26). The difference is statistically significant at the 1% level.

Using a risk index as dependent variable deviates from choice analyzed in the theoretical model because households can cultivate more than two crops. Obviously households could also chose to increase the number of crops in their portfolio as strategy to diversify risk. If there is imperfect correlation between crops' yields, then the risk index would not adequately predict the amount of risk a household is willing to take in his production decisions as it omits the effect of diversification in overall crop risk. However, crop concentration is quite high in the sample: the average household produces only two crops and the baseline Herfindahl index of the crop portfolio is 0.76 in the treatment group. In the control group, the baseline Herfindal index is 0.8 (c.f. Table 2.1).[23]

Agricultural production levels as well as the amount spent on variable inputs (such as seeds, fertilizer and pesticides) are not statistically different between treatment group households and control group households. However, the area cultivated, irrigation levels, the probability to apply fertilizer and to use high yielding variety (HYV) seeds are all higher in the treatment group than in the control group. Table 2.1 also summarizes the occurrence of different shocks in both groups and in both periods. Rainfall deviation and rainfall deviation (lag) describe the deviation of annual cumulative rainfall levels from their long-term average. Finally, Table 2.1 reports the participation status with

survey that gives information about the relative importance of each crop in a household's production.

[21] The distribution of the risk index as well as of the change in this variable between survey rounds is plotted in Figures (2.D.1) and (2.D.2) respectively in the Appendix.

[22] Section 2.B.2 of the Appendix provides more information on how the variable is constructed. The robustness of my findings to the selection of alternative dependent variables, such as the weighted average of the standard deviation of crop returns, but also to different methods of aggregating the risk index is shown in Table 2.D.1 in the Appendix.

[23] I also show in Section 2.5.4 that the introduction of the NREGS has a positive (although not statistically significant effect) on crop concentration.

the NREGS at the time of the baseline data collection. As can be seen, 66% of the households in the treatment districts report having registered with the NREGS in 2007.

2.4 Estimation strategy

The key prediction of the model described in Section 2.2 is that an increase in expected labor market wages in the harvesting period, *ceteris paribus*, increases the share of inputs allocated to risky crops if households were previously constrained in their crop choice by high levels of uncertainty regarding output levels and dysfunctional insurance markets. It is not possible to test this hypothesis directly for two reasons. First, households' expectations with regard to wages depend on a range of individual factors (such as perceived access to the labor market) that would not be captured by observed village-level wages. Second, a range of unobserved village characteristics may change over time and those changes probably influence both expected labor market wages and households' crop choice.

To circumvent the problems mentioned above, I explore the availability of the NREGS as a source of exogenous variation in expected labor market wages during the harvest period. As argued in Section 2.2.4, the introduction of NREGS increases expected wages in the harvest period because employment opportunities through the NREGS do not depend on favorable weather outcomes and hence do not covary with village-level shocks.

It is important to notice here that the NREGS does not only affect households' crop choices through the insurance effect - which is the main focus of this paper. Because increases in available income and wealth due to the NREGS might also influence a household's ability to cope with shocks, their access to credit and their willingness to take risks, it is essential to control for these changes in order to isolate the insurance effect.[24] The outcome equation can be written as follows:

$$R_{ijt} = \beta_1 D_{ijt} + \beta_2 X_{it} + \beta_3 Z_{jt} + u_i + \gamma_j + \delta_t + \upsilon_{ijt}. \tag{2.7}$$

The dependent variable is the risk index of household i's crop portfolio at time t. D_{ijt} represents a household's access to the NREGS. Let X_{it} be a set of time-varying household characteristics that affect preferences and crop choice (such as education,

[24]Table 2.D.2 in the Appendix shows that the results are robust to the omission of these variables.

wealth, income and past experience with shocks) and u_i be time-constant unobserved household characteristics (such as risk aversion, farming ability and land quality). Z_{it} is a set of time-varying village-level characteristics (e.g. weather trends, extension services, prices, etc.), γ_j are time-constant village characteristics (such as the land's suitability for certain crops), δ_t is a time fixed-effect and υ_{ijt} is the error term.

Taking the first difference removes unobserved household and village level characteristics that are constant over time:[25]

$$\triangle R_{ij} = R_{ij,t+1} - R_{ij,t} = \beta_1 \triangle D_{ij} + \beta_2 \triangle X_i + \beta_3 \triangle Z_j + \triangle\delta + \triangle\upsilon_{ij}. \tag{2.8}$$

For β_1 to have a causal interpretation, the differences in the change of the risk index between the treatment and control groups must be entirely due to the NREGS. This assumption could be violated for a number of reasons. First, since the access to the NREGS is non-random, treatment could be correlated with potential outcomes of R_{ijt}. Second, households in the treatment and control group may not be following parallel trends in their crop choices. The remainder of this section discusses how I address these points.

This paper uses four different treatment variables. First, as discussed above, I explore the universal nature of the NREGS by coding as 'treated' those households based in districts where the NREGS was introduced in 2006.[26] Second, I use lagged block-level disbursements under the program as an indicator of the intensity of treatment, arguing that households living in blocks with higher past disbursements expect employment to be more readily available in situations of need. The average lagged disbursement in treatment districts is INR 14.27 Mio. with a standard deviation of 9.64 in 2009/10. Third, following the same logic, I use the lagged annual total of employment person-days generated per job card at the block-level. In 2009/10, the number of person-days generated was 11.15 on average with a standard deviation of 5.58. Fourth, I explore the self-selection of households into the program by comparing the changes in the risk

[25]With two time periods, taking the first differences is essentially the same as estimating the model in fixed effects.

[26]Given the size of the program and the huge awareness campaigns undertaken at the beginning of implementation, it seems valid to assume that households in rural Andhra Pradesh form expectations about income opportunities through the NREGS based on the local availability of the program and not only based on being registered with the program.

index of households who were registered with the NREGS by 2007 with the rest of the sample.

At the district level, the NREGS should have been introduced in the poorest districts first. This could potentially bias the estimates downwards because poorer districts are less likely to have extension services and marketing structures in place that would enable households to seize the opportunity to plant more profitable cash crops. However, in most states - and in Andhra Pradesh in particular - the prioritization of the poorest districts was not systematically implemented. In this sample, the general economic characteristics of treatment and control districts do not differ greatly.[27] The treatment intensity at the block level should also be exogenous to potential outcomes. Estimates could be biased if funds allocated to blocks responded to rainfall shocks and if these rainfall shocks also affected a household's input allocation decision. However, the amount of funds to be sanctioned per block is defined between December and March for the following financial year (April to March). Since I am using lagged values of disbursed funds, these amounts are fixed 14 to 18 months before household's decide on their input allocation.[28] Lastly, I explore differences in crop choices across households who registered with the NREGS or not. Here, the possibility that unobserved shocks or other time-varying variables affect both the decision to register and a household's crop choice cannot be ruled out. I employ matching techniques to reduce selection bias, but this is admittedly not sufficient to rule out non-random assignment.

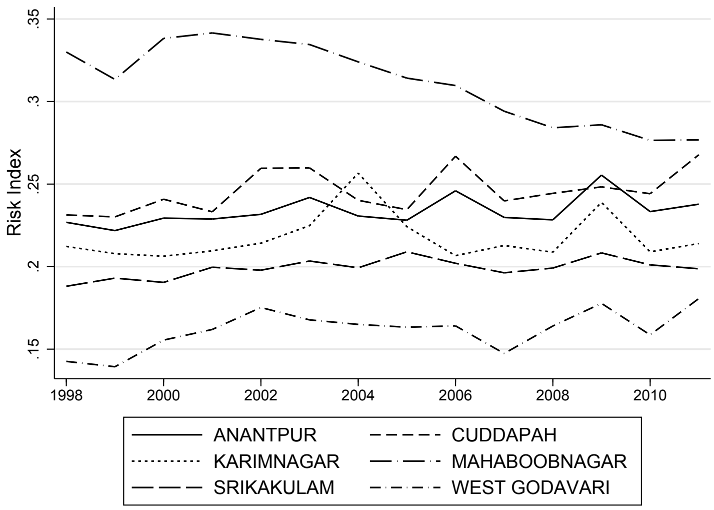
The parallel trends assumption could be violated due to differences in crop productivity which cause the share of certain crops in total input allocation to increase independently of the NREGS. Given the small number of districts in the sample, this could significantly bias the results. District-wise time trends in the risk index of crop production are displayed in Figure 2.1. One of the treatment districts (Mahaboobnagar) displays a decreasing trend in the risk index, while all other districts seem to be following the same trend.

Another - more subtle - violation of the parallel trends assumption could emerge from mean reversion in the dependent variable. Why might households with riskier crop portfolios display a negative change in the risk index? The reason could be effects of

[27]See Section 2.B.4 and Table 2.D.3 in the Appendix for more information.

[28]It is also fixed between 6 and 8 months before the start of the monsoon, which could affect next years input allocation through time-lags in the effect of shocks. For more information on the time line see Appendix, Section 2.B.1.

Figure 2.1: District-wise risk-index of land use



Source: Own estimation based on the Land Use Statistics and District-wise Crop Production Statistics, Ministry of Agriculture, GoI.

lagged shocks on current input choices which are rooted in the non-separability of production and consumption decision of agricultural households (Sadoulet and De Janvry, 1995). In a world with imperfect credit markets and risk, past shocks affect current wealth and therefore also current input allocation decisions. If household wealth is perfectly captured by the data, controlling for changes in wealth should eliminate any bias. If wealth is, however, also reflected in soil nutrition, which is affected by weather shocks and not captured in the data, then controlling for wealth is not sufficient (Foster and Rosenzweig, 2010a).

Assume that the risk index of each household's crop portfolio follows a modified AR(1) process, where - in the absence of a shock - the risk index at time $t+1$, R_{t+1}, is equal to a linear transformation of the risk index of the previous period plus some random noise, e.g $\rho R_t + \epsilon_{t+1}$.[29] In contrast, if a shock occurs, households with higher risk in their crop portfolio also face higher losses in agricultural production. This forces them to choose a more conservative crop portfolio in the following period. Formally, this process can be described as follows:

$$R_{t+1} = \rho R_t + \delta u_t + g(R_t)u_t + \epsilon_{t+1}. \tag{2.9}$$

[29] For expositional purposes, I drop all subscripts except the time subscript.

The shock u_t has expected value zero and $g(R_t)$ is a flexible function of input allocation, which allows shocks to have a differential effect on next seasons crop choice, depending on the level of R_t. In the absence of any program effect, the observed change in crop choice would be the following:

$$\begin{aligned}\triangle R &= R_{t+1} - R_t \\ &= \rho(R_t - R_{t-1}) + \delta(u_t - u_{t-1}) + g(R_t)u_t - g(R_{t-1})u_{t-1} + \epsilon_{t+1} - \epsilon_t \\ &= (\rho - 1)R_t + \delta u_t + g(R_t)u_t + \epsilon_{t+1}. \end{aligned} \tag{2.10}$$

In expectation this change would be $E(\triangle R) = (\rho - 1)R_t$. A placebo treatment effect is zero in expectation only if the process approaches a random walk (e.g. $\rho = 1$) or if the distribution of R_t is equal in treatment and control groups. The placebo treatment effect is even higher if the occurrence of lagged shocks u_t is different in both groups. The low number of districts used in this analysis warrants special attention to this phenomenon. As discussed earlier, baseline levels of risk as well as the occurrence of shocks are substantially different between treatment and control groups. I estimate the importance of mean reversion in the control group only and find estimates of $\rho - 1$, δ and $g(R_t)u_t$ equal to -0.61, 0.03 and -0.24 respectively.[30]

I account for shock induced mean reversion by adjusting equation (2.8) in a way that eliminates sources of correlation between $\triangle D_{ij}$ and $(\upsilon_{ij,t+1} - \upsilon_{ij,t})$. Using equation (2.10) to rewrite eq. (2.8) yields:

$$\begin{aligned}\triangle R_{ij} = \beta_1 \triangle D_{ij} + \beta_2 \triangle X_i + \beta_3 \triangle Z_j + \triangle\delta \\ + (\rho - 1)R_{ijt} + \delta u_{jt} + g(R_{ijt})u_{jt} + \triangle\upsilon_{ij}. \end{aligned} \tag{2.11}$$

Following Chay et al. (2005), I estimate a simplified version, such as:

$$\begin{aligned}\triangle R_{ij} = \beta_1 \triangle D_{ij} + \beta_2 \triangle X_i + \beta_3 \triangle Z_j + \triangle\delta \\ + \beta_4 R_{ijt} + \beta_5 u_{jt} + \beta_6 R_{ijt} u_{jt} + \triangle\upsilon_{ij}. \end{aligned} \tag{2.12}$$

Before proceeding, one last empirical challenge needs to be addressed: within cluster correlation in $\triangle\upsilon_{ij}$. Studies that work with a small number of clusters always face the

[30] I use the level and the square of R_t as approximation for $g(R_t)$. Detailed results can be found in the Appendix, Table 2.D.4.

challenge of adequately adjusting standard errors for potential within cluster correlation of errors. Throughout the paper, I calculate Eicker-White standard errors clustered at the sub-district (block) level or district level depending on the level of aggregation of the regressors. However, since the number of clusters is fairly small, these standard errors are likely to be downward biased (Cameron et al., 2008). In cases of very few clusters, Cameron et al. (2008) suggest to calculate p-values using a wild cluster-bootstrap with Rademacher weights.[31] In a more recent paper, Cameron and Miller (2015) suggest the use of Webb's (2014) weights if the number of clusters is smaller than ten, which seems reasonable when using a district level treatment variable. P-values of the respective treatment variable using both versions of the bootstrap with 4,999 replications are reported at the bottom of Table 2.4.[32]

2.5 Results

This section starts by presenting estimates for an agricultural production function. It proceeds by assessing the extent to which the NREGS can actually support households in this sample in coping with shocks, which is the precondition for expecting any insurance effect. This section then analyzes the effects of the NREGS on households' crop choices and presents a number of robustness checks.

2.5.1 Identifying profitable production strategies

To understand inhowfar households' crop choice can improve their income from agricultural production, I estimate an agricultural production function, linking the total value of agricultural output Q_{ijt} to input allocation K_{ijt}, labor L_{ijt}, plot size A_{ijt} and risky crop choice R_{ijt}. I estimate agricultural output assuming a Cobb-Douglas production function, in which the choice of crops affects output multiplicatively in the following

[31]This approach was applied, inter alia, by Adrianzen (2014) to data clustered in 26 villages and by Akosa Antwi et al. (2013) to 28 quarter-year groups.

[32]The wild cluster-bootstrap calculates t-statistics for each bootstrap sample and estimates rejection rates based on the resulting distribution of t-statistics. Because this method does not calculate standard errors, I report clustered standard errors throughout the text. Implementation of the bootstrap in Stata is done based on the do-file written by Douglas Miller, which can be accessed online: http://www.econ.ucdavis.edu/faculty/dlmiller/statafiles/.

manner:

$$Q_{ijt} = (K_{ijt}^{\beta_1} L_{ijt}^{\beta_2} A_{ijt}^{\beta_3}) e^{g(R_{ijt})}. \tag{2.13}$$

I allow R_{ijt} to affect output non-linearly because it seems very likely that increasing the average risk in a crop portfolio is only beneficial to a certain extent, beyond which risk is simply too high to increase output. The production function described in equation (2.13) can be estimated by log-transforming the data and controlling for shocks Z_{ijt}, unobserved characteristics γ_{ij} and time effects δ_t. Again, I use the level and the square of R_{ijt} as approximation for $g(R_{ijt})$:

$$\begin{aligned} \ln(Q_{ijt}) = \beta_0 + \beta_1 \ln(K_{ijt}) + \beta_2 \ln(L_{ijt}) + \beta_3 \ln(A_{ijt}) + \beta_4 R_{ijt} + \beta_5 R_{ijt}^2 \\ + \beta_6 Z_{ijt} + \gamma_{ij} + \delta_t + \upsilon_{ijt}. \end{aligned} \tag{2.14}$$

The equation is estimated in OLS, random effects and fixed effects. As can be seen in Table 2.2, all models generate similar results. Columns (1) and (2) report OLS estimates for the survey round of 2007. These show that estimates are not affected by the exclusion of labor from the agricultural production function.[33] Columns (3) and (4) show random effects estimates, and column (5) and (6) fixed effects estimates. In columns (4) and (6), I additionally allow the effect of rainfall to vary with the amount of risk in a household's crop portfolio. The estimates in Table 2.2 suggest that households could significantly raise the value of their agricultural production if they were to increase the share of inputs allocated to riskier crops. However, this is only true up to a certain level. The square of the risk index is statistically significant at the 5% level in all specifications that use both rounds of data. Based on the fixed effects estimates, predicted agricultural output reaches its maximum at a risk index of 0.42.[34] Beyond this point, a further increase in risk would reduce total agricultural output. Average risk levels in households' crop portfolios are well below this value; in the survey round of 2007 the average risk index was 0.36 in the treatment group and 0.26 in the control group (c.f. Table 2.1). Other variables, such as the amount of inputs allocated, total cultivated area and labor have the expected sign and are all

[33] I cannot control for labor in the panel data models, because time information was only collected in 2007 and not in 2009/10.

[34] C.f. Figure 2.D.3 in the Appendix.

statistically significant.[35] The interaction term of rainfall and the risk index is positive and statistically significant at the 5% level. At the optimal risk level of 0.42, the marginal effect of rainfall is as high as 0.24 with a standard error of 0.17.

Table 2.2: Agricultural Production Function

	2007 OLS		Random Effects		Fixed Effects	
	(1)	(2)	(3)	(4)	(5)	(6)
Risk index of crop portfolio	5.771*	6.040**	6.665**	6.618**	6.747*	6.353*
	(1.996)	(1.989)	(2.354)	(2.344)	(3.069)	(2.913)
Risk index of crop portfolio (squared)	-7.100**	-6.887**	-8.442**	-8.616**	-7.953+	-8.226*
	(2.312)	(2.240)	(2.818)	(2.769)	(3.770)	(3.545)
Variable inputs (log)	0.876***	0.805***	0.915***	0.913***	0.736***	0.743***
	(0.114)	(0.109)	(0.131)	(0.131)	(0.128)	(0.127)
Area cultivated (acres, log)	0.784**	0.622**	0.872***	0.880***	0.659**	0.673**
	(0.225)	(0.196)	(0.229)	(0.227)	(0.188)	(0.185)
Labour (hours, log)		0.225***				
		(0.056)				
Rainfall (deviation)	-0.347	-0.391	-0.191	-0.683	-0.010	-1.147+
	(0.434)	(0.421)	(0.176)	(0.533)	(0.155)	(0.619)
Rainfall (deviation) × Risk index				1.439		3.311*
				(1.259)		(1.552)
Observations	1088	1088	2176	2176	2176	2176
R^2	0.295	0.318			0.129	0.132

Notes: Dep. var.: Income from agricultural production (log). Additional controls are share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Col. (1) & (2) additionally control for household characteristics: age, sex, and education of household head, and household size. Standard errors (clustered at the the sub-district) in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

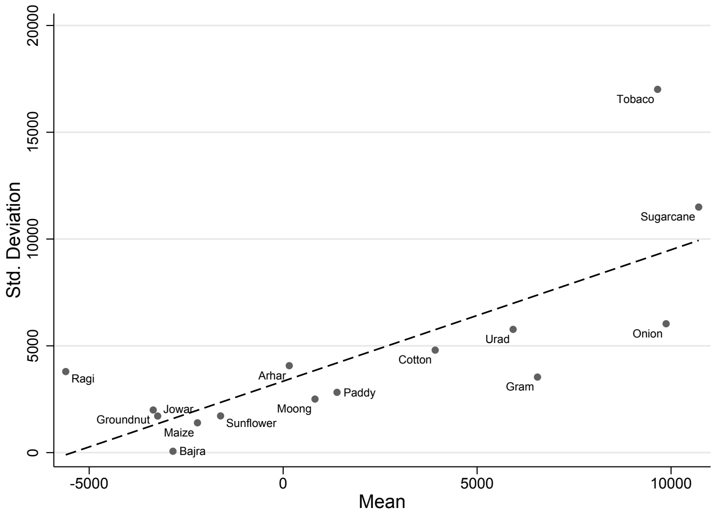
To gauge the robustness of this result, I additionally consider state-level statistics on the returns per hectare for major crops between 1996 and 2009.[36] Figure 2.2 plots the average returns of different crops against the standard deviation of these returns for the years 1996 to 2006 in Andhra Pradesh. The scatter plot shows a clear positive relationship between average returns and their volatility, indicating again that the riskiness of crops is strongly correlated with returns to producing these crops.[37]

[35] Additionally, the share of area under irrigation seems to increase output levels. In contrast, the dummies indicating whether or not a household applied fertilizer or high yielding variety (HYV) seeds are not statistically significant. This might seem somewhat surprising, but since expenditure on fertilizer and seeds is included in variable inputs, one should not attribute too much weight to this finding.

[36] Unfortunately, these statistics are only available at state level and only for very few crops.

[37] Many of these commodities are traded internationally, such that risk-aversion of farmers alone can

Figure 2.2: Returns per hectare of major crops



Source: Own estimation based on the Cost of Cultivation Statistics for Andhra Pradesh, Ministry of Agriculture, GoI.

2.5.2 Does the NREGS support households in coping with shocks?

Next, I estimate to which extent the NREGS helps households in coping with shocks. I argue that the NREGS has an insurance effect only if work provision sufficiently reacts to increasing demand in the case of a shock. Therefore, I test whether deviations from mean rainfall levels, as well as households' self-reported shocks, drive changes in the number of days households work for the NREGS in a fixed effects model. The results are reported in Table 2.3. In the first two columns, the total number of days worked in the past 12 months is the dependent variable; in the last two columns it is the log of this variable. The estimation is also restricted to phase one districts; thus only households who had access to the NREGS in both survey rounds are considered.

The results suggest that the number of days worked for the NREGS changes considerably with variation in rainfall levels. The greatest change is observed for lagged rainfall levels - that is, cumulative rainfall in the agricultural year preceding the period of reference. The coefficient of the lagged rainfall variable is negative 63.1, which im-

probably not explain the observed correlation between the riskiness of crops and their returns. Other reasons could be differences in the concentration of supply or demand between crops. Analyzing the reasons for the apparent relationship between risk and returns in crop portfolios is beyond the scope of this paper.

Table 2.3: Number of days worked with NREGS

	(1)	(2)	(3)	(4)
	NREGS days		NREGS days (log)	
Rainfall (deviation, lag)	-65.423**	-63.104**	-2.945**	-2.873**
	(18.461)	(19.151)	(0.887)	(0.845)
Rainfall (deviation)	-28.798**	-31.004**	-0.800	-0.876
	(8.700)	(8.935)	(0.554)	(0.497)
Self-reported shock	1.543	1.437	0.174*	0.179*
	(3.631)	(3.899)	(0.069)	(0.077)
Observations	1490	1490	1490	1490

Notes: Estimation in fixed effects. Dep. var.: No. of days a household worked for the NREGS in the past 12 months. Time trend and region-time trends included, but not reported. Col. (2) and (4) additionally control for area cultivated (acres, log), wealth index of the household, and if household benefits from credit/training program. Standard errors (clustered at the sub-district) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

plies that households worked 6.3 more days for the NREGS if lagged rainfall levels were 10% below average. This supports the assumption that the NREGS helps households in coping with shocks, because households use the program to smooth income ex post - for instance, after harvest and after agricultural products have been sold.[38]

Table 2.3 also shows how important maturation of the program is. A large share of the variance in the number of days worked for the NREGS can be explained by time alone. In contrast, wealth levels do not seem to influence the dependent variable, and the size of the cultivated area is only statistically significant in one specification. This is probably due to the limited variation of this variable over time.[39] Self-reported shocks also seem to increase the number of days worked for the NREGS.[40]

To quantify the contribution of the NREGS to households' risk coping, I compare agricultural losses due to rainfall shortages with income gains through the NREGS. The agricultural production function estimated in Section 2.5.1 (col. 6) suggests that a deviation from average annual rainfall by negative 25%, would reduce agricultural

[38]Similar evidence is provided by Johnson (2009), who finds that the number of days households work for NREGS increases if rainfall levels are lower than average.

[39]A positive coefficient could indicate program capture by wealthier households. But a further investigation of this issue is beyond the scope of this paper.

[40]The variable is coded as one of a household reported any of 12 self-reported shocks related to agricultural production.

output by 5.9% at the optimal level of the risk index. For the average household, this implies a nominal loss of about INR 1,740 (or US$ 37.5 in constant July 2006 values). The same deviation in lagged rainfall would lead households to work about 15.8 more days for the NREGS, which would generate an additional income of INR 1,020 (US$ 22) at mean wages observed in the sample. The NREGS thus allows households to compensate about 58% of agricultural production losses caused by rainfall shortages. Since rainfall fluctuations are among the most important sources of risk for rural households, these results suggest that the NREGS could indeed have an insurance effect in Andhra Pradesh.

2.5.3 The effects of the NREGS on households' crop choices

In this section I estimate the effect of the NREGS on households' input allocation decisions. Table 2.4 reports the effects of the NREGS on the risk index of a household's crop portfolio. As described in Section 2.4, I estimate all equations in first differences and control for initial condition in columns (2), (4) and (6). To isolate the insurance effect described in Section 2.2.4, I also control for variables that might be affected by the NREGS and might influence a household's crop choice through effects other than the insurance effect. These variables include household off-farm income and wealth, as well as key farming characteristics, such as the size of cultivated land, irrigation and total value of variable inputs allocated.[41] In all specifications I also control for self reported shocks, access to other government programs and rainfall levels (current and lagged). Additionally, a time dummy is included to control for state-wide changes in input and output prices, weather trends that are not captured by rainfall data and other changes at the state level that could influence a household's crop choice.

The results show a positive effect of the NREGS on the riskiness of households' production decisions. Consistent with the higher prevalence of shocks in the treatment districts and higher initial values of the risk index, controlling for mean reversion increases the estimated effect of the NREGS. Given the low number of clusters, inference should be based on the p-values obtained from the wild-cluster bootstrap. The effect

[41] Household off-farm income consists, inter alia, of income generated through the NREGS in the past 12 months. Optimally, this should be a lagged value because input allocation decisions are taken at the beginning of the season, while the income variable refers to the time period shortly after these allocative decisions were taken. Unfortunately, the survey does not include this information. Table 2.D.2 in the Appendix shows that the results are not influenced by changes in income or changes in total input allocation.

Table 2.4: Effect of the NREGS on risk index of crop portfolio

	(1)	(2)	(3)	(4)	(5)	(6)
D.NREGS introduced in district	0.038*	0.072**				
	(0.011)	(0.017)				
D.Cumulative expend., NREGS (log, lag)			0.007**	0.014***		
			(0.002)	(0.003)		
D.Employment per JC generated, NREGS (lag)					0.002^{+}	0.002
					(0.001)	(0.002)
Rainfall (deviation) at baseline		0.121^{+}		0.109		0.175*
		(0.050)		(0.072)		(0.080)
Risk index at baseline		-0.603***		-0.582***		-0.512***
		(0.061)		(0.058)		(0.068)
Risk index × Rainfall (deviation)		-0.109		-0.152		-0.296^{+}
		(0.142)		(0.132)		(0.156)
Bootstrap p-value of main treatment variable						
Rademacher weights:	0.107	0.047	0.072	0.015	0.326	0.388
Webb weights:	0.099	0.045	0.062	0.013	0.315	0.391
Observations	1088	1088	1088	1088	1088	1088
R^2	0.067	0.443	0.066	0.435	0.058	0.400

Notes: Estimation in first differences. Dep. var.: Risk index of a household's crop portfolio. Additional controls are variable inputs (log), area cultivated (log), share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, annual income, off-farm activities (log), if household benefits from credit/training program, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Standard errors (clustered at the district in cols. (1) and (2) and at the sub-district in cols. (3) to (6)) in parentheses. P-values are obtained by performing a wild cluster-t bootstrap with 4,999 replications and two alternative weights. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

of the introduction of the NREGS at district level is statistically significant at the 10% and 5% level in columns (1) and (2), respectively (using Webb weights). In columns (3) to (6), I also test for the effects of cumulative expenditure and total employment generated per job card under the NREGS. These variables are lagged by one year, to avoid correlation of the treatment intensity with past shocks. The coefficient on cumulative spending in the sub-district is statistically significant at the 10% and 5% level, depending on the specification considered. Only the amount of employment generated in the sub-district does not yield statistically significant effects when inference is based on the wild-cluster bootstrap.

Results presented in column (2) suggest that the risk index in households' crop portfolios increased by 7.2 percentage points due to the introduction of the NREGS at the district level. Given that the risk index in the treatment group was 0.36 at baseline, the introduction of the NREGS raised the average risk index to 0.43 (absent any shock induced mean reversion), which is remarkably close the optimal risk index of 0.42 identified in Section 2.5.1.

In terms of economic relevance, the results suggest that per additional day of employment generated in the block, each household would increase the risk index by 0.15 percentage points (col. 5). One standard deviation increase in the number of person-days generated per job card (6.9) would increase a household's risk index by 1.07 percentage points and raise net income from agricultural production, *ceteris paribus*, by about INR 480 (or US$ 10.4 in constant July 2006 values). This is particularly interesting from a cost-benefit perspective, since these net income gains per household are slightly higher than the wage cost (evaluated at the sample average of observed NREGS wages) of creating 6.9 days of employment under the NREGS, e.g. INR 467 (US$ 10). Of course, wage costs make up for only a part of overall program costs and not all of the NREGS participants own their own land, but nevertheless the magnitude of this effect is striking.

2.5.4 Robustness checks

This section presents a number of robustness checks. The first robustness check is intended to rule out the possibility that the observed effects is not due to the NREGS. The second set of robustness checks is intended to rule out potential alternative mechanisms through which the NREGS could affect crop choices.

Table 2.5: Effect of registration with the NREGS on risk index of crop portfolio

	(1)	(2)	(3)	(4)	(5)
D.NREGS registered (2007)	0.019^{+}	0.035**	0.034**	0.026*	
	(0.010)	(0.010)	(0.010)	(0.010)	
D.NREGS registered (2009/10)					0.007
					(0.006)
Rainfall (deviation) at baseline		0.206**	0.194**	0.205**	0.227***
		(0.055)	(0.055)	(0.061)	(0.056)
Risk index at baseline		-0.515***	-0.572***	-0.477***	-0.500***
		(0.062)	(0.059)	(0.073)	(0.061)
Risk index × Rainfall (deviation)		-0.343*	-0.276*	-0.373*	-0.342*
		(0.129)	(0.128)	(0.146)	(0.136)
Observations	1088	1088	839	1088	1088
R^2	0.057	0.414	0.459	0.387	0.395

Notes: Estimation in first differences. Dep. var.: Risk index of a household's crop portfolio. Cols. (1), (2) & (5) present results for the full sample without matching. Col. (3) restricts the sample to households who have registered with the NREGS by 2009/10. Col. (4) matches households based on baseline characteristics. Additional controls are variable inputs (log), area cultivated (log), share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, annual income, off-farm activities (log), if household benefits from credit/training program, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Standard errors (clustered at the sub-district) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

As a first robustness check, I test whether households that registered with the NREGS change their input allocation more strongly than households who are not registered with the NREGS. To account for potential self-selection bias, I match households on their probability to register with the NREGS by using entropy balancing, a method developed by Hainmueller (2012).[42] Table 2.5 reports the effects of registering with the NREGS on the risk index of households' crop portfolios. I find that households that already registered with the NREGS in 2007 are more likely to grow a higher share of risky crops in the follow-up period. Five different specifications are presented: Columns (1) and (2) show estimates without matching, where column (2) additionally controls for initial conditions. Column (3) excludes all households that did not register with the NREGS by 2009/10.[43] Column (4) shows estimates for the matched sample. As we can see, the effects are only slightly smaller when matching households on their probability to register with the NREGS. Overall, the effects are of a similar size in most specifications though somewhat lower than the estimates presented in Table 2.4, column (2). Column (5) shows the estimation results for the full sample without matching. Here, being registered by 2009/10 is the main explanatory variable. As we would expect, households that registered with the NREGS only shortly before or even after deciding on their crop portfolio, did not alter their input allocation in a meaningful way.

As mentioned before, the NREGS can affect household decisions via different mechanisms. The next set of robustness checks seeks to understand if the observed effect of the NREGS is indeed an insurance effect and not due to alternative mechanisms such as the increase in income of participating households or the change in agricultural wages. If, for example, risky crops are also more capital intensive, then observed outcomes could also be driven by increases in income and wealth or better access to credit of participating households. Likewise, if risky crops are also less labor intensive, then observed outcomes could be driven by wage changes due to the NREGS instead of its insurance effect. Finally, the observed effect in the risk index could be due to an overall change in the production strategy, where the observed change in the risk index is due to a higher or lower diversification of the household's crop portfolio not due to a switch towards riskier crops.

[42] More details on the matching strategy can be found in the Appendix, Section 2.C.

[43] This is to exclude all households from the sample that - either because they consider it socially undesirable or because they have other means of risk coping - would probably never register with the NREGS.

Table 2.6: Effect of NREGS on labor intensity, cost intensity and crop specialization

	Labor intensity		Cost intensity		Concentration	
	(1)	(2)	(3)	(4)	(5)	(6)
D.NREGS introduced in district	0.011*		1766.336+		0.024	
	(0.004)		(997.043)		(0.031)	
D.NREGS registered (2007)		0.003		774.037		0.004
		(0.005)		(523.065)		(0.021)
D.Area cultivated (acres, log)	-0.021***	-0.020**	-1574.977**	-1472.847*	-0.133***	-0.132***
	(0.005)	(0.005)	(538.583)	(564.904)	(0.020)	(0.020)
D.Irrigated area (% of total)	-0.009	-0.008	46.274	81.716	-0.108**	-0.107**
	(0.008)	(0.008)	(863.949)	(866.096)	(0.029)	(0.029)
D.Fertilizer (dummy)	0.008	0.010	2134.392+	2366.757*	0.008	0.012
	(0.009)	(0.009)	(1056.735)	(950.514)	(0.025)	(0.021)
D.HYV seeds (dummy)	-0.002	-0.002	666.174	667.709	-0.033*	-0.033*
	(0.006)	(0.006)	(547.474)	(552.124)	(0.015)	(0.014)
D.Participated in labor sharing (dummy)	-0.005	-0.005	-764.135	-830.174	0.004	0.003
	(0.006)	(0.006)	(514.038)	(525.498)	(0.015)	(0.015)
D.Hh benefits from credit/training program	0.004	0.004	123.802	127.960	0.004	0.005
	(0.003)	(0.003)	(363.984)	(368.670)	(0.011)	(0.011)
D.Self-reported shock	0.004	0.005	345.951	444.132	-0.009	-0.008
	(0.003)	(0.003)	(424.723)	(418.309)	(0.011)	(0.011)
D.Rainfall (deviation)	0.002	-0.001	604.210	253.652	0.011	0.003
	(0.007)	(0.007)	(936.822)	(901.360)	(0.026)	(0.024)
D.Rainfall (deviation, lag)	-0.010	-0.005	-715.053	95.161	-0.032	-0.018
	(0.007)	(0.008)	(1169.059)	(941.590)	(0.036)	(0.031)
Constant	-0.001	0.001	-2.115	354.866	-0.009	-0.003
	(0.003)	(0.004)	(812.034)	(844.377)	(0.018)	(0.018)
Observations	1012	1012	1012	1012	1088	1088
R^2	0.033	0.030	0.028	0.024	0.099	0.098

Notes: Estimation in first differences. Dependent variable in cols. (1) & (2) is labor intensity of crop portfolio, in cols. (3) & (4) cost intensity of crop portfolio, and in cols. (5) & (6) Herfindahl index of crop portfolio. Variable definitions and sources are described in the Appendix, Section B. Standard errors (clustered at the district in cols. (1), (3) and (5), and clustered at the sub-district in cols. (2), (4) and (6)) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

I start by testing if the NREGS has effects on the labor intensity, cost intensity or degree of concentration of households' crop portfolios (c.f. Table 2.6). Labor intensity per crop is calculated as the share of expenditures on labor in total production costs. Cost intensity is defined as the total production cost that has to be incurred per hectare for each crop.[44] The amount if concentration is captured by the Herfindahl index of each household's share of inputs allocated to different crops. The coefficient on labor intensity is positive, indicating that the NREGS, if anything, increases the labor intensity of crop portfolios (c.f. cols. (1) and (2)). The coefficient on cost intensity is also positive, suggesting that households are able to spend more on their agricultural production. However, only one out of two specifications is statistically significant at the 10% level (c.f. cols (3) and (4)). This suggests, that the NREGS acts through the insurance effect more than through the wage or income mechanism. The effect of the NREGS on concentration in the crop portfolio is positive but not statistically significant (c.f. cols. (5) and (6)). A positive effect means that households with access to the NREGS tend to further specialize in their crop choices, and not diversify their portfolio. This suggests that the observed change in the risk index indeed reflects a greater amount of risk taking in agricultural production.

The presence of alternative mechanisms through which the NREGS could affect production decisions also means that households might register with the NREGS for different reasons. For some households, consumption needs are a much more important reason for registering with the program than the insurance effect. These households would need to work for the NREGS as much as possible to satisfy their consumption needs - even in good years, and are not likely to cultivate higher risk crops despite working for the NREGS. Other households might already have access to alternative risk coping mechanisms, and do not need the NREGS as risk management strategy. We would thus expect households to react differently to the availability of the NREGS depending on whether the program can contribute to smoothing their incomes in the case of a shock. In Table 2.7, columns (1) and (2) I show that households who registered with the program in 2007 while experiencing a shock to agricultural production (i.e. a rainfall shock), adjust their production portfolio, while households who registered with the NREGS despite experiencing favorable rainfall levels did not alter their production

[44] Both measures are based on the crop-wise Cost of Cultivation Statistics, published by the Ministry of Agriculture. See Appendix, Section 2.B.2 for more details.

Table 2.7: Interaction with previously existing programs and rainfall

	(1)	(2)	(3)	(4)	(5)
NREGS introduced in district	0.069**		0.087**	0.076**	0.083**
	(0.016)		(0.015)	(0.016)	(0.018)
NREGS × Rainfall (deviation, lag)	-0.022				
	(0.033)				
NREGS registered (2007)		0.042***			
		(0.010)			
NREGS × Rainfall (deviation, lag)		-0.040^{+}			
		(0.020)			
NREGS × Crop insurance			-0.033		
			(0.029)		
Crop insurance			-0.013		
			(0.022)		
NREGS × Watershed dev.				-0.029	
				(0.022)	
Watershed dev.				0.005	
				(0.006)	
NREGS × Public works					-0.021
					(0.015)
Public works					0.008
					(0.007)
Controls: Rainfall and risk index at baseline	Yes	Yes	Yes	Yes	Yes
Observations	1088	1088	1084	1084	1084
R^2	0.416	0.391	0.440	0.422	0.418

Notes: Estimation in first differences. Dep. var.: Risk index of a household's crop portfolio. Expl. var. in columns (1) and (3) to (5) is NREGS introduced in district; expl. variable in col. (2) is NREGS registered in 2007. Additional controls are variable inputs (log), area cultivated (log), share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, annual income, off-farm activities (log), if household benefits from credit/training program, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Standard errors (clustered at the district in cols. (1), (3), (4) and (5) and at the sub-district in col. (2)) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

decisions.[45] This suggests that households who registered with the NREGS to cope with a shock are much more likely to adjust their input allocation towards more profitable crops, which is exactly what we would expect in case of an insurance effect. An alternative strategy to separate households along their motivation to register with the NREGS is to condition the treatment effect on the initial presence of other government programs, such as watershed development projects, crop insurance schemes or public works programs other than the NREGS. Columns (3) to (5) of Table 2.7 show that treatment effects are smaller in villages with existing watershed development projects, crop insurance programs and public works schemes, although none of the coefficients is statistically significant at the 10% level. These results again support the hypothesis that the NREGS has an insurance function for households because observed effects on input allocation are smaller if households already have access to other insurance or risk mitigation mechanisms.

2.6 Conclusions

This paper presents theoretical and empirical evidence that an employment guarantee, such as the NREGS in India, improves households' ability to cope with shocks in agriculture by guaranteeing income opportunities in areas where and time periods when they previously did not exist. By improving the risk management of households, the NREGS enables households to switch their production towards riskier but also higher profitability products and to generate higher incomes from agricultural production.

The results of this paper show that public works programs can have welfare effects that go beyond immediate income effects. The insurance effect of the NREGS on agricultural productivity is similar to the effects of rainfall insurance analyzed by Cole et al. (2013), Mobarak and Rosenzweig (2013), and Karlan et al. (2014). But in contrast to purchasing insurance, registration with the NREGS provides little ex ante cost to these households. Since trust-related considerations continue to limit the uptake of insurance products in many countries, providing public works schemes - combined with an employment guarantee - could be an alternative option with which to protect

[45] For better visualization, the marginal effect of registering with the NREGS conditional on lagged rainfall is plotted in Figure 2.D.4 in the Appendix.

households against agricultural production risks and to enable productivity gains in agriculture.

Current discussions regarding the effects of the NREGS on agricultural productivity focus mainly on the trade-off between providing minimum income to poor households, on one hand, and ensuring that production costs in the agricultural sector do not rise too drastically due to increased agricultural wages, on the other hand. As this paper shows, these discussions have failed to consider the following key aspect: because the number of workdays each household is entitled to additionally affects its risk management capacity, the amount of risk each household is willing to take in his own agricultural production - and therewith potential productivity gains - crucially depends on the number of days each household can expect to be able to work in the case of production shocks. Thus, increasing the number of days each household is entitled to work with the NREGS could increase agricultural productivity - an argument that has been largely ignored so far. The assumption that only large-scale farmers can raise agricultural productivity is still a mainstream one. Including in the discussion the effects of the NREGS on households' risk management and the resulting changes in production decisions might change the overall picture.

The findings here contain some lessons for the ongoing debates on the effectiveness of the NREGS and for other countries considering the implementation of such schemes. First, for the insurance effect to unfold, the design of a public works program is crucial. An employment guarantee that is entitled by law and entails adequate grievance redress mechanisms provides households with the necessary protection against agricultural production risks to enable them to take more risks in their production and investment decisions. Additionally, it is crucial not to severely limit the number of workdays, otherwise such a scheme's potential as a risk-coping instrument cannot be realized. Second, implementation matters. The data analyzed in this paper cover only the state of Andhra Pradesh. This is, *inter alia*, because the performance of the NREGS in terms of the number of workdays generated per eligible household varies immensely across states and even across districts in India. Andhra Pradesh is one of the best performing states in the implementation of the NREGS, so it goes without saying that many of the effects captured in this paper might not be found in all Indian states. Third, working for a public works scheme is always associated with opportunity costs.

In countries or regions with well functioning off-farm labor markets, providing public works schemes might not be necessary. A food-for-work program or cash-for-work program is only effective in areas and time periods where labor is in surplus.

2.A Mathematical Appendix

2.A.1 Deterministic Case

In the deterministic case, the Lagrange can be summarized as follows:

$$\begin{aligned}\mathcal{L} =& U_1(C_1) + \delta U_2(C_2) \\ &+ \lambda(w_1(T_1 - l_1^d - l_1^s) - g(i^d + i^s) + B - C_1) \\ &+ \mu[(p - \alpha w_2^r)(Q^d + Q^s) + w_2 T_2 - (1+r)B - C_2] \\ &+ \varphi(B^m - B) \\ &+ \rho(1 - a^d - a^s)\end{aligned}$$

Differentiating the Lagrange with respect to the choice variables, leads to the following first order conditions:[46]

$$\frac{\partial \mathcal{L}}{\partial C_1} = \frac{\partial U_1}{\partial C_1} - \lambda = 0 \tag{2.A.1}$$

$$\frac{\partial \mathcal{L}}{\partial C_2} = \delta \frac{\partial U_2}{\partial C_2} - \mu = 0 \tag{2.A.2}$$

$$\frac{\partial \mathcal{L}}{\partial l_1^d} = -\lambda w_1 + \mu(p - \alpha w_2)\frac{\partial f^d}{\partial l_1^d} = 0 \tag{2.A.3}$$

$$\frac{\partial \mathcal{L}}{\partial l_1^s} = -\lambda w_1 + \mu(p - \alpha w_2)\frac{\partial f^s}{\partial l_1^s} = 0 \tag{2.A.4}$$

$$\frac{\partial \mathcal{L}}{\partial i^d} = -\lambda g + \mu(p - \alpha w_2^r)\frac{\partial f^d}{\partial i^d} = 0 \tag{2.A.5}$$

$$\frac{\partial \mathcal{L}}{\partial i^s} = -\lambda g + \mu(p - \alpha w_2^r)\frac{\partial f^s}{\partial i^s} = 0 \tag{2.A.6}$$

$$\frac{\partial \mathcal{L}}{\partial a^d} = \mu(p - \alpha w_2^r)\frac{\partial f^d}{\partial a^d} - \gamma = 0 \tag{2.A.7}$$

$$\frac{\partial \mathcal{L}}{\partial a^s} = \mu(p - \alpha w_2^r)\frac{\partial f^s}{\partial a^s} - \gamma = 0 \tag{2.A.8}$$

$$\frac{\partial \mathcal{L}}{\partial B} = \lambda - \mu(1+r) - \varphi = 0 \tag{2.A.9}$$

[46] Remember that $Q^d = f^d(a^d, l_1^d, i^d)$ and $Q^s = f^s(a^s, l_1^s, i^s)$.

Rearranging the first order conditions (2.A.1) and (2.A.2) gives:

$$\lambda = \frac{\partial U_1}{\partial C_1} \tag{2.A.10}$$

$$\mu = \delta \frac{\partial U_2}{\partial C_2} \tag{2.A.11}$$

And including (2.A.10) and (2.A.11) into (2.A.3)-(2.A.9) gives our decision rules:

$$w_1 \frac{\partial U_1}{\partial C_1} - (p - \alpha w_2^r)\delta \frac{\partial U_2}{\partial C_2}\frac{\partial f^d}{\partial l_1^d} = 0 \Leftrightarrow \frac{\partial f^d}{\partial l_1^d} = \frac{w_1}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial U_2}{\partial C_2}} \tag{2.A.12}$$

$$w_1 \frac{\partial U_1}{\partial C_1} - (p - \alpha w_2^r)\delta \frac{\partial U_2}{\partial C_2}\frac{\partial f^s}{\partial l_1^s} = 0 \Leftrightarrow \frac{\partial f^s}{\partial l_1^s} = \frac{w_1}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial U_2}{\partial C_2}} \tag{2.A.13}$$

$$g \frac{\partial U_1}{\partial C_1} - (p - \alpha w_2^r)\delta \frac{\partial U_2}{\partial C_2}\frac{\partial f^d}{\partial i^d} = 0 \Leftrightarrow \frac{\partial f^d}{\partial i^d} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial U_2}{\partial C_2}} \tag{2.A.14}$$

$$g \frac{\partial U_1}{\partial C_1} - (p - \alpha w_2^r)\delta \frac{\partial U_2}{\partial C_2}\frac{\partial f^s}{\partial i^s} = 0 \Leftrightarrow \frac{\partial f^s}{\partial i^s} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial U_2}{\partial C_2}} \tag{2.A.15}$$

$$\frac{\partial f^d}{\partial a^d} = \frac{\partial f^s}{\partial a^s} \tag{2.A.16}$$

$$\varphi = \frac{\partial U_1}{\partial C_1} - \delta(1+r)\frac{\partial U_2}{\partial C_2} \tag{2.A.17}$$

Equation (2.A.17) can be rewritten to describe the optimal consumption rule over both periods given credit constraints:

$$\frac{\partial U_1}{\partial C_1} = \delta(1+r)\frac{\partial U_2}{\partial C_2} + \varphi \tag{2.A.18}$$

If the credit constraint is binding, φ is greater than zero and the marginal utility from consumption in the planting period greater than the discounted marginal utility from consumption in the harvesting period. This means that consumption in the planting stage is lower than what could be achieved if the credit constraints were not binding. Including equation (2.A.18) into equation (2.A.14) also reveals the effect of the credit constraint on input allocation:

$$\frac{\partial f^d}{\partial k^d} = \frac{g(1+r)}{(p - \alpha w_2^r)} + \frac{g\varphi}{(p - \alpha w_2^r)\delta \frac{\partial U_2}{\partial C_2}} \tag{2.A.19}$$

If the credit constraint is not binding, $\varphi = 0$, the marginal product of input allocation is lower and input allocation higher. The same effect holds for input allocation to the stochastic crop Q^s, as well as for labor allocation to each of the crops.

2.A.2 Stochastic Case

When introducing uncertainty, the Lagrange becomes the following:

$$\begin{aligned}\mathcal{L} =& U_1(C_1) + \lambda(w_1(T_1 - l_1^d - l_1^s) - g(i^d + i^s) + B - C_1)) \\ &+ E[\delta U_2(C_2) + \mu[(p - \alpha w_2^r)(Q^d + Q^s) + w_2 T_2 - (1+r)B - C_2]] \\ &+ \varphi(B^m - B) \\ &+ \rho(1 - a^d - a^s)\end{aligned}$$

Note here that the household forms expectations not only about the utility he derives from consumption in period 2, but also about the level of consumption that can be achieved. Differentiating the Lagrange with respect to the choice variables, leads to the following first order conditions:[47]

$$\frac{\partial \mathcal{L}}{\partial C_1} = \frac{\partial U_1}{\partial C_1} - \lambda = 0 \tag{2.A.20}$$

$$\frac{\partial \mathcal{L}}{\partial C_2} = E[\delta \frac{\partial U_2}{\partial C_2} - \mu] = 0 \tag{2.A.21}$$

$$\frac{\partial \mathcal{L}}{\partial l_1^d} = -\lambda w_1 + E[\mu](p - \alpha w_2^r)\frac{\partial f^d}{\partial l_1^d} = 0 \tag{2.A.22}$$

$$\frac{\partial \mathcal{L}}{\partial l_1^s} = -\lambda w_1 + E[\mu(p - \alpha w_2^r)\epsilon \frac{\partial f^s}{\partial l_1^s}] = 0 \tag{2.A.23}$$

$$\frac{\partial \mathcal{L}}{\partial i^d} = -\lambda g + E[\mu](p - \alpha w_2^r)\frac{\partial f^d}{\partial i^d} = 0 \tag{2.A.24}$$

$$\frac{\partial \mathcal{L}}{\partial i^s} = -\lambda g + E[\mu(p - \alpha w_2^r)\epsilon \frac{\partial f^s}{\partial i^s}] = 0 \tag{2.A.25}$$

$$\frac{\partial \mathcal{L}}{\partial a^d} = E[\mu](p - \alpha w_2^r)\frac{\partial f^d}{\partial a^d} - \gamma = 0 \tag{2.A.26}$$

$$\frac{\partial \mathcal{L}}{\partial a^s} = E[\mu(p - \alpha w_2^r)\epsilon \frac{\partial f^s}{\partial a^s}] - \gamma = 0 \tag{2.A.27}$$

$$\frac{\partial \mathcal{L}}{\partial B} = \lambda - E[\mu](1+r) - \varphi = 0 \tag{2.A.28}$$

[47] Remember that $Q^d = f^d(a^d, l_1^d, i^d)$ and $Q^s = \epsilon f^s(a^s, l_1^s, i^s)$.

Rearranging (2.A.20) and (2.A.21) gives:

$$\lambda = \frac{\partial U_1}{\partial C_1} \tag{2.A.29}$$

$$E[\mu] = \delta\frac{\partial EU_2}{\partial C_2} \tag{2.A.30}$$

And the optimal consumption rule becomes:

$$\frac{\partial U_1}{\partial C_1} = (1+r)\delta\frac{\partial EU_2}{\partial C_2} + \varphi. \tag{2.A.31}$$

The consumption rule - equation (2.A.31) - changes slightly when introducing uncertainty because for any expected consumption level C_2, expected utility $EU_2(C_2)$ is lower than the utility of the expected value $U_2(E(C_2))$, and marginal expected utility is higher than the marginal utility of the expected value. Since all other variables remain constant, C_2 has to be higher relative to C_1 under uncertainty for the identity to hold. This is equivalent with the well-known argument that risk decreases current consumption levels and enhances savings.

Including (2.A.29) and (2.A.30) into (2.A.22)-(2.A.27) gives our decision rules for l_1^d,

$$\begin{aligned} & w_1\frac{\partial U_1}{\partial C_1} - (p-\alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}\frac{\partial f^d}{\partial l_1^d} = 0 \\ & \Leftrightarrow \frac{\partial f^d}{\partial l_1^d} = \frac{w_1}{(p-\alpha w_2^r)}\frac{\frac{\partial U_1}{\partial C_1}}{\delta\frac{\partial EU_2}{\partial C_2}} \end{aligned} \tag{2.A.32}$$

for l_1^s,

$$\begin{aligned} & w_1\frac{\partial U_1}{\partial C_1} - (p-\alpha w_2^r)\frac{\partial f^s}{\partial l_1^s}\delta E[\frac{\partial U_2}{\partial C_2}\epsilon] = 0 \\ & \Leftrightarrow (p-\alpha w_2^r)\frac{\partial f^s}{\partial l_1^s}\delta[\frac{\partial EU_2}{\partial C_2}E[\epsilon] + cov(\frac{\partial U_2}{\partial C_2},\epsilon)] = w_1\frac{\partial U_1}{\partial C_1} \\ & \Leftrightarrow \frac{\partial f^s}{\partial l_1^s} = \frac{w_1}{(p-\alpha w_2^r)}\frac{\frac{\partial U_1}{\partial C_1}}{\delta\frac{\partial EU_2}{\partial C_2}} - \frac{cov(\frac{\partial U_2}{\partial C_2},\epsilon)}{(p-\alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}} \end{aligned} \tag{2.A.33}$$

for i^d,

$$g\frac{\partial U_1}{\partial C_1} - (p - \alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}\frac{\partial f^d}{\partial i^d} = 0$$
$$\Leftrightarrow \frac{\partial f^d}{\partial i^d} = \frac{g}{(p - \alpha w_2^r)}\frac{\frac{\partial U_1}{\partial C_1}}{\delta\frac{\partial EU_2}{\partial C_2}} \tag{2.A.34}$$

for i^s,

$$g\frac{\partial U_1}{\partial C_1} - (p - \alpha w_2^r)\frac{\partial Q^s}{\partial i^s}\delta E[\frac{\partial U_2}{\partial C_2}\epsilon] = 0$$
$$\Leftrightarrow (p - \alpha w_2^r)\frac{\partial f^s}{\partial i^s}\delta[\frac{\partial EU_2}{\partial C_2}E[\epsilon] + cov(\frac{\partial U_2}{\partial C_2}, \epsilon)] = g\frac{\partial U_1}{\partial C_1}$$
$$\Leftrightarrow \frac{\partial f^s}{\partial i^s} = \frac{g}{(p - \alpha w_2^r)}\frac{\frac{\partial U_1}{\partial C_1}}{\delta\frac{\partial EU_2}{\partial C_2}} - \frac{cov(\frac{\partial U_2}{\partial C_2}, \epsilon)}{(p - \alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}} \tag{2.A.35}$$

for a^d,

$$\delta\frac{\partial EU_2}{\partial C_2}(p - \alpha w_2^r)\frac{\partial f^d}{\partial a^d} = \gamma$$

and a^s,

$$(p - \alpha w_2^r)\frac{\partial f^s}{\partial a^s}\delta E[\frac{\partial U_2}{\partial C_2}\epsilon] = \gamma$$
$$\Leftrightarrow (p - \alpha w_2^r)\frac{\partial f^s}{\partial a^s}\delta\frac{\partial EU_2}{\partial C_2}E[\epsilon] + cov(\frac{\partial U_2}{\partial C_2}, \epsilon) = \gamma$$

resulting in:

$$\frac{\partial f^s}{\partial a^s} = \frac{\partial f^d}{\partial a^d} - \frac{cov(\frac{\partial U_2}{\partial C_2}, \epsilon)}{(p - \alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}}. \tag{2.A.36}$$

The decision rules can be reformulated to include the credit constraint. Then, input allocation to the risky crop is determined as follows:

$$\frac{\partial f^s}{\partial k^s} = \frac{g(1+r)}{(p - \alpha w_2^r)} + \frac{g\varphi}{(p - \alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}} - \frac{cov(\frac{\partial U_2}{\partial C_2}, \epsilon)}{(p - \alpha w_2^r)\delta\frac{\partial EU_2}{\partial C_2}} \tag{2.A.37}$$

We can see from equation (2.A.37) that both risk and credit constraints go in the same direction and reduce the input allocation to the risky crop. More importantly, it also shows that uncertainty reduces input allocation to the risky crop relative to the deterministic crop even if credit constraints are not binding.

2.B Data Description

2.B.1 Young Lives Survey

- Reference periods: In most questions the references period of the YLS are the 12 months prior to the date of interview. However, for all questions on agricultural production, the period of reference is a particular agricultural year. In 2007, the reference period was the agricultural year 2005/06, thus May 2005 to April 2006. In 2009/10, the reference period was the agricultural year 2008/09.

- Wealth index: The wealth index is calculated as a simple average of housing quality, consumer durables and services. Housing quality is the simple average of rooms per person and indicator variables for the quality of roof, walls and floor. Consumer durables are the scaled sum of 12 variables indicating the ownership of items such as radios, fridges, televisions, phones or vehicles. Services are calculated as the simple average of dummy variables indicating households' access to drinking water, electricity, toilets and fuels. For more information on the wealth index refer to the Young Lives data justification documents at http://www.younglives.org.uk.

2.B.2 Crop production

In this paper, the agricultural year refers to the period May to April. Agricultural production in India generally takes place over two seasons: the rainy (Kharif) and the dry (Rabi) season. Most agricultural output is produced during the rainy season, which, in Andhra Pradesh, lasts roughly from June to September. Planting of major crops such a rice and cotton starts in May and needs to be completed before end of July. The most important input allocation decision thus takes place around May and June of every year, which is before the monsoon's rainfall is fully observed.

- Risk index of major crops: The riskiness of crops is calculated from crop- and district-wise yield data in the six survey districts over the period 1998/99 to 2011/12. The data were obtained from the District-wise crop production statistics, Directorate of Economics and Statistics, Ministry of Agriculture, GoI, and are available online: $http://apy.dacnet.nic.in$.

This data is available for 26 crops, which represents about 90% of the crop production in the YLS sample. The risk index for household i given input allocation k to crop m is defined as $R_i = \sum r_m k_m / \sum k_n$, where r_m is the coefficient of variation of the yield of crop m. Note here, that r_m is only available for a subset of all crops n, such that $m \subseteq n$. The way in which I treat these missing crops could potentially affect my results. In all results, I implicitly treat crops with missing risk data as having a risk measure of zero, which obviously biases my results. To reduce this bias, I drop all observations from the sample which have no crop in their portfolio for which risk information is available, e.g. $\sum k_m = 0$ or $R_i = 0$, in one or both of the survey rounds.

In order to gauge the robustness of my results, I recalculate the main results using a range of alternative risk measures, see Table 2.D.1. In columns (1) and (2), I use the standard deviation of returns per hectare as risk measure for each crop. In columns (3) and (4), I first remove a linear time trend and district-level differences in average productivity from the yield data and then compute the standard deviation of the residual. This measure is then divided by the crop's average yield such that the data is on a scale between 0 and 1. For columns (5) and (6), I compute a risk measure that takes into account only those crops for which information is available, e.g. $R_i^{alt} = \sum r_m k_m / \sum k_m$. Here r_m is again the coefficient of variation of the yield of crop m. And finally, columns (7) and (8) report the main results using the risk index described initially. The results do not change when using alternative risk measures.

To calculate the risk-index in district-level land use (Figure 2.1), I merge this information with the district wise land use statistics, which are also available from the Directorate of Economics and Statistics, Ministry of Agriculture, GoI. The risk index is calculated as follows: $R_{jt} = \sum r_m a_{mjt} / \sum a_{mjt}$, where a_{mjt} is the land allocated to crop m in district j at time t and r_m is the coefficient of variation of crop m.

- Cost and Labor intensity: The cost and labor intensity of crops is calculated from the cost of cultivation statistics for Andhra Pradesh from 1995/96 to 2009/10.The data were obtained from the Directorate of Economics and Statistics, Ministry of Agriculture, GoI, and are available online: *http://eands.dacnet.nic.in*.

This data is available for 11 crops, which represents about 80% of the crop production in the YLS sample. I calculate the cost intensity for each crop c_m as the average production cost per hectare indicated by the data. The cost intensity index per household is the $C_i = \sum c_m k_m / \sum k_n$, where k_m are the inputs allocated to crop m. The labor intensity is calculated as the share of labor cost in total production cost as indicated by the same data. The aggregation method is also the same: $L_i = \sum l_m k_m / \sum k_n$. Again, I drop all observations with $\sum k_m = 0$ in one or both of the survey rounds.

- Standard deviation in returns: Standard deviation in returns is calculated as the weighted average of each crop's standard deviation in returns per hectare, as reported in the cost of cultivation statistics for Andhra Pradesh from 1995/96 to 2009/10. The standard deviation is calculated as the standard deviation over all years for which the cost of cultivation statistics provides data. The data were obtained from the Directorate of Economics and Statistics, Ministry of Agriculture, GoI, and are available online: *http : //eands.dacnet.nic.in*.

2.B.3 Rainfall data

The rainfall data used in this paper were compiled by the Directorate of Economics and Statistics, Government of Andhra Pradesh. Rainfall data are available at the block level for the years 2002 to 2011. Rainfall deviation and rainfall deviation (lag) describe the relative deviation of cumulative rainfall over the agricultural year (May - April) from the long-term average, e.g. $devrain^{05/06} = (rf^{05/06} - \overline{rf})/\overline{rf}$. For the 2007 round of interviews, current rainfall uses the 2005/06 rainfall, and lagged rainfall uses rainfall in the agricultural year 2004/05. For the 2009/10 round of interviews, current rainfall uses the rainfall in the agricultural year 2008/09, and lagged rainfall uses data from the agricultural year 2007/08.

2.B.4 NREGS data

The implementation of the NREGS was intended prioritize India's 200 poorest districts, subsequently extending to the remaining districts. India has a total of 655 districts, of which 625 had introduced the NREGS as of 2008. The 30 remaining district were urban districts. In 2003 the Planning Commission of India elaborated clear rules stating which districts should be included in which round of implementation of the NREGS. However, the process of district selection was influenced by political considerations due to the huge size and financial relevance of this program and the rules elaborated by the Planning Commission were not strictly followed.

- NREGS introduced in District: This variable is an indicator which equals 1 if a household has access to the NREGS at the district level at the beginning of the agricultural cycle. Since the period of reference for the 2007 round of interviews is the agricultural year 2005/06 (May 2005 to April 2006) and the introduction of the NREGS started in April 2006, D_{ijt} equals 0 for all households in the baseline. The period of reference for the 2009/10 interviews is the agricultural year 2008/09. By that time, NREGS works had started in the districts Anantapur, Cuddapah, Karimnagar and Mahaboobnagar. In Srikakulam and West Godavari the introduction of the NREGS was in August 2007 and in March 2008 respectively. Since activities started only very slowly in Srikakulam, we treat this district as control district despite the introduction of the NREGS mid 2007.

- Treatment intensity, NREGS: Cumulative expenditure and number of person-days of employment generated at the block level are used to capture the treatment intensity of the NREGS. The amount sanctioned per village depends on a village's list of projects, which has to be approved by the block program officer. The block program officer has to estimate employment demand for the following financial year and consolidate all village lists before submitting the Block Employment Guarantee Plan to the district program coordinator. The district council (zilla parishad) has to approve all plans before transferring them to the state government. Data are retrieved from Government of Andhra Pradesh, Department for Rural Development, *http://www.nrega.ap.gov.in*.

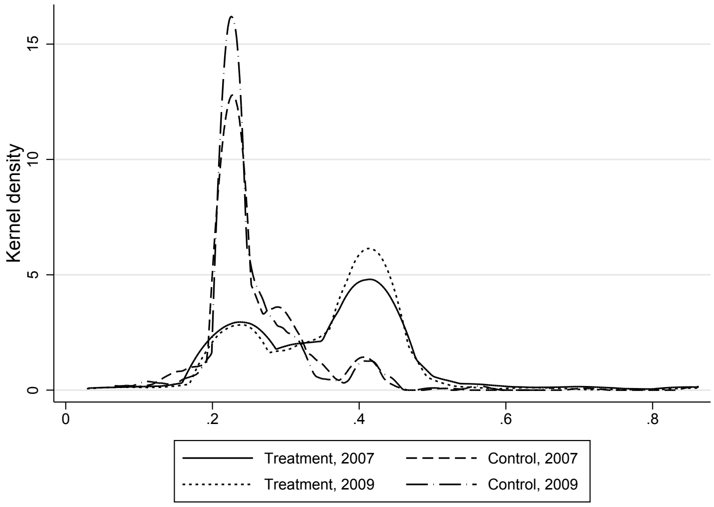
2.C Matching strategy

In this paper, I use entropy balancing as matching strategy. Entropy balancing seems to outperform most existing matching algorithms in terms of the balance reached on the entire set of relevant covariates (Hainmueller, 2012). The matching algorithm assigns weights to all observations in the control group such that the distribution of selected variables matches the observed distribution in the treatment group. These weights can then be used as sampling weights in the estimation. Since I estimate the model on a balanced sample, the same weights can be applied to the 2009/10 round of interviews.

I match households on the mean and the variance of variables that determine a household's registration with the NREGS and potentially influence post-treatment outcomes, such as cost incurred in agricultural production, total cultivated area, percentage of area irrigated, a dummy indicating whether a household participates in labor sharing in agriculture, wealth levels and off-farm income, and household characteristics, e.g. education, age and sex of the household head, indebtedness, and the ability to raise INR 1,000 (US$ 21.6) in one week. The resulting covariate balance is shown in Table 2.D.5. This method focusses on the covariate balance and less on the common support among the treatment and control group. In order to understand how both groups differ in terms of the selected variables, I estimate the propensity score for each household based on the selection variables described above, and plot its distribution in Figure 2.D.5. As can be seen, there is substantial overlap in the estimated propensity scores.

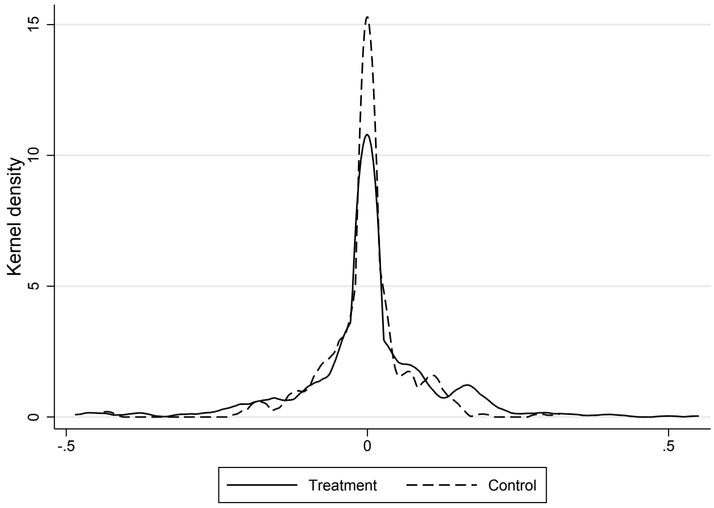
2.D Supplementary Figures and Tables

Figure 2.D.1: Distribution of risk-index



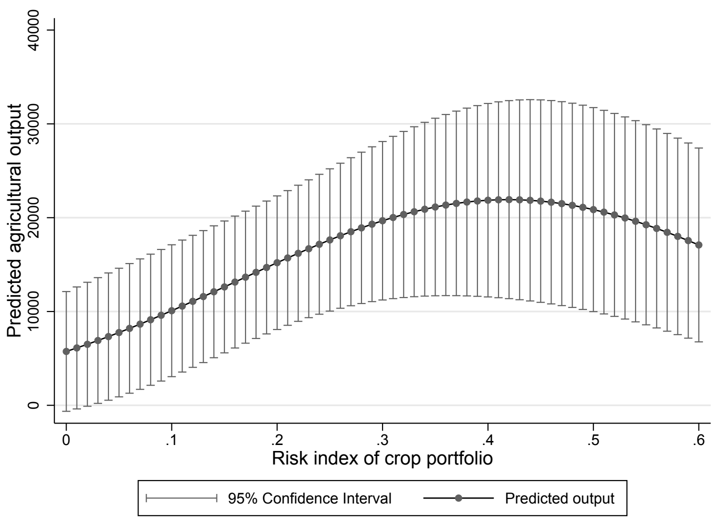
Source: Own estimation based on District-wise Crop Production Statistics, Ministry of Agriculture, GoI, and Young Lives data.

Figure 2.D.2: Distribution of change in risk index



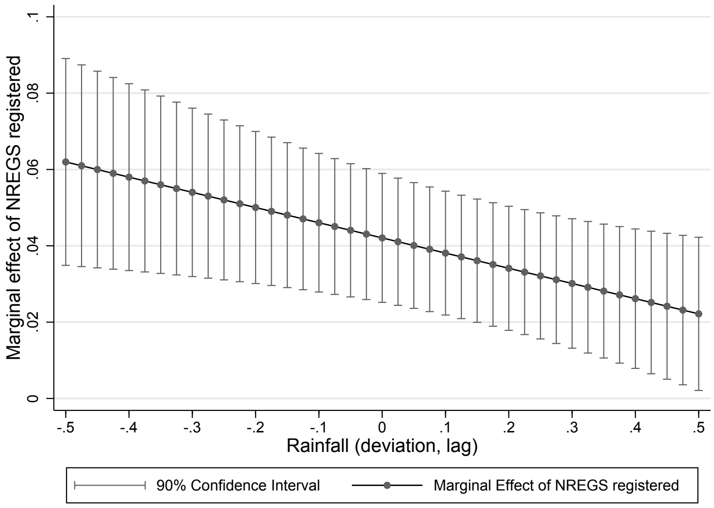
Source: Own estimation based on District-wise Crop Production Statistics, Ministry of Agriculture, GoI, and Young Lives data.

Figure 2.D.3: Agricultural output as function of the riskiness of crops



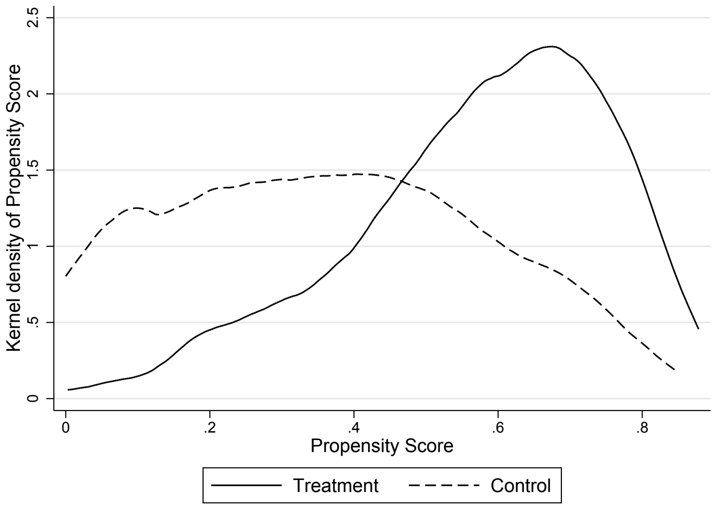
Source: Own estimation based on the Young Lives data.

Figure 2.D.4: Effect of the NREGS on risk index conditional on lagged rainfall



Source: Own estimation based on the Young Lives data.

Figure 2.D.5: Distribution of the propensity score



Source: Own estimation based on Young Lives data.

Table 2.D.1: Sensitivity of results to alternative dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D.NREGS introduced in district	380.968	371.945	0.027*	0.055**	0.029^{+}	0.059**	0.038*	0.072**
	(226.264)	(264.924)	(0.009)	(0.010)	(0.012)	(0.012)	(0.011)	(0.017)
D.Variable inputs (log)	309.106^{+}	321.650*	-0.013^{+}	-0.005	-0.014^{+}	-0.005	-0.016^{+}	-0.005
	(122.889)	(117.817)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)
D.Area cultivated (acres, log)	-320.843*	-342.916*	0.016**	0.010***	0.022***	0.016***	0.017***	0.011**
	(120.242)	(127.542)	(0.003)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
D.Irrigated area (% of total)	-106.320	-135.469	-0.021	-0.014	-0.021	-0.014	-0.033	-0.024
	(151.279)	(161.717)	(0.015)	(0.013)	(0.014)	(0.012)	(0.019)	(0.015)
D.Fertilizer (dummy)	-3.552	-14.517	-0.038	-0.026	-0.039	-0.027	-0.042	-0.028
	(205.011)	(188.155)	(0.027)	(0.014)	(0.032)	(0.018)	(0.030)	(0.016)
D.HYV seeds (dummy)	200.381	187.183^{+}	0.010^{+}	0.019*	0.007	0.016*	0.013^{+}	0.025**
	(106.243)	(86.849)	(0.005)	(0.006)	(0.007)	(0.004)	(0.005)	(0.006)
D.Participated in labor sharing (dummy)	-140.856*	-130.803*	-0.005^{+}	-0.003	-0.001	0.001	-0.003	-0.002
	(39.809)	(41.429)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
D.Annual income, off-farm activities (log)	-11.745	-13.340	-0.000	-0.002*	-0.002	-0.004**	-0.000	-0.003
	(11.255)	(11.648)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
D.Hh benefits from credit/training program	-3.626	-16.001	-0.008	-0.011**	-0.008	-0.010^{+}	-0.011	-0.013*
	(95.964)	(99.109)	(0.006)	(0.002)	(0.008)	(0.005)	(0.008)	(0.004)
D.Self-reported shock	143.001	145.515	0.009^{+}	0.003	0.006	-0.001	0.009^{+}	0.001
	(103.683)	(101.521)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
D.Rainfall (deviation)	-54.154	141.892	0.010	0.033	0.013	0.040^{+}	0.025**	0.050^{+}
	(139.597)	(268.821)	(0.006)	(0.020)	(0.009)	(0.019)	(0.006)	(0.025)
D.Rainfall (deviation, lag)	-341.779	-499.453	-0.035*	-0.036^{+}	-0.040*	-0.044*	-0.044**	-0.042^{+}
	(324.697)	(391.765)	(0.010)	(0.015)	(0.015)	(0.016)	(0.011)	(0.019)
Controls: Rainfall and risk index at baseline	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1012	1012	1088	1088	1088	1088	1088	1088
R^2	0.078	0.080	0.067	0.431	0.055	0.356	0.067	0.443

Notes: Estimation in first differences. Columns (1) and (2) use the standard deviation in returns per hectare as risk measure for the dependent variable. In columns (3) and (4), I first remove a linear time trend and district-level differences in average productivity from the yield data and then compute the standard deviation of the residual. This measure is then divided by the crop's average yield such that the data is on a scale between 0 and 1. For columns (5) and (6), I compute a risk measure that takes into account only those crops for which information is available, e.g.$R_i^{alt} = \sum r_m k_m / \sum k_m$. Columns (7) and (8) report the main results. Standard errors (clustered at the district) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.D.2: Robustness to inclusion of controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D.NREGS introduced in district	0.007	0.027^{+}	0.028^{+}	0.038^{*}	0.059^{+}	0.068^{*}	0.070^{**}	0.072^{**}
	(0.009)	(0.012)	(0.011)	(0.011)	(0.028)	(0.017)	(0.016)	(0.017)
D.Self-reported shock		0.007	0.007	0.009^{+}		-0.002	-0.002	0.001
		(0.004)	(0.004)	(0.004)		(0.003)	(0.004)	(0.004)
D.Rainfall (deviation)		0.026^{*}	0.027^{*}	0.025^{**}		0.047	0.052	0.050^{+}
		(0.007)	(0.007)	(0.006)		(0.031)	(0.030)	(0.025)
D.Rainfall (deviation, lag)		-0.036^{+}	-0.036^{+}	-0.044^{**}		-0.041	-0.045	-0.042^{+}
		(0.015)	(0.016)	(0.011)		(0.026)	(0.025)	(0.019)
D.Annual income, off-farm activities (log)			0.001	-0.000			-0.002	-0.003
			(0.002)	(0.002)			(0.002)	(0.001)
D.Hh benefits from credit/training program			-0.009	-0.011			-0.014^{*}	-0.013^{*}
			(0.007)	(0.008)			(0.005)	(0.004)
D.Variable inputs (log)				-0.016^{+}				-0.005
				(0.006)				(0.005)
D.Area cultivated (acres, log)				0.017^{***}				0.011^{**}
				(0.002)				(0.002)
D.Irrigated area (% of total)				-0.033				-0.024
				(0.019)				(0.015)
D.Fertilizer (dummy)				-0.042				-0.028
				(0.030)				(0.016)
D.HYV seeds (dummy)				0.013^{+}				0.025^{**}
				(0.005)				(0.006)
D.Participated in labor sharing (dummy)				-0.003				-0.002
				(0.002)				(0.002)
Controls: Rainfall and risk index at baseline	No	No	No	No	Yes	Yes	Yes	Yes
Observations	1088	1088	1088	1088	1088	1088	1088	1088
R^2	0.001	0.019	0.021	0.067	0.392	0.408	0.413	0.443

Notes: Estimation in first differences. Dep. var.: Risk index of a household's crop portfolio. Standard errors (clustered at the district) in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.D.3: District-level statistics

		Treatment	Control
GDP per capita in INR (2006/07)		783,487	776,179
Rural population (2001 census)		80.54	84.64
SC/ST population (2001 census)		20.50	18.36
Literacy rate (2001 census)		54.6	64.4
Cropping Intensity (2007/08)		1.238	1.505
Average wage rate of agric. laborers (2007)	Men	70.26	82.92
	Women	54.91	57.23

Source: Districts at a glance, Directorate of Economics & Statistics, Govt. of Andhra Pradesh.

Table 2.D.4: Evidence on mean reversion

	(1)	(2)
Risk index of crop portfolio	-0.608***	-0.220
	(0.024)	(0.353)
Rainfall (deviation)	0.033	-0.089
	(0.080)	(0.443)
Risk index of crop portfolio × Rainfall (deviation)	-0.241	0.484
	(0.276)	(3.167)
Risk index of crop portfolio (squared)		-0.488
		(0.415)
Risk index of crop portfolio (squared) × Rainfall (deviation)		-0.780
		(5.270)
Observations	338	338
R^2	0.404	0.422

Notes: Estimation in OLS. Dependent variable: $\triangle R = R_{t+1} - R_t$. Standard errors (clustered at the sub-district) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.D.5: Weighted summary statistics

	Treatment		Control			
			(not matched)		(matched)	
	Mean	SD	Mean	SD	Mean	SD
Value of variable inputs	12.81	(16.41)	15.88	(55.71)	14.10	(27.61)
Area cultivated (acres)	3.96	(4.53)	3.50	(5.19)	3.88	(3.85)
Irrigated area (% of total)	0.14	(0.29)	0.19	(0.33)	0.14	(0.28)
Participated in labor sharing (dummy)	0.79	(0.41)	0.73	(0.44)	0.79	(0.41)
Annual income, off-farm activities	23.68	(21.48)	22.76	(28.15)	24.45	(26.72)
Male household head	0.96	(0.19)	0.96	(0.20)	0.96	(0.19)
Age of household head	41.20	(12.07)	42.01	(12.01)	41.20	(11.79)
Household head is literate	0.32	(0.47)	0.28	(0.45)	0.32	(0.47)
Wealth index	0.37	(0.11)	0.40	(0.19)	0.37	(0.13)
Household size	6.02	(2.56)	5.88	(2.40)	6.02	(2.51)
Able to raise 1000 rupees in one week	0.56	(0.50)	0.49	(0.50)	0.56	(0.50)
Any serious debts	0.67	(0.47)	0.50	(0.50)	0.67	(0.47)
Observations	496		592		592	

Notes: Data from 2007 round of interviews. All values in constant INR 1,000 (July 2006). One US$ is equivalent to 46.38 INR (July 2006). Variable definitions and sources are described in the Appendix, Section 2.B.

3

Do cows have negative returns? The evidence revisited

with Michael Grimm

3.1 Introduction

Next to the cultivation of food and cash crops, livestock farming is one of the most important activities of rural households in developing countries. It is widely seen as a profitable activity, and hence it is supported by many agricultural policy interventions (see e.g. Swanepoel et al., 2010). The cow should not only supply milk, which can be an important source of nutrition and income to families, but also manure, which is a source of fertilizer for crops and biofuel for cooking (Hoddinott et al., 2014; Mdoe and Wiggins, 1997). Livestock farming, or better livestock accumulation, is also often seen as a reliable savings device for poor households with limited access to formal banking; in particular in a context in which high inflation rapidly dilutes financial assets (Rosenzweig and Wolpin, 1993). Based on such considerations, the government of Rwanda approved, for example, the "One cow per poor family program" in 2006 (Protos et al., 2011).

The common belief in the profitability of cows has recently been shaken by a paper of Anagol et al. (2014).[1] In that paper, the authors estimate that the median annual

[1]The Working Paper was first published in 2013. Throughout the text, we cite the updated version published in December 2014.

return to owning a dairy animal in northern rural India (Uttar Pradesh) is negative by 65% for buffaloes and by 293% for cows (not including the opportunity cost of capital). Even if the authors make the extreme assumption that the opportunity cost of labor is zero, they find a return of only negative 7% for cows and positive 17% for buffaloes. This result is surprising, given the widespread ownership of cows in India and in many other places. Anagol et al. (2014) put forward a number of potential explanations for their findings: measurement error, better quality of home-produced milk, buffer stock savings, labor market failures, time variation in returns, preference for positive skewness in returns, the social and religious value of cows and intra-household conflict over savings. Attanasio and Augsburg (2014) have recently revisited the issue and argue that variation in returns with weather conditions is more likely to explain why Anagol et al. (2014) find such large negative average returns. The authors argue that Anagol et al. (2014) collected data during a drought period in which fodder was scarce and fodder prices high. They recalculate returns using three rounds of data in a different state of India (Andhra Pradesh) and find positive average returns in good years (in terms of rainfall) and negative returns in bad years.

In this paper, we revisit the paradox between widespread support of cattle farming through agricultural policy interventions and negative returns to cattle. We also use data from Andhra Pradesh but from alternative years. On one hand, we want to see how generalizable the findings of Anagol et al. (2014) as well as Attanasio and Augsburg (2014) are; on the other hand - and more importantly - we want to explore in more detail the economic choices households face when herding cattle in India. In order to do so, we complement the accounting approach proposed by Anagol et al. (2014), which basically serves to calculate the average profitability of holding cows, with an analysis of marginal returns, of economies to scale and of returns to different varieties of cows and buffaloes.

The results of this paper suggest that cattle farming might well generate positive returns for households in rural India, but that most households seem to operate at unprofitable levels. Similarly to Anagol et al. (2014), we find negative average returns to cattle. If we set the opportunity cost of labor to the average market wage for women for unskilled labor observed in the sample, average returns are in the order of -8% at the mean and vary between negative 53% (in the lowest quintile of cattle value) and positive 2% (in the highest). Similarly to Attanasio and Augsburg (2014), we also find that returns are considerably higher in times of favorable weather conditions than in periods with low rainfall levels. When exploring the evolution of returns with increasing cattle value we find substantial variation in the profitability of cattle holding. The empirical pattern suggest a non-convex production technology, and can be best explained with the existence of substantial economies of scale and higher returns of modern breeds. A detailed analysis of the cost structure shows that decreasing labor costs are one of the main drivers of profitability gains associated with increasing herd sizes. Our analysis also hints at entry barriers which consist in overcoming high upfront expenses of acquiring modern breeds; in particular modern variety cows. Wealthier households and households with lower costs to access veterinary services are more likely to overcome these barriers. These findings are in contrast to estimates from the off-farm sector, where entry barriers do not seem to play a role (see e.g. Banerjee and Duflo, 2014; De Mel et al., 2008; Dodlova et al., 2015; Fafchamps et al., 2014; Grimm et al., 2011; Kremer et al., 2010; McKenzie and Woodruff, 2006); and entail very important implications for the support of cattle farming in development policy interventions.

The remainder of this paper proceeds as follows. Section 3.2 provides some background information on livestock production in India and presents the data used in the analysis. Section 3.3 explores returns to cattle holding. Section 3.4 puts forward potential explanations for observed non-convexities in marginal returns with a focus on economies of scale, the profitability of different breeds, and entry barriers. Section 3.5 concludes.

3.2 Context and Data

India is the second largest cow-milk producer in the world (FAOSTAT, 2015).[2] Cows play an important role in the lives and livelihoods of rural households in India. They are considered sacred in the Hindu religion, and cattle slaughter is prohibited in most states of India. At the same time, dairy products are widely consumed in India, as they are the main source of animal proteins of many households (GoI-NSSO, 2013).

[2] The FAO estimates the total production of cow milk in 2012 to be around 54 million metric tons. If buffalo milk and cow milk are considered jointly, India is the largest producer in the world, with 110 million metric tons produced in 2012. The largest cow milk producer worldwide is the United States, with about 91 million metric tons of fresh milk produced in 2012.

Households might own cattle for a number of reasons: in order to generate income, as source of social status, to accumulate savings, because they prefer home produced milk and dairy products etc. In order to estimate returns to cattle value, we need to understand what outputs can be generated from cattle farming and what costs are associated with it.

The main outputs from cattle farming are milk and dairy products as well as calves. Calves can either be sold shortly after birth, be raised and sold later, or be kept by the household for future dairy production. Households can also sell dung, which is used for manure as well as a cooking fuel in rural India. But since the survey used in this paper does not collect any information about dung, we are not able to account for it.[3]

Paid-out costs associated with cattle farming are mainly expenditures on fodder, veterinary services and insemination. Furthermore households invest time in cattle farming, such that the opportunity costs of time have to be valued appropriately. Lastly an important source of costs in cattle farming is the depreciation of cattle over time, given that the animals only produce milk in their fertile age.

The data used in this paper are the Young Lives Survey (YLS) data for Andhra Pradesh. Andhra Pradesh is the third largest milk producer in the country; only Uttar Pradesh and Rajasthan have higher milk production per year (GoI-DAHD, 2012). The slaughter of cows and calves has been prohibited since 1977, and bulls and bullocks can only be slaughtered upon permission, e.g. if owners can prove that these cannot be used for reproductive purposes or in agricultural production (GoI-DAHD, 2002).

The YLS is part of a long-term research project that seeks to understand the changing patterns and long-term consequences of childhood poverty. For that, it collects panel data in Ethiopia, India (Andhra Pradesh), Peru and Vietnam. The data is intended to cover a time span of 15 years upon completion of the project. The dataset on Andhra Pradesh consists of 3,019 households living in six different districts. The selection of districts under the YLS ensured that all three geographical regions were represented in the survey as well as poor and non-poor districts of each region. Classification of districts was done along economic, human development and infrastructure indicators (Galab et al., 2011). This sample design ensures that the YLS is broadly representative

[3] Anagol et al. (2014) include this source of revenue in their estimation, and estimate that the revenue from the sale of dung makes up 14-15% of total revenue.

for the population of Andhra Pradesh.[4] Four rounds of interviews have been conducted so far: in 2002, 2007, 2009/10 and 2013. Panel attrition is relatively low: 2,910 households could be revisited in 2009/10, which gives an attrition rate of 3.6% (Galab et al., 2011). For reasons of comparability and availability, only the second (2007) and third (2009/10) rounds are considered in the current analysis.[5]

Although the main focus of the survey is on child development, it also collects information on households' characteristics, their income sources, ownership of assets and production strategies. It also contains a section on livestock, which inquires about the type, number and current value of different animals; and about households' expenses for fodder, veterinary services and other items. Households are also asked to report on the revenue they generated in the past 12 months from the sale of milk and dairy products and on the production costs incurred.

Because we are interested in the productivity of cattle, we restrict the sample to households living in rural areas. Although it is still common to see cattle being held in Indian cities, the profitability of farming cattle and of producing dairy is likely to be very different in cities as opposed to rural areas. Furthermore, the sample is restricted to households that lived in the same locality in 2007 and 2009/10 because we assume that livestock is one of the fist things to be sold when a household decides to move. This results in a sample of 2,080 households (4,160 observations). Out of these, 678 households owned cattle (either cows or buffaloes) in either one or both of the survey rounds. The sample of cattle owners contains 975 observations (463 observations in 2007 and 512 in 2009/10) distributed across 80 villages, 15 sub-districts and 6 districts.[6] Finally, we exclude influential outliers from our analysis, as discussed in Section 3.3.2. This eliminates three observations, reducing the final sample to 972 observations.

[4]This is in reference to the State of Andhra Pradesh in 2013, prior to its division into the states of Andhra Pradesh and Telangana.

[5]In the 2002 data, households were only asked about the number of cattle owned, not its value. Also, the questionnaire does not distinguish between buffaloes and cows, modern and traditional breeds or between adult animals and calves. The 2013 data has not been released at the time of writing the article.

[6]A total of 26 observations were excluded from the sample of cattle owners because households owned cattle as well as goats or sheep. Since we cannot distinguish between the revenues of the sale of dairy from cattle and dairy from goats or sheep in the dataset, we decided to drop these observations from our analysis.

Table 3.1: General household characteristics

	2007					2009				
	No cattle		Cattle owners			No cattle		Cattle owners		
	Mean	SD	Mean	SD	p-value*	Mean	SD	Mean	SD	p-value*
Male household head	0.92	0.27	0.95	0.22	0.05	0.92	0.27	0.96	0.20	0.00
Age of household head	39.32	10.81	42.89	12.65	0.00	40.05	9.08	42.05	10.92	0.00
Household size	5.31	1.98	6.23	2.70	0.00	5.24	2.07	6.01	2.73	0.00
Household head is literate	0.31	0.46	0.34	0.48	0.14	0.31	0.46	0.32	0.47	0.60
Wealth index	0.38	0.16	0.43	0.15	0.00	0.45	0.16	0.48	0.14	0.00
Housing quality index	0.48	0.29	0.55	0.26	0.00	0.54	0.28	0.58	0.25	0.00
Consumer durables index	0.16	0.16	0.23	0.17	0.00	0.26	0.17	0.32	0.17	0.00
Housing services index	0.50	0.18	0.52	0.16	0.00	0.55	0.18	0.56	0.15	0.53
Income, non-farm activities	25.22	25.30	27.37	28.79	0.15	34.23	42.80	34.31	39.94	0.97
Income, agricultural production	5.49	25.90	13.06	37.53	0.00	5.14	22.54	17.83	43.43	0.00
Income, savings	0.16	1.80	0.74	5.75	0.03	0.21	1.24	0.35	2.35	0.21
Total area owned	1.55	2.97	3.70	4.81	0.00	1.84	2.81	3.74	4.76	0.00
Observations	1617		463			1568		512		

Notes: * T-test on the equality of means. All values in constant INR 1000 (July 2006). One US$ is equivalent to 46.38 INR. The wealth index is the simple average of the housing quality index, the consumer durable index and the housing services index. Housing quality is the simple average of rooms per person and indicator variables for the quality of roof, walls and floor. Consumer durables are the scaled sum of 12 variables indicating the ownership of items such as radios, fridges, televisions, phones or vehicles. Services are calculated as the simple average of dummy variables indicating households' access to drinking water, electricity, toilets and fuels. For more information on the wealth index refer to the Young Lives data justification documents at http://www.younglives.org.uk.

Summary statistics of general household characteristics are presented in Table 3.1. We split the sample by cattle ownership and by year. As we can see, cattle owners are significantly different from non-cattle owners. In particular, cattle-owning households are more likely to be headed by males and have older household heads. Households that own cattle are also larger on average. Both groups are not statistically different in the proportion of literate household heads, which is very low (around 30%) in both groups. Households with cattle are also significantly wealthier than non-cattle owners: Households with cattle have on average more land, which is the primary indicator for wealth in rural India. But cattle owners have also better-quality houses (in terms of the structure of roof, walls and floor), more consumer goods (such as television, radio, refrigerator, etc.) and are more likely to have access to electricity, water and sanitation.[7] The difference in wealth is more pronounced in 2007 than in 2009/10, although it is statistically significant in both periods. The income structure is also very different between both groups. Although both have similar incomes from non-agricultural activities (equality of means cannot be rejected), cattle owners have much higher incomes from crop production than the rest of the sample.

Table 3.2 presents some household-level information about revenue and costs associated with cattle farming and dairy production. As we can see, the total value of owned cattle increased between 2007 and 2009/10 from INR 12,150 to INR 13,600 (US$ 262 to US$ 293, in constant July 2006 values).[8] This increase is partly reflected in a slight increase in the quantity of cattle and partly in the increase of the average value of the cows and buffaloes owned by these households. The composition of animals owned also changed between 2007 and 2009/10. We find a considerable increase in the number of cows in the sample: the average number of modern-variety cows - thus, European breeds and their crossbreeds - owned by each household increased from 0.17 to 0.26, and the number of traditional cows increased from 0.78 to 0.83. In contrast, the number of buffaloes seems to have decreased over time, for both modern and traditional varieties.

[7]This information is summarized in three indices: housing quality index, consumer durables asset and housing services index. The wealth index reports the simple average of these three indices.

[8]This is the total beginning-of-period value of all grown female cows and buffaloes owned by the household. Households were asked to report the end-of-period value in the survey, which we multiply with the inverse of one minus the depreciation rate to reflect the beginning-of-period value. How we derive the depreciation rate is discussed in Section 3.3.1. Two households reported the value of their animal to be zero. In order not to lose any information, we replaced the value of these cows with the 5th percentile of cattle value observed in the sample (INR 437). We used official exchange rates from July 2006 to convert INR to US$.

Table 3.2: Farming characteristics

	2007		2009	
	Mean	SD	Mean	SD
Total value: cattle	12152.9	12524.4	13624.5	13064.1
Quantity: cattle	1.97	1.47	1.96	1.47
Quantity owned: Cow (modern)	0.17	0.68	0.26	0.61
Quantity owned: Cow (traditional)	0.78	1.18	0.83	1.28
Quantity owned: Buffalo (modern)	0.19	0.62	0.11	0.53
Quantity owned: Buffalo (traditional)	0.82	1.29	0.75	1.23
Average cattle value	6079.8	4117.7	6832.1	3442.0
Total value: calves	815.8	1369.5	2244.5	3627.8
Quantity: calves	0.95	1.14	1.35	1.34
Total revenue from sale of dairy products	3599.2	9003.3	6187.1	10780.7
Total cost from sale of dairy products	1380.2	4562.9	2243.4	4281.5
Expenditure on cattle: veterinary	205.5	471.7	232.5	624.3
Expenditure on cattle: fodder	1801.8	4138.7	1550.1	2994.0
Expenditure on cattle: other cost	32.4	148.4	74.8	195.4
Total expenditure on cattle	2039.7	4449.5	1857.4	3224.9
Time spent on cattle (hours per year)	450.2	190.3	486.4	199.6
Shock affected livestock	0.13	0.33	0.19	0.39
Rainfall (deviation)	-0.26	0.24	0.000070	0.24
Observations	463		509	

Notes: All values in constant INR (July 2006). One US$ is equivalent to 46.38 INR.

The number of calves per cow or buffalo was 0.48 in 2007 and 0.69 in 2009/10, which gives an average annual reproduction rate of about 58.9%. The value of calves over both rounds corresponds to about 17.4% of the total value of adult female cattle.[9]

In the survey, households were asked to report the total revenue from the sale of milk and dairy products in the past 12 months (including the value of their own consumption).[10] As we can see in Table 3.2, the annual revenue from the sale of dairy products increased between 2007 and 2009/10, from INR 3,600 to INR 6,200 (US$ 77 to US$ 133).

Expenditures on cattle are also collected in the survey and comprise fodder, but also veterinary costs, insemination costs and labor. The cost variables are obtained from two sections. In the livestock section, households were asked to report their total expenditure on fodder, veterinary services and other expenses incurred for all animals owned in the last 12 months.[11] In order to derive from this information the expenditure incurred for cattle, we divide these variables by the total value of all animals owned by the household. We then multiply it by the reported value of cattle (mother cows, mother buffaloes and calves) in the household. Expenditure on fodder makes up almost 90% of total paid-out cost. In 2007, spending on fodder was INR 1,800 (US$ 39) on average, whereas households spent only INR 200 (US$ 4) on veterinary services and INR 30 (US$ 1) on other items.[12] Another source of information about the costs associated with cattle farming is the income section in which households were asked about the total costs associated with producing and selling dairy products in the last 12 months. Households were asked to also include expenditures on fodder and veterinary

[9]The value of calves reported by the household reflects current ownership, and hence excludes all calves that were sold before the survey took place. Revenue from the sale of calves was not included explicitly in the survey, which implies that we probably underestimate the reproduction value of cattle. Again, a few households reported the value of their calves to be zero, and it is not clear from the data whether this information was simply not known or misreported. Therefore, we replaced the value with the 5th percentile observed in the sample: INR 95.

[10]With the data used in this paper, we cannot assess as to whether households correctly account for the value of their own milk consumption. It is likely that there is non-random measurement error in this variable because households that operate at a lower scale presumably consume a higher share of produced milk within the household.

[11]The survey question explicitly asks for purchased fodder only. It is therefore likely that we are underestimating the true expenses for fodder, as households with land ownership might let cattle forage on their fields. We therefore control for land ownership in our estimations.

[12]Total expenditure on cattle is the sum of these three variables. Expenditure on fodder was only multiplied by the value of adult female cattle, hence we assume fodder expenses for calves to be zero. We show in Section 3.3.3 that our results do not change if we relax this assumption.

services for those animals that produce dairy.[13] This information is captured by the variable "Total cost from sale of dairy products". Cost estimates from both sections are somewhat different, which is why we estimate returns based on both cost estimates for additional robustness. We do not find that the selection of cost estimates affect our results substantially (c.f. Section 3.3.3).

In order to account for labor allocated to caring for the animals and for dairy production, we construct a time variable based on the 2007 survey information. In the 2007 survey, all household members (incl. children) were asked about their three most important activities and about the number of hours per day, days per week and weeks per month they spent on this activity. From this question, we compute an aggregate variable that captures the total hours per year that households spent on livestock farming. To obtain the hours worked in cattle farming, we divide this value by the number of adult equivalent animals owned by the household and multiply it by the number of cattle (both calves and adult cows/ female buffaloes).[14] This gives an estimate of total hours per year that households spent on caring for their cows, female buffaloes and calves. Because the 2009/10 questionnaire did not include the same information, we have to impute this data. In order to do so, we use the 2007 data and run a simple OLS regression of the number of hours spent on cattle per year on the number of currently owned cattle. Because we observe that the number of hours that households spent on their animals increased with the number of owned animals, but at a decreasing rate (due to complementarities), we also include the square of this variable.[15] From this regression, we can predict for each observation the hours per year spent on cattle farming and dairy production. This predicted time variable for 2007 and 2009/10 is reported in Table 3.2.

[13]According to personal communication of the survey team, the variable also includes wage costs of the household for caring for the animals and marketing the product. But when comparing this variable with the costs variable computed from the livestock section, it does not seem to be much higher, which it would have to be if labor costs were adequately accounted for.

[14]The adult equivalent of cattle is 0 for poultry and birds, 0.2 for pigs, and 1 for bullocks, bulls, cows, buffaloes and calves. We assume it equals 1 for calves in order to account for increased labor input when cattle is being milked.

[15]The coefficient of the square of that variable is statistically significant at the 1% level (p-value 0.002).

3.3 Returns to cattle holdings

3.3.1 Empirical specification

In order to understand how profits from cattle farming develop with cattle value, breed and input allocation, we specify a profit function and then estimate both average and marginal returns to cattle. In the absence of experimental data, this will remain a rather descriptive analysis, yet we will conduct several robustness checks to get a better idea regarding the potential margin of variation of these parameters.

We assume that $Q_{it} = f(K_{it}, L_{it}, F_{it})$ is the production of milk, other dairy products and calves, with capital (current value of cattle), labor and fodder as inputs. The sales revenue of cattle products can be summarized by pQ_{it}, where p is the price vector of all outputs. Opportunity costs of time can be captured by w, the price of fodder by g, and c summarizes all other costs associated with cattle farming (i.e. veterinary services and insemination). We assume that land enters the production function only through the fodder it provides and therefore do not include it explicitly here.[16] We also assume the opportunity cost of capital to be zero, but have to account for the fact that cattle depreciates over time. A profit function (net of depreciation) can thus be written as follows:

$$\pi_{it} = pQ_{it} - cK_{it} - wL_{it} - gF_{it} - \delta K_{it}. \tag{3.1}$$

We value the total time a household allocates to cattle at an hourly wage of INR 5 (US$ 0.10) in 2007 and INR 8 (US$ 0.17) in 2009. This is equivalent to average daily female wages for unskilled work reported in our data, and since caring for livestock is mostly in the responsibility of women and children, it seems reasonable to impute opportunity costs of time in that range.[17] Obviously opportunity costs of time could be very different for skilled vs. unskilled workers. We show in Section 3.3.3 that our results do not change if we impute different wages according to the educational level of the household member that is mainly responsible for caring for livestock.

The depreciation rate δ reflects the change in value of cattle from the current period t to the next period $t+1$, and is simply $-(K_{t+1} - K_t)/K_t$. The value of cattle depreciates

[16] However, we control for land owned (in logs) in all our estimations.

[17] The observed daily wage for is INR 48 in the 2007 round and INR 75 in the 2009/10 round (in constant July 2006 values). We assume a workday consists of 10 hours on average in India.

strongly over time because cows produce milk only as long as they are fertile.[18] The depreciation rate cannot be estimated with the data used in this paper because the survey does not contain any information about the age of the animals.[19] We thus have to rely on secondary sources for this value. Data from the Animal Husbandry Department of the Government of Andhra Pradesh suggest that a fertile cross-breed cow cost about INR 10,500 (US$ 227) in Andhra Pradesh in financial year 2008/09 (AHD-GAP, 2009).[20] Cows enter reproductive age at about 2.5 years (buffaloes after 3 years) and are expected to calve about five times during their lifetime (Ruvuna et al., 1984).[21] Given an average reproduction rate of around 0.59 per year (as observed in the sample), we can assume cows and buffaloes to be productive for about 8.5 years after entering reproductive age. Cows would thus be fertile up to the age of about 11 years, buffaloes up to the age of 12.[22] As explained earlier, we expect the real value of a cow or buffalo to be zero once it reaches that age. Assuming linear depreciation of cattle, this would imply that each animal depreciates by around INR 1,240 (US$ 27) per year. If we assume that the depreciation is declining with increasing age, an annual relative decrease in cattle value of 20% would imply that animals have an end-of-fertility value of INR 1,400 (US$ 30), which is slightly more than 1/10 of the initial value. This depreciation rate is also used by government entities in their project reports (see e.g. GHP-AHD, 2014) and produces more conservative estimates than a linear depreciation.[23]

Based on the profit function specified above, average and marginal returns to cattle

[18] And since cattle cannot be sold for slaughter, this implies that the value of a cow will be zero once it is no longer of reproductive age. Of course, reports exist throughout the country of unproductive animals being sold off to other states in which cattle slaughter is not prohibited. But in this paper, we assume that the market value of a cow approaches zero with the end of its fertility.

[19] We are also not able to account for potential increases in animal value after the cow or buffalo has first calved.

[20] In July 2006 prices. The average price of a Graded Murrah buffalo was roughly the same. These prices vary between districts, however.

[21] Of course these are rough averages; reproduction rates and number of calves per animal vary across breeds. Crossbreeds seem to have higher reproduction rates than traditional varieties (Mukasa-Mugerwa, 1989).

[22] Some studies even refer to 12 years of productive life for crossbred cows (Ghule et al., 2012).

[23] A relative depreciation seems more appropriate here because we do not know the initial value but only the current value of each animal. We would otherwise introduce the rather unrealistic assumption of equal initial value across breeds and animals.

value would be:

$$\frac{\pi_{it}}{K_{it}} = p\frac{Q_{it}}{K_{it}} - c - \frac{wL_{it}}{K_{it}} - \frac{gF_{it}}{K_{it}} - \delta \tag{3.2}$$

and

$$\frac{\partial \pi_{it}}{\partial K_{it}} = p\frac{\partial Q_{it}}{\partial K_{it}} - c - \delta. \tag{3.3}$$

In contrast to average returns, our estimates of marginal returns to cattle strongly depend on assumptions concerning the functional form of the production function. In order to get a better idea of the pattern of marginal returns to cattle, we try both parametric and semi-parametric approaches. In the parametric approaches, we allow the production function $f(K_{it}, L_{it}, F_{it})$ to be linear, quadratic or constant elasticity of substitution (CES) type.

We start with a linear production function, where estimating marginal returns is straightforward. We estimate profits (net of depreciation) as a function of cattle value and account for a number of control variables x_{it}, such as household characteristics, period effects and shocks. We also control for fodder expenses and land ownership (both in logs).[24] We estimate:

$$\pi_{it} = \beta_0 + \beta_1 K_{it} + \beta x_{it} + \epsilon_{it}. \tag{3.4}$$

Alternatively, we also include the square of cattle value, which allows marginal returns to increase or decrease with cattle value. In a CES type production function, such as $f(K_{it}, L_{it}, F_{it}) = K_{it}^{\alpha} L_{it}^{\eta} F_{it}^{\chi}$, the functional form imposes decreasing marginal returns, as long as $0 < \alpha < 1$, $0 < \eta < 1$ and $0 < \chi < 1$. These would be $\partial\pi_{it}/\partial K_{it} = \alpha p Q_{it}/K_{it} - c - \delta$. Calculating marginal returns to cattle under CES functional form assumptions requires an estimate of α, which we obtain by estimating the log-transformation of the production function:[25]

$$log(pQ_{it}) = \beta_0 + \alpha log(K_{it}) + \eta log(L_{it}) + +\chi log(F_{it}) + \beta x_{it} + \epsilon_{it}. \tag{3.5}$$

[24] We do not control for labor as this variable is imputed for all households. We show in Section 3.3.3 that the results are not affected by the omission of labor, nor by the omission of fodder expenses from the estimation.

[25] We again control for land ownership in logs to account for self-produced fodder.

To compute marginal returns, we then multiply α with representative values of the sales revenue of milk products and calves divided by the value of cattle and subtract the depreciation rate and marginal costs. And finally, we leave the functional form unrestricted and estimate marginal returns in a semi-parametric framework, such as:

$$\pi_{it} = \beta_0 + g(K_{it}) + \beta x_{it} + \epsilon_{it}. \tag{3.6}$$

Here, $g(K_{it})$ is a non-parametric function of cattle value. In order to isolate the non-parametric part of the equation, we follow Robinson (1988) in removing conditional expectations given K_{it} from the dependent variable and all regressors. Formally, we estimate $\pi_{it} - E(\pi_{it}|K_{it}) = \beta(x_{it} - E(x_{it}|K_{it})) + \epsilon_{it}$ to obtain an estimate of βx_{it}. We then estimate the following equation non-parametrically:

$$\pi_{it} - \hat{\beta} x_{it} = \beta_0 + g(K_{it}) + \epsilon_{it}. \tag{3.7}$$

We estimate equation (3.7) with locally weighted mean smoothing using Cleveland's (1979) tricube weighting function and a bandwidth of 0.5. Estimates of 90% confidence intervals are obtained from 5,000 bootstrap replications. To obtain estimates of the marginal returns, we predict the slope of the nonparametric fit at all values of K_{it}. We then smooth the pointwise slope estimates and calculate confidence intervals using the same approach as described above.

Estimating returns to cattle with observational data is challenging for a variety of reasons. First, we have good reasons to assume that unobservable household characteristics, such as farming ability, are correlated with the observed value of cattle and profits. Estimating returns in a fixed effect model can partly remedy this problem by accounting at least for time-constant unobservable characteristics. The drawbacks associated with the fixed effects estimation are: 1) that the panel is rather short and therefore offers only limited possibilities to use within household variation in the independent variable to estimate the parameters of interest; and 2) that the panel is not fully balanced because many households seem to have changed their production strategies by moving in or out of cattle farming and by increasing their herds with other livestock, such as sheep or goats.[26] Second, capital stocks are usually measured with

[26] Remember that we have to exclude all households with other milk-producing animals because the questionnaire does not distinguish between these different sources of milk production.

high imprecision, and the current value of a farmer's cattle is probably no exception. This could lead to attenuation bias, where estimates of marginal returns would be biased toward zero due to measurement error in the explanatory variable. Third, reverse causality - i.e., the fact that in the presence of capital market imperfections, higher profits lead to faster capital accumulation - might cause an upward bias in estimated returns.

Given these limitations, our estimate cannot be interpreted as causal, nevertheless given the detail of the data, the panel dimension and the various robustness checks we provide, we believe that our estimates provide at least a useful description of the rough pattern of average and marginal returns. A more accurate estimation of these parameters must be left to future work that can draw on experimental data.

3.3.2 Estimates of average and marginal returns

Following closely Anagol et al. (2014) we estimate average rates of return to cattle by calculating profits (net of depreciation) for each household and dividing it by K_{it}, e.g. the cattle value at the beginning of the period. As reported in Table 3.3, average rates of return are negative at the mean of cattle value (INR 12,900 or US$ 279) by roughly 8% annually.[27] Furthermore, we find that average returns are lower at lower quintiles of investment in cattle. They range from -53% annually (in the lowest quintile) to positive 2% annually (in the highest).[28] Only households in the fifth quintile, with animals worth INR 33,000 (US$ 711) on average, are able to generate positive average returns. And even in this range, returns are well below the estimates of returns to capital in micro and small non-agricultural enterprises in India (see e.g. Banerjee and Duflo, 2014).

[27] These estimates draw on the costs reported in the income section of the survey and are our preferred estimates. However, to check robustness, we calculate returns also based on the costs reported in the livestock section (see also Section 3.2). Results can be found in the Appendix, Table 3.A.1. Both approaches provide fairly similar results.

[28] To reduce bias from influential outliers, returns are calculated by using the the mean of cattle value, revenue and cost in each group, instead of calculating the group mean of rates of return calculated at the household level.

Table 3.3: Average returns by quintile of cattle value

Quintiles of cattle value	Average cattle value*	Quantity: cattle	Total value: cattle	Revenue: dairy	Value: calves	Cost: dairy	Labor cost	Depre-ciation	Profits	ROR
	A	B	C	D	E	F	G	H	I=D+E -F-G-H	J=I/C
1	2862	1.12	3002	1705	539	949	2293	600	-1599	-0.53
2	5174	1.20	5733	2433	810	1029	2284	1147	-1217	-0.21
3	6833	1.51	8946	2780	1294	1101	2836	1789	-1653	-0.18
4	8015	2.18	14776	5289	1704	1586	3496	2955	-1045	-0.07
5	9574	3.91	33081	12949	3566	4646	4700	6616	552	0.02
Total	6474	1.97	12924	4954	1564	1832	3110	2585	-1008	-0.08

Notes: Cells report the mean value of the variable of each column for a given quintile. All values are constant INR (July 2006). One US$ is equivalent to 46.38 INR. Rate of return (ROR) calculated as follows: $(\pi_t - (K_t - K_{t+1}))/K_t$, where K_t is the value of the animal at time t.

* Average value of all cows and buffaloes owned by household.

One of the drawbacks to the accounting approach is that it is difficult to understand the circumstances under which observed profits come about. By calculating averages, we also completely ignore external factors that might be driving observed results. Marginal returns are presumably better able to inform about the different options households face at different levels of cattle value and should thus shed further light on the question why households would own cattle if average returns are found to be negative.

In order to estimate marginal returns, we rely first on a parametric approach and consider three types of production functions: linear, quadratic and CES. All three functional forms seem to fit the data fairly well.[29] We then also estimate marginal returns semi-parametrically, leaving the functional form of the production function unspecified. As mentioned earlier, we drop three observations based on the DFITS statistic and cutoff values recommended by Belsley et al. (1980) in order to reduce the influence of outliers.[30] Shocks, a time dummy and socio-economic household characteristics are included as controls in all estimations. We also control for fodder expenses and land ownership (in logs).

The estimates of marginal returns assuming a linear or quadratic production function are reported in Table 3.4. Column 1 reports pooled OLS estimates for the full sample. The point estimate of cattle value is 0.086, suggesting marginal returns to cattle of about 8.6% annually. In column 2, we add the square of cattle value to explore potential non-linearities in returns. The coefficient of the squared term is close to zero and not statistically significant at the 10% level. When accounting for unobserved heterogeneity in random effects models (column 3), the estimates of marginal returns remain exactly the same. In the fixed effect model, however, estimates drop in size considerably (column 4). The most probable reason for this strong reduction is the fact

[29] We regress revenue from cattle farming on the value of cattle and correlate predicted revenue with actual revenue to get an impression of how well each functional form fits the data. The square of the correlation coefficient then gives the R-squared. The quadratic production function seems to fit the data best with an R-squared of 0.42. However, there is not much difference in the R-squareds of all three regressions: The R-squared using a CES function is 0.40 and is 0.39 using a linear functional form.

[30] We calculate the DFITS statistic in our estimation of marginal returns assuming a quadratic production function. We choose the quadratic production function for this procedure instead of the linear because it leads us to drop three observations, as compared to two observations in a linear production function framework. Furthermore, two of the three observations would have to be dropped in the linear function as well. The recommended cutoff value is 2/sqrt(k/N), with k being the degrees of freedom plus one and N the number of observations.

Table 3.4: Marginal returns to cattle: linear production function

	OLS		RE	FE
	(1)	(2)	(3)	(4)
Total value: cattle	0.086$^+$	-0.019	0.086$^+$	0.010
	(0.044)	(0.059)	(0.044)	(0.061)
Total value: cattle (squared)		0.000		
		(0.000)		
Total land owned (acres, log)	-356.316	-212.663	-356.316	319.958
	(303.658)	(275.897)	(303.658)	(938.559)
Expenditure on cattle: fodder (log)	258.994***	272.849***	258.994***	32.522
	(62.159)	(61.229)	(62.159)	(108.434)
Income, non-farm activities (log)	798.665***	814.274***	798.665***	990.622**
	(136.688)	(135.453)	(136.688)	(325.271)
Household size	-218.000*	-205.153*	-218.000*	486.370
	(97.150)	(97.093)	(97.150)	(353.868)
Age of hh head	-28.489	-28.911	-28.489	-109.072*
	(18.837)	(18.917)	(18.837)	(54.329)
Highest grade: hh head	80.688	77.962	80.688	51.220
	(75.274)	(72.654)	(75.274)	(143.561)
Male household head	401.121	440.127	401.121	-2150.610
	(895.848)	(887.593)	(895.848)	(2544.307)
Wealth index	1433.428	1214.291	1433.428	355.809
	(1503.474)	(1511.515)	(1503.474)	(4438.892)
Shock affected livestock	-92.046	33.062	-92.046	457.196
	(711.934)	(700.110)	(711.934)	(1194.771)
Rainfall (deviation)	2099.687**	1917.049**	2099.687**	3852.322***
	(688.923)	(673.906)	(688.923)	(928.651)
Year 2009 (dummy)	626.140	733.796$^+$	626.140	597.189
	(399.126)	(394.633)	(399.126)	(767.993)
Observations	972	972	972	972
R^2	0.111	0.121		0.120

Notes: Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Standard errors (clustered at the household) in parentheses. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

that the panel is relatively short and unbalanced, meaning that quite some households own cattle only in one of the two survey periods. For those households who own cattle in both survey rounds, we find very little changes in cattle value over time, which biases the coefficient towards zero.

Alternatively, we estimate marginal returns to cattle assuming a CES functional form of the production function. For that, we start by estimating equation (3.5) to get an estimate of α. Results are reported in Table 3.5. The first column presents results of pooled OLS estimation. The second and third columns present results of random effects and of fixed effects estimations, respectively. Estimates of α are large and statistically significant for most specifications. The size of the coefficient is similar throughout specifications 1 and 2, ranging between 0.72 and 0.75. The coefficient on cattle value is again lowest in the fixed effects estimation, i.e. 0.16. We calculate marginal returns to cattle at all values of α (e.g., the coefficients on the cattle variable) and for the median as well as at the mean of cattle value, revenue and cost. Results are reported at the bottom of Table 3.5. As we can see, estimated marginal returns are positive for all α except the fixed effects estimate. At the highest value of α (0.75), marginal returns at the mean of cattle value (INR 12,900 or US$ 279), revenue (INR 6,500 or US$ 141) and cost (INR 1,800 or US$ 40) are 16% annually. At the lower estimate of α (0.72), estimated marginal returns at the mean are 14%. This is slightly higher than the return calculated in levels (cf. Table 3.4). At the median of cattle value (INR 9,000 or US$ 194), revenue (INR 2,800 or US$ 60) and cost (INR 170 or US$ 4), the marginal return to cattle is close to zero.

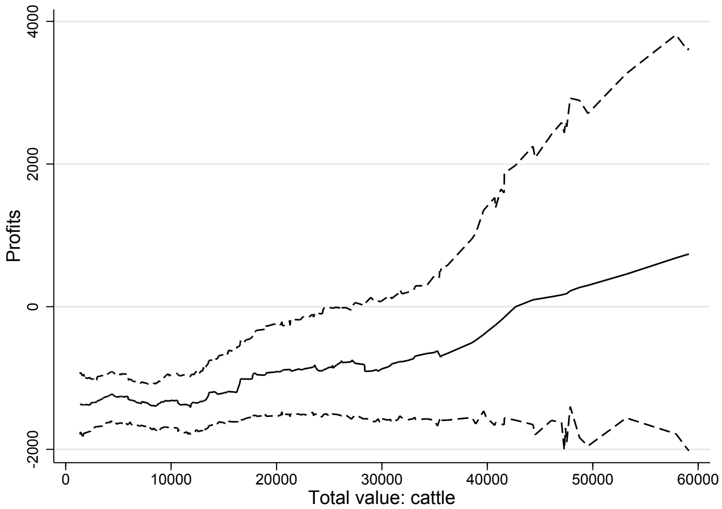
As mentioned before, imposing a particular functional form might not be appropriate if the functional form is a priori unknown. We therefore proceed with estimating marginal returns semi-parametrically, as discussed in Section 3.3.1. Results are shown in Figures 3.1 and 3.2. Interestingly, marginal returns seem to follow a U-shape at cattle values below INR 13,000 (US$ 280), being quite high at very low levels of cattle value and falling with increasing cattle value. The minimum seems to lie at cattle values of around INR 7,000 (US$ 108). At higher levels, marginal returns increase again, reaching their maximum at cattle values of roughly INR 13,000 (US$ 280), which is just above the sample mean. After that, marginal returns seem to remain fairly constant at about 10% per annum, before decreasing again at cattle values above INR 30,000

Table 3.5: Marginal returns to cattle: CES production function

	OLS (1)	RE (2)	FE (3)
Total value: cattle (log)	0.754***	0.724***	0.162
	(0.120)	(0.121)	(0.249)
Total land owned (acres, log)	0.139	0.164	0.128
	(0.142)	(0.144)	(0.398)
Expenditure on cattle: fodder (log)	0.196***	0.191***	0.072
	(0.031)	(0.032)	(0.057)
Income, non-farm activities (log)	0.358***	0.344***	0.197
	(0.097)	(0.097)	(0.139)
Household size	-0.069+	-0.066	0.133
	(0.040)	(0.040)	(0.138)
Age of hh head	-0.011	-0.012	-0.030
	(0.009)	(0.009)	(0.021)
Highest grade: hh head	0.051	0.046	-0.028
	(0.037)	(0.037)	(0.053)
Male household head	-0.093	-0.006	0.970
	(0.413)	(0.428)	(0.939)
Wealth index	-0.350	-0.313	1.894
	(0.708)	(0.717)	(2.277)
Shock affected livestock	0.668**	0.680***	0.650+
	(0.208)	(0.206)	(0.363)
Rainfall (deviation)	0.329	0.309	0.253
	(0.388)	(0.385)	(0.520)
Year 2009 (dummy)	1.718***	1.721***	1.569***
	(0.274)	(0.271)	(0.381)
Marginal returns to cattle at:			
Median	0.02	0.01	-0.16
Mean	0.16	0.14	-0.14
Observations	972	972	972
R^2	0.209		0.240

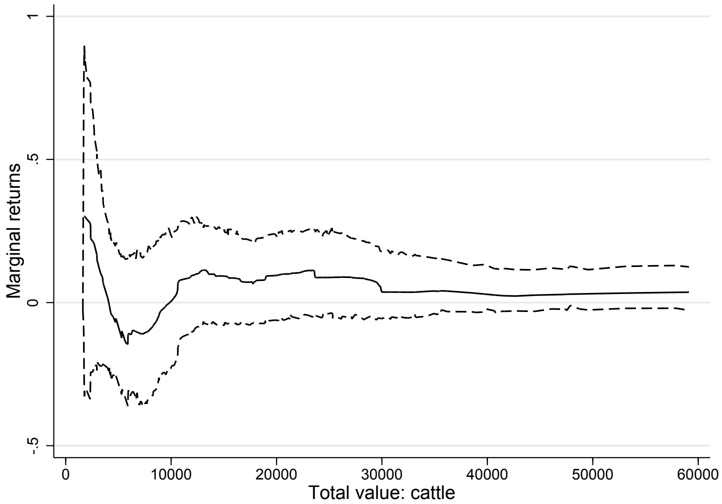
Notes: CES production function assumed. Dep. var: Revenue from sale of dairy products and calves (log). Standard errors (clustered at the household) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 3.1: Semiparametric estimation of profits from cattle farming



Notes: Locally weighted mean smoothing, bw= 0.5, 90% confidence intervals obtained through 5000 bootstrap replications.
Source: Own estimation based on YLS data.

Figure 3.2: Semiparametric estimation of marginal returns to cattle



Notes: Locally weighted mean smoothing, bw= 0.5, 90% confidence intervals obtained through 5000 bootstrap replications.
Source: Own estimation based on YLS data.

(US$ 647). Because confidence intervals are very large, we cannot reject the possibility of constant marginal returns at all levels of cattle value.

The fact that marginal returns seem to increase with cattle value (at least over a certain range of cattle value) could be an indication for the existence of non-convexities in the production technology. Why these non-convexities exist and what they imply for policy is ex-ante less clear. We will turn to this question later and present some robustness checks first.

3.3.3 Robustness checks

Not only the point estimate of marginal returns can be biased due to a number of reasons, also the observed pattern in marginal returns could be due to omitted variables, rather than the specificities of cattle farming. This section seeks to gauge the robustness of the point estimates of average and marginal returns as well of the observed pattern.

Attanasio and Augsburg (2014) stress the importance of adequately accounting for the effects of shocks on the productivity of animals. Below average rainfall, for example, will likely affect the productivity of cattle because fodder is less accessible, and this will affect both the nutritional status of animals as well as their milk production. When we split the sample by round of interview, we find that average returns at the mean are higher in 2009/10 (-3%) than in 2007 (-14%). This could be due to the fact that most households had faced severe rainfall shortages during the period of reference of the 2007 interviews, whereas rainfall levels were close to the long-term average in the 2009/10 reference period (c.f. Table 3.A.2 in the Appendix). We also allow marginal returns to vary with rainfall conditions to get more explicit evidence on the role of weather conditions. We find that returns to cattle increase with higher rainfall levels, and approach zero when rainfall is lower than normal. At zero rainfall deviation (hence at the 10-year average of annual rainfall), marginal returns are 11.6% (c.f. Table 3.A.3 in the Appendix). Obviously, rainfall shocks could also have another effect on returns: if cattle prices drop during a rainfall shock because many farmers want to sell their animals simultaneously and reported cattle values reflect the drop in animal prices, we would probably overestimate returns in drought years. However, we could not find any evidence for this: reported cattle values seem to be largely unrelated to current rainfall levels.[31]

Another reason why we might get biased estimates of returns to cattle are unobserved household characteristics such as farmer ability, that affect both the value of cattle as well as returns to cattle. To get an impression of how severe such a bias could be, we allow marginal returns to vary with the educational level of the household head as well of the main person responsible for taking care of the livestock. We find that marginal returns to cattle farming are indeed higher if the household member responsible for livestock farming has completed primary education (c.f. Table 3.A.4 in the Appendix). We also find evidence that better educated households own higher value animals: the mean cattle value is IRN 12,200 for households whose main person responsible for livestock has not completed primary education, as opposed to INR 15,000 for households with higher educated members. Returns are also higher for wealthier households (c.f. Table 3.A.4 in the Appendix). This suggests that wealthy households are able to operate at higher cattle values, which were shown to have higher marginal returns.

Given the questionnaire design, we have to make a number of assumptions in the calculation of average and marginal returns. Our assumptions regarding the opportunity costs of time could well influence both point estimates as well as the observed patterns in average and marginal returns. If we set, for instance, the opportunity costs of time to zero, average returns are positive throughout all quintiles and highest at the lower quintiles (c.f. Table 3.3). We also observe that the average time allocated to each animal decreases with the number of animals owned (c.f. Section 3.2). This implies higher labor costs per animal for households operating on a smaller scale. But if opportunity costs of time are lower for poorer and less educated households and these households tend to operate at lower cattle values, then this would change our results. We therefore allow imputed wages to depend on the educational status of the household member that is mainly responsible for taking care of cattle as additional robustness check. We set wages for individuals with less than 12 years of schooling to equal the observed average wage for herding cattle in the data and set wages for individuals with 12 years of education and more to equal the observed average wage

[31]The correlation coefficient of total cattle value and the deviation of rainfall from the long-term average is -0.066, and the correlation coefficient of the average value of cattle in the household with rainfall is 0.0025. Also Figures 3.A.1 and 3.A.2 show that reported average and total cattle values do not vary systematically with the date of interview.

of a teacher. As shown in Table 3.A.5 in the Appendix, this does not affect our point estimates of marginal returns.[32] It also does not affect the observed non-convexity in marginal returns (c.f. Figure 3.A.3 in the Appendix).

Also, the manner in which we compute fodder expenses could affect our results. The main results presented so far use cost estimates from the income section (c.f. Section 3.2), as these estimates do not require us to make any additional assumptions. We show in the Appendix, Tables 3.A.1 and 3.A.6 that we get the same results for average and marginal returns, respectively, if we use cost estimates from the livestock section. Table 3.A.6 additionally shows that the results are robust to allowing calves to need purchased fodder too.

Finally, the assumptions we make regarding the production function and its parameters might be influencing estimated outcomes. Allowing fodder inputs to affect output directly, for example, could change our results. The underlying assumption would be that if a household fails to adequately nourish its dairy animals during the pregnancy and milking period, then this is likely to influence the returns on that animal. In the main specifications, we control for fodder expenses in logs. However controlling for expenses in levels might be more adequate when assuming a linear production function. We show in the Appendix, Table 3.A.6 that returns are only marginally higher when excluding fodder as regressor. Also the coefficients on the level and the square of fodder (in levels) are not statistically significant. When including fodder in levels, the coefficient on cattle value drops considerably, from 8.6% to 3.4%. One problem in correctly measuring both returns, is that fodder expenses are highly correlated with animal value, and animal value is likely to reflect the feeding practices of households (well-nourished animals with high milk output have a higher current value than undernourished animals). The correlation coefficient of both variables is about 0.52.

3.4 Explaining the non-convexities: Returns to scale and returns to modern-variety cows

The results presented so far suggest that returns to cattle vary considerably over the distribution of cattle value. While average returns to cattle increase continuously, and

[32]Table 3.A.5 also shows that the results are robust to the omission of time (both reported and imputed values) from the empirical specification.

are only positive in the highest quintile of cattle value, marginal returns vary over the range of cattle value. Looking at the overall pattern, four different stages can be distinguished. In the range of cattle values up to INR 6,000 (US$ 129), marginal returns are positive but strongly falling and average returns are significantly negative. At cattle values between INR 6,000 and 10,000 (US$ 216) both marginal and average returns to cattle value are negative. In the range of INR 10,000 to 34,000 (US$ 733), marginal returns to cattle value are positive again, reaching their maximum just above the sample mean of cattle value (at INR 13,000 or US$ 280). Average returns continue to be negative. At cattle values above INR 34,000, finally, both average and marginal returns are positive, although marginal returns are somewhat lower than in the third stage.

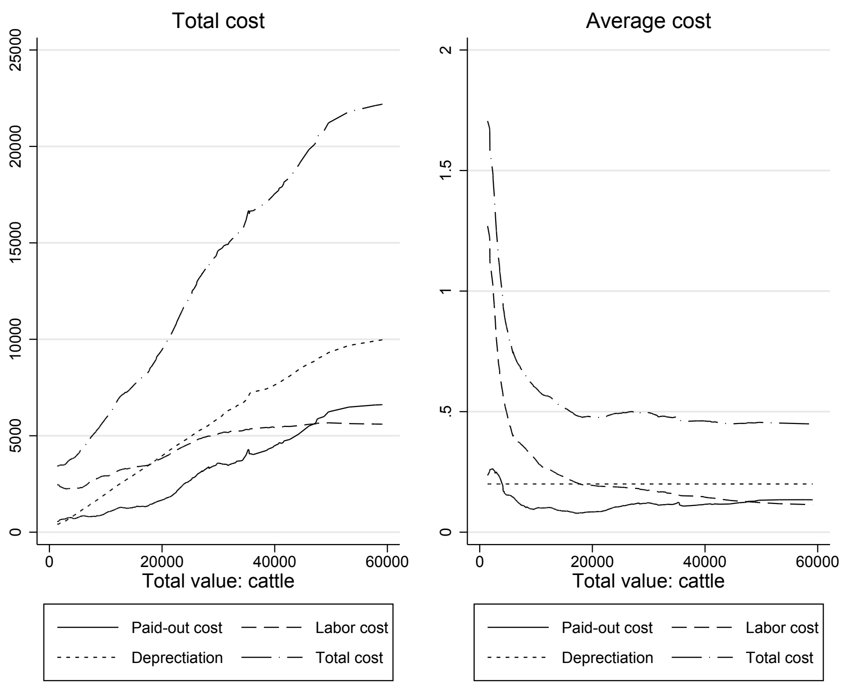
There are different potential explanations for the observed variation in returns to cattle value, and this section seeks to explore some of them. The first reason for observed non-convexities could be returns to scale: minimum thresholds of the number of animals a household needs to own in order to be able to operate at profitable levels. The second reason could be heterogeneity in the profitability of different cattle breeds. Upgrading traditional breeds by cross-breeding them with European varieties has a long tradition in India (Turner, 2004). If only high-value animals generate positive returns, then this would generate additional entry-barriers to cattle farming.

3.4.1 Returns to scale

To get a better understanding of the importance of returns to scale, we estimate different costs and plot them against cattle value (Figure 3.3). These curves are fitted non-parametrically using locally weighted mean smoothing with a bandwidth of 0.2. We do not make any specific assumption about the substitutability between cattle and other inputs, and we just use the observational data. Remember also that we assume a constant depreciation rate of 20% per annum. The first graph shows predicted absolute costs, whereas the second graph shows predicted average costs.

Total costs are roughly INR 3,200 (US$ 69) at the minimum of cattle value, which is more than twice the corresponding cattle value. In this range, labor costs make up the main part of total costs, whereas paid-out costs are about one-third of total costs. Depreciation is negligible at this level due to the low value of cattle. Average costs decrease pronouncedly with cattle value up to a cattle value of roughly INR 20,000

Figure 3.3: Cost structure



Notes: Locally weighted mean smoothing, bw= 0.2.
Source: Own estimation based on YLS data.

(US$ 431). The starkest decrease can be observed at very low cattle values. Beyond INR 20,000, average costs increase again, reaching 50% of cattle value at cattle values of INR 27,000 (US$ 582), and then decrease again. With increasing cattle value, the cost structure changes. Average labor costs decrease steadily with cattle value, reaching 11% at the maximum of cattle value. Average paid out costs, in contrast, reach their minimum (8%) at cattle values of INR 20,000 (US$ 431), increasing again beyond this value to 12% at the maximum of cattle value. Returns to scale thus seem to play an important role in cattle farming, as average costs decrease over most of the cattle value distribution. These cost reductions come to substantial extend from falling average labor cost, and are due to economies of time associated with increasing herd sizes.

To test if there is a threshold regarding the minimum number of animals required to operate at profitable levels, we allow returns to cattle to vary with the number of animals owned. In the sample, roughly half of the households (51.0%) own only one cow or buffalo, the vast majority of them (78.9%) no more than two animals. To identify potential thresholds we create a set of dummies; for households who own more than two, more than three animals etc. Because we are worried that unobserved household characteristics, such a farming ability, might affect both the number of animals owned, as well as the returns to cattle farming, we present estimates in OLS and in fixed effects. However, even fixed effects cannot rule out the possibility that unobserved time-varying variables are driving the observed results. Yet, it is reassuring to see that, controlling for household fixed effects does at least not substantially affect the results.

Table 3.6: Marginal returns to cattle by number of animals owned

	OLS (1)	FE (2)	OLS (3)	FE (4)	OLS (5)	FE (6)	OLS (7)	FE (8)	OLS (9)	FE (10)
Total value: cattle	0.029	-0.090	0.107^{+}	-0.092^{+}	0.045	-0.046	0.029	-0.054	0.044	-0.111^{*}
	(0.096)	(0.113)	(0.056)	(0.052)	(0.049)	(0.050)	(0.050)	(0.045)	(0.055)	(0.055)
Two cows and more	-1701.516	-2339.702								
	(1162.667)	(1528.806)								
Two cows and more × Total value: cattle	0.084	0.128								
	(0.110)	(0.113)								
Three cows and more			-3053.640^{+}	-3853.956^{+}						
			(1680.555)	(2299.881)						
Three cows and more × Total value: cattle			0.035	0.174						
			(0.085)	(0.122)						
Four cows and more					-4531.020^{*}	-9525.026^{*}				
					(2016.108)	(4399.162)				
Four cows and more × Total value: cattle					0.126	0.211^{+}				
					(0.084)	(0.127)				
Five cows and more							-5200.643^{*}	-15109.662^{**}		
							(2164.513)	(5345.212)		
Five cows and more × Total value: cattle							0.158^{+}	0.324^{*}		
							(0.089)	(0.126)		
Six cows and more									-7313.548^{**}	-19480.318^{**}
									(2591.213)	(6330.690)
Six cows and more × Total value: cattle									0.175	0.469^{*}
									(0.106)	(0.198)
Marginal return at minimum farm size:	0.113	0.038	0.142^{+}	0.082	0.170^{+}	0.165	0.187^{*}	0.269^{*}	0.219^{*}	0.359^{*}
	(0.073)	(0.085)	(0.083)	(0.105)	(0.086)	(0.127)	(0.086)	(0.130)	(0.099)	(0.171)
Average return at minimum farm size:	-0.044		-0.035		-0.006		0.037		0.044	
Observations	972	972	972	972	972	972	972	972	972	972

Notes: Pooled OLS. Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Controls are those of main model. Std. errors (clustered at the household) in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Interacting cattle value with these dummies reveals that marginal returns become statistically significant for households who own three or more animals (c.f. Table 3.6, col. 3). The difference in returns between both groups becomes statistically significant at a threshold of four animals: households who own four animals or more have marginal returns that are between 12.6 and 21.1 percentage points higher than households who own less than four animals. As the herd size increases further, returns to cattle continue to increase, reaching 36% for households with 6 cows or female buffaloes and more (cols. 9 and 10). Average returns by farm size are displayed at the bottom of Table 3.6. As can be seen average returns become positive at herd sizes of five and higher. These results suggest that cattle farming is associated with considerable returns to scale and becomes profitable only beyond the threshold of five to six animals. In our sample, the average farmer with that many animals operates at cattle values of around INR 43,000 (US$ 927). However, economies of scale cannot explain why we find marginal returns that reach their maximum at cattle values around INR 13,000 (US$ 280) and then decrease again.

3.4.2 Returns to modern variety cows

As mentioned before, an additional explanation for the observed non-convexities in marginal returns could be related to the differences in productivity across animal breeds and value. Investing in cattle not only implies acquiring more cattle, but typically also implies exchanging animals for more productive breeds. Hence, positive and increasing marginal returns may be found even in farms where economies of scale are not being fully exploited.

In order test this hypothesis more systematically, we re-estimate average and marginal returns but now split the sample by animal breed. Table 3.7 reports estimates of marginal returns for different cattle breeds. Our pooled OLS estimates suggest that modern-variety cows and buffaloes - thus, imported European breeds and their cross-breeds - have by far the highest returns (43% annually), whereas traditional breeds and buffaloes have marginal returns close to zero. Fixed effect estimates are somewhat different, but again we find that modern variety cows have returns of 10% while all other breeds seem to generate negative or zero marginal returns. Estimates of average returns by cattle breed support this finding. While modern-variety cows have average

Table 3.7: Marginal returns to cattle by cattle breed

	(1) N	(2) Av. cattle value	(3) Av. return	(4) OLS	(5) FE
Mixed	109	7480.1	-0.03	0.144^{+}	0.183
				(0.079)	(0.125)
Cow (modern)	119	7471.2	0.04	0.427***	0.098
				(0.065)	(0.101)
Cow (traditional)	363	5351.5	-0.20	-0.047	0.013
				(0.062)	(0.133)
Buffalo (modern)	63	7419.3	-0.06	-0.016	-0.670**
				(0.089)	(0.255)
Buffalo (traditional))	318	6849.2	-0.05	0.016	-0.019
				(0.051)	(0.083)
Other than cow (modern)	827	6227.2	-0.11	0.023	-0.005
				(0.036)	(0.072)
At least 1 cow (modern)	145	7879.7	0.04	0.220*	0.066
				(0.105)	(0.128)

Notes: Column 2 reports the mean of the average value of all cows or buffaloes owned by a household. Column 3 reports the average returns of each subgroup. Columns 4 & 5 report estimates of marginal returns to cattle. Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Controls are those of main model. Std. errors (clustered at the household) in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

returns of 4%, all other varieties have negative average returns.[33] To gauge the robustness of our findings, we also split the sample by the ownership of at least one modern variety cow. Again, we find that returns are roughly 20 percentage points higher for households who own at least one modern variety cow.

Given that modern-variety cows are also among the most expensive animals (c.f. Table 3.7), these results suggest that returns to modern-variety cows (i.e. returns to acquiring a more productive animal) are at least as important as returns to scale in order to explain the found non-convexities in cattle holding.

3.4.3 Combining returns to scale and returns to quality

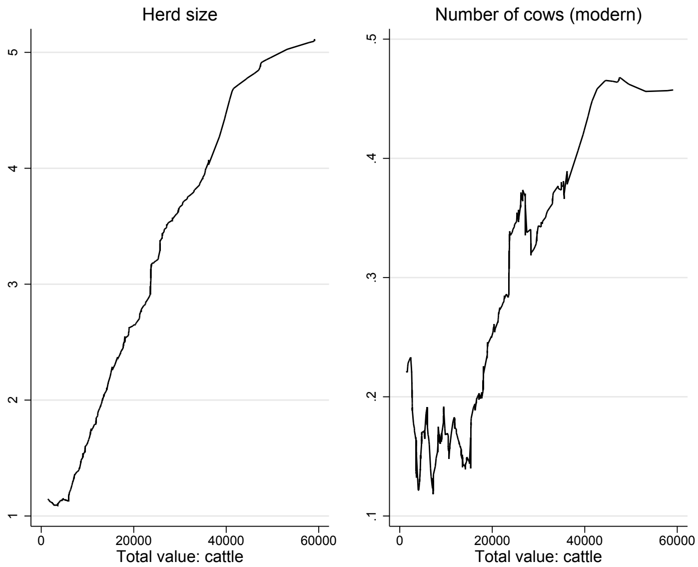
To see if the specificities of cattle farming described above are sufficient to explain the observed pattern in average and marginal returns, we plot the predicted farm size and the predicted number of modern variety cows as function of cattle value in Figure 3.4. The predicted outcomes are fitted non-parametrically using locally weighted mean smoothing with a bandwidth of 0.2. As can be seen, most farms own only one animal up to cattle values of INR 6,000 (US$ 129). Beyond this value the predicted herd size increases continuously with cattle value. The predicted ownership of modern variety cows increases most sharply in the range of cattle values from INR 14,000 to INR 26,000 (US$ 302 - 561). Given these ownership patterns, the existence of both returns to modern-variety breeds and returns to scale could explain the observed non-convexities in average and marginal returns.

At very low levels of cattle value (up to INR 6,000 or US$ 129), we find negative average returns and positive but falling marginal returns. In this range, households tend to own only one animal, and increasing cattle value probably means exchanging that animal for a more productive one. In these low levels of cattle value, average costs fall drastically with small increases in cattle value. Increasing the value of cattle thus raises costs, but only marginally when compared to the productivity gains of increasing cattle value. Increasing cattle value thus increases profits in this range, as reflected in positive marginal returns.

As cattle value increases further (INR 6,000 to 10,000, or US$ 129 to 216), the reduction in average costs with increasing cattle value slows down, leading to a decline

[33]Detailed estimates of average returns by cattle breed can be found in the Appendix, Table 3.A.7.

Figure 3.4: Predicted herd size and composition as function of cattle value



Notes: Locally weighted mean smoothing, bw= 0.2.
Source: Own estimation based on YLS data.

and eventually negative marginal returns. At cattle values above INR 6,000 the predicted herd size starts to increase with cattle value. But still very few farmers own two animals or are increasing cattle value by acquiring a modern variety cow.

At cattle values of INR 10,000 most farmers already own two animals. Also the predicted number of modern variety cows owned increases most pronouncedly in the range of cattle values between INR 14,000 and 26,000 (US$ 302 - 561). Thus most farmers are increasing the herd size from one to at least two animals, and many of them seem to be doing this by acquiring at least one modern variety cow. Given that modern variety cows are more productive, this probably explains why we find positive and increasing marginal returns in this range. Marginal returns reach their maximum at cattle values around INR 13,000 (US$ 280), while average returns continue to be negative.

Beyond cattle values of INR 13,000 marginal returns decrease continuously, but remain positive. Average returns increase, and become positive at cattle values around INR 34,000 (US$ 733). In this range, most farmers own at least one modern variety cow, and increasing cattle value means increasing herd size as shown in Figure 3.4. With increasing herd size average cost continue to decrease, particularly due to economies of time (c.f. Figure 3.3).

3.4.4 Entry barriers

There seem to be two important thresholds involved in cattle farming in India. The first threshold seems to lie at cattle values around INR 10,000 to 14,000 (US$ 216 - 302): households generally own more than one animal and shift their production towards modern variety cows. This allows them to explore some returns to scale and the higher returns of owning modern variety cows. Wile average returns in this range continue to be mostly negative, marginal returns are positive. And for those households who successfully shifted towards modern variety cows and own at least two animals average returns are also positive. The second threshold seems to lie at herd sizes greater than five and cattle values around INR 34,000 (US$ 927). Beyond this value, not only marginal returns but also average returns are positive for all farmers.

Both thresholds need to be overcome to reach herd values that generate positive average returns. It is likely that different obstacles are at play for these two thresholds. Identifying those obstacles with observational data is virtually impossible, and with this caveat in mind, we explore a number of potential explanations of how to overcome both thresholds. Table 3.8 displays a simple OLS estimation of potential determinants of modern cow ownership and of herd sizes of five animals and greater.

As can be seen, the educational level of the household head seems to be associated a higher probability of owning a modern cow. This could suggest that modern cows require better farming ability, but could also just mean, that more educated farmers are wealthier and have therefore better possibilities to finance these more expensive animals (either through self-financing or through credit). This would be in line with the finding that wealthier households (measured by the housing services index) are more likely to own these animals. Interestingly, the self-reported access to a number of government programs seems to increase the probability of owning a modern variety cow, which suggest that knowledge and access to information about new breeds are important, as is the cost of accessing veterinary services and insemination facilities (see negative coefficient on the distance to veterinary hospitals). Another reason for the limited ownership of modern variety cows could be that fodder expenses are usually higher for these animals (Turner, 2004). Supportive evidence for higher fodder expenses

Table 3.8: Determinants of ownership of modern variety cows and minimum herd size

	Modern cow ownership			Herd size ≥ 5		
	(1)	(2)	(3)	(4)	(5)	(6)
Age of hh head	0.000	-0.001	-0.003	0.000	0.000	0.001
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
Male household head	-0.002	-0.023	0.093	-0.038	-0.033	-0.016
	(0.053)	(0.051)	(0.118)	(0.043)	(0.046)	(0.067)
Household head is literate	0.042^{+}	0.041	0.000	0.024	0.023	0.000
	(0.025)	(0.025)	(.)	(0.021)	(0.019)	(.)
Housing quality index	-0.019	-0.047	0.025	-0.059^{*}	-0.063^{*}	0.033
	(0.056)	(0.060)	(0.096)	(0.029)	(0.030)	(0.057)
Consumer durables index	-0.116	-0.106	-0.245	-0.038	-0.035	-0.180
	(0.080)	(0.085)	(0.157)	(0.058)	(0.072)	(0.131)
Housing services index	0.250^{*}	0.214^{*}	0.301^{*}	0.062	0.015	-0.053
	(0.102)	(0.095)	(0.149)	(0.075)	(0.047)	(0.068)
Income, non-farm activities (log)	-0.011	-0.005	0.005	0.018^{***}	0.018^{***}	0.002
	(0.009)	(0.009)	(0.015)	(0.005)	(0.005)	(0.009)
Value of agr. production (log)	0.000	-0.003	0.001	0.003	0.002	0.002
	(0.004)	(0.004)	(0.008)	(0.002)	(0.002)	(0.003)
Total land owned (acres, log)	0.003	0.026	0.066	0.043^{**}	0.054^{**}	0.083^{*}
	(0.018)	(0.021)	(0.050)	(0.016)	(0.017)	(0.031)
Household registered with NREGA	-0.084^{*}	-0.071	-0.147^{*}	-0.016	-0.027	0.020
	(0.039)	(0.044)	(0.067)	(0.030)	(0.029)	(0.037)
Hh benefits from DWCRA	0.001	-0.040	-0.134	-0.015	-0.013	0.037
	(0.049)	(0.058)	(0.082)	(0.025)	(0.023)	(0.046)
Hh benefits from IKP	-0.011	0.015	0.021	0.044	0.029	-0.009
	(0.027)	(0.029)	(0.042)	(0.028)	(0.023)	(0.029)
Hh benefits from PMRY	-0.040	-0.036	0.273^{*}	-0.022	-0.064	-0.052
	(0.037)	(0.074)	(0.128)	(0.047)	(0.113)	(0.077)
Hh benefits from CMEY/Rajivy	0.479	0.478	0.694^{**}	-0.039^{*}	0.019	0.019
	(0.319)	(0.296)	(0.220)	(0.018)	(0.029)	(0.031)
Hh benefits from SGSY	-0.049	-0.054	-0.008	-0.070^{+}	-0.087^{*}	0.012
	(0.080)	(0.061)	(0.105)	(0.036)	(0.038)	(0.039)
Hh benefits from other program	0.171^{*}	0.189^{*}	0.169^{*}	-0.012	0.013	0.010
	(0.065)	(0.073)	(0.082)	(0.035)	(0.036)	(0.051)
Distance to veterinary hospital	-0.008^{***}	-0.004^{*}	-0.002	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Availability: Cattle development program	0.064	-0.036	0.000	-0.063^{**}	-0.049^{+}	0.000
	(0.061)	(0.052)	(.)	(0.020)	(0.026)	(.)
Availability: Dairy development program	0.101^{+}	-0.054	0.000	-0.029	-0.079	0.000
	(0.054)	(0.048)	(.)	(0.023)	(0.069)	(.)
Availability: Free Veterinary camp	0.071^{*}	0.039	0.000	0.020	0.039^{+}	0.000
	(0.032)	(0.035)	(.)	(0.025)	(0.021)	(.)
Year 2009 (dummy)	0.139^{**}	0.124^{*}	0.120^{*}	-0.020	-0.008	0.024
	(0.052)	(0.053)	(0.056)	(0.016)	(0.015)	(0.020)
Fixed Effect	No	Sub-district	Household	No	Sub-district	Household
Observations	969	969	969	969	969	969
R^2	0.123	0.170	0.153	0.068	0.108	0.050

Notes: Pooled OLS. Cells report estimates of average marginal effect. Standard errors (clustered at the village) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

is the fact that modern variety cows have higher returns to fodder expenses than other breeds (as reported in the Appendix, Table 3.A.8). Finally, potential differences in the variability of returns of modern variety cows as compared to other breeds could explain limited adoption of new varieties. However, this cannot be assessed with the data at hand.

In contrast, only income and wealth related variables seem to predict the probability of owning five cows or buffaloes and more: only non-agricultural income and land ownership have positive and statistically significant coefficients.[34] This suggests that credit constraints are among the most important obstacles to reach minimum herd sizes of five animals.

3.5 Conclusions

This paper addresses the apparent puzzle of widespread support of cattle farming through agricultural policy interventions vis-à-vis largely negative returns to cattle, as stressed in recent works. To get a more in-depth impression of the profitability of cattle farming, we explore average and marginal returns to cattle at different levels of cattle value and for different breeds in Andhra Pradesh, India. The results of this paper are as follows. We find that average returns to cattle are negative by 8% at the mean of cattle value and vary between large negative rates at low cattle values and positive rates at high cattle values. Similar to Attanasio and Augsburg (2014), we find that returns increase considerably with favorable weather conditions. In contrast to average returns, marginal returns to cattle are found to be positive on average. At the mean of cattle value, marginal returns range between 9% and 16% annually, depending on the specification considered. Whereas average returns increase with cattle value, marginal returns seem to follow a U-shaped pattern, with the highest returns materializing at extremely low and above-average cattle values.

These estimates are quite substantial and indicate that investing in cattle could be a viable strategy for households in rural areas of Andhra Pradesh. But we also find strong evidence that herd size and quality matter. The fact that only households

[34] Interestingly the existence of cattle development programs at village level seems to reduce the probability of owning five animals and more. As does the self-reported access to SGSY (a credit program targeted at self-help groups). But these could both be selection effects rather than 'treatment' effects.

operating on a larger scale as well as households with the highest-value animals have positive average returns suggests that high entry costs prevent many households from operating at profitable levels. These entry barriers would also explain the observed non-convexities in marginal returns.

Two types of entry barriers were identified: first, economies of scale associated with the substantial cost savings of owning more than one animal; second, differences in animal prices and productivity across cattle breeds. This paper also discusses a few potential explanations of how these entry barriers could be overcome. We find suggestive evidence that access to information, to veterinary services and to adequate fodder seem to matter for the adoption of modern variety cows. But the most important obstacle to overcoming both entry barriers seem to be credit constraints: wealth and income seem to explain both the adoption of modern variety cows and the probability to operate at herd sizes larger than five.

This is not surprising: As we saw in Section 3.3.1, the average market value of a fertile crossbred cow is about INR 10,500 (or US$ 226), and in many cases considerably higher. In contrast, the average value of cows and buffaloes in the sample is roughly INR 6,500 (US$ 140), thus just over half this value. That the cattle value in the sample is consistently below reported market prices suggests that most households in our sample might face difficulties in raising the resources to finance the investment in a high-value animal. The average household income of non-cattle farmers in the sample is INR 30,700 (US$ 662) per year: this is not even three times the market value of a crossbred cow. We also saw that average returns to cattle become positive at cattle values above INR 34,000 (US$ 733), which is more than the total annual income of these households.

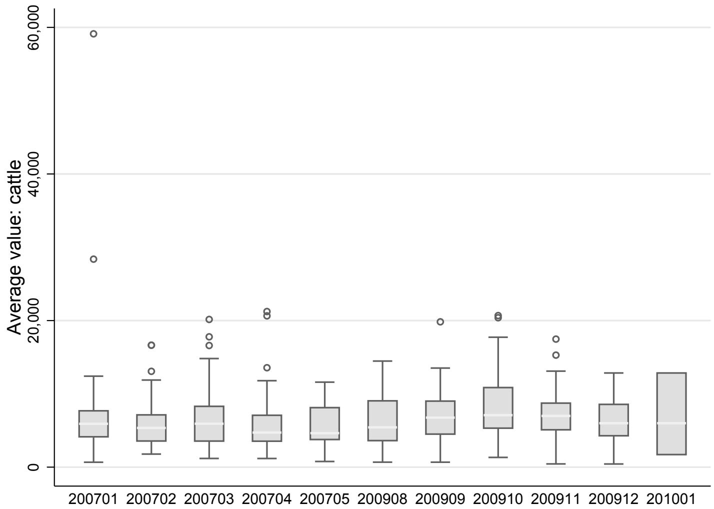
The results of this paper suggest that non-convexities in returns to livestock farming trap poor households in low-productivity asset levels. This finding explains why policy interventions to increase investments in cattle seem to fail in rural India, as stipulated by Morduch et al. (2013). Households can only reach a level of positive average returns to cattle and start on a beneficial accumulation path if they overcome considerable entry barriers. This is obviously harder for poorer households, which would be the potential beneficiaries of asset-based anti-poverty policies.

In terms of policy implications, the results of this paper suggest that policies such as the "One cow per poor family" and "Targeting the ultra poor", can only have lasting impacts on poverty if beneficiaries are enabled - for instance through credits - to invest enough in the quantity and quality of their cattle, thereby ensuring the profitability of the investment.

There could obviously be other reasons beyond financial profits that motivate poor households to hold low values of cattle despite the negative or at least very low returns. This may have to do with a preference for own milk products or because households use cattle as a intertemporal savings device. Exploring these motivations is beyond the scope of this paper and is left for future research.

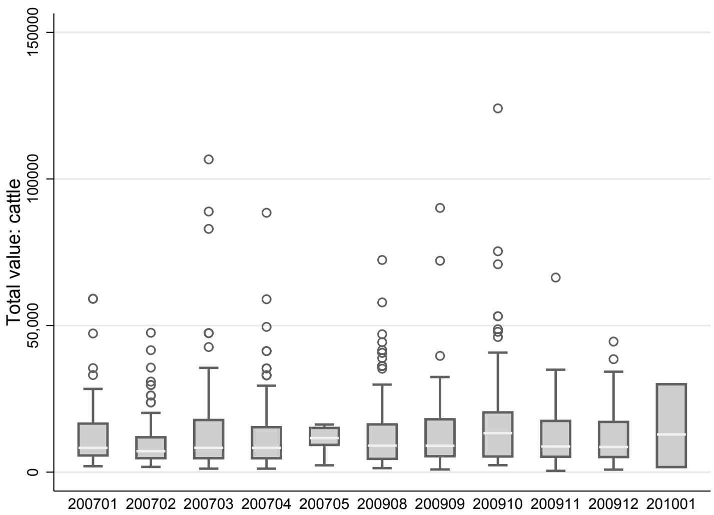
3.A Supplementary Figures and Tables

Figure 3.A.1: Average value of cattle by month of interview



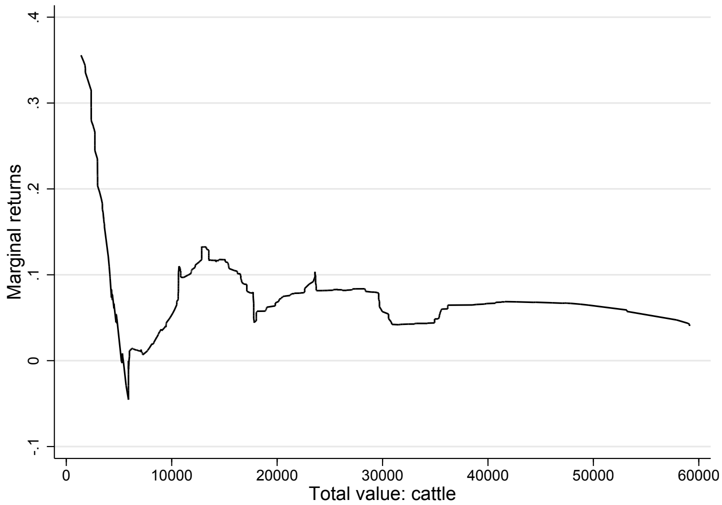
Source: Own estimation based on YLS data.

Figure 3.A.2: Total value of cattle by month of interview



Source: Own estimation based on YLS data.

Figure 3.A.3: Semiparametric estimation of marginal returns to cattle with heterogeneous wages



Notes: Locally weighted mean smoothing, bw= 0.5.
Source: Own estimation based on YLS data.

Table 3.A.1: Average returns by quintile of cattle value (alternative cost estimates)

Quintiles of cattle value	Average cattle value*	Quantity: cattle	Total value: cattle	Revenue: dairy	Value: calves	Fodder cost	Vet. cost	Other cost	Labor cost	Depre-ciation	Profits	ROR
	A	B	C	D	E	F	G	H	I	J	K=D+E-F -G-H-I-J	L=K/C
1	2862	1.12	3002	1705	539	661	86	18	2293	600	-1414	-0.47
2	5174	1.20	5733	2433	810	883	95	37	2284	1147	-1202	-0.21
3	6833	1.51	8946	2780	1294	1121	161	58	2836	1789	-1892	-0.21
4	8015	2.18	14776	5289	1704	1580	235	57	3496	2955	-1332	-0.09
5	9574	3.91	33081	12949	3566	4236	536	105	4700	6616	321	0.01
Total	6474	1.97	12924	4954	1564	1670	220	55	3110	2585	-1120	-0.09
N	972											

Notes: Cells report the mean value of the variable of each column for a given quintile. All values are constant INR (July 2006). One US$ is equivalent to 46.38 INR. Rate of return (ROR) calculated as follows: $(\pi_t - (K_t - K_{t+1}))/K_t$, where K_t is the value of the animal at time t. * Average value of all adult cows and buffaloes owned by household.

Table 3.A.2: Yearwise average returns by quintile of cattle value

Quintiles of cattle value	Average cattle value*	Quantity: cattle	Total value: cattle	Revenue: dairy	Value: calves	Cost: dairy	Labor cost	Depre-ciation	Profits	ROR
	A	B	C	D	E	F	G	H	I=D+E -F-G-H	J=I/C
2007 (N = 463)										
1	2863	1.08	2961	465	333	228	1684	592	-1707	-0.57
2	5147	1.23	5782	1694	564	604	1813	1156	-1315	-0.23
3	6136	1.69	8926	2134	648	696	2084	1785	-1783	-0.20
4	6993	2.45	14364	4159	1091	1206	2658	2873	-1487	-0.10
5	10038	3.78	32123	10564	1596	4543	3256	6425	-2063	-0.06
Total	6080	1.97	12153	3599	816	1380	2251	2431	-1647	-0.14
2009 (N = 509)										
1	2861	1.15	3042	2908	739	1649	2883	608	-1494	-0.50
2	5216	1.15	5656	3605	1200	1704	3030	1131	-1060	-0.19
3	7318	1.39	8961	3229	1743	1383	3360	1792	-1563	-0.17
4	8708	1.99	15055	6055	2119	1843	4064	3011	-745	-0.05
5	9171	4.03	33914	15021	5277	4735	5955	6783	2824	0.08
Total	6832	1.96	13625	6187	2244	2243	3891	2725	-428	-0.03

Notes: Cells report the mean value of the variable of each column for a given quintile. All values are constant INR (July 2006). One US$ is equivalent to 46.38 INR. Rate of return (ROR) calculated as follows: $(\pi_t - (K_t - K_{t+1}))/K_t$, where K_t is the value of the animal at time t. * Average value of all cows and buffaloes owned by household.

Table 3.A.3: Marginal returns to cattle by rainfall conditions

	(1)	(2)
Total value: cattle	0.116^{+}	
	(0.069)	
Total value: cattle × Rainfall (deviation)	0.180^{*}	
	(0.088)	
Rainfall (deviation)	-44.321	
	(1014.967)	
Marginal returns to cattle		
at Rainfall (dev., lag)= -0.5		0.026
		(0.074)
at Rainfall (dev., lag)= -0.4		0.044
		(0.070)
at Rainfall (dev., lag)= -0.3		0.062
		(0.068)
at Rainfall (dev., lag)= -0.2		0.080
		(0.067)
at Rainfall (dev., lag)= -0.1		0.098
		(0.068)
at Rainfall (dev., lag)= 0		0.116^{+}
		(0.069)
at Rainfall (dev., lag)= 0.1		0.134^{+}
		(0.071)
at Rainfall (dev., lag)= 0.2		0.152^{*}
		(0.075)
at Rainfall (dev., lag)= 0.3		0.170^{*}
		(0.079)
at Rainfall (dev., lag)= 0.4		0.188^{*}
		(0.084)
at Rainfall (dev., lag)= 0.5		0.206^{*}
		(0.089)
Observations	972	972

Notes: Pooled OLS. Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Controls are those of main model. Coefficients in first and marginal returns in second column. Std. errors (clustered at the village) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.A.4: Hetergogeneity of marginal returns by household characteristics

	(1)	(2)	(3)	(4)
Total value: cattle	0.028	0.005	-0.204^{+}	-0.254^{+}
	(0.050)	(0.046)	(0.111)	(0.151)
Household head is literate	-806.535			
	(998.389)			
Total value: cattle × Household head is literate	0.103			
	(0.083)			
Person resp. for livestock is literate		-700.820		
		(1034.497)		
Total value: cattle × Person resp. for livestock is literate		0.163^{+}		
		(0.085)		
Wealth index			-6513.209*	
			(2643.811)	
Total value: cattle × Wealth index			0.572*	
			(0.230)	
Housing services index				-6056.276*
				(3038.446)
Total value: cattle × Housing services index				0.530*
				(0.254)
Observations	972	972	972	972
R^2	0.121	0.143	0.135	0.143

Notes: Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Controls are those of main model, except col. 4 does not control for wealth index. The wealth index is the simple average of the housing quality index, the consumer durable index and the housing services index. Housing quality is the simple average of rooms per person and indicator variables for the quality of roof, walls and floor. Consumer durables are the scaled sum of 12 variables indicating the ownership of items such as radios, fridges, televisions, phones or vehicles. Services are calculated as the simple average of dummy variables indicating households' access to drinking water, electricity, toilets and fuels. For more information on the wealth index refer to the Young Lives data justification documents at http://www.younglives.org.uk. Std. errors (clustered at the household) in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.A.5: Sensitivity of results to omission and alternative calculation of labor cost

	Pooled OLS		2007 OLS		Pooled OLS
	(1)	(2)	(3)	(4)	(5)
Total value: cattle	0.086^{+}	0.069	-0.035	0.019	0.093*
	(0.044)	(0.051)	(0.079)	(0.081)	(0.046)
Predicted time spent on cattle (hours per year)		-3.001			
		(6.344)			
Predicted time spent on cattle (squared)		0.004			
		(0.006)			
Time spent on cattle (hours per year)				-5.797***	
				(1.070)	
Time spent on cattle (squared)				0.001*	
				(0.001)	
Total land owned (acres, log)	-356.316	-378.316	-4.902	-355.807	-580.385^{+}
	(303.658)	(301.941)	(435.938)	(418.810)	(329.389)
Expenditure on cattle: fodder (log)	258.994***	256.108***	197.489*	244.674**	200.074**
	(62.159)	(61.920)	(80.697)	(74.893)	(64.260)
Income, non-farm activities (log)	798.665***	803.647***	361.722*	437.177**	717.978***
	(136.688)	(136.631)	(171.243)	(167.003)	(154.772)
Household size	-218.000*	-227.188*	-147.522	-135.337	-193.901^{+}
	(97.150)	(96.863)	(126.415)	(116.599)	(102.212)
Age of hh head	-28.489	-27.974	-29.744	-14.546	-29.680
	(18.837)	(18.933)	(21.042)	(20.279)	(19.377)
Highest grade: hh head	80.688	84.380	45.852	80.824	-98.741
	(75.274)	(75.128)	(75.772)	(71.511)	(85.619)
Male household head	401.121	424.400	545.516	845.043	157.013
	(895.848)	(903.683)	(1009.687)	(1026.958)	(915.638)
Wealth index	1433.428	1482.132	1209.265	990.568	76.309
	(1503.474)	(1494.771)	(2063.193)	(1905.076)	(1757.445)
Shock affected livestock	-92.046	-162.283	-707.349	-728.712	-424.847
	(711.934)	(711.067)	(865.848)	(795.018)	(755.351)
Rainfall (deviation)	2099.687**	2111.613**	-867.956	-1018.930	1829.446*
	(688.923)	(681.215)	(1206.972)	(1126.193)	(722.421)
Year 2009 (dummy)	626.140	621.464			522.635
	(399.126)	(390.594)			(415.272)
Observations	972	972	463	463	972
R^2	0.111	0.112	0.046	0.201	0.096

Notes: Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Cols. 1 & 2 use predicted time allocation to calculate profits and as regressor, cols. 3 & 4 use self-reported time allocation. Column 5 uses alternative wages to calculate profits, as described in Section 3.3. Std. errors (clustered at the household) in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.A.6: Sensitivity of results to omission and alternative calculation of fodder cost

	(1)	(2)	(3)	(4)	(5)
Total value: cattle	0.086^{+}	0.101*	0.034	0.101*	0.088^{+}
	(0.044)	(0.045)	(0.043)	(0.051)	(0.047)
Expenditure on cattle: fodder (log)	258.994***			147.942*	96.262
	(62.159)			(69.271)	(68.055)
Expenditure on cattle: fodder			0.314		
			(0.311)		
Expenditure on cattle: fodder (squared)			0.000		
			(0.000)		
Total land owned (acres, log)	-356.316	-602.031*	-260.433	-221.811	-196.378
	(303.658)	(296.833)	(286.100)	(328.317)	(313.280)
Income, non-farm activities (log)	798.665***	794.134***	814.601***	929.750***	922.924***
	(136.688)	(134.962)	(137.803)	(149.569)	(147.206)
Household size	-218.000*	-228.107*	-185.926^{+}	-240.121*	-227.081*
	(97.150)	(98.728)	(98.491)	(106.598)	(105.666)
Age of hh head	-28.489	-27.915	-31.687^{+}	-29.654	-27.963
	(18.837)	(19.146)	(18.829)	(20.239)	(20.030)
Highest grade: hh head	80.688	89.657	61.599	64.575	58.113
	(75.274)	(76.689)	(70.002)	(82.505)	(79.681)
Male household head	401.121	354.176	127.843	315.041	246.300
	(895.848)	(907.855)	(840.722)	(898.113)	(871.696)
Wealth index	1433.428	2147.341	1183.239	2023.194	1978.738
	(1503.474)	(1509.226)	(1478.330)	(1704.590)	(1679.121)
Shock affected livestock	-92.046	-72.547	-113.549	-622.808	-668.223
	(711.934)	(715.084)	(708.858)	(780.388)	(762.871)
Rainfall (deviation)	2099.687**	2071.893**	2192.682**	3354.262***	3319.264***
	(688.923)	(680.934)	(682.694)	(852.941)	(836.870)
Year 2009 (dummy)	626.140	521.733	620.217	1169.990**	1072.268*
	(399.126)	(396.644)	(381.069)	(440.348)	(428.197)
Observations	972	972	972	972	972
R^2	0.111	0.096	0.134	0.114	0.104

Notes: Pooled OLS. Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Cols. 1-3 use cost estimates from the income section (as described in Section 3.2) to calculate profits. Col. 4 uses cost estimates from the livestock section (as described in Section 3.2) to calculate profits. Column 5 also uses cost estimates from the livestock sections, but assumes that calves also require fodder. Std. errors (clustered at the household) in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.A.7: Average returns by cattle breed

Breeds	N	Average cattle value*	Quantity: cattle	Total value: cattle	Revenue: dairy	Value: calves	Cost: dairy	Labor cost	Depre-ciation	Profits	ROR
		A	B	C	D	E	F	G	H	I=D+E -F-G-H	J=I/C
Mixed	109	7480	3.66	27520	9322	3143	3189	4677	5151	-906	-0.03
Cow (modern)	119	7471	1.45	11200	6669	1405	2547	2846	2192	441	0.04
Cow (traditional)	363	5352	1.74	9448	2183	1717	934	2989	1914	-1912	-0.20
Buffalo (modern)	63	7419	1.68	12374	5699	1060	2667	2419	2646	-802	-0.06
Buffalo (traditional)	318	6849	1.90	12641	5832	1007	1960	2945	2606	-595	-0.05
Total	972	6474	1.97	12924	4954	1564	1832	3110	2585	-1008	-0.08

Notes: Cells report the mean value of the variable of each column for a given cattle breed. All values are constant INR (July 2006). One US$ is equivalent to 46.38 INR. Rate of return (ROR) calculated as follows: $(\pi_t - (K_t - K_{t+1}))/K_t$, where K_t is the value of the animal at time t. * Average value of all cows and buffaloes owned by household.

Table 3.A.8: Heterogeneity in returns to fodder

	OLS	FE	OLS		FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Total value: cattle	0.036	0.018	0.047		0.021	
	(0.044)	(0.066)	(0.043)		(0.067)	
Expenditure: fodder	0.564***	-0.156	0.434		0.140	
	(0.167)	(0.459)	(0.370)		(0.597)	
Cow (modern)			1371.677		2938.965	
			(1373.738)		(2002.134)	
Cow (traditional)			869.671		3331.714*	
			(1042.948)		(1603.183)	
Buffalo (modern)			2704.729*		7025.467**	
			(1329.885)		(2310.671)	
Buffalo (traditional)			739.921		2973.573	
			(1074.054)		(1928.451)	
Cow (modern) × Expenditure: fodder			0.697		-0.722	
			(0.680)		(0.954)	
Cow (traditional) × Expenditure: fodder			-0.110		-0.417	
			(0.420)		(0.657)	
Buffalo (modern) × Expenditure: fodder			-0.585		-2.693***	
			(0.455)		(0.776)	
Buffalo (traditional) × Expenditure: fodder			0.363		0.534	
			(0.401)		(0.677)	
Marginal return to fodder for:						
Mixed				0.434		0.140
				(0.370)		(0.597)
Cow (modern)				1.131*		-0.582
				(0.570)		(0.779)
Cow (traditional)				0.324		-0.277
				(0.241)		(0.278)
Buffalo (modern)				-0.152		-2.553***
				(0.274)		(0.495)
Buffalo (traditional)				0.797***		0.674+
				(0.209)		(0.353)
Observations	972	972	972		972	
R^2	0.131	0.122	0.160		0.271	

Notes: Linear production function assumed. Dep. var: Profits (adj. for labor) - depreciation. Controls are those of main model. Columns 4 & 6 report marginal returns to fodder for each cattle breed. Std. errors (clustered at the household) in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4

Wage risk, labor supply and human capital accumulation in India

with Andrew Foster

4.1 Introduction

Risk affects household behavior in several ways. Previous literature has shown that uncertainty regarding future consumption and the inability of households to insure against shocks affects their current consumption and saving (Rosenzweig and Wolpin, 1993; Udry, 1995), as well as their investment and technology adoption decisions (Dercon, 1996; Karlan et al., 2014; Rosenzweig and Binswanger, 1993).

Only a few papers look at the effect of risk on labor supply. Rose (2001) for example addresses the link between weather risk and off-farm labor supply and shows that variability in rainfall increases labor force participation of households because households need to accumulate savings for later periods. In the context of OECD countries, Pistaferri (2003) links wage risk with the elasticity of intertemporal substitution in labor supply and finds that wage risk increases labor supply in the early years of labor force participation and reduces the intertemporal substitution in hours worked associated with changing wages over the life-cycle.

In India, the context of our study, increased labor supply – especially of women – would be desirable from a policy perspective. Literature has shown that increased own income improves the intra-household bargaining power of women and the social status of women in general (Basu, 2006; Jensen, 2012). Likewise, it has been shown that children in households with working women are generally healthier (Kennedy and Peters, 1992) and better educated (Afridi et al., 2012).

However, wage work is not the only activity of women in India. In fact, the India Time Use survey from 1998/99 revealed that women spend more time than men on activities other than leisure. While women spend 53.4 hours per week on salaried activities and household work combined, men only spend 45.6 hours per week on these activities (Government of India, 2000).[1] One might therefore expect that an increase in labor market work has to go hand-in-hand with a reduction of household-related work for women.

In this paper, we seek to understand the extent to which risk raises labor supply to levels that can become harmful for other members of the household. In the presence of intra-household substitution effects, for instance in the performance of household chores, increased female labor supply might have negative effects on the time allocation of children. If women have less time available for home production and childcare, and such activities can only be foregone at high cost, they might be forced to take older children out of school or to cut down on the time these children study at home in order for them to fill in for these tasks (Ilahi, 2000; Skoufias, 1993). Under such circumstances, risk would not only affect labor supply decisions of adult household members, but potentially also have severe consequences on other members of these households and on their human capital accumulation in particular.

The relationship between labor supply and children's outcomes is unlikely to be linear. Labor supply decisions are very different at the extensive or the intensive margin (Heckman, 1974, 1993), as are likely to be the effects on other household members. Understanding what factors drive labor supply decisions at different margins and how these affect other household members is crucial for the adequate design of policy responses.

[1] Women spend on average 18.7 hours per week on wage work 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.

There is, to the best of our knowledge, no evidence so far on the effects of risk on intra-household substitution in different activities and on the inter-linkages between risk, adult labor supply and the time allocation of children. In this paper, we focus on the effect of risk on the time allocation of children between household work and school, and try to determine to what extent this can be attributed to adults' labor supply. Because we are interested in intra-household substitution of home activities, we concentrate on female labor supply and girls' time allocation. We thereby contribute to existing research on the effects of risk on labor supply (Jayachandran, 2006; Pistaferri, 2003; Rose, 2001), to previous work that estimates labor supply elasticities in developing countries (Abdulai and Delgado, 1999; Bardhan, 1979; Goldberg, 2014) and, lastly, we complement existing evidence on the effects of shocks on children's human capital accumulation (Beegle et al., 2006; Duryea et al., 2007; Gubert and Robilliard, 2008; Jacoby and Skoufias, 1997; Jensen, 2000; Skoufias and Parker, 2006).

In order to address the questions outlined above, this paper develops a model of household time allocation and human capital accumulation that highlights the effect of uncertainty regarding future consumption on a child's school time. Following Jacoby and Skoufias (1997), we model labor supply and schooling decisions in a unitary household. In order to incorporate time allocation to home production, we assume that the household derives utility from two consumption goods, one of which is produced at home and the other on the market. The model is flexible about the degree of substitutability between these goods. In line with previous literature, the model predicts that adults allocate more time to the labor market if they face uncertainty regarding future income and consumption. Likewise, risk leads to a reduction in school time of children because they have to allocate more time to home production.

We test these predictions in the context of rural India, where female labor force participation is higher than in urban areas (24.5% vs. 16.7% in 2009/10)[2] and where time constraints seem to be more important: the average time allocated by women aged 6 and above to salaried activities and household work is 56.5 hours per week in rural India, as opposed to 45.6 hours per week in urban areas (Government of India, 2000). We predict wage risk at the village level as a function of the historical rainfall distribution and a village's share of land that is under irrigation.

[2]According to the NSSO statistics, obtained from Indiastat.

We find that wage risk affects the time allocation of women, increasing their labor supply and reducing the time in home production. We also find that wage risk increases the time girls spend on household chores, and reduces their time in school. We conduct a number of robustness checks to understand if the heterogeneity in observed effects is in line with the model's predictions. Finally, we simulate the effects of a wage-smoothing policy (such as the Indian Employment Guarantee) on household decisions and show that the policy could mediate the effect of risk on the time working women allocate to household chores, allowing girls to spend less time on household chores and more time in school.

The remainder of the paper proceeds as follows. Section 4.2 discusses the theoretical model. Section 4.3 presents the data and Section 4.4 the estimation strategy. Results are discussed in Section 4.5. Section 4.6 simulates the effect of the NREGS on the outcomes of interest, and Section 4.7 concludes.

4.2 A model of household time allocation and human capital investment

4.2.1 General setup

In order to understand the effect of wage risk on labor supply, the intra-household substitution in tasks and time in school, we extend the model developed by Jacoby and Skoufias (1997) to a setting with two consumption goods, one of which is produced at home and the other on the market.

The central assumption in the model is that households derive a positive utility from educating their children. Now, assume that the household has one child of school age, and that the household forms expectations over consecutive agricultural seasons. The beginning-of-period stock of human capital, H_{it}, can be augmented each period by school attendance, S_{it}. We follow Jacoby and Skoufias (1997) in specifying a learning technology that allows the stock of human capital to be larger if school attendance is stable than if school attendance is variable, but with the same mean,

$$H_{it+1} = g(H_{it}, S_{it}; \theta_{it}), \tag{4.1}$$

where g is increasing in H_{it} and S_{it}. θ_{it} represents an education productivity shifter, which reflects the effect of child illness or under-nutrition on human capital accumulation. The marginal rate of transformation in school attendance between both periods is,

$$z_{it} = -\frac{dS_{it}}{dS_{it-1}} = g_{H_t}\frac{g_{S_{t-1}}}{g_{S_t}} \tag{4.2}$$

where $g_{H_t} = \partial g(H_{it}, S_{it}; \theta_{it})/\partial H_{it}$ and so forth.[3] z_{it} thus represents the cost of one unit less education in period $t-1$ in units of period t school attendance. Given that $g_{H_t} > 0$, this cost will always be greater than the ratio of the marginal product of school time in both periods.

The household's value function is the sum of the discounted utility of consumption U of all periods and the utility of end-of-schooling phase education ϕ,

$$V = \sum_{t=1}^{T} \delta^t U(C_{it}^m, C_{it}^h) + \phi(H_{iT+1}) \tag{4.3}$$

where C_{it}^m is the consumption of the market good of household i in period t, C_{it}^h the consumption of the home-produced good, H_{iT+1} the end-of-schooling stock of human capital and δ^t the discount factor. The model ignores leisure both of adults and of the child as unnecessary complication.

The household is subject to two budget constraints in each period: the savings constraint,

$$A_t \leq w_t(T_t^a - h_t^a) + (1+r)A_{t-1} - C_{it}^m, \tag{4.4}$$

and the time constraint on school attendance,

$$S_t \leq (h_t^a + T_t^c) - \frac{1}{\rho}C_{it}^h, \qquad S_t \leq T_t^c. \tag{4.5}$$

In this setup, the only cost of school attendance is the foregone time allocated to home production. In contrast to Jacoby and Skoufias (1997), we assume that there is no market for child labor, i.e. the child can only contribute to the home-produced good,

[3]To see this more explicitly, consider that $H_{it+1} = g(H_{it}(H_{it-1}, S_{it-1}; \theta_{it-1}), S_{it}; \theta_{it})$. Taking the differential and setting it equal to zero, yields $dS_{it}g_{S_t} + dS_{it-1}g_{H_t}g_{S_{t-1}} = 0$.

while the parents can work for wages w_t in order to purchase the market-produced good or allocate their time to home production h_t^a. T_t^c and T_t^a are the total time endowments of the child and the adults respectively, and ρ is the marginal product of time in home production. Assuming that the child cannot work for wages but only in the household is a simplifying assumption, which helps to highlight that uncertainty can affect school time even in the absence of a market for child labor.

In such a framework, the time allocation of adults between labor market time and home production is governed by the equality of the marginal rate of substitution between the two consumption goods to the ratio of the market wage to the marginal product of home production,

$$\frac{U_{C_{it}^h}}{U_{C_{it}^m}} = \frac{w_{it}}{\rho}. \tag{4.6}$$

For any increase in market wages, households will change their consumption bundle towards the market produced good. This can happen by reducing consumption of the home produced good, by increasing consumption of the market good or both. Given the shift in consumption, households will also shift the allocation of adult time away from home production towards the labor market.

4.2.2 Effect of wage risk on each period's school time

In order to highlight the effect of wage risk on each period's school attendance, we introduce uncertainty in wages, and set up the household maximization problem in a world with perfect predictability and in a world with uncertainty and risk aversion. Consider the decision rules of the penultimate period, $T-1$.[4]

In a world with perfect predictability, the household maximizes utility subject to the savings and time constraints. Dropping individual subscripts, the decision rule with respect to school time in period $T-1$ is described by

$$g_{S_{T-1}} = \frac{w_{T-1}\delta^{T-1}U_{C_{T-1}^m}}{\phi_{H_{T+1}}g_{H_T}}. \tag{4.7}$$

[4]Details as well as a more general derivation of decision rules for $t \neq T-1$ can be found in the Mathematical Appendix.

As can be seen, the marginal product of school time is a function of the marginal utility of market-good consumption, of current wages, as well as of the marginal utility of end-of-schooling education and the marginal product of human capital stock on period T education production. g_{H_T} is an increasing function of the entire human capital stock accumulated until the end of period T, which at $T-1$ includes future school time, S_T. The allocation of time to schooling in $T-1$ thus increases in past school time as well as in the time allocated to school in T. School time also increases in the relative value given to human capital stock vis-à-vis market-good consumption. Because parents will shift their time away from home production towards the labor market with increasing wages, and home production can be performed by the parents as well as by the child, school time decreases in current wages.

Now, if the household faces uncertainty regarding period T wages and therefore market-good consumption, the decision rule with respect to school time in period $T-1$ changes to

$$g_{S_{T-1}} = \frac{w_{T-1}\delta^{T-1}U_{C^m_{T-1}}}{E[\phi_{H_{T+1}}]E[g_{H_T}] + cov(\phi_{H_{T+1}}, g_{H_T})}. \tag{4.8}$$

Risk affects the decision regarding how much time to allocate to schooling in two ways. First, because households increase savings in the presence of uncertainty, they will have to reduce school time. To see this, consider the intertemporal marginal rate of substitution in market good consumption,

$$\frac{1}{(1+r)} = \frac{\delta EU_{C^m_T}}{U_{C^m_{T-1}}}. \tag{4.9}$$

Since the marginal expected utility is higher than the marginal utility of the expected value for any risk-averse household, households will have to increase their savings in period $T-1$ in the presence of uncertainty to ensure sufficient consumption of the market good in period T in the presence of a shock. Since this entails increased hours supplied to the labor market but doesn't shift the ratio of consumption between the home-produced and the market good in $T-1$, children will have to reduce their time in school in order to substitute for their parents in home production. Inserting eq. (4.9) into eq. (4.8), yields

$$g_{S_{T-1}} = \frac{w_{T-1}\delta^T(1+r)EU_{C_T^m}}{E[\phi_{H_{T+1}}]E[g_{H_T}] + cov(\phi_{H_{T+1}}, g_{H_T})}. \tag{4.10}$$

Equation (4.10) shows that uncertainty leads to a reduction of school time in period $T-1$ through the savings motive. The difference between $EU_{C_T^m}$ and $U_{C_T^m}$, and thereby in school time S_{T-1} between a scenario with uncertainty and a world with perfect predictability, will be greater the higher a household's risk aversion, the higher the variance in wages and the lower average market-good consumption, i.e. the poorer the household.

Second, because risk in wages makes future investments in school time uncertain, and the returns to current school time increase with future school time, investing in current school time becomes risky. Formally, each period's school time is influenced by the amount of covariance between the marginal utility of end-of-schooling education, $\phi_{H_{T+1}}$, and the marginal product of human capital stock on period T education production, g_{H_T}. g_{H_T} increases in S_T, the final period's school time. The covariance term reflects the fact that risk-averse households will have fewer incentives to invest in their child's current schooling if its returns are risky. The covariance term will always be negative and increase the right-hand side of equation (4.10).[5]

As both effects go in the same direction, this implies that the marginal product of period $T-1$ school time is higher if the household faces uncertainty than it would be in the absence of uncertainty. Everything else being equal, this can only happen if households reduce period $T-1$ school time when facing uncertainty regarding period T wages.

[5] The covariance term is strictly negative, because both terms are influenced by the effect of period T wages on the child's school time: in a bad state of the world, i.e. $w_T \to 0$, market-good consumption can only be ensured if adults increase their time supplied to the labor market and the child will have to take on more household duties. This will reduce the child's school time in period T. If S_T falls, the marginal product of H_T will also fall (which is due to the complementarity of H_T and S_T in the education production function). The same reduction in S_T would increase the marginal utility of end-of-schooling education, $\phi_{H_{T+1}}$.

4.2.3 Heterogeneity in the effect of wage risk on each period's school time

The model predicts that wage risk affects a child's school time through the time allocation of their parents. Equation (4.10) also shows that current wages enter the decision on school time multiplicatively, which means that the effect of risk increases in wages. Since labor supply and time allocated to home production of adults depend on current wages, we can expect the effect of risk to be more important at increasing levels of labor supply, i.e. at the intensive margin. Likewise, we expect the effect of risk to be negligible at low levels of labor supply, i.e. the extensive margin.

A second source of heterogeneity in the effect of wage risk on school time stems from current consumption levels. As discussed earlier, the difference between the marginal expected utility of consumption and the marginal utility of expected consumption is highest at low consumption levels. This would suggest that the effect of uncertainty on adults' labor supply and the child's time allocation between home production and school is likely to be more important for poorer households with lower current consumption. If households face seasonality in wages, such that wages are high in agricultural peak seasons and low in agricultural lean seasons, the model predicts two different effects. To the extent that consumption cannot be perfectly smoothed throughout the year, the model would suggest that the effect of uncertainty is greater in periods with lower wages and thus lower consumption. On the other side, we saw earlier that wages enter the decision rule multiplicatively. This would suggest that uncertainty leads to stronger reductions in school time in periods with relatively high wages. This effect can be thought of as substitution effect: the pressure to accumulate savings is greater in periods with high wages than in periods with relatively low wages. These two effects go in opposite directions, and which of the two effects dominates is essentially an empirical question. We therefore estimate in the data the extent to which seasonality in wages influences the effect of risk on time-allocation decisions.

4.3 Context and data: Risk, labor supply and human capital accumulation in rural India

We test these hypotheses with the 2006 round of the Rural Economic and Demographic Survey (REDS) data. The REDS is the follow-up survey of the Additional Rural

Incomes Survey (ARIS), which was first collected in 1971. The sample was designed to represent the rural population of India across 17 major states. The ARIS covers 4,527 households in 259 villages. Three follow-up rounds were collected in 1982, 1999, and 2006 to re-visit these households. The sample was increased over time by randomly sampling additional households from the same villages. The sample in 2006 consists of roughly 9,500 households in 242 villages.[6] We can only use the 2006 round because of changes in the questionnaire over time.

There are three reasons for using the REDS data to study the questions outlined above. First, the geographic coverage of almost the entire country allows a comparison of households in very different agro-climatic regions and economic conditions. Second, the REDS 2006 survey collects detailed information about time allocated to different activities for all household members. Third, the sample consists mostly of rural households and, as we have seen, time constraints seem to be more important for women in rural areas.

We restrict the sample to households and individuals with complete information on time allocation, income and consumption, and who live in rural areas. This gives a final sample of 8,575 households. These households are distributed across 17 states, 104 districts and 240 villages of India. We create two different subsamples: one for working-age women and one for school-age girls. The subsample of working-age women, e.g. every woman aged 19 to 65, consists of 12,187 individuals. The subsample of school-age girls consists of all girls between 6 and 18 and covers 5,796 individuals.

Table 4.1 reports some general household summary statistics. As we can see, most households in the sample (58%) cultivate their own land. Consistently, income from agricultural production is the most important source of income: Average annual per capita income from agricultural production is INR 5,700, as compared to INR 4,900 from labor-market work and INR 1,900 in non-labor income.

[6]Due to armed conflict no data were collected in Jammu & Kashmir and in Assam in the 2006 round of interviews.

Table 4.1: Household characteristics

	Mean	SD
Household size	5.16	(2.60)
No of children in household	1.65	(1.61)
Age	50.5	(13.3)
Sex	0.89	(0.31)
Married	0.86	(0.34)
Caste: SC/ST	0.25	(0.43)
Religion: Hindu	0.88	(0.32)
Education: no grade	0.39	(0.49)
Education: primary	0.24	(0.43)
Education: secondary	0.23	(0.42)
Education: tertiary	0.13	(0.34)
Hh cultivated any land	0.58	(0.49)
Area cultivated p.c. (acres)	0.50	(1.18)
Annual p.c. income: labor	4871.7	(9329.5)
Annual p.c. income: labor (log)	4.92	(4.32)
Annual p.c. income: non-labor	1850.2	(8282.2)
Annual p.c. income: non-labor (log)	2.43	(3.45)
Profits from agr. production per capita	5704.2	(14524.7)
Total liquid assets per capita	14094.6	(16634.8)
Total liquid assets per capita (log)	9.04	(1.03)
Consumption expenditure per capita	9052.7	(6935.6)
Consumption expenditure per capita (log)	8.95	(0.54)
Total annual precipitation (log)	6.88	(0.56)
Std dev. of log annual rainfall (1960 - 2010)	0.27	(0.097)
Share of irrigated land, village average	0.61	(0.37)
Expected log consumption	8.93	(0.26)
Interquartile range of log consumption	0.033	(0.033)
SD of log consumption	0.025	(0.025)
Observations	8575	

Notes: All values in current INR. Age, sex, married, caste, religion and education refer to the household head.

4.3.1 Risk in rural India

In order to predict wage risk, we merge the REDS data with historical rainfall data.[7] In rural India, labor markets are still dominated by casual agricultural employment.[8] As can be seen in Table 4.2, agricultural casual employment is by far the most important labor market activity of women in our sample.

And casual agricultural employment is inherently risky. As previous literature has shown, wages and employment levels in the agricultural sector are strongly influenced by rainfall conditions (see e.g. Jayachandran, 2006). 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. The variability of rainfall combined with the village-level availability of irrigation systems should therefore be good proxies for wage risk in this context.

We compute consumption risk as follows. First, we estimate in the sample by how much current rainfall levels determine a household's consumption per capita (in logs) given a village's share of area that is irrigated.[9] Results are reported in Table 4.3.[10] Rainfall is interacted with the share of agricultural land under irrigation in the village, to capture differential risk exposure. The assumption here is that rainfall shocks will translate less strongly into consumption outcomes the higher the share of irrigated land in a village is. This can be because households are more likely to have irrigation on their own land in villages with a high share of area under irrigation or because casual agricultural employment will be less affected by current rainfall levels (since most farmers in the village do not depend on current rainfall levels for their agricultural output). We also control for irrigation levels separately, and include a number of controls that could proxy for permanent income such as education, caste,

[7]We use precipitation data compiled by the University of Delaware for the period 1960 to 2010. Data are available for 1900 onwards, but the data quality improved a lot over the time period, which is why we prefer working only with more recent data. We merge the data with the geocode of the village center. Since the data are available for grids of 0.5 degrees in latitude and longitude (approximately 50 km), some of the villages fall in the same cell.

[8]Agriculture is the dominant economic sector in rural India, employing 67% of the all male workers and 83% of all female workers in 2004/05 (National Sample Survey Office, 2006).

[9]This approach is inspired by Dercon and Christiaensen (2011).

[10]In this regression, we drop 469 observations in order to reduce the influence of outliers and to obtain realistic predictions, e.g. that the effect of rainfall on consumption goes to zero with increasing irrigation levels but never becomes negative. Outliers are detected using the DFBETA statistic for the share of irrigated land and the usual cutoff value of 2/sqrt(N), with N being the number of observations.

Table 4.2: Individual characteristics of women and girls

	Women			Girls		
	N	Mean	SD	N	Mean	SD
No of children in household	12187	1.91	(1.78)	5796	3.01	(1.85)
Household size	12187	6.17	(3.16)	5796	6.96	(3.08)
Married	12187	0.86	(0.34)	5796	0.028	(0.17)
Age	12187	38.8	(13.0)	5796	12.2	(3.66)
Caste: SC / ST	12187	0.23	(0.42)	5796	0.26	(0.44)
Religion: Hindu	12187	0.89	(0.32)	5796	0.88	(0.33)
Education: no grade	12187	0.54	(0.50)	5796	0.13	(0.34)
Education: primary	12187	0.17	(0.37)	5796	0.48	(0.50)
Education: secondary	12187	0.18	(0.38)	5796	0.29	(0.45)
Education: tertiary	12187	0.11	(0.32)	5796	0.10	(0.31)
Years of Schooling	12187	3.77	(4.71)	5796	5.15	(3.58)
Presently enrolled	12187	0.016	(0.13)	5796	0.74	(0.44)
Hrs p year: agr. casual labor	12187	119.8	(359.4)	5796	11.9	(109.8)
Hrs p year: agr. casual labor (ffw)	12187	2.31	(25.6)	5796	0.19	(8.10)
Hrs p year: own agr. production	12187	115.2	(300.0)	5796	5.60	(55.8)
Hrs p year: own livestock production	12187	243.3	(352.7)	5796	59.0	(168.6)
Hrs p year: public works	12187	4.51	(45.5)	5796	0.18	(7.47)
Hrs p year: non-agr. casual labor	12187	20.8	(171.5)	5796	0.93	(37.9)
Hrs p year: migration	12187	1.48	(45.2)	5796	0	(0)
Hrs p year: self-employed	12187	15.8	(144.5)	5796	2.38	(54.4)
Hrs p year: construction (own)	12187	12.0	(57.9)	5796	1.47	(9.29)
Hrs p year: household work	12187	1654.1	(740.6)	5796	413.4	(584.8)
Hrs p year: CPR	12187	2.56	(44.5)	5796	0.24	(10.2)
Hrs p year: other	12187	431.3	(493.3)	5796	1499.9	(848.7)
Hrs p year: permanent employment	12187	26.4	(239.5)	5796	0.26	(19.7)
Hours worked (per year)	12187	2649.5	(844.0)	5796	1995.5	(832.7)
Household chores (hours per year)	12187	1910.1	(842.0)	5796	474.1	(667.4)
Labor supply (hours per year)	12187	306.2	(555.7)	5796	21.5	(147.5)

religion and land ownership. State fixed effects and a linear time trend are also included. As can be seen in Table 4.3, the results are robust to the inclusion of these controls.

Given these estimates, we then simulate the amount of risk faced by each household using the historical rainfall distribution and the current share of area that is irrigated in the village. We use the historical rainfall data to calculate the probability of each rainfall outcome in a village. We then predict a household's log consumption per capita for each rainfall outcome given the current availability of irrigation in a village. Combining the probability of rainfall outcomes with predicted log consumption gives us a probability distribution of consumption outcomes for each household.[11] Finally, two approaches are used to predict wage risk: First, we calculate the interquartile range of each household's predicted log consumption per capita. Second, we use the standard deviation of each household's predicted log consumption per capita.

4.3.2 Labor supply in rural India

As discussed earlier, agricultural casual employment is the dominant source of wage income in our sample. The REDS collects information on time allocation for three seasons of the year (each of which lasts for four months), which are also marked by very different levels of agricultural activity and hence demand for labor in the agricultural sector. Seasons 1 and 2 are the agricultural peak seasons in which most of the agricultural production takes place. Season 3 is the dry season, during which only very few crops are cultivated and agricultural employment is considerably lower (c.f. Table 4.4).

For all household members aged 6 to 65, the questionnaire collects information about the total number of days and hours per day allocated to a number of different activities per season.[12] Consistent with the ILO definition, we compute labor supply as the total number of hours per season worked in paid employment or in self-employment. However, we exclude hours worked in own-agricultural production from this variable.

[11] The probability weights are obtained by dividing the sample rainfall distribution in 0.1 intervals of annual log rainfall. We then calculate the historical probability of village-level rainfall to fall in each of these intervals.

[12] The full list of activities are salaried work, agricultural casual labor, own-crop production, own-livestock production, work for public works programs, non-agricultural casual labor, migration, self-employment in non-farming, construction and maintenance of house, farm and other assets, household work, other household-related activities (collecting fuel, herding cattle, fishing, cutting grass) and other activities (schooling, unemployment, leisure).

Table 4.3: Determinants of consumption per capita (log)

	(1)	(2)
Total annual precipitation (log)	0.253**	0.249***
	(0.086)	(0.033)
Share of irrigated land, village average	1.831***	1.616***
	(0.343)	(0.215)
Total annual precipitation (log) × Share of irrigated land	-0.249***	-0.219***
	(0.050)	(0.031)
Average of annual rainfall (1960 - 2010), log	-0.089	-0.100**
	(0.078)	(0.032)
Religion: Hindu		0.103***
		(0.017)
Caste: SC/ST		-0.179***
		(0.011)
Education: primary		0.036**
		(0.012)
Education: secondary		0.107***
		(0.012)
Education: tertiary and higher		0.220***
		(0.015)
Area cultivated per capita (acres, log)		0.107***
		(0.005)
Observations	8106	8106
R^2	0.267	0.354

Notes: OLS estimation. State fixed effects and linear time trend included in all specifications but not reported. Influential outliers excluded using the the DFBETA statistics on the share of irrigated land and the usual cutoffs. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4: Time allocation of women and girls

	Women			Girls		
	N	Mean	SD	N	Mean	SD
Season 1						
Labor force participation	12187	0.17	(0.37)	5796	0.018	(0.13)
Labor supply (hours per season)	12187	70.8	(179.8)	5796	5.91	(49.0)
Own agr. production (hours per season)	12187	58.1	(136.8)	5796	2.82	(26.1)
Household chores (hours per season)	12187	634.2	(285.6)	5796	153.9	(222.5)
Average wage (season)	1877	8.43	(11.4)			
Predicted wage	12187	4.09	(1.53)			
Predicted wage (log)	12187	1.35	(0.31)			
Hours per day: chores				5796	2.08	(2.61)
Hours per day: studying				5796	5.74	(3.76)
Season 2						
Labor force participation	12187	0.17	(0.37)	5796	0.018	(0.13)
Labor supply (hours per season)	12187	69.6	(179.8)	5796	6.05	(50.3)
Own agr. production (hours per season)	12187	43.5	(123.5)	5796	2.19	(23.3)
Household chores (hours per season)	12187	633.4	(288.9)	5796	157.2	(226.1)
Average wage (season)	1835	8.56	(11.5)			
Predicted wage	12187	4.00	(1.48)			
Predicted wage (log)	12187	1.33	(0.31)			
Hours per day: chores				5796	2.12	(2.68)
Hours per day: studying				5796	5.79	(3.77)
Season 3						
Labor force participation	12187	0.13	(0.33)	5796	0.012	(0.11)
Labor supply (hours per season)	12187	50.6	(152.8)	5796	3.91	(40.1)
Own agr. production (hours per season)	12187	13.6	(72.4)	5796	0.59	(14.1)
Household chores (hours per season)	12187	642.5	(300.7)	5796	163.0	(233.2)
Average wage (season)	1355	9.27	(13.3)			
Predicted wage	12187	3.29	(1.34)			
Predicted wage (log)	12187	1.13	(0.32)			
Hours per day: chores				5796	2.17	(2.73)
Hours per day: studying				5796	5.44	(3.70)

Notes: Season 1 & 2 are agricultural peak seasons, season 3 is lean season. Sample of women includes women aged 19 to 65. Sample of girls consists of all girls aged 6 to 18.

The reason for this is that labor supply to own-agricultural production will be affected by rainfall risk in different ways and we want to avoid mixing up different causal mechanisms.[13] Household chores, then, include all activities related to the household: construction and maintenance of the house, farm and other assets, household work and other household-related activities (collecting fuel, herding cattle, fishing, cutting grass). We also count the time allocated to livestock production as home production, arguing that livestock production is mostly a household duty, even though it could also be undertaken for profit.

The labor force participation in our sample of working age women is 17% in the peak seasons and 13% in the lean season (c.f. Table 4.4).[14] In all seasons, labor supply is on average slightly higher than time allocated to own-agricultural production. By far the majority of the time is spent on household chores: 634 hours in season 1, which makes roughly 39.6 hours per week. Hours allocated to labor supply and own-agricultural production together make up for 8.1 hours per week on average in season 1. Total time spent on activities other than leisure is thus 47.7 hours per week in season 1, which is less than the time spent on those activities reported in the Time Use Survey (53.4).[15]

4.3.3 Human capital accumulation in rural India

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. By 2010/11 the gross enrollment ratio reached 114.9% for boys and 116.3% for girls in the classes 1 to 5. The gender gap has also been closing

[13] Labor supply to own-agricultural production should fall with increasing rainfall risk, as this income source becomes increasingly risky. Labor has to be allocated to agricultural production partly before the rainfall realizes, hence before the household can assess how the harvest, and therefore the returns to that labor, will be.

[14] The average labor force participation is 15.6%. This is considerably lower than the 24.5% reported for rural India by the NSSO in 2009/10. But the NSSO classifies own-agricultural production as labor supply, while we do not include it in our analysis. If we include own-agricultural production we get an average labor force participation of 31.3%. In addition, the sample of the NSSO covers women aged 15 and above, while we look at women aged 19 to 65. If we apply the same definition as the NSSO, the labor force participation in our sample is 28.0%.

[15] There is a substantial time lag between the two surveys however. Wealth increases over 8 years could explain the increase in leisure.

in middle school enrollment: by 2010/11, 87.5% of the boys were enrolled in the classes 6 to 10, as compared to 82.9% of the girls (also in gross figures).[16]

Still, literacy rates remain low, particularly for women and in rural areas. According to the Census in 2011, only 50.6% of the women aged 15 and above living in rural India are literate. This is not only an issue among the adult population: according to the Annual Status of Education Report (ASER) for Rural India, only 48.2% of the children in grade 5 could read a grade 2 level text in 2011 (Pratham, 2012).[17]

While the quality of education is an often-cited reason for low learning outcomes in India (Banerjee et al., 2010; Kremer et al., 2005), demand-driven factors play a role, particularly in the inequality of learning outcomes between boys and girls: according to the Status of Education and Vocational Training in India survey conducted in 2011/12, the ratio of children not attending school in the age group 10 to 14 is highest for girls in rural areas (10.1%), as compared to 6.4% for girls in urban areas and 6.7% for boys in rural areas (National Sample Survey Office, 2015). This inequality increases further if we consider the age group 15 to 19: in this group, 44.8% of the rural girls are out of school as compared to 29.7% of the urban girls and to 34.3% of the rural boys.

The REDS questionnaire contains a section that records the number of hours per day allocated to different activities on a typical day for all household members in each of the three seasons. Since this section explicitly differentiates between time in school and leisure it is particularly interesting for analyzing the effect of wage risk on girls' time in school. In the REDS sample, 74% of the girls aged 6 to 18 are currently enrolled in school, as compared to 81% of the boys in the same age group (c.f. Table 4.2)[18] The difference is even more pronounced when the age group 10 to 18 is looked at: 79% of the boys are enrolled, while only 70% of the girls are enrolled. Average time in school is also higher for boys than for girls; in season 1 boys spend on average 6.4 hours per day in school or studying, while girls spend only 5.7 hours on these activities (c.f. Table 4.4). Again, the difference is even more pronounced if we look at the age group 10 to 18. In this group, boys spend on average 6.3 hours studying, while girls do so for only 5.6 hours.

[16]Ministry of Human Resource Development, Government of India. Data retrieved from Indiastat.

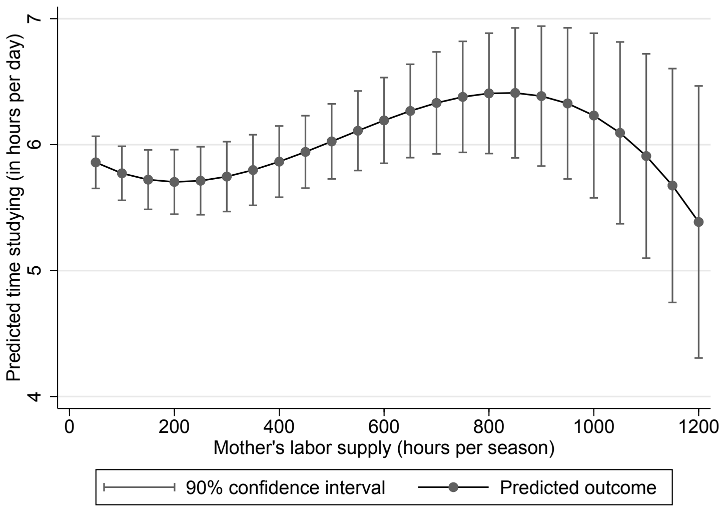
[17]ASER is based on an annual survey that assesses children's schooling status and basic learning levels throughout all rural districts of India. It is facilitated by the Indian NGO Pratham and interviews are conducted by volunteers, which has raised a number of doubts regarding the data quality. Still, it is the only India-wide assessment of learning levels currently available.

[18]Summary statistics for boys are reported in the Appendix, Table 4.B.1.

4.3.4 Female labor supply and the time allocation of girls

That time constraints might be an important explanation for lower school attendance and grade progression of girls was revealed by the NSS Survey on the Status of Education and Vocational Training in India from 2011/12. More than half of the girls aged 5 to 29 years who were currently not in school but had ever attended an educational institution stated that attending domestic chores was the single reason for not being enrolled in any educational institution (National Sample Survey Office, 2015).

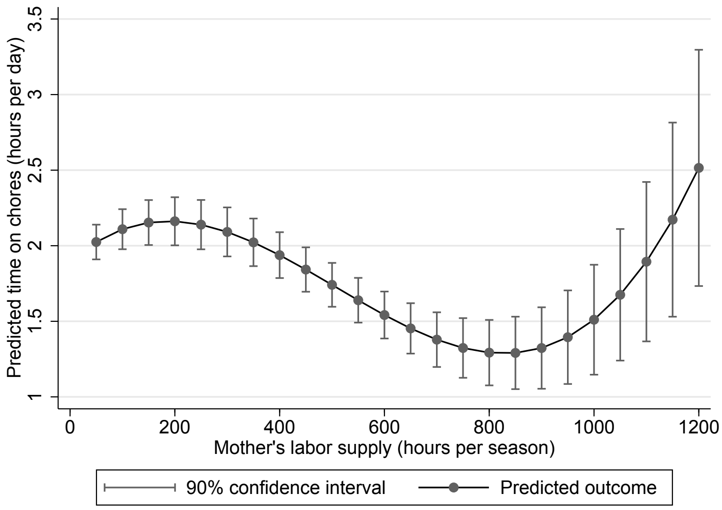
Figure 4.1: Girls' predicted hours in school



Source: Own estimation based on REDS data.

The idea that the time mothers spend on household chores and wage work influences the time allocation of girls in the household is also confirmed in the data used in this paper. With the caveat in mind that female time allocation is endogenous to a number of household decisions, we test if the level, square and cubic of the mother's labor supply and time in home production affect how much time girls allocate to household chores and to studying. Results are reported in Table 4.5. In line with our expectations, mother's labor supply has a substantial effect on their daughters' time allocation. The predicted values plotted in Figures 4.1 and 4.2 show that the effects on girls are non-linear and become quite severe at very high levels of their mother's labor supply. Given the small number of observations in this spectrum, predicted outcomes become fairly imprecise at high levels of hours worked, but the general direction of the results is

Figure 4.2: Girls' predicted hours in home production



Source: Own estimation based on REDS data.

still astonishing. For most levels of a mother's labor supply, a small increase in labor supply increases the time girls spend in school. At very high levels of a mother's labor supply, however, a further increase in labor supply seems to strongly reduce girls' time in school. The same holds for the amount of time girls allocate to household chores: up to a labor supply of 50 hours per week, an increase in mother's labor supply reduces the time girls spend on chores. Above this level, however, a further increase in the mother's labor supply also increases the time girls spend on household chores.

4.4 Estimation strategy

Assume the structural hours function to be estimated is

$$h_{ijt} = \beta_0 + \beta_1 \ln w_{ijt} + \beta_2 N_{ijt} + \beta_3 Y_{ijt} + \beta_4 R_{ijt} + \beta_5 X_{ijt} + \epsilon_{ijt}, \tag{4.11}$$

where the dependent variable is the amount of time individual i, living in village j, allocates to the labor market (or to home production) at time t. The dependent variable will first of all depend on wages w_{ijt}, but also on non-labor income, asset and land ownership Y_{ijt} and other household members' labor income N_{ijt}. We are particularly interested in estimating β_4, e.g. the effect of wage risk R_{ijt} on hours allocated to

Table 4.5: Girls' time allocation as function of mother's time allocation

	(1)	(2)	(3)	(4)
	Home production		Studying	
Mother's labor supply (hours per season)	0.0031***		-0.0030*	
	(0.0008)		(0.0014)	
Mother's labor supply (square)	-0.0000***		0.0000*	
	(0.0000)		(0.0000)	
Mother's labor supply (cubic)	0.0000***		-0.0000**	
	(0.0000)		(0.0000)	
Mother's time on household chores (hours per season)		-0.0002		0.0030
		(0.0012)		(0.0019)
Mother's time on household chores (square)		0.0000		-0.0000
		(0.0000)		(0.0000)
Mother's time on household chores (cubic)		0.0000		0.0000
		(0.0000)		(0.0000)
Age	-0.1829***	-0.1702**	1.7635***	1.7560***
	(0.0544)	(0.0552)	(0.1024)	(0.1027)
Age (square)	0.0237***	0.0235***	-0.0813***	-0.0811***
	(0.0024)	(0.0024)	(0.0044)	(0.0044)
Married	1.8201*	1.9131*	-1.6089+	-1.6388+
	(0.7748)	(0.7710)	(0.9457)	(0.9536)
Household size	-0.1251***	-0.1153***	0.1053**	0.1100***
	(0.0176)	(0.0171)	(0.0329)	(0.0317)
No of children in household	0.2202***	0.2088***	-0.2963***	-0.3055***
	(0.0307)	(0.0294)	(0.0644)	(0.0645)
Caste: SC / ST	-0.0103	0.0775	-0.5233*	-0.5603**
	(0.0964)	(0.0895)	(0.2086)	(0.2048)
Religion: Hindu	0.0828	0.0961	0.9737**	0.9703**
	(0.1534)	(0.1470)	(0.3022)	(0.3078)
Area cultivated per capita (acres, log)	-0.0590	-0.0781+	0.1393+	0.1585+
	(0.0490)	(0.0445)	(0.0836)	(0.0835)
Annual p.c. income: labor (log)	0.0486***	0.0475***	-0.0913***	-0.0883***
	(0.0098)	(0.0094)	(0.0168)	(0.0168)
Annual p.c. income: non-labor (log)	0.1070***	0.0896***	-0.0078	-0.0039
	(0.0171)	(0.0139)	(0.0253)	(0.0243)
Total annual precipitation (log)	-0.4022***	-0.3778***	0.4711*	0.5031*
	(0.0985)	(0.0982)	(0.2203)	(0.2195)
Observations	10669	10669	10669	10669
R^2	0.340	0.352	0.125	0.125

Notes: OLS estimation. Linear time trend included, but not reported. Mother's labor supply includes time in own agricultural production. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

different activities. We assume that preferences for work can be captured by individual and household characteristics X_{ijt} such as age (squared), education, caste, religion, the number of children (in the household), household size and the marital status of individual i. The error term can be decomposed into a permanent and two transitory components such that $\epsilon_{ijt} = \mu_j + \sigma_{jt} + \eta_{it}$, where μ_j is a state-level fixed effect, σ_{jt} are village-level shocks and η_{it} is a mean zero, strictly exogenous, idiosyncratic shock.

When estimating the function described above, we need to adequately control for wages each woman would face if she were in the labor market. But adequately controlling for wages is challenging for a number of reasons. First, observed wages are potentially endogenous to labor supply if wages reflect work effort as well as skills or ability, such that individuals who work longer hours are likely to get higher wages. This is of concern particularly in the labor markets of developed economies (see e.g. Keane, 2011). In the context of low-skill agricultural wage work, however, we feel that this is less of an issue, because wage rates are determined by local conditions rather than individual abilities (Jayachandran, 2006; Rose, 2001; Rosenzweig, 1978).

Second, both wages and other household members' labor income are endogenous to risk. There might be general equilibrium effects of risk on wages: if all households in the village supply more labor due to risk, equilibrium wages should be lower than in the absence of risk. Jayachandran (2006) makes a similar argument for the effect of shocks on wages. We need to be aware that we are estimating only the direct effect of risk on labor supply, and that we control for predicted wages that already account for general equilibrium effects of risk on wages in the village economy.

Third, wages are only observed for individuals who are currently in the labor market, such that we have to deal with missing wages for all individuals who are not participating in the labor market. Potentially the sample of workers is not a random sub-sample of all individuals, such that we have to account for selection bias when imputing wages. Fourth, most individuals in our sample reported working in different activities, at different wages. Fifth, we have to deal with classical measurement error since wages are mostly measured with considerable error in micro data (Keane, 2011).

We address the last three issues by predicting wages for all individuals in our sample. Following Blundell et al. (2007), we predict wages using a Heckman selection correction. Since individuals report working in a number of different activities at different wage levels, we calculate weighted average wages, i.e. calculate total income from all activities and divide it by the total number of hours worked.[19] Assume wages are determined as follows:

$$\ln w_{ijt} = \delta_0 + \delta_1 X_{ijt} + \mu_j + \sigma_{jt} + \upsilon_{it}. \tag{4.12}$$

μ_j and σ_{jt} capture state fixed effects and village-level rainfall shocks respectively. The structural participation equation looks exactly as the structural hours equation, except that we allow the coefficients to be different,

$$p_{ijt} = \alpha_0 + \alpha_1 \ln w_{ijt} + \alpha_2 N_{ijt} + \alpha_3 Y_{ijt} + \alpha_4 R_{ijt} + \alpha_5 X_{ijt} + \epsilon_{ijt}. \tag{4.13}$$

Substitution eq. 4.12 into the structural participation equation, gives the reduced-form participation equation, which is the selection equation for the Heckman selection model,

$$p_{ijt} = a_0 + a_1 N_{ijt} + a_2 Y_{ijt} + a_3 R_{ijt} + a_4 X_{ijt} + \mu_j + \sigma_{jt} + \upsilon_{it}. \tag{4.14}$$

We estimate this reduced-form participation equation first and then estimate the log hourly wage equation (4.12) including the inverse Mills ratio obtained from the estimated participation equation (4.14). In this approach, the wage equation is identified from the exclusion of other household members' income, non-labor income, land ownership and ownership of assets from the wage equation, as well as from the normality assumption. The structural hours equation (4.11) can then be estimated with imputed wages.[20]

4.5 Results

Based on the model and the implications described above, we want to estimate the effect of wage risk on girls' time in school. Our hypothesis is that the effect of risk on school time is due to intra-household substitution effects in home production. We therefore estimate first the effect of risk on women's labor supply and time allocated to

[19] Alternatively one could use marginal wages, hence the lowest wages observed at which individuals supply a positive number of hours. But we are worried that this measure cannot account for necessity-driven labor supply, i.e. cases in which labor supply is high because wages are low in all activities that are performed.

[20] In the selection model, we augment the participation equation by the square of non-labor income and in other household members' income as it improves the model fit.

home production, and then estimate the effect of risk on girls' time allocated to home production and on girls' time in school.

The model predicts that the effect of risk might be different between periods with high wages and periods with relatively low wages. We therefore estimate the effect separately for agricultural peak and agricultural lean seasons, arguing that labor market opportunities and wage levels are very different between these seasons.

4.5.1 Risk and labor supply

The first hypothesis we want to test is whether wage risk increases labor supply of women. We use two variables to proxy wage risk, i.e. the interquartile range as well as the standard deviation of log consumption. As discussed above, we control for predicted wages and for individual and household-level socio-economic characteristics in all our specifications. In order to account for the effect of non-labor income on labor supply, we include a measure of all other household members' labor income as well as total household non-labor income per capita. We also control for rainfall shocks, state fixed effects, a linear time trend and a lean-season fixed effect. Standard errors are clustered at the village level; i.e. the level of variation of the main explanatory variable.

Because the model predicts that the effect of risk on labor supply is very different at different levels of labor supply, we estimate extensive and intensive margin responses to risk separately. First, we estimate the participation frontier (eq. 4.13) using a probit model. We then estimate the hours equation (eq. 4.11) for the sample of working women in OLS.

Estimates of the effect of wage risk on labor force participation of women are presented in Table 4.6.[21] In line with our expectations, the marginal effect of both risk measures, i.e. the interquartile range and the standard deviation of log consumption, is close to zero and very imprecisely estimated. We also cannot find any statistically significant difference between the peak and the lean seasons. We thus cannot find any evidence that risk in wages increases labor supply at the extensive margin. The coefficient on log wages is statistically significant at the 1% level and has a point estimate of 0.24. This implies an extensive-margin labor supply elasticity of 1.57, which is slightly

[21] In the absence of specific questions on work-seeking behavior of individuals in the questionnaire, we classify every individual as being in the labor force who supplied non-zero amount of hours to the labor market at any time in the current season.

Table 4.6: Female labor force participation

	(1)	(2)	(3)	(4)
Interquartile range of log consumption	-0.229			
	(0.253)			
SD of log consumption			-0.079	
			(0.338)	
_in peak season		-0.201		-0.025
		(0.253)		(0.339)
_in lean season		-0.259		-0.179
		(0.249)		(0.354)
Predicted wage (log)	0.217***		0.236***	
	(0.042)		(0.040)	
No of children in household	0.028***		0.027***	
	(0.005)		(0.005)	
Household size	-0.023***		-0.022***	
	(0.004)		(0.004)	
Married	-0.092***		-0.087***	
	(0.017)		(0.016)	
Age	-0.000		-0.000	
	(0.000)		(0.000)	
Caste: SC / ST	0.055***		0.051***	
	(0.013)		(0.013)	
Religion: Hindu	0.009		0.009	
	(0.014)		(0.014)	
Education: primary	-0.034**		-0.034**	
	(0.011)		(0.011)	
Education: secondary	-0.062***		-0.061***	
	(0.012)		(0.011)	
Education: tertiary and higher	-0.111***		-0.114***	
	(0.013)		(0.012)	
Annual p.c. income: non-labor (log)	0.002		0.002	
	(0.001)		(0.001)	
Annual p.c. income: other hh members' labor (log)	0.017***		0.016***	
	(0.002)		(0.002)	
Area cultivated per capita (acres, log)	0.012*		0.012*	
	(0.005)		(0.005)	
Total annual precipitation (log)	-0.018		-0.014	
	(0.016)		(0.016)	
Lean season	-0.001		0.003	
	(0.009)		(0.009)	
Observations	36561	36561	36561	36561

Notes: Probit estimation. Cells report average marginal effects. State fixed effects and linear time trend included, but not reported. Col. (3) and (4) report marginal effect of risk variable in peak and lean seasons, obtained from add. including an interaction term in the regression. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

above the elasticities found for married women in OECD countries in the 1960s and 1970s (Blundell and MaCurdy, 1999; Keane, 2011).[22] Interestingly, the number of children in the household seems to increase the probability of a woman being in the labor force, which stands in contrast to findings from OECD countries (Angrist and Evans, 1998). Women belonging to scheduled castes and tribes are also more likely to be in the labor force as well as women without education. This suggests that poorer women are more likely to be in the labor force. However, the coefficients on land ownership as well as on other household members' labor income are also positive, suggesting that wealthier women tend to be more likely to work. Those effects are probably non-linear though, which is consistent with previous literature that suggests that mostly poor and very rich women tend to work (c.f. Klasen and Pieters, 2015). Transient shocks do not seem to affect the probability of participation in the labor force.

In contrast to the extensive margin, we find evidence that wage risk increases the number of hours worked conditional on being in the labor force (Table 4.7). The estimated effect is positive and statistically significant for both risk measures. The estimated effect is also considerably higher in the agricultural peak season than in the agricultural lean season; and not statistically significant in the agricultural lean season. Our estimates suggest that a one standard deviation increase in the interquartile range of log consumption (0.033) increases labor supply by roughly 33.6 hours per season (or 8.4 hours per month) at the intensive margin in the peak season. Using the alternative explanatory variable, we find that a one standard deviation increase in the standard deviation of log consumption (0.025) would increase hours worked by 26.3 hours per season and 6.6 hours per month in the agricultural peak season. The marginal effect of log wages on hours worked is 388.8 (column 3), so for a 1% increase in wages labor supply in a given season increases by roughly 3.87 hours. With an average labor supply of 360 hours per season, this results in a labor supply elasticity of 1.08 at the intensive margin. Labor supply is thus very elastic also for women who are in the labor force. At the intensive margin, other household members' labor income seems to decrease hours worked, while land ownership still has a positive effect on hours worked.[23] Household

[22] If we estimate the participation equation in OLS, we obtain a point estimate on log wages of 0.136, which gives an extensive margin elasticity of 0.88. This value is more in line with previous literature.

[23] Remember that we excluded time allocated to own-agricultural production from our labor-supply variable. Thus, this does not reflect greater time allocated to own-farm activities with increasing land ownership.

Table 4.7: Female labor supply in hours per season

	(1)	(2)	(3)	(4)
Interquartile range of log consumption	811.459**			
	(281.517)			
SD of log consumption			814.642*	
			(349.840)	
_in peak season		1018.880**		1052.562*
		(326.361)		(415.433)
_in lean season		480.310		430.843
		(348.212)		(439.876)
Predicted wage (log)	412.893***		388.766***	
	(99.256)		(101.579)	
No of children in household	-8.506		-7.747	
	(6.009)		(6.143)	
Household size	18.248***		17.305***	
	(4.883)		(4.977)	
Married	13.755		10.271	
	(18.295)		(18.689)	
Age	-0.107		-0.013	
	(0.666)		(0.680)	
Caste: SC / ST	-24.519		-21.809	
	(17.644)		(17.852)	
Religion: Hindu	-25.510		-26.011	
	(16.271)		(16.419)	
Education: primary	-2.125		-2.329	
	(14.106)		(14.114)	
Education: secondary	52.319**		50.875**	
	(17.363)		(17.419)	
Education: tertiary and higher	-41.617		-29.573	
	(58.432)		(59.397)	
Annual p.c. income: non-labor (log)	1.997		2.068	
	(1.949)		(1.952)	
Annual p.c. income: other hh members' labor (log)	-5.347***		-5.368***	
	(1.577)		(1.592)	
Area cultivated per capita (acres, log)	28.303***		28.040***	
	(6.208)		(6.150)	
Total annual precipitation (log)	-13.837		-18.392	
	(16.100)		(16.126)	
Lean season	-36.323+		-40.668*	
	(19.134)		(19.444)	
Observations	6462	6462	6462	6462
R^2	0.238	0.239	0.236	0.237

Notes: OLS estimation. Cells report average marginal effects. State fixed effects and linear time trend included, but not reported. Col. (3) and (4) report marginal effect of risk variable in peak and lean seasons, obtained from add. including an interaction term in the regression. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

size is also positively associated with hours worked, while the coefficient on children is negative but not statistically significant. Finally, women with secondary education seem to supply the highest number of hours.

4.5.2 Risk and home production

Given that risk seems to increase women's labor supply, we proceed by testing if this translates into less time allocated to household chores by working women. We use the same controls as in the estimation of hours worked. Again, we test if the effect is different in agricultural peak and lean seasons.

Table 4.8 shows estimates of the effect of wage risk on hours allocated to home production. The results suggest a negative effect of risk on time allocated to household chores (conditional on women being in the labor market). Again, the effect is greater in the agricultural peak seasons than in the lean season. In the peak season, a one standard deviation increase in the interquartile range of log consumption (0.033) reduces the time allocated to home production by roughly 26.5 hours per season and by 6.6 hours per month. A one standard deviation increase in the standard deviation of log consumption (0.025) would translate into a reduction of time allocated to household chores by 6.1 hours per month or 20 minutes per day.[24] The size of the effect corresponds very closely to the observed increase on labor market work due to risk. Interestingly, the coefficient on wages is considerably smaller for home production than for intensive margin labor supply. The marginal effect of log wages is -184.2 (column 3), thus a 1% increase in wages would reduce time allocated to household chores by 1.83 hours per season. In our sample, working women allocate on average 503 hours to household chores, which gives a wage elasticity of home production of -0.37. That the wage elasticity of home time is so much smaller than the wage elasticity of labor supply, suggests that women cut down on their leisure or on time in own-agricultural production as well when increasing labor supply due to increasing wages. The remaining controls have the expected signs: being married increases the time allocated to household chores as does the number of children in the household. Women with no education and from scheduled castes and tribes seem to spend least time on chores, although the coefficients are not statistically significant. The coefficients on income and land ownership are also statistically zero. Again, rainfall shocks do not seem to affect the time allocated to home production.

[24] Assuming that this time is distributed over five working days per week.

Table 4.8: Female home production in hours per season

	(1)	(2)	(3)	(4)
Interquartile range of log consumption	-783.847*			
	(329.477)			
SD of log consumption			-940.047*	
			(412.819)	
_in peak season		-802.228*		-976.881*
		(332.371)		(418.319)
_in peak season		-754.501*		-880.629*
		(354.698)		(443.926)
Predicted wage (log)	-194.899**		-184.244**	
	(66.603)		(65.834)	
Married	103.418***		105.019***	
	(14.945)		(14.910)	
Age	-0.146		-0.186	
	(0.580)		(0.572)	
Household size	-33.597***		-33.153***	
	(4.806)		(4.772)	
No of children in household	30.470***		30.048***	
	(6.094)		(6.073)	
Education: primary	26.314		26.561	
	(17.109)		(17.147)	
Education: secondary	10.624		11.672	
	(18.899)		(18.953)	
Education: tertiary and higher	37.320		31.456	
	(37.625)		(37.000)	
Caste: SC / ST	-10.616		-12.202	
	(14.948)		(14.955)	
Religion: Hindu	1.138		1.509	
	(19.498)		(19.480)	
Annual p.c. income: non-labor (log)	-1.219		-1.287	
	(1.901)		(1.900)	
Annual p.c. income: other hh members' labor (log)	-1.997		-2.001	
	(1.610)		(1.614)	
Area cultivated per capita (acres, log)	4.408		4.492	
	(8.685)		(8.645)	
Total annual precipitation (log)	8.896		11.032	
	(23.236)		(22.895)	
Lean season	-1.142		0.777	
	(12.175)		(11.962)	
Observations	6462	6462	6462	6462
R^2	0.337	0.337	0.336	0.336

Notes: OLS estimation. Cells report average marginal effects. State fixed effects and linear time trend included, but not reported. Col. (3) and (4) report marginal effect of risk variable in peak and lean seasons, obtained from add. including an interaction term in the regression. Standard errors (clustered at the village level) in parentheses. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.5.3 Girls' time allocation

If women have less time available for home production, these tasks might have to be performed by someone else. Home production often entails tasks that cannot be delayed: caring for younger children, livestock production, preparing food etc. In a household in which all adults are already in employment, there is a high risk that parents are forced to reduce the time their children spend in school or on leisure activities in order that the children can undertake such tasks. These tasks are furthermore typically assigned to the girls in the household, which is why we conduct all estimations in this section for girls only.

The data used in this estimation come from a different section. As mentioned in Section 4.3.3, the questionnaire has one section in which all household members are asked to report on their activities on a typical day in each of the three seasons. This data also records hours per day spent in school and studying. Unfortunately, this variable does not inform about periodic school drop-outs, which are expected to happen in periods in which labor demand increases dramatically (such as during sowing or the main harvest).

Using the information on time allocation on a typical day, we find that risk increases the hours per day girls spend on household chores. Since the distribution of the dependent variable (hours per day allocated to household chores) is best approximated by a negative binomial distribution, we use the appropriate count data model to estimate this relationship. The effect is slightly higher in the agricultural peak seasons. Using the estimates from column (4) in Table 4.9, we find that a one standard deviation increase in the standard deviation of log consumption (0.025) increases the time spent on home production by 0.13 hours or roughly 8 minutes per day in the peak season. In the sample, there are roughly 2.7 school-age girls per working woman. Thus, at the household level the effect on girls corresponds almost exactly to the observed effect of risk on working women.[25] We do not find any statistically significant effect of mother's wage on girls' time allocation, which is why we exclude this variable from our estimation.[26] Instead, we control for total non-labor and labor income in the household

[25] As shown in Section 4.5.2, the same change in risk would reduce the time a working woman spends on chores by 20 minutes on average in the peak season. If 2.7 girls increase their time in home production by 8 minutes daily, this results in an increase by 21.6 minutes per household.

[26] This also allows us to use the time information of all school-age girls in the sample. When controlling for mothers wages the sample drops by 1503 observations, for which we could not identify the mother

Table 4.9: Girls' time in home production in hours per day

	(1)	(2)	(3)	(4)
Interquartile range of log consumption	3.426^{+}			
	(1.836)			
SD of log consumption			5.091^{*}	
			(2.468)	
_in peak season		3.548^{+}		5.193^{*}
		(1.811)		(2.446)
_in lean season		3.177^{+}		4.881^{+}
		(1.928)		(2.565)
Age	0.451^{***}		0.451^{***}	
	(0.024)		(0.022)	
Married	1.099^{***}		1.101^{***}	
	(0.141)		(0.139)	
Household size	-0.146^{***}		-0.146^{***}	
	(0.024)		(0.024)	
No of children in household	0.142^{***}		0.142^{***}	
	(0.033)		(0.033)	
Caste: SC / ST	0.099		0.101	
	(0.093)		(0.092)	
Religion: Hindu	0.315^{*}		0.315^{*}	
	(0.161)		(0.160)	
Area cultivated per capita (acres, log)	-0.019		-0.019	
	(0.043)		(0.043)	
Annual p.c. income: labor (log)	0.040^{***}		0.040^{***}	
	(0.011)		(0.011)	
Annual p.c. income: non-labor (log)	0.060^{***}		0.061^{***}	
	(0.016)		(0.016)	
Total annual precipitation (log)	-0.011		-0.009	
	(0.202)		(0.201)	
Lean season	0.083^{***}		0.083^{***}	
	(0.019)		(0.018)	
Observations	17388	17388	17388	17388

Notes: Negative binomial estimation. Cells report average marginal effects. State fixed effects and linear time trend included, but not reported. Col. (3) and (4) report marginal effect of risk variable in peak and lean seasons, obtained from add. including an interaction term in the regression. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and find that labor income increases the time girls spend on household chores. This is most likely due to the intra-household substitution in home production. Interestingly, non-labor income also seems to increase time spent on chores. Again, we cannot find that rainfall shocks affect the time allocation of girls.

Coming back to the central hypothesis of our paper, we proceed by estimating the effect of wage risk on school time. Table 4.10 shows that risk considerably reduces time in school. Again, the effect of risk is considerably higher in agricultural peak seasons than in the lean season. According to our estimates, a one standard deviation increase in the standard deviation of log consumption (0.025) reduces the time girls spent in school or studying by 0.25 hours or roughly 15 minutes every day. The remaining coefficients have the expected sign: as girls get older or when they marry, they spend less time in school. A household's labor income seems to be negatively associated with time in school, which again points at the importance of intra-household substitution effects. Non-labor income, in contrast, seems to increase the average time spent in school.

The effect of risk on time allocated to school is almost twice as high as the observed effect on time allocated to household chores. There are two potential explanations for this finding. First, as shown in Section 4.2.2, the theoretical model predicts that wage risk affects time allocated to schooling through two channels: the first one being the intra-household substitution in home production and the second being the uncertainty regarding future time allocation to school, which makes the returns to current school investment risky. This would explain why the effect of risk is greater for school time than for home time. The other potential explanation for observing a greater effect on time in school could be non-divisibility in school attendance: if girls drop out of school for too long, they might not be able to catch up and be forced to repeat the year. With the data used here we cannot assess how important the second explanation is for our findings.

Our results thus suggest a strong relationship between risk at the household level and girls' time allocation, both to household chores and to studying. These effects could potentially be very harmful to human capital accumulation and future earnings of these girls. We will discuss the potential role for public policy later and present some robustness checks first.

in the household.

Table 4.10: Girls' time in school in hours per day

	(1)	(2)	(3)	(4)
Interquartile range of log consumption	-8.669*			
	(4.031)			
SD of log consumption			-10.122+	
			(5.191)	
_in peak season		-9.658*		-11.370*
		(4.164)		(5.402)
_in lean season		-6.692+		-7.627
		(3.872)		(4.919)
Age	-0.204***		-0.205***	
	(0.019)		(0.019)	
Married	-3.038***		-3.036***	
	(0.236)		(0.236)	
Household size	0.106***		0.106***	
	(0.029)		(0.029)	
No of children in household	-0.161**		-0.164**	
	(0.058)		(0.058)	
Caste: SC / ST	-0.603***		-0.605***	
	(0.175)		(0.176)	
Religion: Hindu	1.038***		1.059***	
	(0.251)		(0.251)	
Area cultivated per capita (acres, log)	0.075		0.073	
	(0.068)		(0.069)	
Annual p.c. income: labor (log)	-0.072***		-0.073***	
	(0.014)		(0.014)	
Annual p.c. income: non-labor (log)	0.043+		0.042+	
	(0.022)		(0.022)	
Total annual precipitation (log)	0.147		0.188	
	(0.310)		(0.317)	
Lean season	-0.326***		-0.326***	
	(0.040)		(0.040)	
Observations	17388	17388	17388	17388
R^2	0.216	0.216	0.215	0.215

Notes: OLS estimation. Cells report average marginal effects. State fixed effects and linear time trend included, but not reported. Col. (3) and (4) report marginal effect of risk variable in peak and lean seasons, obtained from add. including an interaction term in the regression. Standard errors (clustered at the village level) in parentheses. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.5.4 Robustness checks

The estimates provided so far are only averages and probably hide considerable variation across households. To gauge the robustness of our findings, we test the extent to which observed effects vary along household characteristics, such as household composition and income structure, as well as with individual characteristics, such as the age of the girls.

We expect the effect of rainfall risk on wages and consumption to be strongest for households who depend on casual agricultural employment as their main income source. Typically, these would be landless households. We therefore test if the effect of risk on girls' time allocation is different depending on whether they live in households with land or not. In line with our expectations, the marginal effect of the standard deviation of log consumption on time allocated to household chores is higher for girls in landless households than for girls in land-owning households (c.f. Table 4.11). However, this does not seem to translate into less time in school: here the effect of risk is greater in land-owning than in landless households. This might be because other factors than intra-household substitution in home production affect time in school. In particular, girls might have to work on the farm as well as in the household in land-owning households.

We would also expect households in which at least one member has permanent employment to be less dependent on current agricultural production conditions. Accordingly, rainfall risk should affect them to a lesser extent than households with no member in permanent employment. We find that the marginal effect on the standard deviation of log consumption on the time girls spend on household chores is smaller in households with a permanently employed member (c.f. Table 4.11). The same is true for the effect on time in school: the marginal effect of our risk measure is considerably smaller and statistically zero if at least one household member has permanent employment.

With regards to household composition, the pressure on girls' time allocation is presumably greater in households with no other female household member who could take on household work. Indeed, we find that the effect of risk on girls' time allocated to household chores is almost halved if at least one women above the age of 50 lives in

Table 4.11: Heterogeneity in effect of risk on girls' time allocation

Marginal effect of SD of log consumption on:	Chores (1)	School (2)
Landless households	7.802**	-4.308
	(2.939)	(5.496)
Land owners	3.340	-13.113*
	(2.853)	(5.706)
No hh member permanently employed	5.589*	-10.219+
	(2.468)	(5.304)
Any hh member permanently employed	-1.142	-5.741
	(3.931)	(7.495)
No female hh member aged > 50	5.826*	-12.026+
	(2.649)	(6.246)
Any female hh member aged > 50	3.556	-6.004
	(3.118)	(4.638)
Girl's age = 10	0.374	-6.613
	(1.616)	(5.756)
Girl's age = 16	14.131**	-16.795**
	(5.210)	(5.314)
Observations	17388	17388

Notes: OLS estimation. Cells report average marginal effects obtained from add. including an interaction term in the regression. Controls are those of main model. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the same household. As shown in Table 4.11, the same is true for the effect of risk on time in school.

We would also expect older girls to be under more pressure to perform household tasks and to drop out of school occasionally or permanently. We therefore test if the effect of risk varies with the age of the girls. Again, we find that risk influences time-allocation decisions more strongly as girls get older. While risk does not seem to affect the time allocation of girls aged 6 to 10, the marginal effect of the standard deviation of log consumption on time in household chores increases to 14.1 at the age of 16; and to -16.8 for time in school (c.f. Table 4.11).

4.6 Simulating the effect of the NREGS on wage risk, labor supply and human capital accumulation

Given the magnitude of the effects of risk observed above, it seems worthwhile to explore potential policy tools to mediate these. Obviously, any policy tool 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 (wage) 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. Therefore we analyze the extent to which a wage-smoothing policy, such as the Indian National Rural Employment Guarantee Scheme (NREGS), can mitigate wage risk, reduce labor supply of women at the intensive margin, and allow children to spend more time in school.

The National Rural Employment Guarantee Act (NREGA) 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. This scheme can affect labor supply through two effects: first, it provides employment at higher wages than casual agricultural wages, which could affect both total labor supply and the amount of labor supplied to the private sector, and therewith equilibrium wages in the private sector. Second, it reduces risk in wages in rural areas, because it provides a minimum amount of employment at a fixed wage level independently of rainfall shocks.

To estimate the importance of the risk-reduction effect of the NREGS on labor supply, time at home and at school, we use the Socio-Economic Profiles of Rural Households in India (SEPRI) data for 2014. It is a follow-up survey to the REDS that was collected in 8 states of India: Andhra Pradesh, Bihar, Chhattisgarh, Haryana, Jharkhand, Madhya Pradesh, Tamil Nadu and Rajasthan. It samples the entire population of the REDS survey villages, but applies a questionnaire that is considerably shorter than in the REDS.

Table 4.12: Determinants of consumption per capita (log) SEPRI

	(1)	(2)
Total annual rainfall (log)	0.241^{+}	0.286^{*}
	(0.127)	(0.130)
Share of irrigated land, village average	0.890	1.146
	(0.751)	(0.727)
Total annual rainfall × Share of irrigated land	-0.111	-0.146
	(0.116)	(0.112)
Average of annual rainfall (1960 - 2010), log	-0.211	-0.213
	(0.135)	(0.128)
Employment generated per capita in 2011-12, NREGA (log)		0.678^{*}
		(0.300)
Total annual rainfall × Employment generated per capita		-0.105^{*}
		(0.044)
Observations	50979	50979
R^2	0.374	0.376

Notes: OLS estimation. State fixed effects and linear time trend included in all specifications but not reported. Influential outliers excluded using the DFBETA statistics on the share of irrigated land and the usual cutoffs. Standard errors (clustered at the village level) in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Using the SEPRI data, we test the extent to which the presence of the NREGS, or more specifically, the amount of employment generated within the NREGS in a given year, mediates the effect of rainfall on household consumption. Formally, we estimate the same equation as in Table 4.3, but now add the log of employment per capita (in person-days) generated in a given village within the NREGS and its interaction with rainfall to the estimation. Results are reported in Table 4.12.[27] We find that

[27]The rainfall data covers the agricultural year 2012/13, which is presumably the period that deter-

a 1% increase in employment per capita generated, reduces the effect of rainfall on consumption by 0.11 percentage points. We use these results to predict the standard deviation of log consumption at different levels of employment per capita generated.

Before we can proceed, we need another estimate, namely the effect of the NREGS on wages. Obviously, the NREGS has by far more effects than solely on risk management and wages. But we concentrate on those two effects, as these are the ones through which the NREGS should mainly affect labor supply and time-allocation decisions. Assuming that labor supply of all other household members remains constant, we can use the change in wages also to augment reported labor income of other household members. Obviously, the assumption that wage changes affect only income levels but not labor supply of other household members is unlikely to hold. But given the small effects of other members' labor income on time-allocation decisions observed in this paper, this assumption should not affect our results in a meaningful way. Due to the observational nature of the data, we have to treat our coefficients on wages and labor income with caution. But as they are in line with previous literature, we think they present reasonable approximations to the true effect sizes.

The most well documented paper on the effect of the NREGS on wages is by Imbert and Papp (2015). Using data from 2004/05 and 2007/08, the authors estimate that the NREGS increased daily wages by 4.73% in the dry season and by 2.87% in the rainy season.[28] This gives an average effect of 3.8%. In the year 2006/07 the total amount of employment-days created within the NREGS was 905,056,000. In that year, the NREGS covered a population of 627,369,270.[29] The average number of person-days of employment generated per capita was thus 1.44 in the implementing districts.

In the SEPRI data, average employment creation per capita within the NREGS is 0.94, with a standard deviation of 1.30 in 2011/12. At the mean of irrigation and NREGS employment creation, an increase in NREGS employment by 1.44 days per capita would reduce the standard deviation of log consumption by 0.011 on average. This corresponds to 31.2% of the sample mean of this variable.

mines consumption outcomes in 2014. Due to data limitations we have to use the amount of employment generated in the financial year 2011/12. The employment data of 2012/13 is incomplete in the survey.

[28] Because the authors define the dry season from January to June and the rainy season from July to December, the seasonal estimates are not comparable with our data.

[29] This is the population of Phase I and Phase II districts, thus all districts that implemented the NREGS by May 2007. To get the population estimates, we take the simple average of the Census 2001 and Census 2011 data.

With these estimates we can simulate the effect of the NREGS on labor supply, time allocated to household chores and to school activities. In our sample, the average predicted wage for women aged 19 to 65 is INR 3.79 per hour. An increase by 3.8% would raise hourly wages by INR 0.14. According to the estimates presented in Table 4.7, an increase in wages by 3.8% would raise labor supply by 14.5 hours per season or 3.6 hours per month. The reduction in risk that can be attributed to the NREGS reduces the standard deviation of log consumption by 0.011. Using the estimates from Table 4.7, this change would reduce labor supply by 8.7 hours per season. Finally, the increase in wages would also affect other household members' income and therefore labor supply. This effect is negligible though; increasing other household members' income by 3.8% would reduce labor supply by merely 0.2 hours per season. The net effect of the NREGS on labor supply at the intensive margin would thus still be positive, implying that the average woman would increase labor supply by 5.6 hours per season.

As discussed previously, the wage elasticity of time allocated to home production is considerably smaller than the wage elasticity of labor supply. As reported in Table 4.8, a 1% increase in wages would reduce time allocated to household chores of working women by 1.8 hours. An increase in wages by 3.8%, as attributed to the NREGS, would reduce time on chores by 6.9 hours per season. The corresponding change in the standard deviation of log consumption, the risk reduction effect of the NREGS, would increase the time allocated to chores by 10.1 hours. The effect through other household members' labor income is again negligible at -0.1 hours. The net effect of the NREGS on time allocated to household chores is thus positive by 3.8 hours per season.

We excluded mother's wages from the estimation of the time allocation of girls because we could not find a statistically significant effect. Instead, we control for all household members' labor income. Consistent with intra-household substitution in chores, the family's labor income seems to increase the time girls spend in home production, and decrease the time girls spend studying. Using the estimates of Table 4.9, we find that the risk reduction associated with the NREGS would reduce the time girls allocate to household chores by 0.05 hours per day. The associated change in wages and increase in other members' labor income increases time allocated to chores by 0.001 hours. We thus find a net effect of the NREGS on the time girls allocate to household chores on a typical day by negative 0.05 hours, or roughly 3 minutes daily. This corresponds to half the effect of a one standard deviation reduction in the risk variable. Using the estimates presented in Table 4.10, column (3), the reduction in risk exposure that can be attributed to the NREGS increases time allocated to studying by 0.109 hours per day. Through other household members' labor income, the NREGS reduces the time girls spend studying by 0.003 hours per day. The net effect of the NREGS would thus be an increase in time spent studying by 0.106 hours or roughly 6 minutes daily.

Our simulation suggests that an employment guarantee such as the NREGS could have positive effects on girls' human capital accumulation by increasing the time they spend in school and studying. The results suggest that at least half of the effect size can be attributed to fewer obligations within the household, as their mothers can spend more time on household chores. The average effect is admittedly fairly small, but as documented in Section 4.5.4 there is considerable variation in the effect of risk on the time allocation of girls. This suggests that the NREGS could have substantially higher effects on home time and time in school for girls in poorer households.

4.7 Conclusions

This paper develops a model that highlights the effect of wage risk on labor supply, the intra-household substitution in tasks and girls' time in school. Based on the model predictions, we test, in the context of rural India, whether wage risk affects female labor supply and time allocation to home production. We further test whether wage risk affects girls' time spent on household chores and in school.

Our results suggest that wage risk due to rainfall fluctuations increases female labor supply and reduces the time women allocate to home production. This seems to go hand in hand with an increase in the time girls in these households spend on household activities, and with a reduction of their time in school. We also conduct a number of robustness checks that support the idea that the observed effect on girls' time allocation is due to the effect of risk on the time working women can spend in home production.

We also find that the effect of risk on time in school is greater than the observed effect of risk on girls' time on household chores. This can be due to two reasons: first, because school investment becomes risky as future time allocations to school are uncertain, or second, because of the non-divisibility of school attendance.

What is the role for public policy in such a context? We argue that a public works program could offset some of the negative effects of risk and simulate the effect of the National Rural Employment Guarantee Scheme on the time allocation of women and girls. We concentrate on two effects of the program: first, it increases wages, and second, it reduces wage risk as it provides employment independently of agricultural production shocks. The simulated effect of the NREGS on time-allocation decisions is as follows: it increases the time women allocate to the labor market at the expense of leisure and of time in own-agricultural production. However, it also increases the time working women spend on household chores. This leads to a reduction in the time girls spend on chores and increases their school time. Based on this simulation, we conclude that the NREGS could benefit girls' human capital accumulation by reducing the pressure of wage risk on female labor supply.

What we cannot assess with the data used in this paper is the extent to which wage risk and, conversely, a wage-smoothing policy affect girls' school attainment, and future earnings. These are tasks for future work.

4.A Mathematical Appendix

4.A.1 Deterministic Case

In the deterministic case, the Lagrange can be summarized by

$$\begin{aligned} L = &\sum_{t=1}^{T} \delta^t U(C_t^m, C_t^h) + \phi(H_{T+1}) \\ &+ \sum_{t=1}^{T} \mu_t [g(H_t, S_t; \theta_t) - H_{t+1}] \\ &+ \sum_{t=1}^{T} \eta_t [w_t(T_t^a - h_t^a) + (1+r)A_{t-1} - A_t - C_t^m] \\ &+ \sum_{t=1}^{T} \upsilon_t [\rho(h_t^a + T_t^c - S_t) - C_t^h]. \end{aligned} \tag{4.A.1}$$

Assuming interior solutions, the first-order conditions for each period include

$$\delta^t U_{C_t^m} = \eta_t, \tag{4.A.2}$$

$$\delta^t U_{C_t^h} = \upsilon_t, \tag{4.A.3}$$

$$\mu_t g_{S_t} = \rho \upsilon_t, \tag{4.A.4}$$

$$w_t \eta_t = \rho \upsilon_t, \tag{4.A.5}$$

$$\mu_{t-1} = \mu_t g_{H_t}, \tag{4.A.6}$$

and for the last period,

$$\phi_{H_{T+1}} = \mu_T. \tag{4.A.7}$$

In the penultimate period $T-1$, equation (4.A.4) can be rearranged to

$$g_{S_{T-1}} = \frac{\rho \upsilon_{T-1}}{\mu_{T-1}}. \tag{4.A.8}$$

Inserting equations (4.A.5), (4.A.2), (4.A.6) and (4.A.7) leads to the decision rule with respect to school time

$$g_{S_{T-1}} = \frac{w_{T-1} \delta^{T-1} U_{C_{T-1}^m}}{\phi_{H_{T+1}} g_{H_T}}. \tag{4.A.9}$$

To see how the decision rule changes when going further back in time, consider that we can use (4.A.6) and (4.A.7) to get

$$\begin{aligned}\mu_{T-2} =&\mu_{T-1} g_{H_{T-1}}\\ =&\phi_{H_{T+1}} g_{H_T} g_{H_{T-1}}.\end{aligned} \tag{4.A.10}$$

The decision rule with respect to school time in $T-2$ thus changes to

$$g_{S_{T-2}} = \frac{w_{T-2}\delta^{T-2}U_{C^m_{T-2}}}{\phi_{H_{T+1}} g_{H_T} g_{H_{T-1}}}. \tag{4.A.11}$$

Equation (4.A.11) shows that each period's investment in school time depends on all future school time decisions.

4.A.2 Stochastic Case

When introducing uncertainty, the Lagrange becomes

$$\begin{aligned}L =&E_t \sum_{t=1}^{T} \delta^t U(C_t^m, C_t^h) + E_t\phi(H_{T+1})\\ &E_t \sum_{t=1}^{T} \{\mu_t[g(H_t, S_t; \theta_t) - H_{t+1}]\}\\ &E_t \sum_{t=1}^{T} \{\eta_t[w_t(T_t^a - h_t^a) + (1+r)A_{t-1} - A_t - C_t^m]\}\\ &E_t \sum_{t=1}^{T} \left\{\upsilon_t[\rho(h_t^a + T_t^c - S_t) - C_t^h]\right\}.\end{aligned} \tag{4.A.12}$$

Assuming interior solutions, the first-order conditions for each period include

$$E_t\left\{\delta^t U_{C_t^m}\right\} = E_t\left\{\eta_t\right\}, \tag{4.A.13}$$

$$E_t\left\{\delta^t U_{C_t^h}\right\} = E_t\left\{\upsilon_t\right\}, \tag{4.A.14}$$

$$E_t\left\{\mu_t\right\} g_{S_t} = \rho E_t\left\{\upsilon_t\right\}, \tag{4.A.15}$$

$$E_t\left\{\eta_t\right\} w_t = \rho E_t\left\{\upsilon_t\right\}, \tag{4.A.16}$$

$$\mu_{t-1} = E_{t-1}\left\{\mu_t g_{H_t}\right\}, \tag{4.A.17}$$

and additionally for the last period,

$$E_t\left\{\phi_{H_{T+1}}\right\} = E_t\left\{\mu_T\right\}. \tag{4.A.18}$$

In the penultimate period $T-1$, equation (4.A.15) can be rearranged to

$$g_{S_{T-1}} = \frac{\rho E_{T-1}\left\{\upsilon_{T-1}\right\}}{\mu_{T-1}}. \tag{4.A.19}$$

Inserting equations (4.A.16), (4.A.13), (4.A.17) and (4.A.18) leads to the decision rule with respect to school time in the presence of uncertainty

$$g_{S_{T-1}} = \frac{w_{T-1}\delta^{T-1}U_{C^m_{T-1}}}{E_{T-1}\left\{\phi_{H_{T+1}}\right\}E_{T-1}\left\{g_{H_T}\right\} + cov(\phi_{H_{T+1}}, g_{H_T})}. \tag{4.A.20}$$

To see how the decision rule in the presence of uncertainty changes when going further back in time, consider that we can use (4.A.17) and (4.A.18) to get

$$\begin{aligned}\mu_{T-2} =& E_{T-2}\left\{\mu_{T-1}g_{H_{T-1}}\right\}\\ =& E_{T-2}\left\{\phi_{H_{T+1}}g_{H_T}g_{H_{T-1}}\right\}.\end{aligned} \tag{4.A.21}$$

The decision rule with respect to school time in $T-2$ thus becomes

$$g_{S_{T-2}} = \frac{w_{T-2}U_{C^m_{T-2}}}{E_{T-2}\left\{\phi_{H_{T+1}}g_{H_T}g_{H_{T-1}}\right\}}. \tag{4.A.22}$$

Equation (4.A.22) shows that each period's investment in school time depends on all future school time decisions as well as on the covariance between the marginal utility of end-of-schooling education and each future marginal product of human capital stock. The more periods remain until the end-of-schooling, the smaller is each individual covariance term is, but the more uncertain overall future human capital accumulation becomes.

4.B Supplementary Tables

Table 4.B.1: Time allocation of boys

	N	Mean	SD
No of children in household	6700	2.72	(1.75)
Household size	6700	6.58	(2.98)
Married	6700	0.0052	(0.072)
Age	6700	12.2	(3.61)
Caste: SC / ST	6700	0.26	(0.44)
Religion: Hindu	6700	0.87	(0.33)
Education: no grade	6700	0.083	(0.28)
Education: primary	6700	0.48	(0.50)
Education: secondary	6700	0.32	(0.47)
Education: tertiary	6700	0.11	(0.32)
Years of Schooling	6700	5.55	(3.50)
Presently enrolled	6700	0.81	(0.39)
Labor force participation (season 1)	6700	0.052	(0.22)
Labor supply (hours, season 1)	6700	25.9	(121.2)
Own agr. production (hours, season 1)	6700	11.3	(60.3)
Household chores (hours, season 1)	6700	86.6	(136.4)
Hours per day: chores (season 1)	6700	1.32	(1.80)
Hours per day: studying (season 1)	6700	6.35	(3.50)

Notes: Sample of boys consists of all boys aged 6 to 18. Season 1 is the agricultural peak season.

References

Abdulai, A. and C. L. Delgado (1999). Determinants of nonfarm earnings of farm-based husbands and wives in Northern Ghana. *American Journal of Agricultural Economics 81*(1), 117–130. 119

Adrianzen, M. A. (2014). Social Capital and Improved Stoves Usage Decisions in the Northern Peruvian Andes. *World Development 54*, 1 – 17. 34

Afridi, F., A. Mukhopadhyay, and S. Sahoo (2012). *Female Labour Force Participation and Child Education in India: The Effect of the National Rural Employment Guarantee Scheme*. IZA Discussion Paper No. 6593. 118

Akosa Antwi, Y., A. S. Moriya, and K. Simon (2013). Effects of Federal Policy to Insure Young Adults: Evidence from the 2010 Affordable Care Act's Dependent-Coverage Mandate. *American Economic Journal: Economic Policy 5*(4), 1–28. 34

Anagol, S., A. Etang, and D. Karlan (2014). *Continued Existence of Cows Disproves Central Tenets of Capitalism?* revised version of NBER Working Paper No. 19437. Cambridge, USA. 69, 70, 72, 83

Angrist, J. D. and W. N. Evans (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review 88*(3), 450–477. 142

Attanasio, O. and B. Augsburg (2014). *Holy cows or cash cows?* NBER Working Paper No. 20304. Cambridge, USA. 70, 90, 103

Bandiera, O. and I. Rasul (2006). Social networks and technology adoption in Northern Mozambique. *The Economic Journal 116*(514), 869–902. 4

Banerjee, A. V. (2009). Big Answers for Big Questions: The Presumption of Growth Policy? In J. Cohen and W. Easterly (Eds.), *What works in development?: Thinking big and thinking small*, pp. 207–232. Washington DC: Brookings Institution Press. 1

Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster, and S. Khemani (2010). Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy 2*(1), 1–30. 134

Banerjee, A. V. and E. Duflo (2011). *Poor economics: A radical rethinking of the way to fight global poverty.* PublicAffairs. 2, 8

Banerjee, A. V. and E. Duflo (2014). Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program. *The Review of Economic Studies 81*(2), 572–607. 7, 71, 83

Bardhan, P. K. (1979). Labor supply functions in a poor agrarian economy. *The American Economic Review*, 73–83. 119

Barrett, C., S. Holden, and D. C. Clay (2005). Can food-for-work programmes reduce vulnerability? In S. Dercon (Ed.), *Insurance against poverty*, pp. 361–386. New York: Oxford University Press. 16

Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics 51*(2), 193–215. 7

Bartel, A. P. and F. R. Lichtenberg (1987). The comparative advantage of educated workers in implementing new technology. *The Review of Economics and statistics*, 1–11. 5

Basu, K. (2006). Gender and Say: a Model of Household Behaviour with Endogenously Determined Balance of Power. *The Economic Journal 116*(511), 558–580. 118

Beegle, K., R. H. Dehejia, and R. Gatti (2006). Child labor and agricultural shocks. *Journal of Development Economics 81*(1), 80–96. 119

Belsley, D. A., E. Kuh, and R. E. Welsch (1980). *Regression diagnostics: Identifying influential observations and sources of collinearity.* New York, USA: John Wiley and Sons. 85

BenYishay, A. and A. M. Mobarak (2014). *Social learning and communication.* NBER Working Paper No. 20139. Cambridge, USA. 9

Berg, E., S. Bhattacharyya, R. Durgam, and M. Ramachandra (2012). *Can rural public works affect agricultural wages? Evidence from India.* CSAE Working Paper No. 2012-05. Oxford, UK. 18

Besley, T. and A. Case (1993). Modeling technology adoption in Developing Countries. *The American Economic Review 83*(2), 396–402. 4

Bhalla, S. S. (1979). Farm Size Productivity and Technical Change in Indian Agriculture. In R. Berry and W. Cline (Eds.), *Agrarian structure and productivity in developing countries*, pp. 141–193. Baltimore: Johns Hopkins University Press. 6

Bhargava, A. K. (2014). *The impact of India's rural employment guarantee on demand for agricultural technology.* IFPRI discussion paper No. 1381. Washington DC. 18

Bianchi, M. and M. Bobba (2013). Liquidity, risk, and occupational choices. *The Review of Economic Studies 80*(2), 491–511. 6

Binswanger-Mkhize, H. P. (2012). Is there too much hype about index-based agricultural insurance? *The Journal of Development Studies 48*(2), 187–200. 16

Blundell, R., P.-A. Chiappori, T. Magnac, and C. Meghir (2007). Collective labour supply: Heterogeneity and non-participation. *The Review of Economic Studies 74*(2), 417–445. 138

Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. *Handbook of Labor Economics 3*, 1559–1695. 142

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics 90*(3), 414–427. 34

Cameron, A. C. and D. L. Miller (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources 50*(2), 317–372. 34

Carter, M. R., A. de Janvry, E. Sadoulet, and A. Sarris (2014). *Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling.* FERDI Working Paper No. P111. 16

Chay, K. Y., P. J. McEwan, and M. Urquiola (2005). The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools. *American Economic Review 95*(4), 1237–1258. 33

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 74*(368), 829–836. 82

Cohen, J. and W. Easterly (2009). Introduction: Thinking big versus thinking small. In J. Cohen and W. Easterly (Eds.), *What works in development?: Thinking big and thinking small*, pp. 1–23. Washington DC: Brookings Institution Press. 1

Cole, S., X. Gine, J. Tobacman, R. Townsend, P. Topalova, and J. Vickery (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics 5*(1), 104–135. 16

Cole, S., X. Gine, and J. Vickery (2013). *How does risk management influence production decisions? Evidence from a field experiment.* World Bank Policy Research Working Paper No. 6546. Washington DC. 6, 15, 47

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in Ghana. *The American Economic Review 100*(1), 35–69. 4

De Mel, S., D. J. McKenzie, and C. Woodruff (2008). Returns to capital in microenterprises: Evidence from a field experiment. *Quarterly Journal of Economics 123*(4), 1329–1372. 7, 8, 71

Deolalikar, A. B. (1981). The inverse relationship between productivity and farm size: A test using regional data from India. *American Journal of Agricultural Economics 63*(2), 275–279. 7

Dercon, S. (1996). Risk, crop choice, and savings: Evidence from Tanzania. *Economic Development and Cultural Change 44*(3), 485–513. 117

Dercon, S. and L. Christiaensen (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics 96*(2), 159–173. 5, 18, 128

Dodlova, M., K. Goebel, M. Grimm, and J. Lay (2015). Constrained firms, not subsistence activities: Evidence on capital returns and accumulation in Peruvian microenterprises. *Labour Economics 33*, 94–110. 6, 71

Duflo, E., M. Kremer, and J. Robinson (2008). How high are rates of return to fertilizer? Evidence from field experiments in Kenya. *The American Economic Review 98*(2), 482–488. 3, 4, 15

Duflo, E., M. Kremer, and J. Robinson (2011). Nudging farmers to use fertilizer: theory and experimental evidence from Kenya. *The American Economic Review 101*(6), 2350–2390. 8

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. 119

Dutta, P., R. Murgai, M. Ravallion, and D. Van de Walle (2012). *Does India's Employment Guarantee Scheme Guarantee Employment?* World Bank Policy Research Working Paper No 6003. Washington DC. 16, 25

Eswaran, M. and A. Kotwal (1986). Access to capital and agrarian production organisation. *The Economic Journal 96*(382), 482–498. 7

Fafchamps, M. (1993). Sequential labor decisions under uncertainty: An estimable household model of West-African farmers. *Econometrica 61*(5), 1173–1197. 18

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana. *Journal of Development Economics 106*(C), 211–226. 9, 71

Fafchamps, M. and J. Pender (1997). Precautionary Saving, Credit Constraints, and Irreversible Investment: Theory and Evidence From Semiarid India. *Journal of Business & Economic Statistics 15*(2), 180–194. 6, 7

FAOSTAT (2015). *Production statistics: Livestock Primary.* Food and Agriculture Organization of the United Nations, Statistics Division. http://faostat3.fao.org (accessed February 18, 2015). 71

Feder, G. (1985). The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints. *Journal of development economics 18*(2), 297–313. 7

Feder, G., R. E. Just, and D. Zilberman (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic development and cultural change 33*(2), 255–298. 3

Foster, A. D. and M. R. Rosenzweig (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy 103*(6), 1176–1209. 4

Foster, A. D. and M. R. Rosenzweig (1996). Technical change and human-capital returns and investments: Evidence from the green revolution. *The American Economic Review 86*(4), 931–953. 5

Foster, A. D. and M. R. Rosenzweig (2010a). *Is There Surplus Labor in Rural India?* Yale University, Department of Economics Working Paper No. 85. 32

Foster, A. D. and M. R. Rosenzweig (2010b). Microeconomics of technology adoption. *Annual Review of Economics 2*(1), 395–424. 3, 5, 9, 15

Galab, S., S. Vijay Kumar, P. Prudvikhar Reddy, R. Singh, and U. Vennam (2011). *The Impact of Growth on Childhood Poverty in Andhra Pradesh: Initial Findings from India.* Young Lives Round 3 Survey Report: India. 26, 72, 73

GHP-AHD (2014). *Dairy Project Report.* Government of Himachal Pradesh (India), Animal Husbandry Department. Shimla, India. 80

Ghule, A. K., N. K. Verma, A. K. Cahuhan, and P. Sawale (2012). An economic analysis of investment pattern, cost of milk production and profitability of commercial dairy farms in maharashtra. *Indian Journal of Dairy Science 65*, 329–336. 80

Gine, X. and S. Klonner (2008). Credit Constraints as a Barrier to Technology Adoption by the Poor: Lessons from South- Indian Small-Scale Fishery. In M. Nissanke and E. Throbecke (Eds.), *Globalization and the Poor in Asia: Can Shared Growth be Sustained?* Basingstoke: Palgrave Macmillan. 7

GoI-DAHD (2002). *Report of the National Commission on Cattle*. Government of India, Ministry of Agriculture, Department of Animal Husbandry, Dairying and Fisheries. New Delhi, India. 72

GoI-DAHD (2012). *Basic Animal Husbandry Statistics - 2012*. Government of India, Ministry of Agriculture, Department of Animal Husbandry, Dairying and Fisheries. New Delhi, India. 72

GoI-NSSO (2013). *Key indicators of household consumer expenditure in India, 2011-12*. Government of India, Ministry of Statistics and Programme Implementation, National Sample Survey Office. New Delhi, India. 71

Goldberg, J. (2014). *Kwacha gonna do? Experimental evidence about labor supply in rural Malawi*. Unpublished manuscript, University of Michigan, Ann Arbor, MI. 119

Government of India (2000). *Report of the Time Use Survey 1998-99*. Central Statistical Organization. Ministry of Statistics and Programme Implementation. 118, 119

Grimm, M., J. Krueger, and J. Lay (2011). Barriers To Entry And Returns To Capital In Informal Activities: Evidence From Sub-Saharan Africa. *Review of Income and Wealth 57*(Special Issue), 27–53. 8, 71

Grimm, M., J. Lay, and S. Lange (2012). *Credit-constrained in Risky Activities? The Determinants of the Capital Stocks of Micro and Small Firms in Western Africa*. GIGA Working Paper No. 185. Hamburg. 6

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. 119

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis 20*(1), 25–46. 60

Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica 42*(4), 679–694. 118

Heckman, J. J. (1993). What Has Been Learned About Labor Supply in the Past Twenty Years? *The American Economic Review 83*(2), 116–121. 118

Hoddinott, J., D. Headey, and M. Dereje (2014). *Cows, missing milk markets and nutrition in rural Ethiopia.* Ethiopia Strategy Support Program Working Paper No. 63. 69

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. 118

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. 18, 154

Jacoby, H. G. and E. Skoufias (1997). Risk, financial markets, and human capital in a developing country. *The Review of Economic Studies 64*(3), 311–335. 119, 120, 121

Jayachandran, S. (2006). Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. *Journal of Political Economy 114*(3), 538 – 575. 18, 21, 24, 119, 128, 138

Jensen, R. (2000). Agricultural Volatility and Investments in Children. *The American Economic Review 90*(2), 399–404. 119

Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India. *The Quarterly Journal of Economics 127*(2), 753–792. 118

Jha, R., R. Gaiha, and M. K. Pandey (2012). Net transfer benefits under India's rural employment guarantee scheme. *Journal of Policy Modeling 34*(2), 296–311. 18

Johnson, D. (2009). *Can Workfare Serve as a Substitute for Weather Insurance? The Case of NREGA in Andhra Pradesh.* Centre for Micro Finance Working Paper No. 32. 38

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014). Agricultural Decisions after Relaxing Credit and Risk Constraints. *Quarterly Journal of Economics 129*(2), 597–652. 6, 15, 47, 117

Kaur, S. (2014). *Nominal Wage Rigidity in Village Labor Markets*. Columbia University Working Paper. 24

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature 49*(4), 961–1075. 138, 142

Kennedy, E. and P. Peters (1992). Household food security and child nutrition: the interaction of income and gender of household head. *World development 20*(8), 1077–1085. 118

Klasen, S. and J. Pieters (2015). What Explains the Stagnation of Female Labor Force Participation in Urban India? *The World Bank Economic Review 29*(3). 142

Klonner, S. and C. Oldiges (2014). *Safety Net for India's Poor or Waste of Public Funds? Poverty and Welfare in the Wake of the World's Largest Job Guarantee Program*. University of Heidelberg, Department of Economics Working Paper No. 0564. 18

Kremer, M., N. Chaudhury, F. H. Rogers, K. Muralidharan, and J. Hammer (2005). Teacher Absence in India: A snapshot. *Journal of the European Economic Association 3*(2-3), 658–667. 134

Kremer, M., J. Lee, J. Robinson, and O. Rostapshova (2013, May). Behavioral Biases and Firm Behavior: Evidence from Kenyan Retail Shops. *American Economic Review 103*(3), 362–68. 9

Kremer, M., J. N. Lee, J. M. Robinson, and O. Rostapshova (2010). *The return to capital for small retailers in Kenya: Evidence from inventories*. Unpublished, Harvard University. 71

Krishna, R. (1964). Theory of the firm: Rapporteurs report. *Indian Economic Journal 11*, 514–525. 3

McKenzie, D. and C. Woodruff (2008). Experimental evidence on returns to capital and access to finance in Mexico. *The World Bank Economic Review 22*(3), 457–482. 8

McKenzie, D. J. and C. Woodruff (2006). Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises. *Economic Development and Cultural Change 55*(1), 3–42. 8, 71

Mdoe, N. and S. Wiggins (1997). Returns to smallholder dairying in the Kilimanjaro region, Tanzania. *Agricultural Economics 17*(1), 75–87. 69

Ministry of Rural Development, Government of India (2012). *MGNREGA Sameeksha: An Anthology of Research Studies on the Mahatma Gandhi National Rural Employment Guarantee Act, 2005. 2006-2012*. Edited and compiled by Mihir Shah, Neelakshi Mann and Varad Pande. New Delhi: Orient BlackSwan. 16

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. 6, 15, 47, 152

Morduch, J. (1990). *Risk, production and saving: Theory and evidence from Indian households.* Harvard University, Unpublished. 5

Morduch, J., S. Ravi, and J. Bauchet (2013). *Substitution Bias and External Validity: Why an Innovative Anti-poverty Program Showed no Net Impact.* Institute of Economic Research Working Paper No. 2013-03. Hitotsubashi University. 104

Mukasa-Mugerwa, E. (1989). *A review of a productive performance of female Bos indicus (zebu) cattle.* International Livestock Centre for Africa (ILCA). 80

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics 73*(1), 185–213. 4

National Sample Survey Office (2006). *Employment and Unemployment Situation in India (2004-05).* Ministry of Statistics and Programme Implementation. Government of India. 128

National Sample Survey Office (2014). *Employment and Unemployment Situation in India (2011-12)*. Ministry of Statistics and Programme Implementation. Government of India. 9

National Sample Survey Office (2015). *Status of Education and Vocational Trainingi in India (July 2011 - June 2012)*. Ministry of Statistics and Programme Implementation. Government of India. 134, 135

Pistaferri, L. (2003, July). Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply. *Journal of Labor Economics 21*(3), 729–754. 117, 119

Pratham (2012). *Annual Status of Education Report (Rural) 2011*. www.pratham.org (accessed February 19, 2016). 134

Protos, J., F. Morandini, M. Sheikh, and C. Kanyandekwe (2011). *One Cow per Poor Family: Reaching the most marginalized in Rwanda*. UNICEF. New York, USA. 69

Rao, V. and T. Chotigeat (1981). The inverse relationship between size of land holdings and agricultural productivity. *American Journal of Agricultural Economics 63*(3), 571–574. 7

Ravallion, M. (2009). Comment to "The New Development Economics: We Shall Experiment, but How Shall We Learn?". In J. Cohen and W. Easterly (Eds.), *What works in development?: Thinking big and thinking small*, pp. 51–54. Washington DC: Brookings Institution Press. 1

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 931–954. 82

Rodrik, D. (2009). The New Development Economics: We Shall Experiment, but How Shall We Learn? In J. Cohen and W. Easterly (Eds.), *What works in development?: Thinking big and thinking small*, pp. 24–47. Washington DC: Brookings Institution Press. 1

Rodrik, D. and M. R. Rosenzweig (2010). Development policy and development economics: An introduction. In D. Rodrik and M. Rosenzweig (Eds.), *Handbooks in*

Economics, Volume 5 of *Handbook of Development Economics*, pp. xv – xxvii. Amsterdam: Elsevier. 2

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. 117, 119, 138

Rosenzweig, M. R. (1978). Rural Wages, Labor Supply, and Land Reform: A Theoretical and Empirical Analysis. *American Economic Review 68*(5), 847–861. 138

Rosenzweig, M. R. and H. P. Binswanger (1993). Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *Economic Journal 103*, 56–78. 5, 117

Rosenzweig, M. R. and K. I. Wolpin (1993). Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investment in Bullocks in India. *Journal of Political Economy 101*(2), 223–244. 6, 69, 117

Ruvuna, F., I. Mao, R. McDowell, and M. Gurnani (1984). Environmental and genetic variation in milk yield of native cattle and crosses with Brown Swiss in India. *Journal of Animal Science 59*(1), 74–85. 80

Sadoulet, E. and A. De Janvry (1995). *Quantitative development policy analysis.* Johns Hopkins University Press: Baltimore. 3, 32

Singh, I., L. Squire, and J. Strauss (1986). *Agricultural household models: Extensions, applications, and policy.* Baltimore: Johns Hopkins University Press. 2, 3

Skoufias, E. (1993). Labor market opportunities and intrafamily time allocation in rural households in South Asia. *Journal of Development Economics 40*(2), 277–310. 118

Skoufias, E. (1994). Using shadow wages to estimate labor supply of agricultural households. *American Journal of Agricultural Economics 76*(2), 215–227. 7

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. 119

Suri, T. (2011). Selection and Comparative Advantage in Technology Adoption. *Econometrica 79*(1), 159–209. 3, 15

Swanepoel, F., A. Stroebel, and S. Moyo (2010). *The Role of Livestock in Developing Communities: Enhancing Multifunctionality.* The Technical Centre for Agricultural and Rural Cooperation (CTA). Wageningen, The Netherlands. 69

Temple, J. R. W. (2010). Aid and conditionality. In D. Rodrik and M. R. Rosenzweig (Eds.), *Handbooks in Economics*, Volume 5 of *Handbook of Development Economics*, pp. 4415 – 4524. Amsterdam: Elsevier. 1

Turner, R. L. (2004). *Livestock Production and the Rural Poor in Andhra Pradesh and Orissa States, India.* Pro-Poor Livestock Policy Initiative Working Paper No. 9, Food and Agriculture Organization. 93, 101

Udry, C. (1995). Risk and Saving in Northern Nigeria. *American Economic Review 85*(5), 1287–1300. 117

Van Den Berg, M. (2002). Do public works decrease farmers' soil degradation? Labour income and the use of fertilisers in India's semi-arid tropics. *Environment and Development Economics 7*(3), 487–506. 18

Vasilaky, K. and K. L. Leonard (2013). *As Good as the Networks They Keep? Improving Farmers Social Networks via Randomized Information Exchange in Rural Uganda.* Working Paper, University of Maryland, College Park. 9

Webb, M. D. (2014). *Reworking Wild Bootstrap Based Inference for Clustered Errors.* Queen's University Working Paper No. 1315. 34

Welch, F. (1970). Education in Production. *Journal of Political Economy 78*(1), 35–59. 5

World Bank (2016). *World Development Indicators.* http://wdi.worldbank.org/tables (accessed June 15, 2016). 10