social networks, identity and economic behavior: empirical evidence from India

by Annemie Maertens

SOCIAL NETWORKS, IDENTITY AND ECONOMIC BEHAVIOR: EMPIRICAL EVIDENCE FROM INDIA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Annemie Maertens

August 2010

SOCIAL NETWORKS, IDENTITY AND ECONOMIC BEHAVIOR: EMPIRICAL EVIDENCE FROM INDIA

Annemie Maertens, Ph.D.

Cornell University 2010

We know from economic theory and the sociological literature that (non-productive) identity and social networks matter to economic behavior. The dynamic and endogenous nature of social networks, identity and the economic decisions, as well as the absence of detailed panel data well suited to analyzing evolving social networks and identity, have made empirical analysis of the effects of social networks and identity a challenging task. As a result, the literature has been limited in its ability to empirically identify when and where social networks and identity matter to key economic behaviors.

This dissertation aims at filling this void by providing a better understanding on how identity and social network affect educational investment in children and agricultural technology adoption, arguably two key behaviors that crucially affect the welfare trajectories of rural households in developing countries.

In the first paper I look at the role of social networks in the adoption process of Bt cotton, a new type of cotton available on the Indian market since 2002. The results demonstrate the importance of knowledge about the profitability of a new technology in the adoption decision. This knowledge is established through experimentation, observing other farmers' past inputs and outputs and talking to informed parties such as company representatives and input dealers.

In addition, I find strong evidence of farmers imitating successful farmers and of social pressures inhibiting the adoption process of this new technology.

In the second paper I investigate the role of social customs and norms in educational aspirations. I find that the aspirations that parents have for their children are a complex function of wealth, time preferences, the perceived costs and returns to education, and the prevailing social norms, customs and pressures with regard to age of marriage and old age care. I do not find any evidence of social pressures directly related to the level of education.

In the third paper, we examine farmers' attitudes towards risk and find evidence that credit constraints in combination with a production set shaped by multiple technologies can increase the willingness to take up risks as farmers gamble their way out of poverty.

BIOGRAPHICAL SKETCH

Annemie Maertens was born in Dendermonde, Belgium to Toon Maertens and Trees Verhaeghe, and brought up in Waasmunster together with her sister Kristien. At the age of 17, she left Waasmunster for the city of Leuven to pursue a Master's degree in Engineering in Agriculture from the University of Leuven. For her Master thesis she spent some time in Southeast Tanzania working with rice farmers. This experience had a profound impact on her and she decided to change careers and study development economics. She subsequently obtained a Master's in Development and International Economics from the University of Namur. After a few years of working of for various NGOs, government agencies and international organizations in Belgium, Japan and Vietnam, she moved to Cornell to conduct her doctoral dissertation research under the guidance of Chris Barrett, Kaushik Basu, George Jakubson and Stephen Coate. Her research interest brought her to the semi-arid tropics of India where she collected data among 600 households with the goal of studying the role of identity and social networks in educational investment and agricultural technology adoption decisions. Annemie is married to Amalavoyal V. Chari, whom she met in the office of Kaushik Basu at Cornell University upon her return from India.

ACKNOWLEDGMENTS

I would like to thank my advisor, Chris Barrett, for guiding me with patience and enthusiasm during the six years of my graduate study at Cornell University, and the other members of my committee, Kaushik Basu, Stephen Coate and George Jakubson, for their comments and feedback during various stages of my research. I am also grateful to Jeff Prince for generously agreeing to take Kaushik's place in his (Kaushik's) absence. I am grateful to the AEM department at Cornell University for the financial support they provided, and to Linda Morehouse for administrative and moral support throughout my time in the PhD program. I would also like to acknowledge the outstanding work done by my numerous research assistants: Sanjit Anilesh, Shraavya Bhagavatula, Pramod Bangar, Sana Butool, V.D. Duche, Shital Duche, Madhav Dhere, Anand Dhumale, Nishtha Gosh, Meenal Inamdar, Shilpa Indrakanti, Navika Harshe, Sapna Kale, Jessica Lebo, Labhesh Lithikar, Nishita Medha, Ramesh Babu Para, Abhijit Patnaik, Gore Parmeshwar, Amidala Sidappa, Nandavaram Ramakrishna, K. Ramanareddy, P.D. Ranganath, Arjun Waghmode and Yu Qin. I am extremely grateful to the agencies and departments that financially or logistically supported my fieldwork in India: the American Institute for Indian Studies, the International Crop Research Institute of the Semi-Arid Tropics, the National Science Foundation (Grant No. 0649330), the Agricultural and Applied Economics Association, and Cornell's Mario Einaudi Center for International Studies, Graduate School and International Student and Scholar Office. Finally, I would like to thank all my friends at Cornell, my parents and sister in Belgium, and my cat Afro: their constant good cheer has been the backbone of my academic enterprise.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Seventy percent of the world's poor and three-quarters of the world's extremely poor people live and work in rural areas. Sustainably increasing the rural poor's income implies increasing the productivity of their main assets, i.e., land and labor. We know that education is the primary means of increasing labor productivity (Schultz 1989; Huffman 2001) and that adopting new agricultural technologies increases both labor and land productivity among smallholder farmers (Sunding and Zilberman 2001). Poverty reduction strategies in rural areas therefore often and appropriately concentrate on education and agricultural technologies.

Most of these poverty reduction strategies utilize price incentives, direct (conditional or unconditional) cash transfers, or in the case of education, improve the quality of education, to increase the uptake of new technologies and stimulate investment in education. This choice of development tools is consistent with the results of the earlier empirical research in development economics which concludes that prices, income, wealth and the attributes of the new technology or quality of schools matter.

This dissertation takes an alternative approach and argues that in some settings, identity and the structure of social interactions are likely important determinants of investment in education and agricultural technology adoption. This approach does not necessarily imply that we need to replace the traditional policy instruments with newer ones. Instead, by illuminating how individuals

make choices within their broader social context, it enriches our understanding of how the same policy interventions can generate different outcomes. In some instances, I will suggest some new policy interventions which might complement the existing ones.

The recent economics literature on identity can be roughly organized into two strands. One strand, based on Arrow's (1972) theory of statistical discrimination, argues that non-productive dimensions of identity (e.g., race) can nonetheless correlate with the underlying productive characteristics in equilibrium (Akerlof 1976; Basu 2006). A second, more recent strand, inspired by the literature on identity and categorization in sociology and psychology, goes beyond statistical discrimination and links the concept of identity in different ways to standard micro-economic concepts of beliefs (Fryer and Jackson, 2003), constraints (Loury 2002; Sen 2006) and preferences (Akerlof and Kranton 2000, 2002; Kirman and Teschl 2006).

A separate line of economic research has started to integrate social networks into economic analysis (see Jackson 2008 for an overview). While standard economic theory acknowledges the indirect effects of other agents' actions through the price mechanism and direct effects through preferences and production possibilities (e.g., externalities, public goods), recent studies of behavior on networks include the direct effects of decision makers' social connections on their expectations, constraints and preferences, mainly through information sharing (social learning), informal credit and insurance, bilateral trade and pecuniary scale effects.

These recent theoretical contributions in economics on identity and social networks have only quite recently found their way into empirical development economics. The dynamic and endogenous nature of social networks and identity, in addition to the lack of detailed social network data, has made empirical analysis of their effects a challenging task. As a result, the limited empirical research that exists on the topic typically limits itself to studying some particular channel of influence, e.g., social learning in agricultural technology adoption (Foster and Rozenzweig 1995, Bandiera and Rasul 1996, Conley and Udry 2010), or the role of the social custom of patrilocal exogamy (i.e., marriage into families outside of the village) in investment in education (Foster and Rosenzweig 2001).

This dissertation aims at (partially) filling this void by providing a better understanding of the various roles that identity and social networks play in agricultural technology adoption and educational investment in children.

In the second chapter of this dissertation I show how it is possible to empirically separate out the effects of social learning, imitation and social pressures on the decision to adopt a new genetically modified crop. In the third chapter, I empirically distinguish between the roles of various identity-based social customs, norms and pressures regarding marriage and old age care in the decision to invest in upper secondary and higher education of one's children.

The fourth chapter is a methodological contribution to the development literature on experimentally elicited risk preferences and investment behavior. Here, we start from the observation that the farmer often has various technologies that are available to him. Some of these new technologies, for instance an

irrigation system, might require a substantial investment. In this chapter, we look at how these investment opportunities influence forward-looking risk taking attitudes of farmers when they are credit constrained.

I use two primary sources of data for this dissertation. The first is the ICRISAT-VLS data, covering cropping seasons from 2001 to 2007. These data are collected by the International Crop Research Institute of the Semi-Arid Tropics (ICRISAT) in Hyderabad as part of their Village Level Studies (VLS) program. The second source is the data I collected in six ICRISAT-VLS villages during 2007-2008. In this supplementary survey, I elicited, in addition to the more standard modules on income, wealth, household composition, education, agricultural input and output, detailed information on social networks, expectations with regard to education and new agricultural technologies, risk and time preferences, and aspirations and future plans.

I chose India as the setting for my dissertation due to the complexity of rural life in India with its variety of jatis (local sub-castes), social classes, patron-client relations, factions and religions, which cumulatively make the existence of a completely integrated society unlikely. In short, this is a setting where identity and social networks are particularly salient to economic decisions and outcomes in rural India.

The second chapter of the dissertation builds on a longstanding literature on technology adoption in agriculture. This literature has established that prices, income and individuals' attributes, such as risk aversion and education, are important determinants of adoption behavior. A more recent literature looks at

the role of social interaction effects, a general term encompassing both the pecuniary and non-pecuniary effects of individuals' decisions on each other, in technology adoption. Most of these studies focus on one single type of social interaction effect, primarily social learning, and conclude that farmers learn from each others' experimentation and delay their adoption decision strategically. In terms of policy, this implies that subsidies for early adopters might counteract the free-rider effect and increase the adoption rate, but gives little guidance on who should be targeted by agricultural extension and private agribusiness agents in order to increase technology adoption rates.

In this chapter, I look at the role of social networks in the adoption process of Bt cotton, a new type of (genetically engineered) cotton available on the Indian market since 2002. Unlike existing studies, this particular setting combined with the data that I collected allow me to parse out the various channels through which social interaction effects operate. I explore these qualitatively and narrow the field to the ones that seemed to matter: social learning, social pressures and imitation. Then I develop a simple theoretical model and an econometric identification strategy that enables me, for the first time, to distinguish among these mechanisms.

The results demonstrate the importance of knowledge about the profitability of a new technology in the adoption decision. This knowledge is established by experimentation, observing other farmers' past profits and talking to input dealers, company representatives and government extension agents. For first-time adopters, the last channels, i.e., talking to non-farmer sources, is the most important channel for learning about the profitability of Bt cotton and the main

driver of Bt cotton adoption, followed by learning from a handful of successful farmers in the village. I find strong evidence of social pressures inhibiting the adoption of Bt cotton as well as of imitative behavior, i.e., farmers adopting Bt cotton on the basis of having observed the behavior of certain other farmers but without having observed the yield outcomes or profitability of these other farmers. In addition, I find some evidence of farmers free-riding on the experience of successful farmers.

The lesson that emerges from this analysis is that multiple mechanisms matter, but not equally. And because these mechanisms are correlated, omitting one tends to exaggerate the importance of others. Omitting the effects of non-farmer sources from the analysis or omitting the most successful farmers from the sample overestimates the importance of social learning. And by considering only the past experiences of farmers in one's network or focusing only on the optimal input decision conditional on adoption, one neglects the role of social pressures which can be a serious drag on adoption, whether these are due to (erroneous) beliefs about the new technology (such as about Bt cotton's health and environmental effects) or simply reflect status quo bias.

From a policy perspective this implies that providing information to the most successful farmers in the village is likely to be the most cost-effective strategy. To mitigate the free-rider effect, a small (temporary) subsidy might be given to these early adopters. However, where social pressures are strong, as is the case in one of the villages I investigate, a large-scale information campaign may be required to shift the equilibrium.

In the third chapter of this dissertation, I contribute to the literature on education in developing countries. The literature to date has shown that in addition to parental education, work opportunities, village and regional development, school quality and costs, investments in education are significantly stratified by gender and social groups. Using a detailed child level dataset from three villages in India, I dissect the roles of various identity-based social norms, customs and pressures in the education decision, taking into account the differences in (perceived) returns to education. Given that the large majority of the children in these villages routinely complete elementary education and lower secondary education, I instead focus on the parents' plans to invest in upper secondary education and higher education, which I will refer to as educational aspirations.

Using regression analysis, I explain the variation in aspirations for each child by the variation in (perceived) costs and returns, credit constraints, time preferences, and social norms, customs and pressures. I first document that educational aspirations are much lower for girls (compared to boys) and for lower caste groups (compared to upper caste groups). Lower aspirations for girls can be partially explained by social norms concerning the ideal age of marriage and the provision of old-age care. The ideal age of marriage (as per social norms) is lower for girls (compared to boys). After marriage, the bride moves in with her husband or his parents, a practice known as patrilocality. Once the bride has switched homes, it is often socially unacceptable to continue to support her natal family, either financially or in terms of physical care. Sons are the principal providers of old-age care for the parents, and are therefore more likely to be the recipients of higher education.

Lower aspirations for lower caste groups can be partially explained through lower expectations with regard to the returns to higher education, and different social norms with regard to the ideal age of marriage (which is lower) and the patrilocal system (more is expected from girls compared to the higher castes).

I do not find any evidence of social pressures directly related to the level of education and cannot confirm a negative relationship between the price of education (in terms of dowry price) and aspirations. On the contrary, increasing the dowry associated with a certain education level, increases the chances of that level being aspired to. This might point to social status effects dominating the price effect of an increase in dowry.

The results of this study suggest that subsidizing education will not immediately induce greater educational investments. There is a complex set of social norms that determines the equilibrium value of education for different groups in society, and the dissolution of these norms is likely to be a slow process.

In the fourth chapter, I investigate, together with David Just, how forward-looking investment behavior can influence attitudes towards risk. These attitudes were elicited via their evaluations of hypothetical but realistic farm alternatives involving various risky outcomes. The motivation for our work is the anomalous finding that, using a marginal measure of risk aversion, i.e., based on the difference in willingness-to-pay between a less and a more risky realistic gamble, we find that most individuals are risk loving.

I this chapter, we test a simple model of forward-looking dynamic risk behavior, inspired by Lybbert et al. (2010). This model distinguishes between

the familiar static concept of risk aversion and forward-looking dynamic risk responses. The latter takes into account the fact that in the absence of credit markets, a lucky draw might enable the farmer to make a large fixed-cost investment which would allow the farmer to move to a higher level equilibrium characterized by a higher annual income, and consequently higher standard of living in future periods.

We find that 85% of the farmers are willing to pay more for a distribution which is second-order stochastically dominated by the base line distribution, indicating risk loving behavior and that, across distributions, the farmers pay a disproportionate amount of attention to the probability of best yield outcome. The (estimated) effect of increasing the probability of the best outcome by 10% is about 4 times the (estimated) effect of decreasing the probability of the worst outcome by 10%.

We find evidence in the data that supports the model proposed. For farmers who are credit constrained, increasing the dryland owned by one acre increases the difference in willingness-to-pay between a high variance distribution and a low variance distribution (with the same mean). This indicates that farmers who own more dryland are more risk loving compared to farmers who own less dryland or farmers who own only irrigated land. Also, having one more school going child over the age of 15 years in the household, increases the difference in WTP. This indicates that farmers who might want to invest in higher education of their children (but are credit constrained) are more willing to take on risks compared to farmers who have no school going children in this age range.

This implies that our standard measures of risk aversion which look only at contemporaneous decisions, whether they are inferred from observed economic behavior using structural models of behavior combined with econometric techniques, or elicited via experimental techniques, might not be able to capture and predict actual investment behavior under uncertainty.

To conclude, this dissertation makes a substantial contribution to our understanding of investment behavior in developing countries. From this dissertation we know that various social interaction effects matter, but that they are not all equally important. In addition, we learned that attitudes towards risk depend on more than just static risk preferences.

BIBLIOGRAPHY

[1] Akerlof, George A. 1976. "The Economics of Caste and the Rat Race and other Woeful Tales"' *The Quarterly Journal of Economics,* 90:4, pp. 599–617.

[2] Akerlof, George A. and Rachel E. Kranton. 2000. "Economics and Identity". *The Quarterly Journal of Economics,* 115:3, pp.715-753.

[3] Akerlof, George A. and Rachel E. Kranton. 2002. "Identity and Schooling: Some Lessons for the Economics of Education." *Journal of Economic Literature,* 40:4, pp.1167–1201.

[4] Arrow, Kenneth, J. 1972. "Models of Job Discrimination," in *Racial Discrmination in Economic Life*, A. H. Pascal (ed.). DC Heath: Lexington Books.

[5] Bandiera, Oriana and Imran Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique." *Economic Journal*, 116:514, pp. 869-902.

[6] Basu, Kaushik. 2006. "Participatory Equity, Identity, and Productivity: Policy Implications for Promoting Development." Working Paper Department of Economics, Cornell University.

[7] Conley, Timothy G and Christopher Udry. 2010 "Learning About a New Technology: Pineapple in Ghana." *American Economic Review,* 100:1, pp. 35–69.

[8] Foster, Andrew and Mark R. Rosenzweig. 2001. "Missing Women, the Marriage Market, and Economic Growth." Working Paper.

[9] Foster, Andrew D. and Mark R. Rosenzweig. 1995. Learning by Doing and Learning From Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy,* 103:6, pp. 1176–1209.

[10] Fryer, Roland and Matthew O. Jackson. 2004. "A Categorical Model of Cognition and Biased Decision-Making." Working Paper Department of Economics, Harvard University and Stanford University.

[11] Huffman, Wallace E. 2001. "Human Capital: Education and Agriculture," in *Handbook of Agricultural Economics*. Bruce L Gardner and Gordon C Rausser eds: Elsevier Science, pp. 333-81.

[12] Jackson, Matthew O. 2008. *The Economics of Social Networks*. Princeton University Press.

[13] Kirman, Alan and Miriam Teschl. 2006. *Changing Identities*. GREQAM Working paper, Aix-en-Provence and Marseille.

[14] Lybbert, Travis J., David R. Just, and Christopher B. Barrett. 2010. "Estimating Risk Preferences in the Presence of Bifurcated Wealth Dynamics: Do We Misattribute Dynamic Risk Responses to Statis Risk Aversion?" Working Paper Cornell University and UC Davis.

[15] Loury, Glenn C. 2002. *The Anatomy of Racial Inequality*. Cambridge: Harvard University Press.

[16] Schultz, Paul T. 1989. "Education Investments and Returns," in *Handbook of Development Economics*. Hollis B. Chenery and T.N. Srinivasan eds: Elsevier Science, pp. 544-630.

[17] Sen, Amartya. 2006. *Identity and Violence: The Illusion of Destiny*. New York and London: W.W Norton and Company.

[18] Sunding, David and David Zilberman. 2001. "The Agricultural Innovation Process: Research and Technology Adoption in a Changing Agricultural Sector," in *Handbook of Agricultural Economics*. Bruce L Gardner and Gordon C. Rausser eds: Elsevier Science, pp. 207-61.

CHAPTER 2

WHO CARES WHAT OTHERS THINK (OR DO?) SOCIAL LEARNING, SOCIAL PRESSURES AND IMITATION IN COTTON FARMING IN INDIA

2.1 Introduction

The introduction of new agricultural technologies in developing countries has the potential to significantly improve agricultural incomes and living standards thereby transforming the rural economy. However, when new agricultural technologies are introduced, adoption often does not occur immediately. Farmers instead appear to follow a complex pattern of gradual adoption, dis-adoption and often non-adoption (Besley and Case 1993, Feder et al. 1985, Sunding and Zilberman 2001).

So, what determines the pattern of adoption? Conditional on what an individual knows and with whom he interacts, it is plausible that prices, incomes and the attributes of the individual (such as risk-aversion, education, etc.) are important predictors of adoption behavior and the earlier literature on technology adoption indeed emphasized these factors.[1] To understand the adoption process completely however, we need to understand how an individual learns about the profitability of a new technology and how his social relations might influence his decision.

[1]For an overview of the earlier literature see Feder et al. (1985). This literature tells us that prices, incomes, risk aversion and education matter, however as these studies do not control for knowledge or social connections, some of this correlation might be spurious.

A significant share of the recent literature on agricultural technology adoption therefore has placed the individual's adoption decision in the context of his social relations. An interesting example of how social interactions can mediate technology adoption is the phenomenon of strategic delay: because knowledge is to some extent a public good, an individual may prefer to strategically delay his own experimentation, and instead 'free-ride' on the experimentation of his peers. In the aggregate this may result in a poor equilibrium characterized by extremely low rates of adoption. This learning from the experimentation of others, often referred to in an agricultural context as social learning, can be conceptually distinguished from the phenomenon of imitation, whereby an individual copies the behavior of another, without having observed the results of the latter's experimentation. This kind of behavior may arise from a desire to 'keep-up-with-the-Joneses' (I will refer to this phenomenon as behavioral imitation). Or it may reflect a kind of learning in which the adoption behavior of the individual being copied is itself taken to be a signal about the profitability of the new technology (I will refer to this phenomenon as learning imitation). Finally, there are settings where new technologies are viewed with suspicion or may be (correctly or incorrectly) associated with harmful externalities, in which case social pressures may inhibit the adoption process of the new technology (Appadurai 1989, Moser and Barrett 2006, Rogers 1965 and Vasavi 1994).

Identifying the relative importance of these various channels of influence is obviously important from a policy perspective, but has turned out to be elusive in practice. In the first place, obtaining an accurate picture of an individual's social network from the information contained in a limited sample is not straight-

forward, and existing methods of doing so are not without pitfalls. Even if social networks are well-measured, there remains the thorny issue of inferring causal effects from correlations in individuals' behavior.[2] First, one needs to separate the social interaction effects from the so-called correlated effects, i.e., the unobservables that either coordinate the actions of individuals through similar constraints or directly influence network formation. Second, one needs to deal with the simultaneity problem, i.e., the problem of estimating the causal effect of changing an individual's action on another individual when these actions are jointly determined in equilibrium. Last, but not least, one still needs to find identifying assumptions that will make it possible to separate out the various kinds of social interaction effects, namely, social learning (and associated free-rider effects), social pressures and imitation.

Using a unique household-level dataset I collected for this purpose, I try to separately quantify these three channels of influence. I study the adoption process of *Bacillus thurigiensis* (Bt) cotton, a new type of cotton that requires fewer pesticides and can increase expected yields, in three villages in India (Qaim 2003). In addition to social learning and imitation effects, there is reason to believe that social pressures are important to the adoption process because of the

[2]Manski (1993) shows that in a regression model of behavior in large groups in which individual behavior is permitted to vary linearly with mean behavior in the group (expressing what he terms endogenous social interaction effects), with the mean values of exogenous attributes of group members (expressing what he terms contextual interactions effects), and with personal characteristics that may be similar across group members (expressing what he terms correlated effects), one is - assuming rational expectations - unable to identify endogenous social interaction effects. Prospects of identification however improve if one relaxes one or more underlying assumptions of Manski's model. One could consider non-linear interaction effects instead of linear effects (Bandiera and Rasul 2006), impose time sequencing on the endogenous effects instead of considering contemporaneous endogenous effects (Conley and Udry 2010, Moser and Barrett 2006), assume correlated and contextual effects are absent or move away from the assumption of global interaction and consider network-based interactions.(Bandiara and Rasul 2006, Bramoullé et al. 2009, Conley and Udry 2010).

commonly-held (erroneous) belief in these villages that the cultivation of Bt cotton is harmful to animals (as well as humans and the environment).[3] Using a combination of experiments and questionnaires I elicited detailed information on cotton production, (current) perceived bio-safety hazards of Bt cotton, (current) beliefs regarding the profitability of Bt cotton and social networks. These data are then combined with six years of a panel survey on the same households that was previously collected by the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), containing detailed agro-climatic, price and wealth data which allow me to control for potentially confounding correlated effects.

Using information on farmers' beliefs about the profitability of Bt cotton I directly distinguish social learning (including learning imitation) from (behavioral) imitation and the effects of social pressures - essentially, adoption decisions that are inconsistent with beliefs about the profitability of Bt cotton (after controlling for credit constraints) must reflect the operation of the latter two effects.

To further distinguish between social pressures and imitation, I use information on the behavior of a set of 'progressive' farmers. These generally successful farmers, who were identified at the beginning of the survey, are the earliest adopters of new technologies, play a central role in the diffusion of information and are often an object of imitation. Equally important, their adoption decisions are, by their own account, not influenced by social pressures or imitation,

[3]The scientific evidence available to date has shown that Bt cotton poses no significant risk to the environment or human/animal health (see Menselsohn et al. 2005, Shelton et al. 2002).

which implies a uni-directional imitation effect from the 'progressive' farmers to the 'non-progressive farmers'.

Using information on farmers' beliefs about the profitability of Bt cotton, in conjunction with a simple model of social pressures and data on farmers' beliefs about the bio-safety of Bt cotton, and information on the expected behavior of the progressive farmers, I identify the social pressure and imitation effects. I find strong evidence of social pressures, but no evidence of behavioral imitation. Apparently, technology choice is not a potentially expensive fashion statement.

Using the panel dimension of the data, I investigate the role of social learning in the formation of beliefs about the profitability of Bt cotton. I construct, for each farmer at each point in time, an aggregate measure of the information available to him on (historical) inputs and outputs of other farmers who cultivated Bt cotton, differentiating between information available on the experiences of progressive farmers from that on non-progressive farmers. Information on the contemporaneous behavior of progressive and non-progressive farmers is used to identify the operation of social pressures and imitation. Social pressures are assumed to operate through the contemporaneous behavior of non-progressive farmers, while imitation effects are captured by the contemporaneous behavior of progressive farmers. Because contemporaneous actions of non-progressive farmers are jointly determined, I use an instrumental variable strategy to identify these effects.

The results indicate that for first-time adopters, talking to informed parties such as company representatives and input dealers is the most important driver

of adoption. The second most important channel is learning from progressive farmers. The learning effect is concave, suggesting decreasing returns to new information. Consistent with this learning effect, I find evidence of farmers free-riding on the experimentation of the progressive farmers, delaying their adoption when they know more progressive farmers currently adopting. In addition, I find strong evidence of farmers imitating the progressive farmers, i.e., adopting Bt cotton without having observed the yield outcomes of these farmers and social pressures inhibiting the adoption process of this new technology.

The separate identification of different social interaction effects has been attempted in two other papers. Moser and Barrett (2006) look at the effects of social learning and social pressures on the adoption of a new technique for rice cultivation in Madagascar. Their identification strategy is based on the assumption that a farmer's probability of initial adoption of a new technology is a concave function of the cumulative experience of village farmers. If so, the effect of the cumulative experience of village farmers up to $t-1$ should be larger than the effect of experience of village farmers at time t. They find this not to be the case and attribute this difference to social pressures. Bandiera and Rasul (2006) look at the effects of social learning on the decision to adopt sunflower cultivation in Mozambique. They incorporate imitation effects by including the behavior of a key individual in the village that could act as the focal point as a regressor. However, they find no evidence of imitation using this strategy.

Among papers that study social learning alone, Conley and Udry (2010) look at how farmers learn about the appropriate use of fertilizers from the experimentation of others in pineapple farming in Ghana. They capture social

learning through the past profits of the contact of each farmer and control for correlated effects by using information on soil and other location-specific characteristics. They conclude that social learning is important in the diffusion of knowledge regarding pineapple cultivation in Ghana.

Foster and Rosenzweig (1995) investigate the adoption of high-yielding seeds during India's green revolution. As they have no measure of individuals' social networks, a problem that is fairly common in this literature, they use the total number of plots cultivated with improved seeds in the village up to the previous year to capture the effects of social learning. They also find evidence of social learning and learning spillovers in the shape of strategic delays of experimentation.

In terms of policy, the importance of social learning and free-riding implies that subsidies for early adopters might be required to increase the adoption rate. However, the existing literature gives no guidance on who should be targeted by agricultural extension officers and private agribusiness agents. The contribution of the present study is to show that multiple mechanisms matter (although not equally). This is potentially important because to the extent that these mechanisms are correlated, which I show they are, omitting one tends to exaggerate the importance of others. Omitting non-farmer sources of information or the most successful farmers from the sample overestimates the importance of social learning. And by considering only the past experiences of farmers in one's network or focusing only on the optimal input decision conditional on adoption, one neglects the role of social pressures which can be a serious drag on adoption, whether these are due to (erroneous) beliefs about externalities (as in this study)

or simply reflect status quo bias (as in Moser and Barrett 2006). From a policy perspective, the results of this paper imply that providing information to the most successful farmers in the village is likely to be the most cost-effective strategy. In India, the identity of these successful farmers is generally well-known by the head of the village and the government extension agents who visit the village every 2-3 months to provide information on subsidies, prices and new technologies and collect land and irrigation taxes. As the private agribusiness often have a better knowledge of the most recent technologies available, this gives scope for public-private partnerships with the public side identifying the key individuals and the private side providing information on the technology. To mitigate the free-rider effect, a small (temporary) subsidy might be given to early adopters. However, where social pressures are strong, as is the case in one of the villages I investigate, a large-scale information campaign may be required to shift the equilibrium. Adoption rates that remain below 5-10% as well as strong bio-safety or other concerns with regard to the new technology among the progressive farmers might be indicators of social pressures being at play.

The remainder of this article is structured as follows. The next section provides some background information on the Bt cotton technology in India. Section 3 introduces the data. To set the stage for the analysis, Section 4 discusses some selected descriptive statistics and provides an informal description of the adoption process in the three villages, based on conversations with respondents. Section 5 outlines a simple theoretical model and the empirical identification strategy. Section 6 presents the results and Section 7 concludes.

2.2 Bt cotton in India

India produces almost twice as much cotton as the USA (see Table 2.1). In terms of yield however, India is at the tail-end of the global distribution. The average cotton yield in India is only one third of China's average yield. The main cotton producing states in India are Gujarat, Maharashtra, Punjab and Andhra Pradesh, producing over 80% of the total cotton production. In these regions, cotton is one of the main cash crops of farmers and supports a significant section of the population through the processing industries and trade.

Table 2.1: Basic cotton statistics of China, US and India

	Cotton production[1]					Cotton exports[1]	Average cotton yield[2]
	1985	1990	1995	2000	2008	2008	2008
China	4,137	4,507	4,768	4,420	7,947	16	1,325
USA	2,924	3,376	3,897	3,742	2,985	2,830	951
India	1,964	1,989	2,885	2,380	5,443	1,328	579

Source: United States Department of Agriculture, PSD Online. Updated 10/10/2008. Notes: [1]in thousand metric tons; [2]in kg/ha (ginned) cotton

Losses in cotton production in India are mainly due to its predominant cultivation under rainfed conditions and its susceptibility to 166 species of insects, pests and diseases. Today, nearby 50% of pesticides used in India are used on cotton (ISAAA 2005). The major pests affecting cotton are jassids, aphids, white fly and bollworms.[4]

[4]The cotton bollworm complex encompasses the American bollworm (*Helicoverpa armigera*), pink bollworm (*Pectinophora gossypiella*), spiny bollworm (*Earias insulana*) and spotted bollworm (*Earias vittella*) (ISAAA 2005, Asia-Pacific Consortium of Agricultural Biotechnology 2006).

As a response to bollworm pest problems, Monsanto, a US agricultural company, developed the Bt genetically modified technology during the 1980s. In collaboration with the Maharashtra Hybrid Seed Company (Mahyco), the technology was then introduced into several of Mahyco's breeding lines during the 1990s. In 2002, the Genetic Engineering Approval Committee (GEAC), an Indian government body, approved the commercial release of three Bt cotton cultivars[5] of Mahyco. As of August 2008, 225 Bt cotton cultivars had been approved by GEAC.

Since its introduction in India, the Bt technology has been surrounded by a lot of controversy and debate largely centered around bio-safety of genetically modified crops. These bio-safety concerns may lead to social pressures that inhibit the adoption of Bt cotton, i.e., a farmer might be accused of endangering the health of animals and people in the village as well as generating negative impacts on the soil fertility and water quality of neighboring plots. The scientific evidence available to date has shown that Bt cotton poses no significant risk to the environment or human/animal health (Menselsohn et al. 2005, Shelton et al. 2002).

The Bt cultivars contain a gene sourced from the soil bacterium *Bacillus thuringiensis* in their DNA sequence.[6] This gene produces a protein that is toxic to the bollworms. The Bt gene does not effectively control against all bollworms

[5]A cultivar is a particular variety of plant species that is being cultivated.

[6]This protein, when entering the gut of the insect in the larvae phase, meets a receptor protein, binds with it and punctures the wall of the intestine, which leads to paralysis and eventually death of the insect. This receptor protein is only found in insects of the Lepidoptera order. This implies that Bt cotton has no impacts on humans or other mammals.

and provides no protection against other pests and diseases.[7] Also, when a Bt gene is inserted in the DNA of a plant it only affects its pest resistance. It does not affect its duration, drought resistance, or fiber length, etc. These properties are determined by the genetic properties of the cultivar in which the gene was inserted.

Data from trials on Bt cotton and its isogenic non-Bt counterpart in India show that in high bollworm pressure years, independent of soil and climatic conditions, profits from Bt cotton are higher than profits from non-Bt cotton.[8] If the bollworm pressure is low (and hence few pesticides are needed) and the price of the Bt seed is high this result might not hold. Pemsla et al. (2004), using farm data from non-isogenic Bt and non-Bt cultivars in Karnataka during 2002-03, a low bollworm pressure year, conclude that, irrespective of soil and climatic conditions, non-Bt cotton outperforms Bt cotton.[9]

In the three villages I consider in this paper, Aurepalle in the Mahbubnagar district in Andhra Pradesh and Kanzara and Kinkhed in the Akola district in Maharashtra, profits from Bt cotton are higher than profits from non-Bt cotton, irrespective of bollworm pressure, soil and climatic conditions.[10]

[7]The protein is toxic to several insects of the Lepidoptera order, among others the American bollworm, the spiny bollworm, the spotted bollworm and to a lesser extent, the pink bollworm.

[8]Data from Qaim and Zilberman (2003) from trials conducted in 25 districts on 175 farms in Maharashtra, Tamil Nadu and Madhya Pradesh during 2001-02, a high bollworm pressure year, show that the cumulative distribution function of the profits of Bt cotton first-order stochastically dominates the non-Bt function. I am grateful to Matim Qaim for sharing the data with me.

[9]More specifically, the cumulative distribution function of the profits of non-Bt cotton first-order stochastically dominates the Bt cotton function.

[10]In particular, in 2006-07, a low bollworm pressure year in all the villages, the cumulative distribution function of the profits of Bt cotton first-order stochastically dominates the cumulative distribution function of the profits of non-Bt cotton in Aurepalle and Kanzara. Little can be said about Kinkhed as there are so few Bt cotton farmers. In 2007-08, a high bollworm pressure year in the Akola villages, the cumulative distribution function of the profits of Bt cotton first-order

2.3 Data collected

I use two primary sources to construct the dataset for this paper. The first is six rounds of the ICRISAT-VLS data, covering cropping seasons from 2001 to 2007. These data are collected by the International Crop Research Institute of the Semi-Arid Tropics (ICRISAT) in Hyderabad as part of their Village Level Studies (VLS) program.[11]

To obtain information on social networks and beliefs regarding Bt cotton, I resurveyed the 246 ICRISAT-VLS households in Aurepalle, Kanzara and Kinkhed in 2007-08.[12] In this supplementary survey I also elicited each house-

stochastically dominates the cumulative distribution function of the profits of non-Bt cotton in Kanzara. Nothing can be said about Aurepalle as no cotton farmers cultivated non-Bt cotton in 2007-08. To get an idea of the magnitude of these profits, in 2007-08, in Aurepalle, the average profit of Bt cotton farming stood at 7,760 Rs/acre (st. dev 6,055 Rs/acre). In Kanzara the average profit of Bt cotton farming stood at 7,760 Rs/acre (st. dev.5,299 Rs/acre) versus 1,367 Rs/acre (st. dev. 3,056 Rs/acre) for non-Bt cotton. In Kinkhed the average profit of Bt cotton farming stood at 1,176 Rs/acre (st. dev.3,842 Rs/acre) versus 153 Rs/acre (st. dev. 2,762 Rs/acre) for non-Bt cotton. To check whether Bt cotton farmers are operating in different conditions compared to non-Bt cotton farmers I run a farmer/year fixed effect regression using farmer-level profits of 2001-08; the coefficient on Bt is 3,276 Rs/acre and is significantly different from zero at the 1% level.

[11] In this program, ICRISAT followed 300 households from six villages during the period 1975-1985 every three weeks. This dataset, known as the first generation VLS, contains detailed household and plot level data. In 2001, ICRISAT restarted the panel, revisiting 185 of the first generation VLS households and their split-offs, in addition to 261 newly added households, to make the sample representative for each village in terms of land-holding size. For an overview of the goals, methods and outcomes of, respectively, the first and second generation VLS see (i) Singh et al. (1985) and Walker and Ryan (1990), (ii) Bantilan et al. (2006) and Rao and Charyulu (2007). Of the ICRISAT-VLS data, I use the modules on household composition, landholding, prices and wealth.

[12] Of the 199 households covered in 2001-02 by the ICRISAT-VLS, 92% were still in the sample in 2007-08. The households that dropped out were, on average, smaller in terms of household size, higher educated, with less total land, but more irrigable land compared to the household that remained in the sample. I interviewed these 184 households plus some of their split-offs and newly added households, a total of 246 households. This sample is representative for the village in 2007-08 (see Bantilan et al. 2006 and Rao and Charyulu 2007). Of these, 68% have data for all seven cropping years. Of the 32% of the households who are included in the sample from a later date onwards, 30% are households that have split off from sample households during 2001-2008. These split-off households are included from their date of split-off. See also Appendix A.

hold's cultivation plans for the year 2008-09, past cotton production and marketing decisions covering the period 2001-08 and included questions on household composition, landholding, wealth and per-plot agricultural inputs and outputs all pertaining to the 2007-08 season.

Table 2.2 introduces the three villages. Aurepalle, with 925 households, is the largest of the three villages. It is located in the drought-prone, poor, Telangana region of Andhra Pradesh and in terms of average income situated between the richer Kanzara and poorer Kinkhed. Kanzara and Kinkhed, with respectively 319 and 189 households, are located in the less drought-prone Akola district of West Maharashtra. The ICRISAT-VLS sample includes 128, 63 and 55 households in Aurepalle, Kanzara and Kinkhed, respectively. The average education level of the respondent (i.e., the main decision-maker with regard to agriculture) is low, especially in Aurepalle (2.31 years). The average size of a household is between 4 and 5 members in all three villages.

Table 2.2: Basic descriptive statistics of Aurepalle, Kanzara and Kinkhed

	Aurepalle	Kanzara	Kinkhed
Number of households in village	925	319	189
Number of households in sample	128	63	55
Median rainfall (mm/year)[1]	434	748	745
Distance to nearest town (km)	10	9	12
Average education level of respondent (in years)	2.31	6.61	6.89
Average number of household members	4.23	4.87	4.5
Average yearly income (Rs)[2]	43,543	53,720	38,087

Notes: [1]2001-2007; [2]2004-2005

In addition to the ICRISAT-VLS households, I also collected data among 21 additional 'progressive' farmers. These generally successful farmers play a central role in the dispersion of information and are often an object of imitation.

Omitting these progressive farmers from the analysis could cause spurious correlation among farmers who are connected to the same progressive farmers. Of the ICRISAT-VLS farmers, 22 farmers were also labelled progressive farmers, totalling 43 progressive farmers. I identified this set of generally successful farmers through group discussions with the village leaders, the junior VLS investigators who have been living in the village since 2001 and the senior VLS investigators who have known the villagers since the beginning of the VLS survey in 1975. The tables in Appendix B show how these progressive farmers are different from the remainder of the sample. On average, the progressive farmers are more educated and own more land, of which a larger share is irrigated. The progressive farmers are far more likely to adopt Bt cotton and, conditional on adoption, are more profitable and adopt the new technology earlier compared to the remainder of the sample. Appendix C shows that the VLS respondents indeed approach these progressive farmers for advice in case of agricultural problems.[13]

Finally, I completed a village questionnaire, including information on climate[14], bollworm pressure[15] and village infrastructure, with the assistance of

[13]Note that there is a possibility that I might have omitted progressive farmers who have migrated from the village since 2001. I do not expect this to be a serious issue as, looking at the characteristics of the households who have left the VLS sample since 2001, only one was a larger farmer household that used to cultivate cotton.

[14]I use three different daily rainfall series: from June 2005 onwards, I use the daily rainfall data as measured by the VLS resident investigators in the villages; before June 2005, I use the daily rainfall data of Amangal (gauge data) as provided by the Tehsil office at Madgul for Aurepalle and the daily rainfall data of Murtizapur (gauge data) as available from the Maharastra State government website http://www.mahaagri.gov.in/.

[15]I obtained the bollworm pressure data from various AICIP and GEAC reports and double-checked this information with the perceived bollworm pressure as recorded in the village questionnaires.

the ICRISAT-VLS investigators, *sarpanch*[16], three knowledgeable people in each village, the *Mandal/Tehsil* Revenue Office[17] and the District Collector's Office.

I measured social networks using three different methods. The first method asked the respondent how many farmers he knew in each year since 2001-02 in different social groups (total, village, relatives) that adopted Bt cotton in that year and what the experience of these farmers was on average with Bt cotton. Recognizing measurement errors due to recall, this method has the advantage of capturing most links of the respondent, but provides little information on the nature of this link. Ideally, one would like to take into account certain aspects of the relationship between the respondent and his contact. However, due to time constraints, one cannot ask the respondent for information regarding his relationship with every one of his peers.

This implies that one needs to think about sampling the social network of each respondent. I opt for a technique called the 'random-matching-within-sample' method based on Conley and Udry (2001) and Santos and Barrett (2007).[18]

[16] A *sarpanch* is a democratically elected head of a village level statutory institution of local self-Government called the *Gram* (village) *Panchayat* in India.

[17] *Mandal* refers to the third-level administrative area in Andhra Pradesh, below state and district. The equivalent in other states is *tehsil* (or *taluka*).

[18] The two most common techniques are respondent-driven "snowball sampling" (i.e., existing study subjects recruit future subjects from among their acquaintances) and taking the "network of a sample". The first technique is useful when one is interested in properties of the network itself, but as it results in a non-representative sample of the households, it is not a useful technique for the economic analysis of the effects of social networks on behavior. The second technique artificially truncates the network and is not representative of the "network of the population". As such, this technique might result in biased estimates of micro-economic behavior in the presence of structured networks as unobservables influence both the probability of a link and, independently, the behavior of interest (Santos and Barrett 2007). Note that both the "network of a sample" and the "random matching within sample" will provide biased estimates of the effects of social networks in the presence of star-shaped network structures.

Through the random-matching-within-sample method, I elicit the details of the relationship between two randomly drawn respondents from the VLS sample. Each respondent is matched with six randomly drawn VLS respondents and four ex-ante identified fixed progressive farmers.[19] The questions in the random-matching-within-sample method include a set of questions regarding the relationship between the respondent and the match and the knowledge that the respondent has about the match's farming activities in terms of inputs and outputs.[20]

Note that two important assumptions are underlying the random-matching-within-sample method. First, by using measures based on elicitation in 2008, I assume that the nature of the relationship between two randomly drawn individuals has not changed in the last seven years. [21] Second, I assume that the

The reason for this bias is that if a "source" in the network, i.e., someone who has many links compared to others, is not sampled, its absence generates an omitted variable bias.

[19]Appendix C comments on the selection of these four fixed progressive farmers.

[20]One can use the information elicited through the random matching within sample game in two different ways. First, one can extrapolate the information obtained through the random matching within sample game to the population level as the percentage of links in the game provides an unbiased estimate of the percentage of links in the population. Assuming the absence of unobservables which influence both the probability of a link and the probability to adopt Bt cotton, one can then use this estimate to examine the effect of these links on the adoption decision. Second, one can regress the probability of a link on the attributes of the respondent and the match, and use the results of this regression to predict out-of-sample (i.e., out of the random matching within sample game) the probability of a link between any two individuals in the sample. the advantage of the latter method is that it strips the independent variable "link" of its unobservable part and as such avoids the potential bias. In addition, it makes use from a larger set of farmers each farmer is connected to. However, if the set of attributes poorly predict these links, there might be too little variation left in the predicted link variable to asses the effects of the links on the adoption decision. Note also that using the results of the random matching within sample method directly implies that I have no link for the pairs which involve a match that did not cultivate cotton in 2007-08. These links were set at zero, i.e., "no link". As 25% of the respondents who have farmed cotton in the last 7 years did not farm cotton in 2007-08, this might be a potential source of bias. This is not a problem for the regressions using the predicted links.

[21]During the trial round, I tested this assumption, and went back in time with the set of questions regarding the relationship between the respondent and the match, i.e., how long have you known X? How frequently do you talk to X?, etc. Only a couple respondents mentioned a recent change in the relationship, in all cases caused by a quarrel between families. I could not

relevant social network in terms of social learning, social pressures and imitation does not exceed the village boundaries.[22]

The third method of measuring social networks asks the cotton farmer who he would approach for information if he had problems with his cotton crop, eliciting some characteristics of that individual's farming activities and the relationship between the respondent and that individual.

In order to be able to take into account the information that the respondent receives from non-farmer sources, I added a section on the information obtained since 2001-02 from contacts with extension agents, NGOs, company agents, input dealers and ICRISAT.[23] The majority of the information (from non-farmer sources) is received through unannounced visits of company agents to the village and chats with the input-dealers in the nearest urban hub.

test the set of questions regarding the knowledge the respondent has about his match's farming activities as the respondent could not recall the yield, pesticide use and cultivar choice of the match several years ago. However, using the predicted links based on the results of the random matching within sample method, I constructed an estimate of the year when the respondent heard first about Bt cotton and compared this with the year the respondent stated he heard first about Bt cotton. For the respondent who farm cotton in 2007-08, I correctly predict this year in all cases using the link "know the cultivar of X" (i.e., link no. 6 in Table 2.5), but have an estimate which is too early in 37.5% of the cases using the link "know the cultivar, pesticide use and yield of X" (i.e., link no. 7 in Table 2.5). This issue might cause the social learning coefficient to be biased downwards in the regressions based on the results of the random matching within sample method.

[22]The descriptive statistics suggest that this assumption is relatively innocious in the case of Aurepalle and Kanzara where, respectively, each year, on average, over 85% of the Bt farmers known to the respondent live in the village. However, in Kinkhed, over 40% of the farmers known to the respondent who cultivate Bt cotton do not live in the village. The results of the random matching within sample experiment, however, suggest that one is unlikely to learn from these farmers about the profitability of Bt cotton as the two largest determinants of "knowing the pesticide use and yield" conditional on "knowing the cultivar" are living in the same neighborhood in the village and having a field neighboring the match's field.

[23]Not all respondents could recall the month in which they received the information. I reconstructed the month (before/after June) based on the adoption history of the farmer and knowledge of visits of extension agents etc. from the village questionnaire.

In addition to information on social networks, I elicited information on current beliefs regarding the profitability of Bt and non-Bt cotton and bio-safety hazards of Bt cotton. For the former, I used a yield distribution game, based on Lybbert et al. (2007). In this game, I first elicited the minimum and maximum yield (per acre) of two Bt and non-Bt cultivars of the respondent's choice conditional on the respondent's soil characteristics, irrigation status and expected input use. Then I made five boxes, evenly distributed between this minimum and maximum and I asked the respondent to use 20 stones (each stone representing a 5% probability) to form a yield density function. After each yield distribution game, I asked the respondent how much he expects to pay i) for the seed (per acre), and ii) for pesticides and for other inputs (per acre).[24]

For the latter, I collected data on the farmer's view and the perceived views of different social groups on the effects of Bt cotton on animal health, human health and the environment. More specifically, I asked the respondent "To what degree do you think (or that "others" think) that Bt cotton is hazardous for (1) animal health, (2) for human health, (3) for the environment", where the 'others' refers to, in several sub-questions, other farmers, relatives, company agents, government extension agents, input dealers and ICRISAT. Each of these three questions has five possible answers, ranging from 'strongly disagree' to 'strongly agree' and a 'don't know' option.

[24]During the interview, I explicitly linked the number of stones to percentages. See also Delavande et al. (2008) for various other methods that can be used.

2.4 A first look at the data

Table 2.3 presents selected descriptive statistics regarding the Bt cotton adoption process and the (current) knowledge and beliefs about Bt cotton. Figure 2.1 plots the number of farmers cultivating Bt cotton as a percentage of the total number of cotton farmers (the "adoption curve"). In this section, I discuss this table and figure and tell the story of the adoption process using qualitative interview evidence, thereby setting the stage for the analysis.

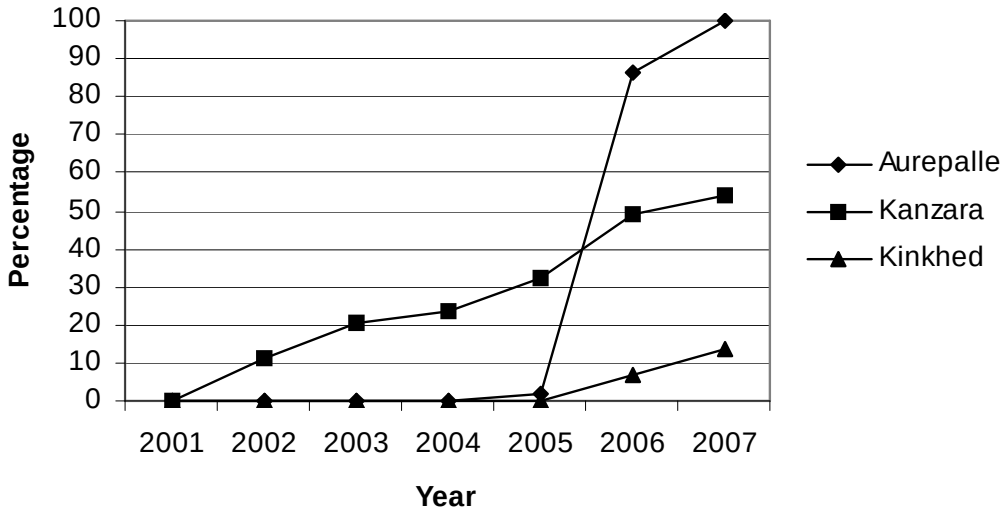

Figure 2.1: Adoption of Bt cotton [as % of cotton farmers]

In Andhra Pradesh, where Aurepalle is situated, one out of three of the Bt cotton cultivars was not on the market during the first two cropping seasons (2002-04). In addition, during the first three years NGOs and the Andhra Pradesh government challenged the decision of GEAC with regard to the approval of Bt cotton, resulting in a discontinuation of the permission for com-

Table 2.3: Bt cotton adoption process and current beliefs 2001-08

	Aurepalle	Kanzara	Kinkhed
% of households that farm cotton, 2001-08	60	84	82
% of cotton farmers that adopt Bt cotton, 2001-08	77	51	11
% of cotton farmers that adopt Bt cotton, 2007-08	100	54	14
% of cotton acreage under Bt cotton, 2007-08	100	55	15
% of cotton acreage under Bt cotton, 2002-03	0	9	0
% of Bt cotton farmers that disadopt Bt cotton, 2007-08	0	14	0
% of Bt cotton farmers that partially adopt Bt cotton, 2007-08	0	31	40
Never heard of Bt cotton (as % of cotton farmers)	4	0	0
Av. perceived mean of Bt cotton distribution (Q/acre)[1]	7.6	6.82	5.45
Av. perceived variance of Bt cotton distribution (Q/acre)[1]	1.57	2.12	1.49
Av. perceived mean of non-Bt cotton distribution (Q/acre)[1]	4.79	3.88	3.72
Av. perceived variance of non-Bt cotton distribution (Q/acre)[1]	0.64	0.79	0.87
Av. perceived number of pesticide sprays needed for Bt cotton	3.4	4.7	2.9
Av.perceived number of pesticide sprays needed for non-Bt cotton	7.6	2.1	1.24
% of respondents with safety concerns regarding Bt cotton	13	51	53

Notes: [1]1 quintal=100kg; Av.=average; calculated assuming a step-wise distribution with the minimum and maximum of the distribution as specified by the respondent; the differences between Bt cotton and non-Bt cotton in perceived mean and perceived number of sprays needed is significant in all three villages

mercial cultivation of several Mahyco cultivars in 2005.[25] The legal battle and controversy around GM crops in the early years in Andhra Pradesh affected the farmers in Aurepalle as company agents did not actively go around the country side to promote the new Bt technology. Only one of the progressive farmers in Aurepalle learned about Bt cotton from non-farmer sources as early as 2002 and started to cultivate Bt cotton in 2003. The majority of the progressive farmers heard first about Bt cotton from the media and each-other in 2003-2004 and the other cotton farmers heard about Bt cotton from the progressive farmers and company agents visiting the village in 2006-07. The large majority of the cotton farmers, 91%, adopted Bt cotton in the same year that they heard about Bt cotton for the first time, displaying great confidence (by their own account) in the judgement of the progressive farmers and relying heavily on the advice of company agents and input dealers.

[25]Sources: The list of approved hybrids as released by GEAC and confirmed through personal communication with Mahyco, dated 7 September 2009.

The few farmers who delayed adoption mentioned high seed cost, unavailability of the seed with their dealer and the need to observe the results for themselves as the main reasons. Farmers who did not adopt up to 2007-08 mentioned lack of land, family labor, credit, knowledge and high seed costs as the main reasons. No disadoption occurred in Aurepalle. Two farmers adopted the technology partially in 2006, cultivating both Bt and non-Bt cotton. The current adoption rate stands at 100%.

As can be seen from Table 2.3, the Aurepalle respondents currently have little to no bio-safety concerns regarding Bt cotton. Only 13% of the respondents 'agreed' or 'strongly agreed' with (at least) one of the following statements: 'Bt cotton is hazardous for animal health: they might get sick or die when they eat it', 'Bt cotton is hazardous for human health: if you touch it too much, you might get sick' and 'Bt cotton is hazardous for the environment: it damages crops and soils'.

In Kanzara, the Bt adoption process has been smooth, characterized by partial adoption, strategic delays and disadoption. Disadoption of Bt cotton is between 14% and 27% each year and thepercentage of Bt farmers who are partial adopters ranges from 31% to 80%. The progressive farmers in the villages learned about Bt cotton from company agents, media and each other as early as 2001 and started cultivation as early as 2002. The other village farmers heard first about Bt cotton from company agents, input dealers and progressive farmers. Most farmers, including the progressive farmers, did not adopt Bt cotton in the year that they first heard about it.

The reasons mentioned for this delay are lack of experience, including the experience of others in their network, lack of irrigation facilities, high seed cost and lack of credit availability. The current adoption rate stands at 54%. Presently, 51% of the respondents have bio-safety concerns regarding Bt cotton, but none of the progressive farmers reported any of these concerns.

Kinkhed is located only a few of miles from Kanzara. Despite the fact that the progressive farmers in the village heard about Bt cotton from the media, company agents, input dealers and progressive farmers outside of the village as early as 2002, they displayed a 'wait and see' attitude and, as a result, the adoption process started only in 2006. The other village farmers heard first about Bt cotton from input dealers, farmers outside of the village and the progressive farmers. The current adoption rate stands at 14%. In fact, only one farmer classified as a non-progressive farmer has adopted the technology.

The reasons for non-adoption mentioned by the farmers are: lack of information, irrigation, and land, the high price of the seed, lack of credit availability and the danger Bt cotton poses to farm animals.

Currently, 53% of the respondents have bio-safety concerns and these safety concerns are mentioned as one of the reasons for non-adoption by several farmers. Unlike in Kanzara, even the progressive farmers in the village attach some belief to these concerns.

While in all villages, on average, respondents currently believe that Bt cotton provides a higher yield compared to non-Bt cotton, bio-safety concerns, the process which formed these beliefs, and resulting adoption process are markedly

different in the three villages. What are the underlying factors driving these different adoption processes? First, there are the usual constraints: credit, land and irrigation which are different across villages. For instance, Aurepalle farmers, using groundwater, have typically access to irrigation during the rainy season, while Kanzara and Kinkhed farmers, using surface water, do not. The physical environment is different in terms of climate, expected bollworm pressure and soils and while the distance to the first urban hub is similar in all three villages, the different in state policies and attitudes with regard to genetically modified crops implied that Aurepalle was not visited by company agents until 2005. The education level is higher in the Akola villages, and the Kanzara farmers have, on average, better access to media sources.

This difference in education level and information received from non-farmer sources not only results in different beliefs with regard to the profitability but also different beliefs with regard to bio-safety. A regression analysis of the perceived bio-safety concerns of other village farmers on various household characteristics, information received from outside sources, information received from village farmers, and the perceived relationship between animal deaths and Bt cotton indicates that education significantly increases these concerns and information received from non-farmer sources decreases these concerns.[26] The data do not allow me to check whether the perceived bio-safety concerns of other village farmers are correct, i.e., whether surrounding village farmers' concerns are

[26]This regression analysis includes the following independent variables: number of times information received from non-farmer sources, education, value of livestock, cumulative number of farmers known who cultivated Bt cotton, the perceived relationship between animals died and number of Bt farmers (measured as the individual-level estimated slope of this relationship also interacted with how certain one is of this perceived relationship), land owned, number of adult and total number of family members.

indeed as the respondent perceives them. They do reveal a very high correlation between own concerns and perceived concerns of others (with a correlation coefficient of 0.92).

Last but not least, the different underlying network structures of who-communicates-with-whom also plays a role in generating these different adoption processes. In 2007-08, the respondents in Aurepalle, Kanzara and Kinkhed (on average) knew, respectively, 85, 38 and 34 farmers who cultivate Bt cotton in 2007-08.[27] In Aurepalle, Kanzara and Kinkhed (on average), respectively, 78%, 85% and 57% of the Bt farmers known to the respondent live in the village. Table 2.4 presents selected results from the random-matching-within-sample method. Recall that each respondent draws 6 name cards of VLS respondents and is given a set of 4 fixed cards with names of progressive farmers. Denote the individual on the card by X. One can see that in a small village like Kinkhed, literally everyone knows everyone. In Aurepalle, only 87.8% of the cards are known. Conditional on knowing cotton farmer X, the respondent thinks he knows the pesticide use, cultivar choice and yield of farmer X in 2007-08 in, respectively, 21.9%, 27.8% and 63.6% of the cases in Aurepalle, Kanzara and Kinkhed.[28]

This implies that the average Aurepalle farmer, by sheer virtue of living in a larger village, knows more farmers and as such has more opportunities to learn by observing other farmers' experimentation. The average Kinkhed farmer, on

[27] In case that a range was given by the respondent, the mean was taken. If the respondent answered "don't know" I assumed they knew no Bt cotton farmers.

[28] Seperating these results out by progressive farmers status of the match, does not change these results drastically. Comparing progressive farmer matches with non-progressive farmer matches, in Kanzara and Kinkhed, the respondent is up to 10% more likely to know the cultivar, yield and pesticide use conditional on knowing the farmer. In Aurepalle, the respondent is up to 6% less likely to know the cultivar, yield and pesticide use.

Table 2.4: Knowledge about others' activities (based on the results of the random matching within sample method)

	Aurepalle	Kanzara	Kinkhed
1. Know X? (%)	87.8	99.2	100
2. Does X farm? (% of 1)	82.3	83.7	91.6
3. Does X farm cotton? (% of 2)	57.2	70.2	90
4. Know X's yield? (% of 3)	30.2	39.1	68.6
5. Know X's pesticide use? (% of 3)	29.5	31.1	75.9
6. Know X's cultivar? (% of 3)	69.3	85.8	75.4
7. Know X's yield, pesticide use and cultivar? (% of 3)	21.9	27.8	63.6
8. X's yield correct (% of 4)	31.4	21.2	16.3
9. X's pest correct? (% of 5)	14.6	25.1	61.1
10. X's cultivar correct? (% of 6)	86	81.9	77.3
11. X's yield, pesticide use and cultivar correct (% of 7)	7.4	5.7	12.4

Note: In (4), (5), (6) and (7) "knowing" means that the respondent was able to name the cultivar, the amount of pesticides used, the yield per acre obtained etc. of match X. Knowledge of yield and pesticide use was considered correct if the believed value was within a 10% range of the actual value. If X cultivated multiple cultivars, the believed value of the average yield of Bt and non-Bt was compared with the actual average. In case of pesticide use the discrete decision was often known (whether X uses pesticides or not) but not the exact number of sprays. In this case, knowledge was considered incorrect.

the other hand, knows fewer farmers but knows these farmers in a more profound manner, by cultivar, yield and pesticide use and as such is more likely to learn something useful conditional on knowing one Bt cotton farmer.

In, respectively, 7.4%, 5.7% and 12.4% of the cases in Aurepalle, Kanzara and Kinkhed the knowledge about yield, pesticide use and cultivar choice is correct. Incorrect knowledge about other farmers' activities might be one of the factors underlying dis-adoption, i.e., a farmer could adopt Bt cotton based on the perceived knowledge that x number of sprays are needed to obtain, on average, y quintal/acre, but once the farmer has adopted Bt cotton, he experiences a lower yield than expected and decides to disadopt. While the figures on yield in the Akola villages of Table 2.4 should be taken with a grain of salt, as the respondent had to convert the yields from an intercropping system to quintal/acre,

note that Kanzara, where knowledge is most frequently incorrect, is also the village with the highest disadoption rate.

Note that in the random-matching-within-sample method one only obtains information about the links between the respondent and 6 randomly selected VLS farmers and 4 fixed progressive farmers. In order to obtain a link variable for any two randomly drawn VLS respondents, I use a probit regression of the link variable of interest on a set of attributes of the respondent and match. I then use the results of this regression to predict out-of-sample, i.e., out of the random-matching-within-sample method but within the VLS sample.[29]

The results of this regression suggest that the probability of knowing X's yield, pesticide use and cultivar choice increases when either respondent or match are a progressive farmer, the respondent and match belong to the same caste, either respondent or match are of the male gender, a higher education level of respondent or match and the respondent and match live in the same sub-village. Unfortunately, I cannot include the two variables with the largest predictive power, 'living in the same neighborhood' and 'having a field next to each other' as I only have this information on the respondents and matches drawn in the random-matching-within-sample method. However, as one can

[29]Concretely, for each village, I run a probit regression with two-way clustering (the first cluster is the respondent and the second cluster is the match) and each variable entered up to the third degree. Denote an arbitrary correlate with x, then the suffix "add" and "min" refer, respectively, to $x_{respondent} + x_{match}$ and $x_{respondent} - x_{match}$. The list of correlates included is: "being a progressive farmer" (PFadd, PFmin), "being a laborer" (laboreradd, laborermin), "belonging to the same caste", "having the same family name", "acreage of land owned" (landadd, landmin), "years of education obtained by the decision-maker" (educationadd, educationmin), gender (genderadd, gendermin), "income per capita in kharif 2007-08" (incomeadd, incomemin), "number of children" (childrenadd, childrenmin), "value of agricultural machinery" (machineryadd, machinerymin) and "living in the same (sub)-village".

see from Table 2.5, the model still performs reasonably well with between 72% and 88% of the links correctly predicted.

Table 2.5: Percentage of links correctly predicted

	Link no. 5 Know cultivar	Link no. 7 Know cultivar, pesticide and yield
Aurepalle	77	88
Kanzara	80	83
Kinkhed	74	72

2.5 Modeling the decision to adopt Bt cotton

In this section I present a simple model of the decision to cultivate Bt cotton, conditional on having decided to cultivate cotton, and derive a reduced-form solution which I can then directly estimate. Following Besley and Case (1993) and Moser and Barrett (2006) I abstract from the details of the learning process by including a state variable denoted $K_{i,t}$ representing the knowledge the farmer i has about about Bt cotton. In addition, I refrain from characterizing the equilibrium of the strategic interaction between individuals underlying the free-rider and social pressure effects among non-progressive farmers.[30] In the empirical specification, I use an instrumental variable strategy to deal with this issue of simultaneity.

Regarding notation, three kinds of subscripts are employed, the first subscript denotes the individual i or the contacts of individual i, denoted $-i$; the

[30] Finding an analytical solution for the fixed point problem such an equilibrium would imply is challenging due to the overlapping network structure. As such, I leave such a characterization to future work.

second subscript denotes the crop (c refers to cotton, b to Bt cotton and n to non-Bt cotton) and the third subscript denotes time t. Vectors are indicated in bold.

Assuming a time-separable utility function and discrete time steps, $\tau \in \{t, t+1, ...\}$, where τ is the 'integration dummy', the farmer maximizes at each time period t the discounted flow of instantaneous utility over an infinite horizon:

$$V(K_{i,t}) = \max_{\{A_{i,b,\tau}, \mathbf{x}_{i,b,\tau}, \mathbf{x}_{i,n,\tau}\}_{\tau=t}^{\infty}} E\left[\begin{array}{c} \sum_{\tau=t}^{\infty} \delta^{t+\tau} u_i(\pi_{i,b,\tau}(\cdot) + \pi_{i,n,\tau}(\cdot)) + \\ s(A_{i,b,\tau}, \{A_{-i,b,\tau}\}_{\forall -i \in N_{P_i}}, \{A_{-i,b,\tau}\}_{\forall -i \in N_{I_i}}) \end{array} \right] \tag{2.1}$$

In (2.1),$V(\cdot)$ is the value function of the maximization problem, $\delta \in (0,1)$ is the discount rate, summarizing preferences over time and E is the expected value operator (over ϵ and $\{A_{-i,b,\tau}\}_{\forall -i, \forall \tau}$). In terms of choice variables, $A_{i,b,\tau}$ denotes the acreage under Bt cotton and $\mathbf{x}_{i,b,\tau}$ and $\mathbf{x}_{i,n,\tau}$, respectively, denote the vector of inputs for Bt cotton and non-Bt cotton.

The farmer's per-period well-being is a function of the farmer's consumption, as captured by the standard individual-specific $u_i(\cdot)$ Bernoulli utility function and a non-material satisfaction term, denoted $s(\cdot)$ included to capture the influence of social pressures and the (behavioral) imitation. The set of fellow, i.e., non-progressive, farmers relevant to farmer i in terms of social pressures is denoted N_{P_i} and the set of progressive farmers relevant to farmer i in terms of imitation is denoted N_{I_i}.

The profit of Bt cotton ($\pi_{i,b,\tau}$) and non-Bt cotton ($\pi_{i,n,\tau}$) are, respectively, defined as:

$$\pi_{i,b,\tau} = p.F_b(A_{i,b,\tau}, \mathbf{x}_{i,b,\tau}, \epsilon_\tau | K_{i,t}) - \mathbf{p}_x.x_{i,b,\tau} \tag{2.2}$$

$$\pi_{i,n,\tau} = p.F_n(A_i - A_{i,b,\tau}, \mathbf{x}_{i,n,t}, \epsilon_\tau) - \mathbf{p}_x.x_{i,b,\tau} \tag{2.3}$$

where p denotes the price of cotton[31], $\mathbf{p}_x$denote the vector of prices of inputs, ϵ_t denotes a random variable capturing unexpected shocks caused by weather fluctuations and bollworm pressure (distributed according to H_ϵ) and F_b and F_n denote, respectively, the production function of Bt cotton and non-Bt cotton. Note that I assume that the profitability of non-Bt cotton[32] and the prices of all inputs and outputs are well known.[33]

Denote the realization of F by f. The production function of Bt cotton is a function of the the knowledge the farmer i has about about Bt cotton, $K_{i,t}$, and is subject to the following law of motion:

$$K_{i,t+1} = K\left(K_{i,t}, f_{b,i,t}, \{f_{-i,b,t}, \mathbf{x}_{-i,b,t}\}_{\forall -i \in N_{L_i}}, \{A_{-i,b,t+1}\}_{\forall -i \in N_{l_i}}, O_{i,t}\right) \tag{2.4}$$

where N_{L_i} denotes the set of farmers whose inputs and output choices farmer i has observed in period t and $O_{i,t}$ the information farmer i has received from non-farmer sources, such as company agents and input dealers in period t. Thus, according to (2.4) farmer i updates his knowledge regarding the profitability of

[31]Note that the price of cotton is assumed to be the same for Bt cotton and non-Bt cotton. Using the 2007-08 input-output data, I show in Appendix D that this assumption is correct.

[32]This is a reasonable assumption as the large majority of the farmers who switch to Bt cotton have been farming cotton for several years. Only ten farmers make the switch from no cotton to Bt cotton directly, but even in these cases it is possible that these farmers have farmed cotton before 2001. At the minimum, these ten farmers have observed other farmers for minimum of ten years, which should be sufficient to establish their beliefs with regard to these well-known non-Bt cotton cultivars.

[33]This might be a strong assumption, especially with regard to the seed prices of Bt and non-Bt cotton. Even though I elicited expected price of seeds in 2008 and have the actual price paid in 2007-08, I cannot compare these two directly as the former was elicited on a per acre basis while the latter is in kg or bag, and the majority of the farmers did not stick to the recommended 1 bag per acre (most farmers use 2 to 8 bags per acre). Extrapolating their 2007-08 use to 2008-09, about 40 percent of the farmers who plan to grow cotton in 2008-09 correctly knew the price of the Bt cultivar. I however expect this figure to underestimate the knowledge of the farmers in this regard as the prices change from year to year, and the farmers, according to their own account, are generally well aware of these new prices. In addition, the prices are publicly advertised on boards in front of the shops in the mandal capital which is frequently visited by most farmers.

Bt cotton through experimentation, observing other farmers' input and output choices, observing the progressive farmers input choices, i.e., learning imitation, and interactions with non-farmer sources.

The control variables in this dynamic optimization problem are subject to a non-negativity constraint and the state variables to a set of initial conditions:

$$A_i \geq A_{i,b,t} \geq 0 \qquad \forall t \tag{2.5}$$

$$\mathbf{x}_{i,b,t}, \mathbf{x}_{i,n,t} \geq 0 \qquad \forall t \tag{2.6}$$

$$K_{i,0} = K_0 \tag{2.7}$$

Equations (4.4) - (2.7) jointly specify a dynamic optimization problem. Without imposing any learning model or preferences, write the reduced form solution for the choice of interest $A_{b,i,t}$ at time period t as:[34]

$$A_{b,i,t} = A(A_i, K_{i,t}, p, \mathbf{p}_x, \{A_{-i,b,\tau}\}_{\forall\tau,\forall -i\in N_{P_i}}, \{A_{-i,b,\tau}\}_{\forall\tau,\forall -i\in N_{I_i}}, H_\epsilon, \delta, R_i) \tag{2.8}$$

where R_i captures preferences with regard to risk. Plugging (2.4) in (2.8) this yields:

$$\begin{aligned} A_{b,i,t} \quad = \quad & A(A_i, K_0, \{f_{b,i,\tau}\}_{\tau=0}^{\tau=t-1}, \left\{\{f_{-i,b,\tau}, \mathbf{x}_{-i,b,\tau}\}_{\forall -i\in N_{L_i}}\right\}_{\tau=0}^{\tau=t-1}, \\ & \left\{\{A_{-i,b,\tau}\}_{\forall -i\in N_{I_i}}\right\}_{\tau=0}^{\tau=t}, \{O_{i,\tau}\}_{\tau=0}^{\tau=t-1}, \\ & p, \mathbf{p}_x, \{A_{-i,b,\tau}\}_{\forall\tau,\forall -i\in N_{P_i}}, \{A_{-i,b,\tau}\}_{\forall\tau,\forall -i\in N_{I_i}}, H_\epsilon, \delta, R_i) \end{aligned} \tag{2.9}$$

In order to bring (2.9) to the data, one needs to determine, for each farmer, the set of farmers N_{P_i}, N_{I_i} and N_{L_i} and decide on a aggregation rule to aggregate

[34]Note the abuse in notation with regard to $\{A_{-i,b,\tau}\}_{\forall\tau,\forall -i\in N_{P_i}}$ and $\{A_{-i,b,\tau}\}_{\forall\tau,\forall -i\in N_{I_i}}$, i.e., the adoption decision might depend on the entire perceived joint distribution of the current and future actions of others.

the experiences and actions of the different farmers $-i$.[35] Ideally, one would like to include all the information contacts of farmer i in N_{P_i}, N_{I_i} and N_{L_i}. This is exactly what the first social network question does, asking the number of farmers farmer i knew each year since 2001-02 who cultivated Bt cotton. One can take the total number of adopters farmer i has known up to $t-1$ as a proxy for what farmer i has learned about the production function of Bt cotton and the current number of adopters known as a measure of the aggregate effects of (behavioral) imitation, learning imitation, social pressures, and free-rider effects. As I expect the role of social pressures to be different in the acreage decision, conditional on adoption, let's focus on the binary decision adoption decision, denoted $ADOPT$ in (2.9)):

$$P(ADOPT_{i,t}) = \Phi \left[\begin{array}{c} A_i, K_0, \sum_{\tau=1}^{t-1} ADOPT_{i,\tau}, \sum_{\forall -i \in N} \sum_{\tau=1}^{t-1} ADOPT_{-i,\tau} \\ \sum_{\tau=1}^{t-1} O_{i,\tau}, p, \mathbf{p}_x, \sum_{\forall -i \in N} ADOPT_{-i,t}, H_\epsilon, \delta, R_i \end{array} \right] \tag{2.10}$$

where N denotes the total set of farmers farmer i interacts with, i.e., $N = N_{P_i} \cup N_{I_i} \cup N_{L_i}$.In the empirical specification, instead of $A_{i,t}$, I use the acreage of land owned and the Mills ratio predicted by a regression modelling the decision to cultivate cotton to capture selection effect. In addition, I add education (as education might help in processing new information) and measures of soil fertility, irrigation and credit constraints.

Using the results of the random-matching-within-sample experiment, one can take into account certain aspects of the relationship between farmer i and

[35]Moving from (2.9) to the empirical specification I make two simplifications. First, I do not include the past cultivar decisions of the progressive farmers in the imitation network. This simplification does not entail any loss of information. As none of the progressive farmers dis-adopt, both current adoption and past adoption (without observing yields) gives the same information to farmer i: 'progressive farmer x believes Bt cotton will increase his profits compare to non-Bt cotton'. Second, I assume that the free-rider effects only work from period $t+1$ to period t, i.e., what a farmer expects his contact to do in period $t+2$ does not matter to him.

his contact $-i$:

$$P(ADOPT_{i,t}) = \Phi \left[\begin{array}{c} A_i, K_0, \sum_{\tau=1}^{t-1} ADOPT_{i,\tau}, \sum_{\forall -i \in N_{L_i}} \\ \sum_{\tau=1}^{t-1} ADOPT_{-i,\tau}, \sum_{\forall -i \in N_{I_i}} ADOPT_{-i,b,t}, \\ \sum_{\tau=1}^{t-1} O_{i,\tau}, p, \mathbf{p}_x, \sum_{\forall -i \in N_{N_i}} ADOPT_{-i,t}, H_\epsilon, \delta, R_i \end{array} \right] \tag{2.11}$$

where N_{L_i} is the set of farmers for whom farmer i thinks he knows the cultivar, yield and number of pesticide sprays, including fellow farmers and progressive farmers. N_{N_i} is the set of fellow farmers for whom farmer i thinks he knows the cultivar and N_{I_i} is the set of progressive farmers for whom farmer i thinks he knows the cultivar. In the econometric specification, I include both the fraction (of total cotton farmers known) and the absolute number of the progressive farmers adopters known at time t. The latter captures the free-rider effects only, while the former captures imitation effects.

I use (2.10) and (2.11) to estimate the effects of social learning, social pressures and imitation using the panel data 2001-08. For the 2008-09 year I can directly use the (subjective) beliefs regarding the profitability of Bt cotton versus non-Bt cotton and the (perceived) strength of the bio-safety concerns of other farmers. I terms of social pressures, what is relevant for the decision of farmer i is the strength of these concerns, as captured by the variable $OTHER\_BIOSAFETY$, relative to the number of people known who plan to adopt Bt cotton as the latter represents the probability of being blamed when something would go wrong in the village. Rewriting (2.8):

$$P(ADOPT_{i,t}) = \Phi \left[\begin{array}{c} A_{i,t}, M_{i,b,t}, V_{i,b,t}, M_{i,n,t}, V_{i,n,t}, \{A_{-i,b,t}\}_{\forall -i \in N_{I_i}}, \\ \frac{OTHER\_BIOSAFETY}{\sum_{\forall -i \in N} ADOPT_{-i,t+1}}, H_\epsilon, \delta, R_i \end{array} \right] \tag{2.12}$$

where $M_{i,b,t}$, $V_{i,b,t}$, $M_{i,n,t}$, $V_{i,n,t}$ denote the mean and variance of the subjective conditional distribution of profits of Bt and non-Bt cotton[36], and $OTHER\_BIOSAFETY$ constructed as the average of the answers to: 'To what degree do you think that other village farmers think that Bt cotton is hazardous for (1) animal health, (2) for human health, (3) for the environment'. Specification (2.12) implies that social pressures can be reduced through two channels: by decreasing the perceived bio-safety concerns of others or by increasing the number of current adopters in the village.

2.6 Results

In this section, I present the main results of the cross-sectional analysis of the future adoption decisions regarding the 2008-09 cropping year and the panel data analysis, encompassing the period 2001-2008. In order not to interrupt the flow of this section, I have relegated most comments regarding data quality, construction of control variables and assumption checks to either footnotes or appendices.

Throughout the analysis, the main population of interest are the cotton farmers, i.e., I will look at the decision to cultivate Bt cotton conditional on cultivating cotton.[37] In order to extrapolate the influences of the social interaction

[36]Note that as the prices are included in the profit calculation, there is no need to include additional prices in (2.12).

[37]As it is only for the cotton farmers that the preference ordering between Bt-cotton and non-Bt cotton is revealed through their choice, and on a more practical level, the profitability of non-cotton crops can be ignored (conditional on controlling for risk aversion). As farmers do not change their intercropping arrangement after switching to Bt cotton, including only the price of cotton as the output price is sufficient after having controlled for the selection effect.

effects to the village population, I will take into account the 'self-selection' of the respondents into this 'cotton-farming' status.[38] Note also that by using the social networks and the beliefs of the respondent, I implicitly assume that the respondent interviewed is also the decision maker of the household in terms of agricultural decisions.[39]

With regard to measurement of risk aversion, one remark is in order. I constructed a measure of risk aversion based on a risk experiment that I conducted in 2007-08.[40] By using the same coefficient across years, I assume that the risk preferences of the decision maker have not changed over time.

[38] According to the respondents' own account the choice between cotton and non-cotton crops depend on many factors, notably, soil quality, weather expectations and access to irrigation, soil fertility/pest management and the need for crop rotation, the need for cash, the availability of labor and how labor intensive each crop is and the prevailing input and output prices. Among these, the need for crop rotation, which is not observable, might potentially bias the social interaction effects of interest when one includes only the cotton farmers in the analysis. An example can make this clear. Imagine that the (reduced form) true model of social learning is the following: for each farmer one has observed adopting Bt cotton, the probability of adoption goes up with 10%. Imagine that a farmer has observed 5 farmers and would like to adopt Bt cotton, however due to crop rotation rules (an unobservable), the farmer delays his decision to adopt with one year, after which he has observed an additional three farmers. If this happens at a large scale, the effect of social learning might be biased downward. As such, in the panel data analysis, I dropped the ten households that according to their own account practice crop rotation and switch in and out of cotton farming.

[39] During the fieldwork, I set up interviews each time with the person who the resident VLS investigators thought makes the agricultural decisions in the household. During the interview, several open questions were asked regarding the decision making process, among others "in year X, who decided on the cultivar". In the few cases that the answers to these questions did not correspond to the person being interviewed, the interview was repeated with, this time, the correct respondent. As such, this assumption can be considered correct. Even though in theory the landlord might have a say in the input choices of a farmer who shares-in land, in practice, this appears not to be the case in the three villages I consider. In 2007-08, 4.5% of the plots were shared-in or shared-out and only 2.3% of the cotton plots were shared-in or shared-out. In only one case, the respondent mentioned that the landlord had a say in choices of inputs, and this household did not farm cotton at any point in time and was not planning to farm cotton in 2008-09.

[40] Using the results of the risk experiment, I calculate a risk aversion measure as the difference in willingness-to-pay for a bag of cotton seeds when moving from a baseline yield distribution to a yield distribution with the same mean but a higher variance (see Maertens and Just 2009). A larger coefficient indicates a higher degree of risk aversion.

Finally, I assume that for all farmers both Bt and non-Bt cotton seeds are available in the market where they usually buy inputs. As only 11 farmers mentioned that they were unable to purchase the cultivar they intended to purchase with their usual input dealer, availability of seeds does not seem to be a problem in the villages.[41]

Starting with the results of the cross-sectional analysis, recall that to identify non-monetary effects I exploit the data on current beliefs regarding the profitability of Bt cotton versus non-Bt cotton and compare these beliefs with the planned Bt cotton cultivation decisions for the 2008-09 season.

Using the results of the yield distribution game, and assuming a uniform distribution within each yield category, I calculated the expected profits and reconstructed the cumulative distribution function (CDF) of the profit of each cultivar (per acre) for each respondent at the time of the interview in 2008.[42]

[41] I deleted the three households from the panel data analysis who could not find a Bt or non-Bt cultivar of their choice and ended up buying, respectively, a non-Bt and Bt cultivar.

[42] The profit is defined as the output price multiplied by the output minus the paid out input costs. These paid out inputs do not include own labor. Using the 2007-08 data I checked that this input does not differ much between Bt cotton and non-Bt cotton. Note that the results of the yield distribution game provide values for all these variables except for the output price of cotton. To construct an individual-level output price, I use the 2007-08 input-output data collected and regress the output price of cotton on village dummies, the number of adult family members, the education level of the decision maker (years), the acreage of owned land, the square acreage of owned land and whether or not the household had a functional irrigation source in 2007-08. I determined the choice of these variables based on the analysis presented in Table 2.18 in the Appendix. By using this predicted price, I avoid the potential bias cause by unobservables correlated with both price and social interaction effects. It must be noted that the results did not change much when using the actual price (in combination with a village-average price for the respondents who did not farm cotton in 2007-08) instead of this predicted price. Note that, in the case of Aurepalle, the total amount expected to be paid for other inputs (all inputs excluding pesticides) was not recorded. I used the total amount paid for other inputs per acre in 2006-07 as a measure for the Bt cultivars and this amount minus 1671 Rs (the average difference in cost between Bt and non-Bt cultivars in 2006-07) as a measure for the non-Bt cultivars. It should be noted that the results presented here are robust with respect to this particular number.

Table 2.6 compares the plans to cultivate Bt cotton in 2008-09 with the subjective beliefs regarding the profitability of Bt cotton.[43] In Aurepalle, all 73 farmers who plan to grow cotton in 2008-09, plan to grow Bt cotton. All of these farmers expect a higher average profit for Bt cotton compared to non-Bt cotton. In Kanzara there are two farmers who plan to grow Bt cotton despite the fact that they expect a lower average profit for Bt compared to non-Bt. On the flipside, in Kanzara and Kinkhed, respectively, five and fifteen farmers plan to grow non-Bt cotton despite the fact that they expect a lower average profit for non-Bt cotton compared to Bt cotton.

Table 2.6: Compare current beliefs with future plans

	Beliefs in 2008		
Plans for 2008-09	E[π-BT]>E[π-NBT]	E[π-BT]<E[π-NBT]	Total
		Kanzara	
Cultivate Bt-cotton	15	2	17
Cultivate Bt-cotton and non-Bt cotton	0	1	1
Cultivate non-Bt cotton	5	0	5
Cultivate cotton, but don't know which cultivar	0	0	0
Do not cultivate cotton	35	3	38
Don't know whether to cultivate cotton	2	0	2
Total observations in Kanzara	57	6	63
		Kinkhed	
Cultivate Bt-cotton	11	0	11
Cultivate Bt-cotton and non-Bt cotton	6	1	7
Cultivate non-Bt cotton	15	0	15
Cultivate cotton, but don't know which cultivar	2	0	2
Do not cultivate cotton	12	3	15
Don't know whether to cultivate cotton	4	1	5
Total observations in Kinkhed	50	5	55
		Aurepalle	
Cultivate Bt-cotton	64	0	64
Cultivate Bt-cotton and non-Bt cotton	9	0	9
Cultivate non-Bt cotton	0	0	0
Cultivate cotton, but don't know which cultivar	1	1	2
Do not cultivate cotton	40	8	48
Don't know whether to cultivate cotton	3	2	5
Total observations in Aurepalle	117	11	128
Total observations			246

[43] In practice, as I have data on 2 Bt cultivars and 2 non-Bt cultivars, the expected profit of one of the Bt cultivars should be larger or equal to the expected profit of the two non-Bt cultivars in order to chose for Bt cotton.

These results might point at non-monetary effects or the fact that other aspects of the distribution, apart from the average, matter to the respondents. As for none of the Kanzara farmers who plan to grow Bt cotton does the profit CDF of non-Bt cotton First Order Stochastically Dominates (FOSD) the profit CDF of Bt cotton, risk preferences cannot be excluded as a reason to opt for non-Bt cotton. However, in Kanzara and Kinkhed, there are, respectively, three and five farmers who plan to grow non-Bt cotton and whose profit CDF of Bt-cotton FOSD the profit CDF of non-Bt cotton. In these cases, risk preferences can be excluded and non-monetary effects, conditional on the absence of credit constraints, are cleary driving these decisions.

Table 2.7 presents the average and variance of selected characteristics of the groups introduced in Table 2.6. The two farmers in Kanzara who plan to grow Bt cotton and expect a lower average profit for Bt compared to non-Bt are surrounded by progressive farmers who all plan to cultivate Bt cotton. The five and fifteen farmers in, respectively, Kanzara and Kinkhed who plan to grow non-Bt cotton and expect a lower average profit for non-Bt cotton compared to Bt cotton are surrounded by farmers who are, on average, more concerned with bio-safety issues and by fewer farmers who plan to cultivate Bt cotton. Disregarding credit constraints, these statistics point at the existence of social pressures and (behavioral) imitation.

Continuing with the analysis of the 2008-09 data, Table 2.8 presents the results of a probit analysis estimating the parameters of (2.12) The dependent variable is the discrete decision to plan to cultivate Bt cotton in 2008-09, conditional

on cultivating cotton.[44] The analysis includes all Kanzara and Kinkhed households[45] who had decided to cultivate cotton in the 2008-09 Kharif season and had decided on a cultivar, except for the progressive farmers. The progressive farmers are excluded as these are, according to their own account, not influenced by social pressures and do not imitate others.[46]

The coefficient on $OTHER\_BIOSAFETY/KNOWN$ is statistically significantly negative, suggesting social pressures. Everything else equal, if one knows sixteen adopters and is surrounded by farmers who have strong bio-safety concerns instead of farmers who have no bio-safety concernsl, decreases the probability of cultivating Bt cotton with about 10%. If one knows only four adopters, this effect is much larger, about 40%. The effect of the fraction of progressive farmer adopters known is not significantly different from zero pointing at the absence of behavioral imitation. Note that the analysis in Table 2.8 controls for credit constraints, risk aversion and the bio-safety concerns a farmer himself might have with regard to Bt cotton. All the other variables have the expected sign, but only few are significantly different from zero. Surprisingly, the bio-safety concerns the farmer has himself about the technology do not seem to matter. This might be due to the fact that most farmers have not a lot of live-

[44]The results of the first stage regression, which in addition includes various measures of wealth, soil fertility, irrigation availability, weather expectations and the predicted prices of all crops, are available upon request.

[45]This analysis does not include the Aurepalle households as the measure of other input use is not comparable across states due to a mistake during the data collection stage.

[46]Note that this assumption cannot be tested directly as there are too few progressive farmers in the dataset. However, re-estimating the regression presented in Table 2.8 including the progressive farmers results in an insignificant social pressure effect.

stock themselves, apart from a bullock pair. The knowledge variables are jointly significantly different from zero.[47]

Moving on with the panel data analysis covering the period 2001-08, one remark is in order. Recall that both the ICRISAT-VLS data and the recall data collected in 2008 provide information on the binary decision to cultivate Bt cotton and the acreage under Bt and non-Bt cotton. I opt to use the recall data for the descriptive statistics and the analysis as the ICRISAT-VLS data do not always contain information on which cultivar is used (in terms of Bt versus non-Bt) and, in addition, the input-output information is sometimes missing for households classified as agricultural laborers and/or households who lease in land.[48]

Table 2.9 presents the results of a probit analysis estimating the parameters of specification (2.10), i.e., based on the total number of adopters known each year. The dependent variable is the discrete decision to cultivate Bt cotton, con-

[47]If, as rational expectations would predict, the expected number of adopters is correct, the constructed social pressure variable might suffer from endogeneity problems. I repeat the analysis using MLE instrumental variable Probit estimation using the number of adopters known in the previous period as an instrument and do not find any significant differences. The Wald test of exogeneity of the instrumental variable shows that there is not sufficient information to confirm an endogeneity problem. This might be due to the small number of observations or the fact that one month before sowing on, in effect, observes an out-of-equilibrium situation. Due to the small number of observations, I opt to report the non-IV results as these will report unbiased coefficient estimates in the absence of endogeneity.

[48]Comparing the two sources of data reveals that in 80% of the cases the binary cotton cultivation decision in the recall data corresponds with the ICRISAT-VLS data, and conditional on cultivating cotton according to the recall data, in 77% of the cases the binary Bt cotton cultivation decision in the recall data corresponds with the ICRISAT-VLS data. As I collected data on the binary cotton and Bt cotton decision in several modules of the questionnaire and these are consistent with each other, I am confident of a high quality of the recall data, as far as these binary decisions are concerned. But as the information on acreage is of lesser quality I opt to exclude the analysis of acreage under Bt cotton, conditional on adoption, from this paper.

ditional on cultivating cotton.[49] The analysis includes all household up to (and including) the first year they adopt Bt cotton, except for the progressive farmers.[50] I opt to focus on the decision to first adopt Bt cotton as one would expect the role of social learning and learning imitation to change after adoption. The control variables included are described in Appendix E.

From Table 2.9, one can see that the coefficients on the total number of adopters known from the village and from outside of the village are not significantly different from zero. The effect of non-farmer information on Bt cotton adoption is significant, positive, concave and large. Hearing from one input dealer, company agent or government extension agent about Bt cotton increases the probability of adoption, on average, with 5%. Recall that the majority of this information is received through unannounced visits of company agents to the village and chats with the input-dealers in the nearest urban hub and few farmers had to travel or pay for this information. As I control for education and other individual-fixed effects such as risk aversion, I do not expect any residual endogeneity problems.[51] Neither is the large magnitude of this effect, compared to the effect of past experiences of one's contacts, due to concavity as in only a

[49] Again, the results of the first stage regression, which contains additional measures of wealth, weather expectations, output prices, soil fertility and irrigation constraints, are available upon request.

[50] In case of split-off households or households that have been recently added to the ICRISAT-VLS, households were considered for inclusion in the analysis from the date of split-off or inclusion in the ICRISAT-VLS onwards. Recall also that the households facing seed availibility or crop rotation constraints are not included in the analysis. In addition, 99 observations were dropped after the inclusion of an Aurepalle*before2005 fixed effect (which I included to capture the difference state level policies in the pre-2005 period). Including these 99 observations does not change the results significantly.

[51] In all specifications, the non-farmer information excluded the information received from input dealers in May/June.

couple of cases did non-farmer information preceeds all observed past experiences of others.

The contemporary social interaction effects including social pressures, imitation and free rider effects have an aggregate significant positive effect. But the effect is small in size. Note, however, that this effect might be biased due to the fact that the farmers in the village make their decision simultaneously and, as such, the current behavior of other farmers is endogenous.

Table 2.10 presents the results a probit analysis estimating the parameters of specification using specification (2.11), i.e., based on the results of the random-matching-within-sample method. Table 2.11 presents the results of the same specification but using the predicted links rather than the actual links of the random-matching-within-sample method. The dependent variable is the discrete decision to cultivate Bt cotton, conditional on cultivating cotton. Again, this analysis includes all cotton farmers up to (and including) the first year they adopt Bt cotton, excluding the progressive farmers.

The first two columns of Table 2.10 and Table 2.11 present the results of a standard probit regression, the last column of Table 2.10 and Table 2.11 shows the results of a MLE probit instrumenting the ”total number of non-PF adopters known at t” with the ”total number of non-PF adopters known at $t-1$”. Theoretically, as I control for learning, these instruments have no effect on adoption behavior and the data reveals that this instrument is strongly correlated with

the endogenous variable in both tables. The Wald test of exogeneity of the instrumental variables, as expected, rejects exogeneity, in both cases.[52]

From third column in Table 2.10 one can see that the coefficient on "non-farmer information" is large and significantly different from zero. In contrast with the results in Table 2.9, the past experience of the farmers one is connected to matters to the farmer's adoption decision. In particular, the farmer learns from the experiences of the progressive farmers he has observed in the village. For every ten progressive farmers' experiences observed in the past, the probability of adoption, on average, increases with about 19%. The past experiences of fellow, non-progressive, farmers seem to matter little. Even though the coefficient is almost significantly different from zero at the 10% level, the size is still small. In all cases, the second derivative has a negative estimated sign, indicating decreasing returns to new information, consistent with, among others, a Bayesian learning model.

The large difference between the effect of the progressive farmers' past experiences and hearing about Bt cotton from a non-farmer source, is most likely due to the fact that from the latter, the farmer receives information about the entire distribution of Bt cotton, i.e., what the farmer can expect on average conditional on a certain number of pesticide sprays, and what can be expected in case of bad weather. In addition, the farmer, in some cases, receives information about

[52]Note that even though I assume that Manski's contextual effects play no role in the decision making process, using the current properties of other farmers instead as instruments is unlikely to work well as (1) the variables which matter to the Bt cotton decision making process, land, irrigation and education level, vary little over time and (2) one has many missing 'instruments' in the case of farmers with zero "know cultivar" or "know cultivar, pest and yield" links.

the bio-safety properties of the new technology. This decreases the perceived safety concerns of other farmers and, as such, decreases social pressures.

The coefficient on the total number of progressive farmer adopters currently known is negative and significant, pointing at free-rider effects with regard to the progressive farmers one knows and observes, i.e., expecting a 'draw' from the production function will decreases the farmer's probability to adopt the new technology. The coefficient on the fraction of progressive farmer adopters currently known is positive, significant and large, pointing at learning imitation effects. This implies that if only, for instance, 2 out 4 progressive farmers known to the farmer are adopting the new technology, the farmer will become suspicious as to why 2 out of 4 are not adopting the technology.

The effect of the current behavior of fellow, non-progressive, farmers is not significantly different from zero in the standard probit, but the total number of fellow, non-progressive, farmers currently known is positive and significant in the IV specification. This results is consistent with the cross-sectional analysis of 2008-09, i.e., conditional on (perceived) bio-safety concerns of others, knowing more current adopters decreases social pressures. Note that free-rider effects, if present, are dominated by these positive social pressure effects.

Table 2.11 presents the results of the same analysis, using the predicted links instead of the actual links. The sign of the coefficients are consistent with the analysis presented in Table 2.10, and the sizes of the coefficients are within the same order of magnitude.

Table 2.7: Characteristics of plans versus beliefs groups in Kanzara and Kinkhed

	Beliefs in 2008	
Plans for 2008-09	E[π-BT]>E[π-NBT]	E[π-BT]<E[π-NBT]
	% of progressive farmers adopters known	
Cultivate Bt-cotton	92.99	100.00
	(9.12)	(0.00)
Cultivate Bt-cotton and non-Bt cotton	77.77	94.44
	(27.21)	(7.85)
Cultivate non-Bt cotton	87.77	NA
	(11.34)	NA
Cultivate cotton, but don't know which cultivar	50.00	NA
	(54.99)	NA
Do not cultivate cotton	94.08	92.59
	(13.07)	(9.07)
Don't know whether to cultivate cotton	72.22	88.88
	(34.24)	NA
	Biosafety concern of others	
Cultivate Bt-cotton	2.70	2.83
	(0.60)	(0.23)
Cultivate Bt-cotton and non-Bt cotton	2.88	2.66
	(0.77)	(0.00)
Cultivate non-Bt cotton	3.00	NA
	(0.69)	NA
Cultivate cotton, but don't know which cultivar	3.16	NA
	(0.23)	NA
Do not cultivate cotton	3.09	3.00
	(0.43)	(0.59)
Don't know whether to cultivate cotton	2.66	2.66
	(0.36)	NA
	Number of adopters known	
Cultivate Bt-cotton	35.94	65.00
	(31.70)	(49.49)
Cultivate Bt-cotton and non-Bt cotton	27.50	52.5
	(6.12)	(38.89)
Cultivate non-Bt cotton	16.95	NA
	(12.22)	NA
Cultivate cotton, but don't know which cultivar	15.00	NA
	(7.07)	NA
Do not cultivate cotton	17.03	14.16
	(18.09)	(14.63)
Don't know whether to cultivate cotton	21.41	5.00
	(22.67)	NA

Notes: standard deviations are reported in parentheses below the means; "% of progressive farmers adopters known" is the number of progressive farmers who plan to cultivate Bt cotton in 2008-09 predicted to be known using the results of the random matching within sample game as a percentage of the total number of progressive farmers predicted to be known (the results are very similar using the actual, and not the predicted links); "Biosafety concern of others" is the average of the answers to "To which degree do you think that other village farmers think that Bt cotton is hazardous for (1) animal health, (2) human health and (3) the environment", this variable ranges from 1 to 5 with higher number referring to an increased concern with biosafety issues; "number of adopters known" is the number of village farmers known that plan to cultivate Bt cotton in 2008-09.

Table 2.8: Effect of social pressures and behavioral imitation on plans to cultivate Bt cotton

Probit regression with dependent variable:	Plan to cultivate Bt cotton in 2008-09	
	Coefficient	Marginal Effect
Mean of perceived profit Bt cotton (in 1,000 Rs)	0.05752	0.00989
	(0.14091)	(0.02394)
Mean of perceived profit non-Bt cotton (in 1,000 Rs)	-0.66237***	-0.11385***
	(0.24764)	(0.03741)
Variance of perceived profit Bt cotton (in 1,000 Rs)	-0.0711	-0.01222
	(0.07374)	(0.01227)
Variance of perceived profit non-Bt cotton (in 1,000 Rs)	0.10084	0.01733
	(0.19297)	(0.03257)
Risk aversion coefficient	-0.00146	-0.00025
	(0.00097)	(0.00016)
Fraction of progressive farmers adopters known	1.93828	0.33317
	(2.11919)	(0.34242)
Biosafety concern of others / number of adopters known	-2.16864***	-0.37277***
	(0.5186)	(0.05880)
Own biosafety concerns	-0.44299	-0.07614
	(0.38351)	(0.06317)
Income (in 1,000 Rs)	0.01178	0.00203
	(0.00964)	(0.00167)
Income * credit constraint	0.05499**	0.00945**
	(0.01963)	(0.00274)
Constant	3.5908	
	(2.73347)	
Control for selection?	yes	yes
Number of observations	47	47

Notes: *** p<0.01; ** p<0.05; * p<0.1; robust standard errors are reported in parentheses below the coefficient and average marginal effects estimates; The mean of the profit of Bt is calculated as the average of the means of the profit of both Bt cultivars (and vice versa for non-Bt); the variance of the profit of Bt is calculated as the average of the variances of the profit of both Bt cultivars (and vice versa for non-Bt); "Risk aversion coefficient" is the difference in willingness-to-pay for a bag of seeds when moving from a baseline yield distribution to a yield distribution with the same mean but a higher variance (see Maertens and Just 2009), a larger coefficient indicates a higher degree of risk aversion; "Fraction of progressive farmers adopters known" is the number of progressive farmers who plan to cultivate Bt cotton in 2008-09 predicted to be known using the results of the random matching within sample game as a fraction of the total number of progressive farmers predicted to be known (the results are very similar using the actual (and not the predicted) links or the total number instead of the fraction); "Biosafety concern of others" is the a to "To which degree do you think that other village farmers think that Bt cotton is hazardous for (1) animal health, (2) human health and (3) the environment, this variable ranges from 1 to 5 with higher number referring to an increased concern with biosafety issues;"Own biosafety concerns" is the average of the answers to "To which degree do you think that Bt cotton is hazardous for (1) animal health, (2) human health and (3) the environment", this variable ranges from 1 to 5 with higher number referring to an increased concern with biosafety issues; "number of adopters known" is the number of village farmers is the number of village farmers known that plan to cultivate Bt cotton in 2008-09; "credit constraint" is a dummy variable referring to whether or not the farmer has access to credit from an input dealer. The knowledge variables are jointly significantly different from zero at the 1% level. Note: using the difference in means and the difference in variance does not change the results significantly but does result in an insignificant effect of knowledge.

Table 2.9: Effect of learning and contemporary social interaction effects on cultivating Bt cotton (based on results of ″how many adopters do you know?″)

Probit regression with dependent variable:	Cultivate Bt cotton	
	Coefficient	Standard Marginal Effect
Non-farmer information up to t	1.72822***	0.04996***
	(0.58657)	(0.02097)
Non-farmer information up to t square	-0.92848***	
	(0.29598)	
Total number of adopters known up to t (village)	-0.01925	-0.00115
	(0.01527)	(0.00081)
Total number of adopters known up to t square (village)	0.00007	
	(0.00013)	
Total number of adopters known up to t (outside of village)	-0.00817	-0.00066
	(0.0335)	(0.00226)
Total number of adopters known up to t square (outside of village)	-0.00008	
	(0.00032)	
Total number of village adopters known at t	0.06118**	0.00395***
	(0.02572)	(0.00098)
Total number of village adopters known at t square	-0.00012	
	(0.00031)	
Risk aversion coefficient	-0.00101	-0.00007
	(0.00079)	(0.00006)
Education decision maker (years)	0.06971	0.00496
	(0.04515)	(0.00307)
Control for total number of people known?	yes	yes
Control for credit constraints?	yes	yes
Control for soil fertility and irrigation?	yes	yes
Control for individual prices?	yes	yes
Control for selection?	yes	yes
Control for state-fixed effects?	yes	yes
Number of observations	408	408

Notes: *** p<0.01; ** p<0.05; * p<0.1; robust standard errors are reported in parentheses below the coefficient and average marginal effects estimates; "non-farmer information up to t" is total number of times the respondent has received information about Bt cotton from a non-farmer source, excluding the information received in May-June, up to the start of season t; "total number of adopters known up to to t" is the, self-reported, total number of adopters know up the start of season t; "total number of vilage adopters known at t" is the, self-reported, total number of adopters known in the village during season t.

Table 2.10: Effect of learning, imitation, social pressures and strategic delays on cultivating Bt cotton (based on results of random-matching-within-sample method, use actual links)

Probit regression with dependent variable:	Cultivate Bt cotton			
	Standard MLE		Instrumental Variable MLE	
	Coefficient	Marginal Effect	Coefficient	Marginal Effect
Non-farmer information up to t	1.61023***	0.05818***	1.30359***	0.09619***
	(0.40241)	(0.01988)	(0.35383)	(0.02800)
Non-farmer information up to t square	-0.82445***		-0.61062***	
	(0.16751)		(0.15545)	
Total number of PF adopters known up to t	0.23055**	0.01360**	0.21929***	0.01899***
	(0.09941)	(0.00579)	(0.07777)	(0.00629)
Total number of PF adopters known up to t square	-0.01542*		-0.01754***	
	(0.00790)		(0.00678)	
Total number of non-PF known up to t	0.00047	0.00001	0.0016	0.00012
	(0.00849)	(0.00059)	(0.00874)	(0.00099)
Total number of non-PF known up to t square	-0.00001		-0.00004	
	(0.00003)		(0.00004)	
Total number of PF adopters known at t	-0.11959**	-0.00933**	-0.18664***	-0.02268***
	(0.05667)	(0.00455)	(0.05509)	(0.00673)
Fraction of PF adopters known at t	3.02467***	0.23597***	2.35255**	0.28588***
	(1.15199)	(0.09080)	(1.06184)	(0.10059)
Total number of non-PF adopters known at t	0.00107	0.00008	0.01624***	0.00197***
	(0.00172)	(0.00013)	(0.00243)	(0.00066)

Notes: *** p<0.01; ** p<0.05; * p<0.1; robust standard errors are reported in parentheses below the coefficient and average marginal effects estimates of the probit and the MLE Instrumental Variable probit; "non-farmer information up to t" is defined as in Table 9; "total number of (non-)PF adopters known up to t" is the cumulative number of times the set of (non) progressive farmers for whom the respondent thinks he knows the cultivar, yield and number of pesticide sprays, has cultivated Bt cotton up to t-1 scaled up to the village level; "total number of (non-)PF adopters known at t" is number of (non) progressive farmers for whom the respondent thinks he knows the cultivar that cultivate Bt cotton at t scaled up to the village level; the "fraction" refers to the ratio of adopters over total number "known"; analysis controls for risk aversion, education decision-maker, total number of people known, credit constraints, soil fertility and irrigation, individual-level prices, and selection; includes state-fixed effects. Number of observations=408.

Table 2.11: Effect of learning, imitation, social pressures and strategic delays on cultivating Bt cotton (based on results of random-matching-within-sample method, use predicted links)

Probit regression with dependent variable:	Cultivate Bt cotton			
	Standard MLE		Instrumental Variable MLE	
	Coefficient	Marginal Effect	Coefficient	Marginal Effect
Non-farmer information up to t	1.81264***	0.06939***	1.53073***	0.07093***
	(0.45756)	(0.02317)	(0.49161)	(0.02431)
Non-farmer information up to t square	-0.98275***		-0.80964***	
	(0.25217)		(0.26759)	
Total number of PF adopters known up to t	0.39862	0.02813	0.71090**	0.05591*
	(0.28820)	(0.01846)	(0.32882)	(0.02919)
Total number of PF adopters known up to t square	-0.02258		-0.04265	
	(0.04230)		(0.04175)	
Total number of non-PF known up to t	0.04561	0.00298	-0.00936	-0.00105
	(0.03314)	(0.00223)	(0.04258)	(0.00342)
Total number of non-PF known up to t square	-0.00079		-0.00034	
	(0.00077)		(0.00066)	
Total number of PF adopters known at t	-0.22432**	-0.01713**	-0.47446***	-0.04083**
	(0.09688)	(0.00749)	(0.15906)	(0.01909)
Fraction of PF adopters known at t	1.74352*	0.13311*	2.26626**	0.19504**
	(0.99742)	(0.07557)	(0.92308)	(0.08506)
Total number of non-PF adopters known at t	0.02515**	0.00192**	0.06250***	0.00538**
	(0.01169)	(0.00086)	(0.02183)	(0.00259)

Notes: *** p<0.01; ** p<0.05; * p<0.1; robust standard errors are reported in parentheses below the coefficient and average marginal effects estimates of the probit and the MLE Instrumental Variable probit; "non-farmer information up to t" is defined as in Table 9; "total number of (non-)PF adopters known up to t" is the cumulative number of times the set of (non) progressive farmers for whom the respondent thinks he knows the cultivar, yield and number of pesticide sprays, has cultivated Bt cotton up to t-1 scaled up to the village level; "total number of (non-)PF adopters known at t" is number of (non) progressive farmers for whom the respondent thinks he knows the cultivar that cultivate Bt cotton at t scaled up to the village level; the "fraction" refers to the ratio of adopters over total number "known"; analysis controls for risk aversion, education decision-maker, total number of people known, credit constraints, soil fertility and irrigation, individual-level prices, and selection; includes state-fixed effects. Number of observations=408.

2.7 Conclusions

This paper looks at the role of social networks in the adoption process of Bt cotton, a new type of (genetically engineered) cotton available on the Indian market since 2002. I take advantage of a panel dataset which contains detailed information on social networks and beliefs with regard to bio-safety and profitability, to, for the first time, identify and estimate the effects of social learning, social pressures and imitation on the technology adoption decision.

The results demonstrate the importance of knowledge about the profitability of a new technology in the adoption decision. This knowledge is established through experimentation, observing other farmers' inputs and outputs and talking to informed parties such as company representatives and seed dealers. For first-time adopters, the last channel is the most important driver to switch to Bt cotton; one conversation with an informed party increases the probability to adopt Bt cotton with 5% to 9%. This finding confirms the results of Conley and Udry (2010) who find that farmers who have received advice from a local extension agent are less likely to change the amount of fertilizers they use on pineapples in Ghana. As farmers, according to their own account, learn about the general properties of the new technology (mean, variance), proper input use and bio-safety issues through these conversations, this result is not surprising and points at the importance of including all sources of information into any analysis of social learning. Note that part of this effect might be due to a decreased perceived bio-safety concern of other's, i.e., decreased social pressures.

The second most important source of information for first-time adopters is the past experience from the progressive farmers to whom one is connected. Consistent with this effect, I also find a (small) free-rider effect of the progressive farmers, i.e., one is more likely to postpone adoption if one knows an additional progressive farmer who plans to cultivate Bt cotton. Surprisingly, the past adoption behavior of the non-progressive farmers one is connected to matters little. This implies that even though, for most farmers, the conditions in which they operate (soil and irrigation) are quite different from the conditions under which the progressive farmers work, they are still more likely to learn from these progressive farmers' experiences than from each other's experiences. In all specifications, the learning effect is concave, suggesting decreasing returns to new information coming in, consistent with, among others, a Bayesian learning model.

Most other studies do not control for different kinds of social interaction effects and , as such, may end up overestimating the extent of social learning. On the other hand, because most studies use a measure of social networks that is likely to be larger than the set of people that a farmer actually learns from, they may end up underestimating social learning effects. Also this effect could depend on the nature of the new technology being introduced such as the difference in expected profit with the old technology. Interestingly, the magnitude of the estimated social learning effects I find is in same range as what Bandiara and Rasul (2006) find in their study of the adoption of sunflower cultivation in Mozambique: knowing one additional farmer, or in this study, knowing

an additional progressive farmer, increases the probability of adopting the new technology by 1.3% to 5.5%.

In addition, I find strong evidence of farmers imitating the progressive farmers they are connected to, i.e., adopting Bt cotton without having observed the yield outcomes of these farmers. As I do not find any evidence of behavioral imitation, I interpret this effect as learning imitation, i.e., a farmer makes inferences about the profitability of the technique if he observes progressive farmers using this technology. Note that while the farmer might not observe the progressive farmer's outputs, he does have some general information about the profitability of this farmer judging by his wealth and income. Compared to the other effects, the magnitude of this effect is large: a farmer who knows only progressive farmers who currently cultivate Bt cotton is about 13% to 28% more likely to adopt the Bt cotton compared to a farmer who knows only progressive farmers who currently cultivate non-Bt cotton.

I find strong evidence of social pressures inhibiting the adoption of Bt cotton. These social pressures find their origin in bio-safety concerns with regard to Bt cotton: a farmer may be accused of endangering the health of the animals and people in the village as well as generating negative impacts on the soil fertility and water quality of neighboring plots. A farmer who is surrounded by farmers who have strong bio-safety concerns instead of farmers who have no bio-safety concerns at all is about 10% less likely to adopt Bt cotton (conditional on knowing sixteen adopters). Note that, technically speaking, I cannot exclude the option that these accusations might lead to financial penalties. While no such case has been observed in any of the villages, it is theoretically possible that

some farmers perceive there to be positive probability of accusations leading to penalties.

A limitation of this study is that the lack of quality profit data prohibits a credible analysis of the determinants of profitability as in Foster and Rosenzweig (1995). Similarly, the lack of quality acreage data means that I cannot further investigate the partial adoption phenomenon, prevalent in one of the village. Qualitative information and an analysis of the cross-sectional data however indicate that farmers who are more sure of the profit and expect a higher profit are more likely to convert all their cotton land to Bt cotton, while farmers with more doubts and who expect a lower profit are more likely to use only a share of their land.

In addition, because social networks were measured at only one point in time, the identification strategy is based on the assumption that the nature of the relationship between two individuals in the village has not changed in the last seven years and that the social network in terms of learning, imitation and social pressures does not exceed the village boundaries. Comparing the results using various measures of social networks point to the importance of identifying the network correctly. While most farmers knew everyone in this village, this did not imply that they knew whether or not a certain village farmer is using the new technology and certainly did not imply that they knew the input choices and outcomes of each village farmer.

Finally, note that the analysis presented in this paper critically hinges upon the selection of the progressive farmers. Ideally one would want to make this

selection endogenous and let the data determine which type of farmers are most influential. One could interacts the social interaction variable of interest with, for instance, the average soil quality of the network contacts, but due to the fact that the social interaction variables are constructed as an aggregate over time and network contacts it is not clear how one could interpret the results of such an analysis. Nevertheless, making this selection endogenous seems to be an important avenue for future research.

Along the same lines, it is important to think through the interpretation of the results in case I have accidently placed some of the progressive farmers into the non-progressive farmers category. In this case the importance of learning from non-progressive farmers would be over-estimated and the importance of social pressures would be under-estimated. This is unlikely to be the case as in this case one would expect the effect of learning from non-progressive farmers to be significantly different from zero, which it is not. If, on the other hand, some of the non-progressive farmers ended up in the progressive farmers category, the importance of learning from progressive farmers and imitation effects would be under-estimated.

To conclude, using a fertile setting, rural India, and detailed quality panel data, this paper takes a first step in distinguishing various social interaction effects. I show that multiple mechanisms matter, but not equally. And as these mechanisms are correlated, omitting one tends to exaggerate the importance of others. In addition, I point to the importance of including non-farmer sources of information and correctly identifying the relevant social network.

APPENDIX A

Of the 199 households covered in 2001-02 by the ICRISAT-VLS, 92% were still in the sample in 2007-08. The households that dropped out were, on average, smaller in terms of household size, higher educated and with less total land, but more irrigable land. I interviewed these 184 households plus some of their split-offs and newly added households, a total of 246 households. This sample is representative for the village in 2007-08 (see Bantilan et al. 2006 and Rao and Charyulu 2007). Of these, 68% have data for all seven cropping years. Of the 32% of the households who are included in the sample from a later date onwards, 30% are households that have split off from sample households during 2001-2008. These split-off households are included from their date of split-off.

Table 2.12 lists the number of households in the sample, by date included in the ICRISAT-VLS.

Table 2.12: Number of households included in the sample, by date included in VLS

	Aurepalle	Kanzara	Kinkhed
From 2001-02 onwards	94	46	29
From 2002-03 onwards	0	0	0
From 2003-04 onwards	5	0	2
From 2004-05 onwards	0	0	0
From 2005-06 onwards	18	13	22
From 2006-07 onwards	11	4	1
From 2007-08 onwards	0	0	1
Total sample	128	63	55

APPENDIX B

In 2007-2008, I interviewed 246 ICRISAT-VLS households and an additional 21 progressive farmers that were not part of the ICRISAT-VLS at the time of the interview. Of these 246 ICRISAT-VLS farmers, I labelled 22 'progressive farmers'. As such, in total, there are 43 progressive farmers in the sample. Table 2.13, Table 2.14 and Table 2.15 show how these progressive farmers diff from the other households. On average, the progressive farmers are more educated and own more land, of which a larger share is irrigated. The progressive farmers are far more likely to adopt Bt cotton and, conditional on adoption, achieve a higher profit per acre and adopt the new technology earlier. The exceptionally low profit per acre of the Kinkhed progressive farmers is due to the crop failure of two progressive farmers in the village due to floods.

Table 2.13: Progressive farmers characteristics in Aurepalle

	Non-progressive farmers			Progressive farmers		
	N	Mean	St. Dev.	N	Mean	St. Dev.
Per acre profit of Bt cotton (Rs)	48	6,588	6,753	11	10,349	3,422
First year of adoption	49	2006.22	0.47	21	2006.2	0.89
Number of household members	116	4.15	1.55	28	4.29	2.66
Education level decision maker (year)	116	2.24	3.98	28	4.2	4.77
Owned dry land (acres)	116	2.1	2.42	28	7.1	8.4
Owned irrigated land (acres)	116	0.64	1.36	28	7.31	8.76
% adopted Bt cotton at any point in time	116	42		28	75	

Note: All variables, except for the first year of adoption and the % adopted Bt cotton at any point in time relate to the year 2007-08.

Table 2.14: Progressive farmers characteristics in Kanzara

	Non-progressive farmers			Progressive farmers		
	N	Mean	St. Dev.	N	Mean	St. Dev.
Per acre profit of Bt cotton (Rs)	15	4,783	5,240	4	6,183	11,691
First year of adoption	23	2004.48	1.56	6	2003.67	1.97
Number of household members	59	4.86	1.78	6	6.67	4.27
Education level decision maker (year)	59	6.34	3.98	6	10.5	2.81
Owned dry land (acres)	59	2.17	4.46	6	0	0
Owned irrigated land (acres)	59	20.06	3.79	6	18.37	6.23
% adopted Bt cotton at any point in time	59	38		6	100	

Note: All variables, except for the first year of adoption and the % adopted Bt cotton at any point in time relate to the year 2007-08.

Table 2.15: Progressive farmers characteristics in Kinkhed

	Non-progressive farmers			Progressive farmers		
	N	Mean	St. Dev.	N	Mean	St. Dev.
Per acre profit of Bt cotton (Rs)	1	3,959	NA	4	-214	2,311
First year of adoption	1	2007	NA	7	2005.85	0.75
Number of household members	49	4.53	1.86	9	4.44	1.24
Education level decision maker (year)	49	6.35	3.61	9	11.56	2.7
Owned dry land (acres)	49	2.2	4.44	9	7.13	6.25
Owned irrigated land (acres)	49	1.49	2.91	9	13.08	7.71
% adopted Bt cotton at any point in time	49	2		9	77	

Note: All variables, except for the first year of adoption and the % adopted Bt cotton at any point in time relate to the year 2007-08.

APPENDIX C

Regarding the selection of the progressive farmers for the random-matching-within-sample method, recall that only 4 progressive farmers were selected for each village out of these 43 progressive farmers. In the case of Aurepalle, 2 different sets of progressive farmers were identified, one for the main village of Aurepalle and one for the sub-village of Nallavaripalli. Each Aurepalle farmer was then matched up with the three main progressive farmers of Aurepalle and the main progressive farmers of Nallavaripalle and vice versa. The idea was to select all progressive farmers who could play a central role in the dispersion of information and potentially be role models for the farmers and as such the object of imitation.

Aurepalle consists of two separate sub-villages: the main village of Aurepalle and the sub-village of Nallavaripalle. I presented the respondent with 4 out of 6 matches from the same village and 2 out of 6 from the other villages. Similarly, for the progressive farmers, 3 out of 4 were from the same village and 1 out of 4 from the other village.

In order to get a sense of how successful this selection was, I compare the result of two social network questions, one from the perspective of the non-progressive farmers, the other from the perspective of the progressive farmers, with the farmers selected for the game.

From the perspective of the non-progressive farmers, Table 2.16 presents the results of the question "If you, today, would have a specific problem with your cotton crop who or where would you go to (up to 5 answers allowed)?"[53] Of all the village farmers mentioned by the respondents, I could retrieve data for 80% to 92% of the cases[54] and 63% to 91% were labelled as "progressive farmer" before the data collection started. In Aurepalle and Kinkhed about 40% of the farmers mentioned by all the VLS respondents of these were selected for the random-matching-within-sample method. The lowpercentage in Aurepalle is due to the fact that several Aurepalle farmers in case of trouble would approach a progressive farmer located in Nallavaripalle. This was confirmed by the ICRISAT-VLS resident investigators in Aurepalle who mentioned that "Aurepalle does not have any real progressive farmers, these are all in Nallavaripalle". The lowpercentage in Kinkhed, on the other hand, is due to the fact that many respondents would approach farmers that were not labelled progressive farmers.

Table 2.16: Selection of PF for random matching within sample as percentage of village farmers mentioned by all VLS respondents

	Data available	Labelled PF	Selected for game
Aurepalle	79.65	76.21	37.29
Nallavaripalle	92.3	90.38	78.85
Kanzara	89.74	78.21	70.51
Kinkhed	89.09	62.73	40

From the perspective of the progressive farmers, Table 2.17 mentions the rank of the 4 progressive farmers selected for the game in terms of the variable

[53]This question was asked to all the respondents who ever cultivated cotton since 2001-02 or who plan to cultivate cotton in the future.

[54]The remainder of the farmers referred to were not part of the ICRISAT-VLS or the set of progressive farmers identified before the data collection started.

"pf_TALK". This variable "pf_TALK" indicates the number of VLS farmers the progressive farmers talks to on an at least monthly basis. Thus the progressive farmer with rank one talks to the largest number of VLS farmers (compared to the other progressive farmers)[55] In the case of Kinkhed the four top farmers were also the ones selected for the game. In the case of Aurepalle, only one of the top five progressive farmers based in Aurepalle was selected for the game and in the case of Nallavaripalle, three farmers from the top four farmers based in Nallavaripalle were selected for the game.

Table 2.17: Rank of 4 progressive farmers selected for the game in terms of pf _ TALK

	Rank	Total number of PF
Aurepalle	7,9,12,17	28
Nallavaripalle	2,5,6,26	28
Kanzara	1,2,5,6	6
Kinkhed	1,2,3,4	9

To conclude, it appears that I did a relatively good job in identifying the most important progressive farmers ex-ante in Nallavaripalle, Kanzara and Kinkhed but failed to include the most important progressive farmers in Aurepalle. As such, I repeat the analysis using two different sets of measures of the imitation and social learning from progressive farmers component. The first set of measures is based on the random-matching-within-sample method. The second set of measures includes all progressive farmers who, according to their own account, talk on a at least monthly basis to the respondent in case of the imitation component and all the progressive farmers who, according to their own

[55] I asked each progressive farmer (both VLS and non-VLS) to tell us, for each VLS respondent, whether he talks daily, weekly, monthly, yearly, twice a year or never to this person. The variable "TALK" was constructed as the sum of the number of individuals each progressive farmer talks to on a daily, weekly and monthly basis.

account, believe that the respondent knows their cultivar use and yield output in case of the social learning component.[56]

[56]With regard to the latter measure - it must be noted that four progressive farmers in Aurepalle could not answer the question ″Do you think the following person (show VLS respondent) knows which cultivars you have sown and how much yield you got from each cultivar this year″ and answered ″don't know″ for all the VLS respondents.

APPENDIX D

Table 2.18 presents the results of a OLS regression of individual prices (in Rs/quintal) received for cotton in 2007-08 on variables representing the who, when and where aspects of the sales of cotton using the 2007-08 data of cotton price (in Rs/Quintal). Note that the price of Bt cotton is not significantly different from the price of non-Bt cotton. The location and time of sale matter and selling in the village yields a lower price compared to selling at the market. Note that the location of sale is responsible for the large majority of the variation in output price. The difference in price received when one sells in the market versus whether one sells in the village reflect the transportation costs between market and village (it would cost about 50 Rs to hire a small three-wheeler truck). When a farmer sells his produce will depend on in the first place on the duration of the cultivar (as cotton is a perishable product), i.e., how many weeks it takes for the crop to mature and in the second place on whether he needs to reimburse a input dealer or money lender urgently. In general a farmer will pick a short-duration cultivar if he plans to cultivate crops during the summer season (March-May), and this in turn will depend on whether he has access to a functional irrigation source during the summer.

Table 2.18: OLS cotton price in 2007-08 N=427, Adj R=0.30

	Coeff.	St. Error
Quantity (quintal)	1.45	1.3
Aurepalle	-194.33***	32.47
Kinkhed	-10.63	30.04
Weeks since 1 October	7.74***	1.34
Dummy Bt cultivar	47.1	34.61
Sold to trader/agent in the village	-43.34**	17
Sold to farmer in the village	-51.70***	15.86
Sold to same person from whom input was bought	3.78	11.16
Constant	2077.41***	33.61

Notes: *** significant at 1% level, ** significant at 5% level, * significant at 10% level

APPENDIX E

The control for soil fertility is the number of plots the farmer has access to (owned, leased/shared in, leased/shared out) larger or equal to one acre of "good" or "very good" quality (as perceived by the farmer). As not all the VLS years contain data on the value of soil in Rs/acre, I construct a soil fertility measure based on the perceived quality of soil in the VLS panel data. As no landholding module was included in the 2002-03 and 2003-04 round, I extrapolated the information of, respectively, 2001-02 and 2004-05.

The measure for credit constraints and risk aversion are constructed in the same way as in the cross sectional analysis, as such I assume that, controlling for wealth (land owned) which does vary over time, these attributes of the individual do not change over time.

The irrigation constraints are captured by an individual-level effect captures whether or not the farmer, in the past seven year, ever has faced a situation where he wanted to give additional irrigation to his cotton crop, but there was insufficient water available to him.

The prices used are individual-level predicted prices for each year. The input-output and Y-modules of the ICRISAT-VLS 2001-07 data contain a year, individual, plot and activity specific price for each input and output, not including the cost of credit. In case no market transaction took place the price provided is an estimate by the enumerator or respondent on the field. Similarly,

the 2007-08 data contains also individual, plot and activity specific prices, but only for the items for which a market transaction took place.

I selected a set of input and output prices, relevant to the cotton cultivation decision. The choice of which output prices to include is complex, as the competing crops change each year. I based my choice on the number of farmers who cultivate a certain crop in 2007-08 and included cotton, blackgram, greengram, sorghum, rice, cowpea, pigeonpea, soybean and castor. As the output prices are not know per crop in 2001-02 in the input-output schedule, I used the production and utilization schedule instead for this year. In terms of fertilizers and pesticide prices, I computed the 'average price' of fertilizers (pesticides) for each individual as follows: the sum of the total value spent on fertilizers (pesticides) divided by the total amount of fertilizers in kg (liter), excluding farm yard manure and the price of labor. Note that these prices are a function of the input choice of farmers in terms of fertilizers (pesticides). Constructing an average price for fertilizers and pesticides is necessary as the available fertilizers and pesticides change each year and for 2004-05 and 2002-03 no information is available on the type of fertilizers and pesticides used. In terms of other input prices, as the Bt plots -as evident from the 2007-08 input-output data - on average receive more inputs, the set of relevant prices must include all other input prices, namely the price of female, male and bullock labor and the rent of a tractor.

I then computed an year/individual specific average using the various plot and activity specific prices of the Kharif season (except for 2001-02 and 2002-03 which use both Kharif and Rabi prices as they cannot be separated out) and pre-

dicted the price that each individual farmer faces each year using the following OLS regression $p_{ijt} = \beta_0 + \beta_1 L_{ijt} + \beta_2^2 L_{ijt}^2 + VT_{jt} + IRRI_{ijt} + \epsilon_{ijt}$, where L stands for the acreage of land owned, $IRRI$ is a dummy variable capturing whether or not the farmer has irrigable land and VT stands for a time-village level fixed effect. In terms of seed costs, as the seed costs are not always known per crop in 2001-04 and the unit is not always recorded in the later years, I used the medium price for non-Bt cotton seeds, excluding the outliers in this calculation, and the official Bt price for Bt seeds.

Finally, I deflated the price series using the State-wise Consumer Price Index Numbers for Agricultural Labourers in India available from INDIASTAT.SECTION 1

BIBLIOGRAPHY

[1] Appadurai, Arjun. 1989. "Transformations in the Culture of Agriculture," in *Contemporary Indian Tradition: Voices on Culture, Nature and the Challenge of Change*. Carla M. Borden ed. Washington and London: The Smithsonian Institution Press, pp. 173-186.

[2] Asia-Pacific Consortium of Agricultural Biotechnology. 2006. *Bt Cotton in India - A Status Report*. New Delhi.

[3] Bandiera, Oriana and Imran Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique." *Economic Journal*, 116:514, pp. 869-902.

[4] Bantilan, M.C.S., P. Anand Babu, G.V. Anupama, H. Deepthi, and R. Padmaja. 2006. "Dryland Agriculture: Dynamic Challenges and Priorities." Research Bulletin No. 20, GT-IMPI, ICRISAT.

[5] Barrett, Christopher B. 2005a. "Smallholder Identities and Social Networks The Challenge of Improving Productivity and Welfare," in *The Social Economic of Poverty: On Identities, Communities, Groups and Networks*. Christopher B. Barrett ed. London and New York: Routledge, pp. 214-43.

[6] Barrett, Christopher B. ed. 2005b. *The Social Economics of Poverty. On Identities, Communities, Groups and Networks*. London and New York: Routledge.

[7] Basu, Kaushik. 2000. *Prelude to Political Economy: A Study of the Social and Political Foundations of Economics*. Oxford University Press.

[8] Besley, Timothy and Anne Case. 1993. "Modeling Technology Adoption in Developing Countries." *American Economic Review*, 83:2, pp. 396-402.

[9] Binswanger, Hans P. 1980. "Attitudes toward Risk: Experimental Measurement in Rural India." *American Journal of Agricultural Economics*, 62:3, pp. 395-407.

[10] Borelli, Sara. 2009. "Social Aspects of Drug Use in Djibouti: The Case of the Leaf of Allah." *Journal of African Economies*, 18:4, pp. 555–591.

[11] Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of Econometrics*, 150, pp. 41-55.

[12] Chamley, Christophe P. 2004. *Rational Herds, Economic Models of Social Learning*. Cambridge University Press.

[13] Conley, Timothy G and Christopher Udry. 2010 "Learning About a New Technology: Pineapple in Ghana." *American Economic Review,* 100:1, pp. 35–69.

[14] Crost, Benjamin, Bhavani Shankar, Richard Bennett, and Stephen Morse. 2007. "Bias from Farmer Self-Selection in Genetically Modified Crop Productivity Estimates: Evidence from Indian Data." *Journal of Agricultural Economics*, 58:1, pp. 24-36.

[15] David, G. Shourie and Y.V.S.T. Sai. 2002. "Bt Cotton: Farmer's Reactions." *Economic and Political Weekly*, November 16, 2002.

[16] Delavande, Adeline, Xavier Giné, and David McKenzie. 2008. "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence." Policy Research working paper World Bank no. WPS 4824.

[17] Dev, S. Mahendra and N. Chandrasekhara Rao. 2006. "Socio-Economic Impact Assessment of Transgenic Cotton in Andhra Pradesh." Working Paper. Center for Economic and Social Studies, Hyderabad.

[18] Doherty, V.S. 1982. "A Guide to the Study of Social and Economic Groups and Stratification in ICRISAT's Indian Village Level Studies." Working Paper GT-IMPI, ICRISAT.

[19] Fafchamps, Marcel and Flore Gubert. 2007. "Risk Sharing and Network Formation." *American Economic Review Papers and Proceedings*, 97:2, pp. 75–79

[20] Feder, Gechon and Sara Savastano. 2006. "The Role of Opinion Leaders in the Diffusion of New Knowledge: The Case of Integrated Pest Management." *World Development*, 34:7, pp. 1287-300.

[21] Feder, Gechon, Richard Just, and David Zilberman. 1985. "Adoption of Agricultural Innovations in Developing Countries: A Survey." *Economic Development and Cultural Change*, 32:2, pp. 255-97.

[22] Floyd, C.N., A.H. Harding, D.P. Paddle, D.P Rasali, K.D. Subedi, and P.P. Subedi. 1999. "The Adoption and Associated Impact of Technologies in the Western Hills of Nepal." Agricultural Research and Extension Network Paper No. 90.

[23] Foster, Andrew D. and Mark R. Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy*, 103:6, pp. 1176-209.

[24] Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica,* 25:4, pp. pp. 501-522.

[25] Hogset, Heidi and Christopher B. Barrett. 2010. "Social Learning, Social Influence, and Projection Bias: A Caution on Inferences Based on Proxy Reporting of Peer Behavior," *Economic Develoment and Cultural Change*, 58:3, pp. 563-589.

[26] Hubbell, Bryan J., Michele C. Marra, and Gerald A. Carlson. 2000. "Estimating the Demand for a New Technology: Bt Cotton and Insecticide Policies." *American Journal of Agricultural Economics*, 82:1, pp. 118-32.

[27] ISAAA, International Service for the Acquisition of Agri-Biotech Applications 2005. "The Story of Bt Cotton in India - Video."

[28] Jackson, Matthew O. and Brian W. Rogers. 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks." *American Economic Review*, 97:3, pp. 890-915.

[29] Krishnan, Pramila and Emanuela Sciubba. 2009. "Links and Architecture in Village Networks." *Economic Journal,* 119:537, pp 917-949.

[30] Liu, Elaine. 2008. "Time to Change What we Sow: Risk Preferences and Technology Adoption Decisions of Cotton Farmers in China." Department of Economics. Princeton University

[31] Lybbert, Travis J. 2006. "Indian Farmer's Valuation of Yield Distributions: Will Poor Farmers Value 'Pro-Poor' Seeds?" *Food Policy*, 31:5, pp. 415-41.

[32] Lybbert, Travis J. and David R. Just. 2007. "Is Risk Aversion Really Correlated with Wealth? How Estimated Probabilities Introduce Spurious Correlation " *American Journal of Agricultural Economics*, 89:4, pp. 839-1224.

[33] Lybbert, Travis J., Christopher B. Barrett, John G. McPeak, and Winnie K. Luseno. 2007. "Bayesian Herders: Updating of Rainfall Beliefs in Response to External Forecasts." *World Development* 35:3, pp. 480–97.

[34] Maertens Annemie and David Just. 2009. "Marginal Versus Average Risk Aversion." Working Paper, AEM, Cornell University.

[35] Manski, Charles F. 1993. "Identification of Social Interaction: The Reflection Problem." *Review of Economic Studies*, 60:3, pp. 531–542.

[36] Mendelsohn, Mike, John Kough, Zigfridais Vaituzis, and Keith Matthews. 2005. "Are Bt Crops Safe?" *Nature Biotechnology*, 21:5, pp. 1003-09.

[37] Moser, Christine M. and Christopher B. Barrett. 2006. "The Complex Dynamics of Smallholder Technology Adoption: The Case of SRI in Madagascar." *Agricultural Economics*, 35:3, pp. 373–388.

[38] Naik, Gopal, Matin Qaim, Arjunan Subramanian, and David Zilberman. 2005. "Bt Cotton Controversy: Some Paradoxes Explained." *Economic and Political Weekly*, April 9 2005, pp. 1514-17.

[39] Narayanamoorthy, A and S. S. Kalamkar. 2006. "Is Bt Cotton Cultivation Economically Viable for India Farmers? An Empirical Analysis." *Economic and Political Weekly*, June 30, 2006, pp. 2716-24.

[40] Pemsla, D., H. Waibela, and J. Orphalb. 2004. "A Methodology to Assess the Profitability of Bt-cotton: Case Study Results from the State of Karnataka, India " *Crop Protection*, 23:12, pp. 1249-57.

[41] Pomp, Marc and Kees Burger. 1995. "Innovation and Imitation: Adoption of Cocoa by Indonesian Smallholders." *World Development* 23:3, pp. 423-31.

[42] Qaim, Matin and David Zilberman. 2003. "Yield Effects of Genetically Modified Crops in Developing Countries." *Science*, 299: 7 February 2003, pp. 900-902.

[43] Qaim, Matin, Arjunan Subramanian, Gopal Naik, and David Zilberman. 2006. "Adoption of Bt Cotton and Impact Variability: Insights from India." *Review of Agricultural Economics*, 28:1, pp. 48-58.

[44] Qaim, Matin. 2003. "Bt Cotton in India: Field Trial Results and Economic Projections." *World Development*, 31:12, pp. 2115–27.

[45] Rao, K.P.C, 2008. "Documentation of the Second Generation Village Level Studies 2001-2004". GT-IMPI, ICRISAT.

[46] Rao, K.P.C. and Kumara D. Charyulu. 2007. "Changes in Agriculture and Village Economies." Research Bulletin no 21, GT-IMPI, ICRISAT.

[47] Rogers, Everett. 1962. *Diffusion of Innovations.* New York: Free Press.

[48] Saha, Atanu, C. Richard Shumway, and Hovav Talpaz. 1994. "Joint Estimation of Risk Preference Structure and Technology Using Expo-Power Utility." *American Journal of Agricultural Economics*, 76:2, pp. 173-84.

[49] Santos, Paulo and Christopher B. Barrett. June 2007. "Understanding the Formation of Social Networks." Working Paper Cornell University.

[50] Singh, R.P., Hans P. Binswanger, and N.S. Jodha. 1985. "Manual of Instructions for Economic Investigators in ICRISAT's Village Level Studies (Revised)." GT-IMPI, ICRISAT.

[51] Shelton, A. M., J.-Z. Zhao, and R. T. Roush. 2002. "Economic, Ecological, Food Safety and Social Consequences of the Deployment of Bt Transgenic Plants." *Annual Review of Entomology*, 47, pp. 845-81.

[52] Stephens, Emma C. August 2007. "Feedback Relationships between New Technology Use and Information Networks: Evidence from Ghana." Working Paper Pitzer College.

[53] Stone, Glenn Davis. 2007. "Agricultural Deskilling and the Spread of Genetically Modified Cotton in Warangal." *Current Anthropology* 48:1, pp. 67-103.

[54] Subramanian, Arjunan and Matin Qaim. 2009. "Village-wide Effects of Agricultural Biotechnology: The Case of Bt Cotton in India." *World Development*, 37:1, pp. 256-267.

[55] Sunding, David and David Zilberman. 2001. "The Agricultural Innovation Process: Research and Technology Adoption in a Changing Agricultural Sector," in *Handbook of Agricultural Economics*. Bruce L Gardner and Gordon C. Rausser eds: Elsevier Science, pp. 207-61.

[56] Vasavi, A. R. 1994. "'Hybrid Times, Hybrid People': Culture and Agriculture in South India." *Man*, 29:2, pp. 283-300.

[57] Walker, Thomas S. and James G. Ryan. 1990. *Village and Household Economies in India's Semi-Arid Tropics*. Baltimore and London: John Hopkins University Press.

CHAPTER 3

HIGH RETURNS, LOW ASPIRATIONS? SOCIAL NORMS AND EDUCATION IN VILLAGE INDIA

3.1 Introduction

Parents in developing countries invest in the education of their children as they expect this will increase their and their children's welfare in the long term. The literature to date has shown that the magnitude of this investment depends on many factors, including parental education, social and economic background, work opportunities, village and regional development, school quality, costs and educational subsidies.[1] Variation in these factors impacts credit constraints, and the costs of and returns to education, and therefore explains a significant amount of the difference often observed in educational attainment between boys and girls and between social groups in developing countries. But it does not explain most of it. In India, girls and children belonging to lower caste groups (controlling for credit constraints, costs and returns to education) are significantly less likely to go to school compared to boys and children from upper caste groups (Drèze and Kingdon, 2000).[2] So what is driving these unexplained gender and caste based differences?

[1]The literature in economics on education is vast. Rosenzweig (1995), Schultz (1961) and Schultz (1989) provide good introductions.

[2]The Indian caste system describes the social stratification and social restrictions in the Indian subcontinent in which social classes are structured along four principle castes (termed *varna*) and defined by thousands of endogamous local groups (termed *jātis* or sub-castes). The formerly called *untouchables* constitute a "fifth caste", those people who fall below the and are, in a sense, outside of the caste system altogether.

A part of the explanation might be group-based heterogeneity in market returns (due to, for instance, labor market discrimination[3]) or imperfect information. Indeed, what will matter to the parents is not the average cost and benefits of education, but what they perceive these cost and benefits to be for their particular child. Parents form their opinion based on the limited information they receive from schools, media and interactions with their children, relatives and friends. As such, due to assortative matching by social groups, opinions might be correlated within groups.[4]

In addition to learning from relatives and friends, the social network in which parents are embedded might also influence their investment in education in a different manner, through social norms, customs and pressures. For instance, in a uneducated family getting a higher education might be frowned upon and alienate family members, while in educated families, stopping one's education prematurely might be criticized.[5]

[3]Studies indicate that women and lower caste groups might suffer from wage discrimination (Ito 2009, Kingdon 1998, Rosenzweig and Schultz 1982). One needs to be careful however in interpreting these results. As women in India tend to be found in very different jobs than men, a difference in wages can not be interpreted as evidence of direct discrimination against women. Once the nature of the job is taken into account, Chakravarty and Somanathan (2007) find no evidence of wage discrimination against women.

[4]Recent literature has pointed to the importance of subjective versus 'objective' beliefs with regard to educational returns. Jensen (2010) finds that measured returns to schooling in the Dominican Republic are high, but the returns perceived by students are extremely low. Nguygen (2008) finds that parents' median perceived return matches the average return estimated from household survey data in Madagascar. In addition, she finds a lot of dispersion in both perception of the average return and perception of the child's own return. She attributes the latter to both heterogeneity in the actual returns and imperfect information. Attanasio and Kaufmann (2008) show that these subjective returns matter and are significant predictors of college and high school attendance choices in Mexico, but only for richer individuals.

[5]Few economists have explicitly studied social norms in the education decision in India. Some work has been done by Chamarbagwala (2009). Outside of the Indian context, Akerlof and Kranton (2005) and Austen-Smith and Fryer (2005) discuss, theoretically, how one's group identity might influence one's investment in education.

Using a detailed child level dataset I collected in three villages in rural India in 2007-08, I dissect the role of these social norms, customs and pressures in the education decision, taking into account the differences in (perceived) returns to education. As the large majority of the children in these villages complete elementary education and lower secondary education, I focus on future plans to invest in upper secondary and higher education. Considering these future plans to invest in education, which I will refer to as educational aspirations, instead of the actual investment in education has some benefits. First, future plans to invest in education determine educational attainment 10-15 years from now, and as education is known to be an important determinant of micro level growth, one is basically examining one of the main drivers of growth directly. Second, looking at aspirations has the methodological advantage that one can directly control for (perceived) costs and benefits of education as it is feasible to elicit both aspirations and expectations during one survey.[6]

The main idea behind the empirical strategy is simple. I check whether the beliefs with regard to the returns to education are consistent with future investment plans with regard to education, meaning if one believes higher education has benefits, net of costs, then one should invest in higher education. If this is not the case, something else is constraining the individual. Using regression analysis, I explain the variation in child-level aspirations by the variation in child-level (perceived) costs and returns, credit constraints and time preferences, and social norms, customs and pressures.

[6]If one were to look at educational attainment instead, one would have to estimate the beliefs parents had several years back with regard to the costs and benefits of education.

In this context, social pressures might directly relate to the level of education and/or indirectly influence aspirations through affecting the (perceived) costs and benefits of education. The latter relates to wedding and marriage customs as practiced in most of rural India. When a girl is in her late teens to early twenties, her parents are often pressured by relatives and friends to get her married. Once a suitable groom is found, the two families negotiate the dowry, i.e., the transfer of gifts and money that the bride brings to the groom's household. This dowry can be perceived as a price for a groom, and will (among others) depend on the level of education of both parties. After marriage, the bride moves in with her husband or his parents, a practice known as patrilocality. Once the bride has switched homes, it is often socially unacceptable to continue to support her natal family, either financially or in terms of physical care. Using child level data on these marriage-related social customs and information on the education obtained by the household's social network, I disentangle the influence of marriage age, dowry, patrilocality and social pressures which relate directly to the level of education.

I find that educational aspirations are much lower for girls and for lower caste groups. Not taking into account the difference in dowry and old age care, the perceived returns to education are significantly lower for girls at all education levels. As expected, the ideal age of marriage is lower for girls (compared to boys) and sons are expected to be the main financial providers at old age. The fraction of the parents' social network that has received some higher education is also much lower when considering the women in the network compared to the men in the network. Similarly, this fraction is lower for lower caste parents.

Aspirations are a complex function of wealth, time preferences and the age of the parent, the perceived costs of and returns to education, and the prevailing social norms, customs and pressures with regard to marriage and old age care. The most important social norm is the one regarding the ideal age of marriage. Increasing the ideal age of marriage significantly increases the probability that parents want higher education for their daughter. The custom of patrilocality through old age care influences the perceived net returns to education from the perspective of the parents and influence aspirations as expected. I do not find any evidence of social pressures directly related to the level of education and cannot confirm a negative relationship between the price of education (in terms of dowry price) and aspirations. On the contrary, increasing the dowry associated with a certain education level, increases the chances of that level being aspired to. This might point to social status effects dominating the price effect of an increase in dowry.

In terms of policy, this study shows how important it is to address the educational gap, as observed between gender and castes in India, in a comprehensive manner. In order to change the aspirations parents have for their female children and that parents of lower castes have for their children, one has to address the various constraints simultaneously. These include, most importantly, credit constraints and information constraints. In addition, one needs to pay attention the role of imperfect land and labor markets in rural India, as these affect the opportunity costs of education in rural areas. Finally, one needs to take into account the current social customs and norms with regard to marriage and old age care.

The identification of the effects of social norms and customs on the education decisions has been attempted in two prior papers. Foster and Rosenzweig (2001), using panel data from India, find that the practice of patrilocal exogamy - marriage into families outside of the village – may make parents reluctant to invest in their girls. Field and Ambrus (2009) directly test the hypothesis that women attain less schooling as a result of what they believe to be social and financial pressure to marry young. Using household data from rural Bangladesh, they are able to isolate the causal effect of marriage timing using age of menarche as an instrumental variable. They find that delaying marriage results in additional schooling. Then, the interaction between dowry and education is discussed, theoretically, by Lahira and Self (2007) who show that a groom specific dowry can result in an additional bias against the female children (resulting is less education) while a bride specific dowry can result in less bias against the female children (resulting in more education).[7]

Among the papers that study social interaction effects in the schooling decision, Bobonis and Finan (2009), using experimental evidence from the Mexican Progresa program, find that enrollment of program-ineligible children increases significantly as more program-eligible children in their peer group start going to school. At a more conceptual level, this study is also linked to the work of Ray (2003) and Appadurai (2004) on aspirations and social groups. Appadurai argues that aspirations are socially determined and that the capacity to aspire is unevenly divided between rich and poor. The instrumental consequence of this

[7]Relatedly, Dasgupta and Mukherjee (2003) show (also theoretically) how parents' preferences for their sons to stay at home after marriage, together with the institution of arranged marriage, can lead to a preference for under-educated brides. According to their analysis, only a switch from arranged marriage to love marriage could raise female literacy.

is that the poor thereby lack the "[aspirational] resources to contest and alter the conditions of their own poverty" [8] Ray further explores the determinants of these aspirations (theoretically) and reasons that people form an "aspiration window" based on the experiences of similar individuals. He then defines the aspirations gap as the difference between the standard of living to which one aspires and the standard of living that one already eperiences, and argues that this gap that drives behavior.

The remainder of this article is structured as follows. The next section provides some background information on the education system in India. Section 3 introduces the data. To set the stage for the analysis, Section 4 discusses some selected descriptive statistics, exploring the variation in plans and expectations with regard to education and provides a description of how dowry, marriage age and old age care interact with the education decision, based on conversations with respondents. Section 5 derives the empirical specification. Section 6 presents the results and Section 7 concludes.

3.2 Education in India

The education system in India comprises school education from 1^{st} to 12^{th} standard, and higher education, beyond 12^{th} standard, also referred to as 12+. A child typically enrolls in school at the age of six. The first eight standards of school education are known as elementary education. Up to the age of four-

[8]Cited from Debraj Ray as a comment on Appadurai's working paper at the World Bank Conference on Culture and Development, June 2002.

teen, school education is (almost) free and compulsory. While the current (gross) enrollment rate for elementary education in India is over 90 %, this aggregate number masks some regional, religious, caste and gender based diversity, with fewer female, Muslim and lower caste children enrolled in school.[9]

The last four standards of school education are known as secondary education. Before entering 11^{th} standard the child has to pass a central or state level administered exam. From the 11^{th} standard onwards, the student chooses three to four subjects in which he or she will specialize. The student completes his or her school education by taking another central or state level administered exam. Gross enrollment figures in India range from about 70 % for lower secondary education to 40 % for higher secondary education.[10]

After 12^{th} standard, there are several options for a student to continue his or her education in colleges, universities and training institutions. One can enroll in a two year diploma course to become, for instance, a teacher or a textile designer. One can opt for technical training at an I.T.I. (Industrial Training Institutions), a two diploma year course, after which one can become electrician, mechanic, painter, welder, etc. Or one can enroll in a three year degree program for a bachelor's degree in sciences, commerce, or arts. A few degrees take four to five years, such as an engineering, law or medicine degree. After finishing a bachelor's degree, a student can opt for a master's degree. In India, about 12 % of the relevant age cohort attend higher education, again with a large, but

[9]See the work of Geeta Kingdon and Anjini Kochar in Basu and Maertens (2010) for an overview of India's school education.

[10]See Geeta Kingdon in Basu and Maertens (2010).

decreasing gap between lower and upper castes and between states. Women make up about 40 % of all enrollments.[11]

Gender and caste affect the cost and returns to education through a complex system of quotas (also called "reservations") in the public sector and government funded educational institutions (for higher education), and caste and gender based subsidies.[12] After independence, the Government of India recognized that the (formerly called) *untouchables* and (aboriginal) tribal groups were on average economically worse off compared to the Hindu majority. The administrative categories of Scheduled Castes (SC) and Scheduled Tribes (ST) were created to refer to these *untouchables* and tribal groups, respectively, and a system of positive discrimination was set up. Today, quotas and subsidies are extended to women, girls, and members of other lower caste groups and religions. The latter two are grouped in a new category termed Other Backward Castes (OBC). The remaining caste groups are referred to as the Upper Castes (UC).

3.3 Data

I collected the data used for this study in 2007-08 in three villages in India. These three villages were selected in 1975 by the International Crop Research Institute

[11]For an overview of India's higher education see Devesh Kapur and Pratab Bhanu Mehta in Basu and Maertens (2010).

[12]For instance, in the three villages I study, free text books are provided to all the children from 6^{th} to 10^{th} standard belonging to Other Backward Castes, Scheduled Castes and Scheduled Tribes. Then, 2 out of 9 Other Backward Caste individuals and 1 out of 1 Scheduled Caste individual (under the age of 26) who pursued higher education obtained a higher education position through the reservation system.

of the Semi-Arid Tropics (ICRISAT) as part of their Village Level Studies (VLS) program. This ongoing program collects detailed household and plot level agricultural data among a sample of households in six villages in semi-arid India.[13]

To obtain information on social networks, expectations and aspirations with regard to education, I resurveyed 339 ICRISAT-VLS households in Dokur, Kalman and Shirapur. In this survey I also elicited, for each individual up to the age of 25, details about their education and activities since age six[14], and included questions on household composition, income, time preferences[15], and current activities and education obtained for household members older than 25 years. I also completed a village questionnaire and school questionnaires among the top five most attended schools in each village with the assistance of the ICRISAT-VLS investigators, the school principals and the *sarpanch.*[16] This included information on village infrastructure, educational programs in the village, the direct cost of education, school infrastructure, facilities and instruction.

Table 3.1 introduces the three villages. Dokur, with 530 households, is the smallest of the three villages. It is located in the drought prone, poor, Telangana

[13]The sample selected is representative for each village in terms of landholding size. For an overview of the ICRISAT-VLS program see Bantilan et al. (2006), Rao and Charyulu (2007), Singh et al. (1985) and Walker and Ryan (1990).

[14]I did not elicit detailed education and employment information for the daughters (and sons)-in-law up to the age of 25 living in the household as they did not receive their education within the household.

[15]As time preferences might be identity dependent and correlated with social norms, one needs to control for them. Implementing a traditional experiment to elicit long-term discount rates appeared infeasible. As such, I followed Loewenstein et al. (2001) to get an estimate of the various dimensions of time preferences (i) impulsivity, (ii) planning/compulsivity, and (iii) behavioral/emotional inhibition. See also Pender (1996) for estimates of traditional discount rates in the ICRISAT villages.

[16]A *sarpanch* is a democratically elected head of a village level statutory institution of local self-Government called the *Gram* (village) *Panchayat* in India.

region of Andhra Pradesh. In terms of income[17], Dokur is situated between the poorer Kalman and richer Shirapur. Kalman and Shirapur, with, respectively, 610 and 580 households, are located in the Solapur district of southwest Maharashtra. At present, Dokur has one public school which offers education up to 10^{th} standard. Kalman and Shirapur have two public schools each offering, respectively, up to 12^{th} and 10^{th} standard (Table 3.2).

Table 3.1: Introducing Dokur, Kalman and Shirapur

	Dokur	Kalman	Shirapur
Distance to nearest town (km)	7	15	12
Number of households in village	530	610	580
Number of households in sample	93	102	144
Number of adults in the sample[1]	294	338	409
Number of young adults in the sample[1]	106	124	145
Number of children in the sample[1]	120	151	189
Number of UC individuals[2]	80	352	430
Number of OBC indivduals[2]	417	128	53
Number of SC individuals[2]	23	32	118
Number of T individuals[2]	0	101	142
Average number of household members	5.59	6.00	5.19
Average *Kharif* income (Rs)[3]	44524	40713	62965
Average education level of respondent (in years)[4]	3.60	4.66	5.76

Notes: [1] An adult is defined as an individual over the age of 25, a young adult is defined as an individual between the ages of 15 and 25 and a child is defined as an individual under the age of 15. Note that the adults also include the daughters-in-law under 25 who did not received their education in the household; [2] FC=Forward Castes (open category), OBC=Other Backward Castes, SC=Scheduled Castes, T=tribes, including Scheduled Tribes; [3] the Kharif season is the rainy season; [4] the respondent is the main decision-maker with regard the education of the children under 25 year in the household .

Table 3.2: Schooling and education

	Dokur	Kalman	Shirapur
Number of schools located in the village	1	2	2
Highest grade offered in the village	10	12	10
Children enrolled in school (%)[1]	90	96	91
Young adults enrolled in school (%)[1]	30	36	19

Notes: [1]See definition in Table 1; as a percentage of all individuals within the relevant age group; children under the age of 6 were excluded from this calculation.

[17]Table 3.1 reports *Kharif* income. The *Kharif* season is the rainy season, and the main agricultural season in the semi-arid tropics of India.

The ICRISAT-VLS sample includes 93, 102 and 144 households in Dokur, Kalman and Shirapur, respectively, a total of 1876 individuals. The average size of a household is between 5 and 6 members in all three villages. In Dokur, 80 % of the individuals are classified as OBC and 15 % as UC.[18] In the Maharashtra villages, 60 % of the individuals belong to UC, and between 16 and 19 % of the individuals are members of the tribal category.[19]

Only one individual was interviewed in each household, the main decision maker with regard to the education of the individuals up to the age 25 years. In many cases, this is the father of the children, in some cases, the mother, grandfather or uncle. In the remainder of this article I will refer to this decision maker as the "parent" of the child.

The average education level of the respondent is low, especially in Dokur (3.6 years). Figure 3.1 shows the distribution of the highest level of education obtained for the individuals who have finished their education.[20] About 36% have not completed first standard and only 6% have done at least one year of higher education. The large majority of the individuals received some school

[18]In Andhra Pradesh, the following categories are recognized of socially and educationally backward classes: the SCs, the STs, Group A (other tribes), Group B (vocational groups), Group C (SC who converted to christianity), Group D (others), and Group E (Muslims). For the purpose of inter-state comparison, I categorised Group C under SC and Group D and the other groups under OBC. I followed the classification as reported by the respondent.

[19]In Maharashtra, the following categories are recognized of backward classes: (1) SC and SC converts to Buddhism, (2) ST including those living outside specified areas, (3) Vimukta Jati & Denotified Tribes, (4) Nomadic Tribes– 1, (5) Nomadic Tribes, (5) Nomadic Tribes– 3, (6) Other Backward Classes. For the purpose of inter-state comparison, I categorized, Group (6) under OBC, Group (1) under SC and the remainder of the groups under T, referring to Tribes. I followed the classification as reported by the respondent.

[20]There are a handful of individuals who have stopped their education but have intentions to re-enroll in the future. These individuals are included in Figure 3.1.

education. It is interesting to note that relatively few individuals are located at the natural stopping points: 8^{th}, 10^{th}, 12^{th} standard, etc.

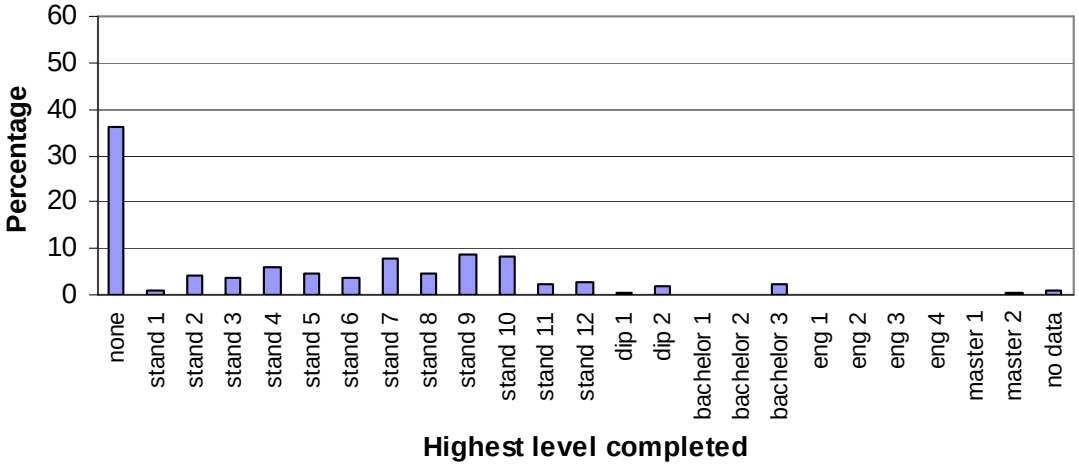

Figure 3.1: Highest level of education completed for individuals who finished their education, excluding children under the age of six, N=1329

The data collection focused on obtaining information on social networks, aspirations and expectations with regard to education and the education obtained by individuals up to the age of 25 years - as the existing ICRISAT-VLS data lacks such information. Of these, 556 individuals are currently enrolled in an educational institute or under the age of six (and, as such, planning to go to school).

The educational aspirations with regard to each child (currently enrolled in school or planning to go to school) were established as answers to the questions: "What is the minimum amount of education you want this particular child to

obtain?" and "What is the maximum amount of education you would allow this particular child to complete?" [21]

One would expect the plans to invest in a child's education to depend on (among other) the expectations the decision maker has with regard to the returns to education. As such, I elicited current beliefs regarding the returns to education, conditional on the child's abilities and other characteristics, but unconditional on the nature of the employment. To obtain a density function of future earnings for each education level and for each child, I used a method inspired by Dominitz and Manski (1996) and Lybbert et al. (2007) [22]

Concretely, I first elicited the minimum and maximum earnings the respondent thought the child would earn when finishing particular schooling milestones, for instance, 12^{th} standard. During this exercise, the respondent was asked to imagine the various options possible, i.e., various types of employment, including self employment, and various locations where the child might live in the future, anticipating migration. Then, I made three boxes, evenly distributed between this minimum and maximum and I asked the respondent to use 20 stones (each stone representing a 5% probability) to form a earnings density function. This question was repeated for the various levels that the child

[21] Note that both questions elicit investment plans with regard to education, and even though they might suffer from measurement error, these measurement errors will not pose any problems as long they are not correlated with any of the other independent variables in the econometric models (see also Bertrand and Mullainathan, 2001). In some cases, the mimimum and maximum are identical, but in many cases the decision maker anticipates receiving more information with regard to the child's ability and the future financial situation of the household, and the minimum and maximum represent, respectively, the worst and best case scenario in the decision maker's mind.

[22] For an overview on how one can elicit subjective expectations see Delavande et al. (2008) and Delavande et al. (forthcoming).

still had ahead of him/her, and could include, 8^{th} standard, 10^{th} standard, 12^{th} standard, diploma, bachelor's, engineering, medical doctor, master's.[23]

In addition to these future earnings, the parents take into account the financial and physical care expected from each child at old age and other monetary aspects of each level of education such as the direct and opportunity cost and the expected dowry. As such, I included an extensive dowry section for the children who were married, a section on the expected direct cost of the various levels of education (in addition to the cost information in the school questionnaires), and asked the decision maker how much he or she expects each child to contribute to their "old age pension".[24]

Finally, I elicited the (perceived) ideal age of marriage, and the education of the members in the parent's social network as both might influence the aspirations through non-monetary social pressures.[25] I used two different methods to elicit the relevant social network. The first method asks the respondent how many people he knows who completed 12^{th} standard, a diploma, a bachelor's, etc., in different social groups (total, relatives,village) and more position-based questions (inspired by Lin, 1999), for instance "How many people do you know who work as a civil servant?"[26] Even though this network might correctly repre-

[23]So for a child currently enrolled in 11^{th} standard, one was asked to reconstruct the density function for 12^{th} standard, diploma, bachelors, engineering, medical doctor, masters, but not for 8^{th} or 10^{th}standard.

[24]As most people in rural areas do not suddenly stop working, but rather gradually reduce the number of hours they work on the farm, in the business and in the household, pre and post retirement remittances were not explicitly distinguished.

[25]The survey includes some questions on the reaction of various peer groups when enrolling the child in a particular standard or removing the child from school.

[26]This question was repeated for a selection of occupations: medical doctor, teacher, engineer, technician and information technology specialist.

sent the information network of the parents, it is likely that some of these distant connections might not be relevant in terms of social pressures.

So the second method asks the respondent to think about his immediate relatives: his family, his parents, his wife's parents, his brothers and sisters and their children and his wife's brothers and sisters and their children. For each of these family members, I elicited the education level, approximate age when education was completed, current occupation and income.

3.4 A first look at the data

The sample includes 835 individuals up to the age of 25 years (Table 3.1). Of the children between the ages of 6 and 15 years, over 90 % are enrolled in school in all three villages (Table 3.2). Of the young adults, between the ages of 15 and 25 years, respectively, 30 %, 36 % and 19 % are enrolled in an educational institute, in, Dokur, Kalman and Shirapur.

There is little gender difference in enrollment up to the age of 15 years, but compared to male young adults, fewer female young adults are enrolled in an educational institute (30 % versus 24 %). In terms of caste, UC children and young adults are, respectively, up to 10 and 20 % more likely to be enrolled in an educational institute compared to other castes.[27]

[27]Detailed descriptive statistics by caste and gender are available on request. Naturally, enrollment does not imply regular attendance. Of all the children and young adults enrolled in an educational institute, 87% does not miss classes more than 10% of the time. Again there is some variation between caste groups and gender, with girls being less likely to miss classes and members of lower caste groups being more likely to miss classes.

For the individuals currently enrolled in an educational institute, Figure 3.2 and Figure 3.3 show, respectively, the distribution of the minimum amount of education the respondent wants this individual to obtain and the maximum amount of education allowed to complete. A first thing to note, comparing Figure 3.2 and Figure 3.3 with Figure 3.1, is the discontinuous nature of these educational plans. While the parents might not plan on stopping their child's education before a program or level is completed, in practice many children do drop out. The (main) reasons mentioned by the parents are: sudden financial problems, discovering the child is of "low ability" (often after having failed) and unexpected marriage prospects.[28] Inspection of the data reveals that drop out is often preceded by failure and low attendance.[29]

The second difference between these two sets of figures, the lower average level of education of Figure 3.1 compared to Figure 3.2 and Figure 3.3, might point to an age cohort effect (due to limited access to schools, credit and income constraints, limited employment opportunities etc. a generation ago) and a migration effect (highly educated individuals are likely to migrate to the cities). This mismatch between aspirations for one's children and what one has achieved oneself is, in effect, a large driver of intergenerational growth at the micro level.

[28]Relatedly, Holla (2007) finds that income volatility, as measured by rainfall shocks, disproportionately affects girls' school drop out.

[29]When asked about attendance, Dokur respondents and lower caste respondents often felt that missing classes to help in the household or on the farm is 'sometimes acceptable'. The difference observed with the Maharashtra villages and higher caste groups could be purely a wealth effect though, as no difference can be found between the sexes, i.e., conditional on being enrolled in an educational institute it is equally acceptable for girls as for boys to miss classes because of illness, help needed in the household help, help needed in the farm, help needed with the preparation of a festival or to attend a festival. Detailed results on attendance are available on request.

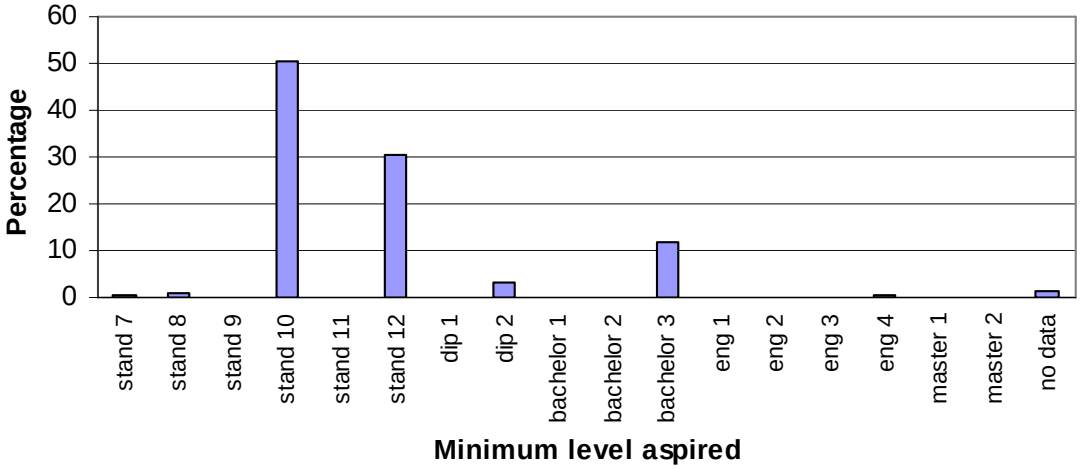

Figure 3.2: Minimum level of education aspired for children who are currently in school, including children under the age of six, N=557

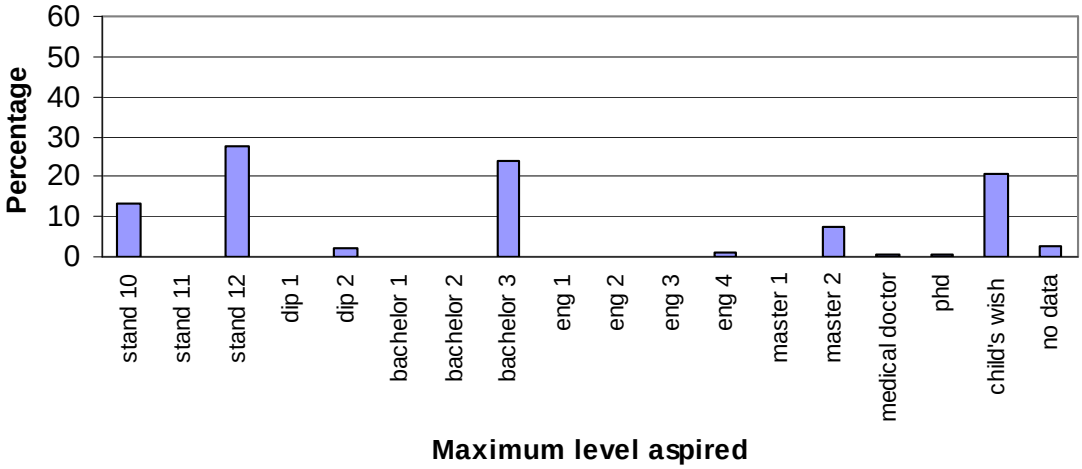

Figure 3.3: Maximum level of education aspired for children who are currently in school, including children under the age of six, N=557

About 80% of the individuals currently enrolled are expected to complete a minimum of 10 or 12 years of education (Figure 3.2) and 45% are allowed to continue for a bachelor's degree or "as long the child wants" (Figure 3.3). These

aggregate numbers however conceal a substantial amount of regional, gender and caste based variation (Table 3.3).

Table 3.3: Minimum and maximum education planned

[in percentage of individuals currently enrolled in school]

By village:	Dokur		Kalman		Shirapur	
	Min	Max	Min	Max	Min	Max
10th or lower	66	32	44	4	50	9
12th	25	20	34	19	30	40
higher than 12th	7	47	20	73	19	49

By gender:	Male		Female	
	Min	Max	Min	Max
10th or lower	40	7	68	21
12th	37	20	22	38
higher than 12th	22	71	8	39

By caste group:	UC		OBC		SC/T	
	Min	Max	Min	Max	Min	Max
10th or lower	36	3	69	28	64	15
12th	38	24	22	24	24	39
higher than 12th	26	72	9	44	8	42

Notes: Totals do not add up to 100% due to a small fraction of people who answered "don't know" or "it depends"; N=557.

In the Maharashtra villages, about 20 % of the individuals are expected to complete a higher education degree; compared to 7 % in Dokur. Consistent with the standards available within the village boundaries, more individuals are expected to complete (at a minimum) 12^{th} standard in Kalman compared to Shirapur. The difference between boys and girls is striking. The large majority of the girls is expected to complete 10^{th} or 12^{th} standard (at a minimum) and only 39 % would be allowed to pursue higher education, compared to 71 % of the boys. A similar difference can be found between UC, OBC and SC/T: 70 %

of the UC individuals are allowed to continue higher education, compared to 42 to 44 % of the backward caste groups.

In qualitative terms, the respondents mention a ”high salary” or a ”government job” as the most important characteristic of the future occupation for about 60 % of the children and young adults. While in Dokur respondents have a strong preference for government jobs, overall, there are few differences between the villages in this regard. Similarly, there are few differences between boys and girls with regard to the ideal characteristics of the future occupation, even though in the case of girls the most important characteristic mentioned is sometimes ”(geographically) close to the relatives”, ”a job through which one helps others” and ”a job which is approved by the spouse”, characteristics which are (almost) never mentioned in the case of boys. In addition to these non-monetary criteria mentioned (”(geographically) close to the relatives”, ”a job through which one helps others”), the UC also consider the ”career potential” of an occupation to be important and SC/T respondents give relatively more importance to a ”stable income” compared to the other castes.[30]

The difference found between villages, gender and castes in terms of aspirations might be (partially) explained by a difference in expectations with regard to the returns to education. Table 3.4 shows that, on average, respondents in Dokur believe that medical doctors, engineers, and ”postgraduates” (i.e., someone with a masters degree) earn, respectively, 31,035 Rs/month, 24,576 Rs/month and 17,634 Rs/month (40,000 Rs is about 1,000 US Dollar). This is significantly higher than to the Maharashtra villages. While this difference is

[30]Detailed descriptive statistics are available on request.

reversed for the bachelors level (i.e., for a "graduate"), it is clear that expectations alone cannot explain the difference in educational plans observed between the villages.[31]

With regard to gender, girls are expected to earn less on a monthly basis, compared to boys, independent of the level of education. Using a simple (unequal variance) *t*-test it appears that this difference is significant for all education levels at the 5 % level, except for 10^{th} standard and the master's level. This is consistent with the difference in educational plans found. Even if there were no systematic difference in perceived ability between girls and boys, this does not imply that the parents anticipate wage discrimination for certain jobs, as these expectations also capture the fact that girls might be expected to do different kinds of work.

The difference in expectations between castes is again somewhat puzzling and inconsistent with the differences in aspirations education. UC respondents, on average, expect a lower return to higher education compared to OBC respondents, but a higher return compared to SC/T respondents. The difference between OBC and UC is significant at the 5 % level for all education levels except for 8^{th} standard, 10^{th} standard and a bachelor's degree. The difference between UC and SC/T is significant for all education levels except for the bachelor's level. Part of the explanation might be the caste based quota system in higher education and government institutions. According to their own account, it is

[31]There are 121 children, mostly in Dokur, for whom the respondent could not answer one or more of the "stone game" questions (to elicit their expectations with regard to the returns to education). These respondents answered "no idea". Comparing these respondents with the other respondents in Dokur, it appears that they are (on average) somewhat lower educated, older and have younger children. Their aspirations for their children (minimum and maximum) are lower compared to the other children in Dokur. These children are omitted from the analysis.

Table 3.4: Expectations with regard to earnings

[in Rs/month]

By village:	Dokur		Kalman		Shirapur	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
8th standard	1,706	724	2,256	2,028	1,422	570
10th standard	3,650	5,108	3,036	1,970	2,082	837
12th standard	3,870	1,366	4,382	3,171	2,924	1,184
diploma	8,953	3,968	8,756	5,174	6,403	1,941
bachelor's	8,063	2,606	9,991	5,621	9,383	3,099
medical doctor	31,035	8,545	22,909	8,511	19,745	8,441
engineer	24,576	7,624	17,904	6,806	15,332	6,565
master's	17,634	13,339	14,324	9,148	12,471	4,725

By gender:	Total		Male		Female	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
8th standard	1,787	1,359	1,921	1,277	1,616	1,436
10th standard	2,793	2,853	2,886	2,496	2,679	3,241
12th standard	3,689	2,276	3,945	2,273	3,370	2,245
diploma	7,794	4,074	8,262	4,355	7,187	3,601
bachelor's	9,402	4,313	10,016	4,652	8,620	3,708
medical doctor	22,496	9,150	23,559	8,810	21,146	9,421
engineer	17,651	7,400	19,080	7,643	15,808	6,658
master's	13,892	8,306	14,238	7,510	13,404	9,242

By caste group:	UC		OBC		SC/T	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
8th standard	1,887	1,753	1,816	889	1,550	785
10th standard	3,069	3,856	2,786	1,360	2,229	1,015
12th standard	3,652	2,337	4,158	2,562	3,212	1,579
diploma	7,773	3,926	9,387	5,141	6,411	2,598
bachelor's	9,309	4,060	10,192	5,586	8,876	3,362
medical doctor	23,076	8,432	26,164	10,464	18,302	8,242
engineer	17,903	6,808	20,457	8,698	14,839	6,773
master's	13,747	6,918	17,893	13,163	11,203	4,913

Notes: For each individual currently enrolled in an educational institute, the expected average monthly earnings is calculated assuming a step-wise distribution with the minimum and maximum of the distribution of earnings as specified by the respondent for each level/program of education completed. These earnings are equivalent to the wage in case of wage employment and are earnings net of costs in the case of self employment. The numbers presented in this Table are the sample averages of these individual averages.

becoming more and more difficult for UC to get admitted to the better colleges and universities, or to secure a government job.

In addition to expectations, social norms and pressures could also influence the planned education levels. Direct elicitation of social pressures with regard to enrollment and drop out reveals little information. When asked how various social groups reacted when a child was enrolled in school, or when a child dropped out of school, very few respondents mention extreme reactions such as ”very negative”, or ”very positive”.

The education decision could also be influenced indirectly by social norms regarding the ”ideal age of marriage”. According to the respondents, a family who has an unmarried older daughter at home is often criticized, the various members’ every move being scrutinized, often compromising the marriage prospects of the other children and resulting in social exclusion of the adult household members. As such, it’s in the parents’ interest to marry their daughter around the ”ideal age of marriage”.

The (average) ideal age of marriage is 18.3 years in the case of girls and 22.7 years in the case of boys. This is a bit higher for both sexes in Dokur compared to the Maharashtra village, and a bit lower for SC/T compared to the other castes.

When a couple gets married, the two families exchange gifts and money, in addition to splitting the cost of the wedding ceremony. The value of these gifts, as well as the financial transfer, is set by both parties through a bargaining process. The final amounts set will depend on the education of the children, their current and future jobs, their age and caste and personal characteristics, such as beauty. Table 3.5 presents the average value of the transfers. On average, the bride’s family spends a total of 87,293 Rs on gifts, dowry and ceremonial ex-

penses (net of the gifts received) and the groom's family spends a total of 3,887 Rs on gifts, ceremonial expenses (net of the dowry and gifts received). Wedding expenses are markedly higher in the Maharashtra villages compare to Dokur and among the UC compared to the OBC and SC/T. According the respondents' own account, the relationship between education and dowry is complex. While a bride's family has to pay more for a more educated groom, a highly educated groom typically does not marry a bride who received little education. And as increasing the education of the bride could lower the dowry, as the bride could financially contribute to the family-in-law's household, the overall effect of increasing the bride's education on the dowry is theoretically ambiguous.[32] In practice, it appears that increasing the education of groom increases the dowry, but only if the groom lives with his parents, and increasing the education of the bride reduces the dowry, but only if the groom lives separately from his parents.[33]

Table 3.5: Transfers at the time of marriage

[Averages, in Rs paid by the respective families]

	Total	Dokur	Kalman	Shirapur	UC	OBC	SC/T
By bride's family	97,293	73,220	128,611	92,459	132,857	82,385	79,000
	(81,587)	(41,122)	(128,611)	(76,088)	(120,820)	(51,205)	(43,646)
By groom's family	3,887	-27,570	5,986	29,336	17,414	-20,105	24,984
	(41,261)	(25,357)	(5,986)	(38,270)	(35,743)	(30,839)	(43,378)

Notes: The standard deviatiations are mentioned under the averages in paranthesis. The bride's family offer gifts and presents to the groom's family and vice versa, both families transfer gifts and cash to the bride and groom and both families contribute to the expenses of the wedding ceremony. The amounts shown in Table 5 are the net amounts paid by the respective families (assuming that the bride keeps her gifts to herself and the groom shares his with his family); N=115.

[32] In addition, education a daughter might have other benefits of the marriage market: for a highly educated daughter, the family does not have to bother looking for candidates themselves, but candidates will present themselves.

[33] The regression included all observations in Table 3.5 and controlled for income, the difference in wealth between the two families, whether the bride was joining a joint or nuclear household and caste and village fixed effects. The regressions results reported are available on request.

Table 3.6 presents the financial support and physical care parents expect from their children in old age. With the exception of a handful of respondents who have access to savings or an old age pension, most respondents reported that they would have to rely 100% on their children. The patrilocal marriage system, as practised in the three villages, in which married couples reside with or near the family of the husband, imply that the (majority) of the returns to education after marriage go to the family in law. Table 3.6 shows that, indeed, in all villages and all caste groups, the majority of the care and support is expected from the male children. The difference between support expected from male versus female children is larger in the case of UC and smaller in Dokur. The latter might be due to the phenomenon of cousin marriage in Andhra Pradesh, making the future earnings of daughters more "accessible" to parents as she stays within the family.[34]

Table 3.6: Support expected from children at old age

[Averages, in percentage support expected]

	Total	Dokur	Kalman	Shirapur	UC	OBC	SC/T
Financial support							
boys	76	70	74	81	80	72	75
girls	36	44	31	35	32	44	33
Physical care							
boys	59	56	63	58	61	59	54
girls	58	59	46	65	58	60	56

Notes: The numbers presented in Table 6 are the sample averages of the percentage care/support expected from male/female children. Note that at a household level, these two percentages add up to 100% in most cases (except for the case where the respondent has access to savings or a pension); N=273 boys and N=238 girls.

[34] In south India it is common for Hindu cross cousins to marry, with matrilateral cross-cousin (mother's brother's daughter) marriages being especially favored. This southern kinship model prevails in the states of Kerala, Tamil Nadu, Karnataka, and Andhra Pradesh. On the cost and benefits of cousin marriages, see the ongoing work of Seema Jayachandran, Mohammed Al Shafaee and Erica Field.

To conclude this section, Table 3.7 presents selected descriptive statistics on the education attained by the members of the parents' social network. One can see that, on average, only 9.3% of the male relatives over the age of 18 (in the extended family) did some higher education, compared to 3.6% of the female relatives. Once again, these numbers are higher for the Maharashtra villages compared to Dokur, and for UC compared to the backward castes.

Table 3.7: The education obtained by the social network

[Averages, in percentage of extended family members who did (some) higher education]

	Total	Dokur	Kalman	Shirapur	UC	OBC	SC/T
Men	9.3	6.3	11.3	9.8	12.9	6.5	5.9
Women	3.6	1.7	4.1	4.4	4.5	2.3	3.3

Notes: Based on the education level of all members of the extended family older than 18 years of age (of the same gender) as reported by the decision maker.

3.5 Empirical specification

I abstract from the details of the dynamic optimization problem of the parent and assume that the parent receives an expected (lifetime) utility $u(e_{ij})$ from educating child i up to education level e_j. Using an additive random utility framework, write:

$$u(e_{ij}) = f_j(r_{ij}, c_{ij,}dowry_{ij}, contr_{ij}, N_j, age\_marr, sex_i, caste, H) + \epsilon_{ij} \qquad (3.1)$$

where r_{ij} denotes the (perceived) monthly earnings of child i having completed education level e_j, c_{ij}, the direct cost of education level e_j for child i, $dowry_{ij}$, the dowry to be expected (which is typically negative for girls and positive for boys), $contr_i$, the contribution expected from the child at old age, N_j the share

of (male/female) relatives who have obtained education level e_j, *age_marr*, the ideal age of marriage according to the parent, and household specific variables denoted H. These include education and age of the decision maker, wealth, and time preferences.

Sex_i, denotes the gender of the child and *caste* denotes the caste group of the family. Note that the opportunity cost of education level e_j are largely captured by the (perceived) monthly earnings of the education level e_{j-1}. What is not being captured though are the forgone value of non-market labor activities such as household work (note that farm work and other household business' work are included in r_{ij}). One would expect this value to depend on the age and gender of the child, and the demand for household work within the household. As such, the coefficient on sex_i partially captures the difference in the forgone value of non-market activities of girls versus boys. In addition, the coefficient on sex_i might also capture the fact that parents tend to be more concerned about the personal safety of a daughter compared to a son. Sending one's daughter off to dubious hostels or on the road to travel to college without a suitable chaperone is a source of concern to parents in addition to a source of social pressures. Note that the latter would (partially) be captured by the coefficient on N_j.

The coefficient on *caste* captures any remaining caste-based aspirations unexplained through the dowry, age of marriage and pratrilocal channel as well as unrelated to the education attained by the members of the extended family. A part of this unexplained caste based variation might be due to a preference for caste specific occupations for one's children, such as, shepherd, blacksmith, barber and toddy tapper. Note that child order is not explicitly taken into ac-

count in this model. The only way a child can affect another child of the same household is through the income and wealth channel.

The parent opts for education level e_j for child i if education level e_j gives the highest utility among the set of education levels available:

$$P(\text{choose } e_j \text{ for child } i) = P(f_j(\cdot) + \epsilon_{ij} > f_s(\cdot) + \epsilon_{is}; \forall s \in B_i) \tag{3.2}$$

$$P(\text{choose } e_j \text{ for child } i) = P(f_j(\cdot) - f_s(\cdot) > \epsilon_{is} - \epsilon_{ij}; \forall s \in B_i) \tag{3.3}$$

which implies B_i is the set of education levels available to child i. Technically, this set is the budget set of child i, but as I have no information on credit constraints, I have to assume that this is the set of all education levels ahead. For instance, for a child enrolled in 12^{th} standard, the set B_i equals $\Big\{$diploma, bachelor's, engineering, medical doctor, master's$\Big\}$.

Assume now that all ϵ_{is} are mutually independent with a log Weibull distribution. In this case, the distribution of each ϵ_{is} is given by:

$$F(t) = \exp(-e^{-t}) \tag{3.4}$$

Under these assumptions, following McFadden (1974), it can be shown that:

$$P(\text{choose } e_j \text{ for child } i) = \frac{\exp(f_j(\cdot))}{\sum_{s \in B_i} \exp(f_s(\cdot))} \tag{3.5}$$

Finally, assuming that $f_s(\cdot)$ is a linear function of the alternative specific variables (denoted X_{ij}) and child/household specific variables (denoted Z_i), let's rewrite (3.5) as:

$$P(\text{choose } e_j \text{ for child } i) = \frac{e^{(X'_{ij} * \beta + Z'_i * \gamma_j)}}{\sum_{s \in B_i} e^{(X'_{ij} * \beta + Z'_i * \gamma_j)}} \tag{3.6}$$

Specification (**??**) together with assumption (3.4) outline a conditional logit model. While this model allows for sufficient education level specific flexibility, it does not impose an order among the various education levels available (such as between and engineering degree and a medical doctor degree). The assumption that all ϵ_{is} are mutually independent implies that the utility levels (conditional on observed characteristics) of any two education levels are independent. So the probability ratio, or odds ratio, between any two alternatives, does not depend on the other alternatives available:

$$\frac{P(\text{choose } e_j \text{ for child } i)}{P(\text{choose } e_k \text{ for child } i)} = \frac{e^{(X'_{ij}*\beta+Z'_i*\gamma_j)}}{e^{(X'_{is}*\beta+Z'_i*\gamma_s)}} \tag{3.7}$$

While it is possible to relax this Independence of Irrelevant Alternatives (IIA) property, this would lead to a more complex model, with more coefficients to be estimated. As, according to their own account, the parents decide on their plans to invest in education in a simultaneous manner, rather than a sequential manner (first decide on whether or not one pursue higher education and then on which type of higher education), it is not clear what the advantages of a more flexible approach would be given that I have only data on a small sample of children and young adults.

3.6 Results

The basic intuition behind the identification strategy is simple. I check whether the beliefs with regard to the returns to education are consistent with the aspirations with regard to education. For instance, if someone believes that obtaining a bachelor's degree could yield a return that more than compensates for the di-

rect and indirect costs involved compared to just finishing high school, then this person should plan to invest in higher education. The discrepancy between beliefs and plans could be explained by a number of factors, among others, credit constraints[35], time preferences and social norms and pressures.

This implies that one could isolate the roles of social norms and pressures by decomposing the variance in educational plans into a part explained by the variance in costs, returns, credit constraints and impatience and an unexplained part which could be attributed to social norms and pressures. Using detailed child level information on the various sources of these social norms and pressures in rural India, I can separately identify the effects of social norms through dowry, expectations with regard to old age care and with regard to the ideal age of marriage. As what is "acceptable" depends on the child's caste and gender, in addition to household characteristics, there is sufficient variation to identify these social norm effects, if present. To check whether there are social pressures present which are directly related to the level of education itself, I include information on the education level of the household's social network members.

The analysis presented in this section takes into account the fact that not all children and young adults are enrolled in school (see Table 3.2). Figure 3.4 shows the percentage of children and young adults enrolled in an educational institute by age group. One can see that aggregate enrollment decreases by age. While 93% of the 6 to 14 years old are enrolled, only 37% of the 17 to

[35]The direct costs of education can be substantial, ranging from 17% of the average annual income per capita for elementary education to over 80% for secondary school education, and over 100% for higher education. While several respondents mentioned they would be willing to borrow money for the education of their sons and daughters, these type of loans are non-existent up to 12^{th} standard. Banks provide loans only for higher education and borrowing from relatives or friends to finance education is not socially acceptable.

18 years old are enrolled. If the children who are enrolled are systematically different from the children who are not enrolled, the estimates of specification (3.6) might suffer from selection bias and using them to say something about the population, i.e., all individuals up to 25 years, would be incorrect. For instance, imagine that less able children are more likely to drop out of school and that ability affects aspirations as more able children are more likely to pass entrance examinations and finish a degree that they started. In this case, the effects of expectations on the educational plans will be biased upwards.

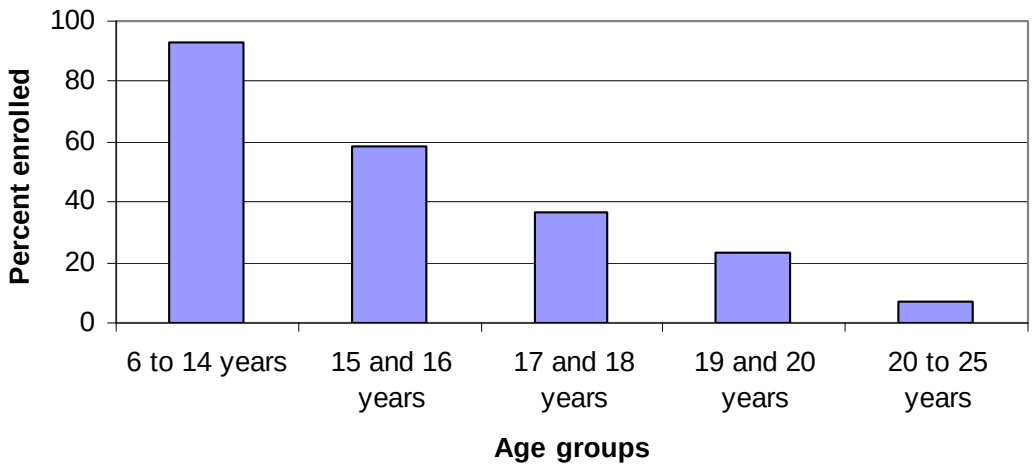

Figure 3.4: Percentage of individuals enrolled in an educational institute, N=683

As the data contain more information than usual with regard to the education decision, including several measures of ability[36], and my information set is therefore very similar to the decision maker's information set, I do not expect this selection bias to be severe. Nevertheless, I used a traditional Heckman

[36]Even though 'true' innate ability is a somewhat of an elusive concept, I can (partially) control for ability using standardized test scores, school grades and the subjective ranking of the children's ability by the parents.

(1979) approach to deal with issues of sample selection . Briefly, whether or not a child is included in the sample mainly depends (non-linearly) on his or her age, gender, caste and village.[37]

Table 3.8 and Table 3.9 present the results of a conditional logit model, estimating (3.6) using, respectively, the minimum and maximum education aspired. From the first section of the table one can see that the minimum education aspired depends on the average and variance of the (perceived) returns to education, the cost of education and the dowry expected. Notice however that the sign on the dowry effect is the opposite of what one would expect: increasing the price of education in terms of the dowry costs, increases the chances of planning to invest in that option. Increasing the expected dowry of a specific education level with 1000 Rs, increases the relative odds ratio of choosing that particular education level, rather than 12^{th} standard, by a factor of 1.02. This might be due to the fact that this dowry includes both the gifts and transfers as well as the wedding expenses, and the latter, being a source of social status, might actual imply that more expenses might be preferred to less expenses. The ideal age of marriage appears to be an important determinant of the (minimum) aspirations. Increasing the age of marriage increases the chances of the child obtaining (at least) a bachelor's degree and decreases the chance of the child ending up with just 10 years of education (at a minimum). The magnitude of this effect is substantial. Increasing the age of marriage with one year, changes the relative odds of aspiring 10^{th} standard, rather than 12^{th} standard, by a factor of 0.86, and the relative odds of aspiring a bachelor's degree, rather

[37]Measures of ability and past attendance are also statistically significant when considering only the children who have ever been to school (results available on request).

than 12^{th} standard, by a factor of 1.30. Gender does not affect the relative odds ratio of aspiring a bachelors' degree rather than 12^{th} standard, but does have an impact on the relative odds ratio of aspiring a diploma or only 10^{th} standard, rather than 12^{the} standard, with lower aspirations for girls. Caste also affects the likelihood of aspiring for 10^{th} standard, rather than 12^{th} standard, with lower aspirations associated with lower castes.

From the first section of the results in Table 3.9 one can see that maximum education level aspired depends on the education of the extended family members, but not in the way I expected. Increasing the percentage of family members who have obtained a particular level decreases the relative odds ratio of aspiring for that particular level, rather than 12^{th} standard, by a factor of 0.97. Thus I find no evidence of direct social pressures to conform in terms of education. The expected contribution from the child at old age affects the maximum level of education aspired. Increasing the child's contribution with 1%, increases the relative odds ratio of choosing a bachelor's degree, rather than 12^{th} standard, by a factor of 1.03 and decreases the relative odds ratio of choosing 10^{th} standard, rather than 12^{th} standard, by a factor of 0.94. The effect is reversed (compared to the bachelor's level) when considering the other higher education options: engineering, medical doctor and master's. Maybe the parents fear that a too highly educated child might leave the region and break contact with the family, and as such they would only allow the children whom they do plan to rely as much on to continue for post-graduate education. Again, increasing the age of marriage increases the chances of the child to be allowed to continue for a bachelor's, master's and engineering degree. Increasing the age of marriage with one

year, changes the relative odds of aspiring a bachelor's degree, rather than 12^{th} standard, by a factor of 1.57, an engineering degree, rather than 12^{th} standard, by a factor of 1.21, and a master's degree, rather than 12^{th} standard by a factor of 1.61. Girls are less likely to be allowed to continue for any of the higher education degrees (except for a diploma) compared to boys, and children of lower caste groups are less likely to be allowed to continue for a master's degree and more likely to be allowed to go to school only up to 10^{th} standard compared to UC children.

Table 3.8: Determinants of aspirations - minimum

Conditional logit regression (base category 12th standard)

	Coefficient		Odds Ratio	
	Coeff.	St.Error	Coeff.	St. Error
Mean of perceived return (in 1,000 Rs)	0.2524***	0.0666	1.2871***	0.0858
Deviation of perceived return (in 1,000 Rs)	-0.7022**	0.3277	0.4955**	0.1624
Percentage of family members (in 1,000 Rs)	-0.005	0.0071	0.995	0.0071
Net contribution at time of marriage (in 1,000 Rs)	0.0199**	0.0088	1.0201**	0.009
Direct cost of education (in 1,000 Rs)	-0.1208***	0.0588	0.8862***	0.0521
Finish 10th standard				
Percentage of financial support expected of the child	-0.0192	0.0176	0.981	0.0172
Interacted with age parent	0.0005	0.0005	1.0005	0.0005
The ideal age of marriage for boys/girls	-0.1442*	0.0823	0.8657*	0.0713
Child is female	2.1273***	0.6614	8.3919***	5.5507
Belong to OBC	1.1275***	0.4266	3.0879***	1.3174
Belong to SC/ST	1.3064**	0.5398	3.6929**	1.9936
Finish a diploma				
Percentage of financial support expected of the child	-0.0232	0.0712	0.9771	0.0696
Interacted with age parent	0.0005	0.0017	1.0005	0.0017
The ideal age of marriage for boys/girls	-0.3464	0.244	0.7073	0.1725
Child is female	-2.889*	1.4392	0.0556*	0.0801
Belong to OBC	-47.031	41010	0	0
Belong to SC/ST	-17.9096	22169	0	0.0004
Finish a bachelor's degree				
Percentage of financial support expected of the child	-0.0088	0.0302	0.9912	0.03
Interacted with age parent	0.0006	0.0008	1.0006	0.0008
The ideal age of marriage for boys/girls	0.2682**	0.1337	1.3076**	0.1748
Child is female	-1.5541	1.1054	0.2114	0.2337
Belong to OBC	-0.6465	1.0125	0.5239	0.5304
Belong to SC/ST	-37.1967	32127	0	0
Finish an engineering degree				
Percentage of financial support expected of the child	-0.2598	1.23	0.7712	0.9486
Interacted with age parent	0.0059	0.0255	1.0059	0.0257
The ideal age of marriage for boys/girls	-0.6592	1.7161	0.5173	0.8876
Child is female	-37.9647	706	0	0
Belong to OBC	-127.2932	3966	0	0
Belong to SC/ST	19.9072	8.64E7	4E+9	4E+16

Notes: *** p<0.01; ** p<0.05; * p<0.1. Includes all children up to the age of 18 currently enrolled in school except for the ones with missing data. Controlled for wealth, credit constraints, time preferences and selection effects. The mean and averages are computed following the procedure outlined in Table 4, and then averaged across the various levels of, respectively, primary and secondary, and higher education. The net contributions at the time of marriage are predicted on the basis of recall data (on dowry, education of bride and groom, joint versus nuclear family status, village and caste) on these contributions from the same villages. Note that as the recall data contains few children who received higher education, these estimates might be imprecise at these higher education levels. The percentage of family members who did no or some higher education is based on the education level of all members of the extended family older than 18 years of age (of the same gender) as reported by the decision maker. N=423. Master's degree, medical doctor degree and 8th standard are included in the analysis but results not shown as the limited number of observations implies that one canno

Table 3.9: Determinants of aspirations - maximum

Conditional logit regression (base category 12th standard)

	Coefficient		Odds Ratio	
	Coeff.	St.Error	Coeff.	St. Error
Mean of perceived return (in 1,000 Rs)	0.0214	0.0149	1.0217	0.0152
Deviation of perceived return (in 1,000 Rs)	-0.1017	0.0904	0.9033	0.0816
Percentage of family members (in 1,000 Rs)	-0.0205**	0.0088	0.9797**	0.0086
Net contribution at time of marriage (in 1,000 Rs)	0.0051**	0.0026	1.0051**	0.0026
Direct cost of education (in 1,000 Rs)	-0.0085	0.006	0.9915	0.0059
Finish 10th standard				
Percentage of financial support expected of the child	-0.0544*	0.0324	0.947*	0.0307
Interacted with age parent	0.001	0.0009	1.001	0.0009
The ideal age of marriage for boys/girls	-0.0188	0.1436	0.9814	0.1409
Child is female	1.0722	0.8477	2.9218	2.4769
Belong to OBC	2.3469***	0.5852	10.4532***	6.1176
Belong to SC/ST	1.055	0.6659	2.8719	1.9123
Finish a diploma				
Percentage of financial support expected of the child	0.0535	0.0521	1.055	0.0549
Interacted with age parent	-0.0014	0.0015	0.9986	0.0015
The ideal age of marriage for boys/girls	0.095	0.2335	1.0996	0.2567
Child is female	-0.9027	1.1841	0.4055	0.4802
Belong to OBC	-0.5764	1.2641	0.5619	0.7104
Belong to SC/ST	-0.6658	1.2703	0.5139	0.6528
Finish a bachelor's degree				
Percentage of financial support expected of the child	0.0385**	0.0216	1.0393**	0.0225
Interacted with age parent	-0.001	0.0006	0.999	0.0006
The ideal age of marriage for boys/girls	0.4516	0.1032	1.5708	0.1621
Child is female	-1.1297**	0.6031	0.3231**	0.1949
Belong to OBC	0.159	0.4843	1.1723	0.5677
Belong to SC/ST	-0.8885	0.6096	0.4113	0.2507
Finish an engineering degree				
Percentage of financial support expected of the child	-0.0759***	0.0291	0.9269***	0.027
Interacted with age parent	0.0016**	0.0008	1.0016**	0.0008
The ideal age of marriage for boys/girls	0.1931*	0.1127	1.213*	0.1367
Child is female	-3.2748***	0.7721	0.0378***	0.0292
Belong to OBC	0.1059	0.5407	1.1117	0.6011
Belong to SC/ST	-1.0623	0.6577	0.3457	0.2274
Finish a medical doctor's degree				
Percentage of financial support expected of the child	-0.0732**	0.0296	0.9294**	0.0276
Interacted with age parent	0.0015**	0.0008	1.0015**	0.0008
The ideal age of marriage for boys/girls	0.1351	0.115	1.1447	0.1316
Child is female	-3.7686***	0.9199	0.0231***	0.0212
Belong to OBC	0.0706	0.5591	1.0732	0.6
Belong to SC/ST	-0.9498	0.6617	0.3868	0.256
Finish a master's degree				
Percentage of financial support expected of the child	-0.062**	0.026	0.94**	0.024
Interacted with age parent	0.001**	0.001	1.001**	0.001
The ideal age of marriage for boys/girls	0.479	0.108	1.615	0.175
Child is female	-2.414***	0.848	0.089***	0.076
Belong to OBC	-0.334	0.561	0.716	0.402
Belong to SC/ST	-1.371**	0.635	0.254**	0.161

Notes: *** p<0.01; ** p<0.05; * p<0.1. Includes all children up to the age of 18 currently enrolled in school except for the ones with missing data. Controlled for wealth, credit constraints, time preferences and selection effects. The mean and averages are computed following the procedure outlined in Table 4, and then averaged across the various levels of, respectively, primary and secondary, and higher education. The net contributions at the time of marriage are predicted on the basis of recall data (on dowry, education of bride and groom, joint versus nuclear family status, village and caste) on these contributions from the same villages. Note that as the recall data contains few children who received higher education, these estimates might be imprecise at these higher education levels. The percentage of family members who did no or some higher education is based on the education level of all members of the extended family older than 18 years of age (of the same gender) as reported by the decision maker. N=424. 8th standard is also included in the analysis but results not shown.

3.7 Discussion

This paper looks at the role of social norms, customs and pressures in the aspirations parents have for their children with regard to education in rural India. I take advantage of a detailed child level dataset which includes information on determinants of these aspirations which are usually unobservable, such as the perceived returns to education and expected contribution of the child to the household income at old age. This detailed child level information allows me, for the first time, to empirically disentangle the various sources of social norms, customs and pressures in the education decision in a developing country.

Consistent with the national level statistics on attained education, I find that the educational aspirations are lower for girls compared to boys, and for the backward castes compared to the upper castes. Only 39% of the girls would be allowed (by the parents) to pursue higher education, compared to 71% of the boys and 70% of the UC individuals are allowed to continue higher education, compared to 42% to 44% of the backward castes.

I find that parent's aspirations are a (complex) function of the landholding, the value of the other assets, time preferences and the age of the decision maker, the perceived costs and returns to education, the education of the extended family members, and the prevailing social norms, customs and pressures with regard to marriage and old age care.

Incomplete markets seem to affect the aspirations of parents in a critical manner. Lack of credit access inhibits investment in higher education for poorer

families and leads older decision makers (who are closer to retirement age) to demand a return to education when the child is younger compared to younger decision makers. Incomplete land and labor markets imply that land cannot be sold easily to finance education and that the opportunity cost of the children's time is sometimes more that what they actually would contribute to the family income as the family cannot always substitute hired labor for family labor.

With regard to the social norms, the minimum amount of education parents aspire for their children is influenced by the ideal age of marriage they have in mind and the effect of education on the expected dowry. Increasing the age of marriage with one year, changes the relative odds of aspiring 10^{th} standard, rather than 12^{th} standard, by a factorof 0.86, and the relative odds of aspiring a bachelor's degree, rather than 12^{th} standard, by a factorof 1.30. Increasing the expected dowry of a specific education level with 1000 Rs, increases the relative odds of aspiring that particular level, rather than 12^{th} standard, by a factorof 1.02. Caste and gender does not affect the relative odds ratio of aspiring a bachelors' degree rather than 12^{th} standard, but does have the expected impact on the relative odds ratio of aspiring 10^{th} standard, rather than 12^{the} standard, with lower aspirations for female and lower caste children.

The maximum amount of education parents aspire for, i.e., the maximum their child would be allowed to complete, depends, in addition also on the financial contribution expected from the child at old age, but in a non-linear manner. Increasing the child's contribution with 1 %, increases the relative odds ratio of choosing a bachelor's degree, rather than 12^{th} standard, by a factorof 1.03 and decreases the relative odds ratio of choosing 10^{th} standard, rather than 12^{th}

standard, by a factorof 0.94. This increase (decrease) is smaller (larger) if the respondent is older, i.e., closer to ”retirement”. The same pattern cannot be found for engineering, medical doctor and a master's degree. The education level of the extended family matters, but in not in the way that one would expect it to matter. Increasing the percentage of family members who have obtained a particular level decreases the relative odds ratio of aspiring for that particular level, rather than 12^{th} standard, by a factorof 0.97. This finding points to educational complementarity rather than social pressures, i.e., an extended family might find it useful to have at least one engineer, or one medical doctor, or one teacher in the family. Again, there is a remaining caste and gender effect for some of the education levels, in particular for all higher education levels in the case of girls, and for master's and standard 10 in the caste of lower castes.

These results (partially) confirm the findings of Foster and Rosenzweig (2001), who find that patrilocal exogamy reduces the investment in girls, and show that in rural India, as in rural Bangladesh (Field and Ambrus ,2009), the young age of marriage (of girls) impedes investment in education. Even though I use a very different method compared to Field and Ambrus, the results are similar, delaying marriage with one year results in (on average) 0.22 to 0.33 additional years of education, versus 0.22 additional years of education following Field and Ambrus. As the dowry depends both on the education of the groom as well as the bride, it is difficult to directly test the predictions of Lahira and Self (2007). It is clear however that even though male education might increase dowry and female education might decrease dowry, dowry has no straightforward effect on the education chosen as social status comes into play.

A limitation of this study is the lack of data on (perceived) direct cost of education. These costs were only elicited for one of the villages, Dokur. While the direct cost of education is largely fixed at the caste level in each village up to 12^{th} standard (as the various prices are fixed by the respective state governments), the direct costs of higher education seem to be institution specific. Even though the analysis controls for the cost of education, by relying on school level cost data up to 12^{th} standard and the direct costs, as recalled by the respondents, for higher education, it is not unlikely that there is some unobservable variation left in this regard. From the Dokur dataset it appears that returns and direct costs are positively correlated at the child level, and hence it is not unlikely that the coefficients on the (perceived) returns to education are somewhat downward biased.

In addition, the fact that for over one hundred children, the respondent had absolutely "no idea" of the returns to education is not only worrisome intrinsically, but also might have affected the analysis. At a minimum, losing one fifth of the sample affects the standard errors of the estimated coefficients., especially because the lost observations all belong to the lower tail of the socioeconomic distribution. One could imagine that these children are in a way different from the children in the analysis, and that these unobservables biased some of the coefficients estimated.

To conclude, using a fertile setting, rural India, and detailed data set, this paper takes a first step in distinguishing between the role of various marriage, old age and education related social norms and customs in the aspirations parents have for their children. I show that aspirations are a complex function of wealth,

(perceived) costs and returns to education, time preferences and age, and that multiple customs and norms matter, but all are equally important.

BIBLIOGRAPHY

[1] Akerlof, George A and Rachel E Kranton. 2002. "Identity and Schooling: Some Lessons for the Economics of Education." *Journal of Economic Literature*, 40:4, pp. 1167-201.

[2] Appadurai, Arjun. 2004. "The Capacity to Aspire: Culture and Terms of Recognition," in *Culture and Public Action*. Vijayendra Rao and Michael Walton eds: Stanford University Press, pp. 59-109.

[3] Attanasio, Orazio and Katja Kaufmann. 2009. "Educational Choices, Subjective Expectations, and Credit Constraints," NBER Working Papers 15087 National Bureau of Economic Research.

[4] Austen-Smith, David and Roland Fryer. 2005. "An Economic Analysis of 'Acting White'." *Quarterly Journal of Economics*, 120:2, pp. 551-83.

[5] Bantilan, M.C.S., P. Anand Babu, G.V. Anupama, H. Deepthi, and R. Padmaja. 2006. "Dryland Agriculture: Dynamic Challenges and Priorities." Research Bulletin No. 20, GT-IMPI, ICRISAT.

[6] Basu, Kaushik and Maertens, Annemie (eds). 2010.*The Concise Oxford Companion to Economics in India*. Oxford University Press.

[7] Bertrand, Marianne and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." Working Paper. Harvard University

[8] Bobonis, Gustavo J and Frederico S Finan. 2009. "Neighborhood Peer Effects in Secondary School Enrollment Decisions." *The Review of Economics and Statistics*, 91:4, pp. 695-716.

[9] Chakravarty, Sujoy and E. Somanathan. 2007. "Non-Discrimination in and Elite Labour Market: Job Placements at the Indian Institute of Management - Ahmedabad." Paper presented at the International Conference on Comparative Development, Indian Statistical Institute, New Delhi, December 18-20, 2007.

[10] Chamarbagwala, Rubiana. 2009. "Social Interactions, Spatial Dependence, and Children's Activities: Evidence from India," *Journal of Developing Areas*, 42:2, pp. 157-178.

[11] Dasgupta, Indraneel and Diganta Mukherjee. 2003. "'Arranged' Marriage, Dowry and Female Literacy in a Transitional Society." CREDIT Research Paper.

[12] Delavande, Adeline, Xavier Giné, and David McKenzie. 2008. "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence." Working Paper.

[13] Delavande, Adeline, Xavier Giné, and David McKenzie. Forthcoming. "Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How sensitive are answers to variations in elicitation design?" *Journal of Applied Econometrics*.

[14] Deshpande, Lalit, Anoop Kumar Stapathy, and Sudha Deshpande. 2007. "Personal Income Distribution and Heterogeneity of Labour Markets in India." *Indian Journal of Labour Economics*, 50:4, pp. 903-26.

[15] Dominitz, Jeff and Charles F Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*, 31:1, pp. 1-26.

[16] Drèze, Jean and Geeta Gandhi Kingdon. 2000. "School Participation in Rural India." *Review of Development Economics*, . 5:1, pp 1-24.

[17] Field, Erica and Attila Ambrus. 2008. "Early Marriage, Age of Menarche, and Female Schooling Attainment in Bangladesh." *Journal of Political Economy*, 116:5, pp. 881-929.

[18] Foster, Andrew and Mark R. Rosenzweig. 2001. "Missing Women, the Marriage Market, and Economic Growth." Working Paper.

[19] Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47:1, pp. 153-161.

[20] Holla, Alaka. 2007. "Missing Students: Risk, Gender, and Human Capital Investment in India." Working Paper Innnovations for Poverty Action.

[21] Ito, Takahiro. 2009. "Caste Discrimination and Transaction Costs in the Labor Market: Evidence from Rural North India." *Journal of Development Economics*, 88, pp. 292-300.

[22] Jensen, Robert. 2010. "The Perceived Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics*, 125:2, pp. 515-548

[23] Kingdon, Geeta Gandhi. 1998. "Does the Labour Market Explain Lower Female Schooling in India?" *Journal of Development Studies*, 35:1, pp. 39-65.

[24] Lahiri, Sajal and Sharmistha Self. 2007. "Gender Bias in Education: the Role of Inter-household Externality, Dowry and other Social Institutions." *Review of Development Economics*, 11:4, pp. 591–606.

[25] Lin, Nan. 1999. "Social Networks and Status Attainment." *Annual Review of Sociology*, 25, pp. 467-487.

[26] Loewenstein, George, Roberto Weber, Janine Flory, Stephen Manuck, and Matthew Muldoon. 2001. "Dimensions of Time Discounting." Presented at conference on Survey Research on Household Expectations and Preferences, Ann Arbor, MI, Nov. 2-3, 2001.

[27] Lybbert, Travis J., Christopher B. Barrett, John G. McPeak, and Winnie K. Luseno. 2007. "Bayesian Herders: Updating of Rainfall Beliefs in Response to External Forecasts." *World Development*, 35:3, pp. 480–97.

[28] Mc Fadden, Daniel. 1974. "The Measurement of Urban Travel Demand" *Journal of Public Economics*, 3:4, pp. 303-28.

[29] Munshi, Kaivan and Mark R. Rosenzweig. 2003. "Why is Social Mobility in India so Low? Social Insurance, Inequality and Growth ." BREAD Working Paper No. 097.

[30] Nguyen, Trang. 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." MIT.

[31] Pender, John L. 1996. "Discount Rates and Credit markets: Theory and Evidence from Rural India." *Journal of Development Economics*, 50:2, pp. 257-96.

[32] Rao, K.P.C, 2008. "Documentation of the Second Generation Village Level Studies 2001-2004". GT-IMPI, ICRISAT.

[33] Rao, K.P.C. and Kumara D. Charyulu. 2007. "Changes in Agriculture and Village Economies." Research Bulletin no 21, GT-IMPI, ICRISAT.

[34] Ray, Debraj. 2003. "Aspirations, Poverty and Economic Change." Working Paper New York University.

[35] Rosenzweig, Mark R. and Paul T. Schultz. 1982. "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India." *American Economic Review*, 72:2, pp. 803-15.

[36] Rosenzweig, Mark. 1995. "Why are there Returns to Schooling?" *American Economic Review*, 85:2, pp. 153–158.

[37] Schultz, Paul T. 1989. "Education Investments and Returns," in *Handbook of Development Economics*. Hollis B. Chenery and T.N. Srinivasan eds: Elsevier Science, pp. 544-630.

[38] Schultz, Theodore W. 1961. "Investment in Human Capital." *American Economic Review*, 51:1, pp. 1-17.

[39] Singh, R.P., Hans P. Binswanger, and N.S. Jodha. 1985. "Manual of Instructions for Economic Investigators in ICRISAT's Village Level Studies (Revised)." GT-IMPI, ICRISAT.

[40] Walker, Thomas S. and James G. Ryan. 1990. *Village and Household Economies in India's Semi-Arid Tropics*. Baltimore and London: John Hopkins University Press.

CHAPTER 4

WHY FARMERS SOMETIMES LOVE RISKS: EVIDENCE FROM INDIA

4.1 Introduction

In developing countries, where the majority of the population depends on agriculture, incomes are considerably more variable due to variations in rainfall, crop diseases, etc. As few insurance possibilities exist, attitudes towards risk are crucial determinants of economic decisions, and investment behavior in particular. As such, it is not surprising that a large empirical literature has emerged with the goal of measuring attitudes towards risk.

In this paper, we investigate how credit constraints and the existence of a production set shaped by the various technologies available to farmers (such as irrigated versus non-irrigated, educated versus non-educated) may influence farmers' attitudes with regard to risk in a way which does not correspond with the traditional measure of risk aversion based on the curvature of the Bernoulli utility function.

We take advantage of a unique individual-level dataset collected among Indian farmers, which contains information on their assets, income in addition to their attitudes with regard to risk. These attitudes were elicited via farmer's evaluations of hypothetical but realistic farm alternatives involving various risky outcomes. Each alternative was presented as a probability distribution over yield outcomes which the farmer evaluated in terms of willingness-to-pay.

Using these data, we test a simple model of forward-looking dynamic risk behavior, inspired by Lybbert et al. (2010). This model distinguishes between the familiar static concept of risk aversion (based on the curvature of the Bernoulli utility function) and forward-looking dynamic risk responses. The latter take into account the fact that, in the absence of credit markets, a lucky draw might enable the farmer to make a large fixed cost investment which would allow the farmer to move to a higher level equilibrium characterized by a higher annual income, and consequently a higher standard of living.

We find that 85 % of the farmers are willing to pay more for a distribution which is second-order stochastically dominated by the baseline distribution, indicating risk-loving behavior. Across distributions, the farmers pay a disproportionate amount of attention to the probability of best yield outcome (which is double the size of the worst yield outcome). The estimated effect of increasing the probability of the best outcome by 10% is about 4 times the estimated effect of decreasing the probability of the worst outcome by 10%.

The results of our analysis support the model proposed. We find that risk taking behavior can be (partially) explained by the need for a large fixed-cost investment, such as, having school going children or dryland. This implies that standard measures of risk aversion which look only at contemporaneous decisions, whether they are inferred from observed economic behavior using structural models of behavior combined with econometric techniques, or elicited via experimental techniques, might not be able to capture and predict actual investment behavior under risk.

Among the studies that look at farmers' risk preferences using experimental data from developing countries, this study relates most closely to the work of Binswanger (1980), Dillon and Scandizzo (1978), Just and Lybbert (2009), Liu (2008) and Yesuf and Bluffstone (2009). Binswanger (1980) measures attitudes towards risk among the same households studied in this paper (about 30 years ago) using two methods: an experimental approach with real and hypothetical payoffs of various magnitudes and an interview method. The results of the experimental method indicate that, at medium sized payoff levels (equivalent to the monthly salary of an unskilled laborer) virtually all individuals are moderately risk averse with little variation according to personal characteristics. He finds no statistical difference in the risk aversion estimates based on real gambles versus hypothetical gambles, but a larger coefficient of partial risk aversion when comparing high stakes gambles to low stakes gambles. Dillon and Scandizzo (1978) assess risk attitudes of samples of small farm owners and sharecroppers in Brazil using hypothetical, high stakes (within the range of the farmers' yearly income), questions involving choices between risky and sure farm alternatives. Results indicate that most but not all peasants are risk averse. Just and Lybbert (2009) use low stakes experimental data from farmers in India with real payoffs. Measuring risk aversion as the change in valuation between two gambles with the same expected value but a different variance, they find that about half of the farmers are risk averse. Liu (2008), using a low stakes real experimental method, finds evidence of both risk and loss aversion among Chinese cotton farmers, in addition to the overweighting of low probabilities. Yesuf and Bluffstone (2009), using medium stakes real payoffs experiments from the

Ethiopia, find that depending on the expected payoff and range, one-third to two-thirds of households are severely or extremely risk-averse.

Among the studies that derive risk aversion econometrically from observed behavior, the results of this study could be compared with Moscardi and De Janvry (1977) and Antle (1987) who also study farmers in developing countries. Moscardi and De Janvry (1977) who derive risk aversion from the observed choice of fertilizers find that risk aversion is high among Mexican farmers. Antle (1987), also using econometric techniques to estimate risk aversion from observed choice of labor allocation and fertilizers among rice farmers in Aurepalle village, one of the villages we investigate as well, finds that farmers are both Arrow-Pratt and downside risk averse. His overall estimates are in the same range as the experimental estimates obtained by Binswanger (1980), but he finds considerably more heterogeneity in the population, ranging from nearly risk neutral to risk averse.

Conceptually, this paper is linked to the work of Lybbert and Barrett (forthcoming) and Lybbert et al. (2010). Lybbert and Barrett theoretically link nonconvex asset dynamics to risk preferences. Nonconvex asset dynamics can lead to the existence of a poverty trap, i.e., a low-level dynamic stable equilibrium in a system of multiple equilibria, and thresholds, i.e., a middle-level unstable equilibrium which separates the asset path dynamics towards the low-level equilibrium from those toward the high-level equilibrium. They show that even when the utility function exhibits decreasing absolute risk aversion, observed behavior may suggest that risk aversion actually increases with wealth near perceived dynamic asset thresholds. Lybbert et al. (2010) continue this analysis and show

via Markov simulations that ignoring this dynamic risk response introduces a bias in static estimates of risk aversion. One underestimates risk aversion for individuals just under the dynamic threshold and overestimates risk aversion for individuals just over the threshold.[1] This is due to the fact that "those with wealth just below the threshold value the gamble more because a lucky draw can push them to a more favorable dynamic path, while those just above the threshold value it substantially less because the gamble threatens their otherwise safe wealth position".

More generally, this paper answers the call of Just and Pope (2003) who argue that many alternative explanations can be offered for observed behavior under risk, only one of which is curvature of the utility function. The observed risk response might be due to preferences (risk aversion), technology, physical constraints, or financial asymmetries and in order to properly infer or measure risk aversion from observed choices, one must carefully isolate the impacts of these factors. So in this paper, we look at the role of technology and financial constraints in shaping the attitudes towards risk.

The remainder of this paper is structured as follows. The next section describes the study site. Section 3 describes the risk experiment in detail, discusses selected descriptive statistics and provides some preliminary results. In section 4 we outline a simple model of behavior under risk which we then test in section 5. Section 6 concludes.

[1]Using a coefficient of relative risk aversion of 3, they find that one could under/over estimate risk aversion by a factor of 3.

4.2 Description of the study site

Table 4.1 introduces the three villages selected for this study. These villages have been followed for over 35 years by the Village Level Studies (VLS) program of the International Crop Research Institute of the Semi-Arid Tropics (ICRISAT).[2] The experimental data were collected by the first author in 2007-2008 among 205 of the 246 ICRISAT-VLS respondents in Aurepalle, Kanzara and Kinkhed.[3]

Table 4.1: Basic descriptive statistics of Aurepalle, Kanzara and Kinkhed

	Aurepalle	Kanzara	Kinkhed
Number of households in village	925	319	189
Number of households in sample	128	63	55
Number of households in the experiment	95	57	54
Median rainfall (mm/year)[1]	434	748	745
Distance to nearest town (km)	10	9	12
Average land owned (acre)	3.39	5.24	5.92
Average number of household members	4.23	4.87	4.50
Average yearly income (Rs)[2]	43,543	53,720	38,087
Average education level of respondent (in years)	2.31	6.61	6.89
Average maximum level of education in HH (in years)	7.08	10.41	10
% of young adults enrolled in educational institute[3]	32	32	27
% of households that farm cotton[4]	60	84	82
Average cotton yield [Q/acre]	8.97	3.5	1.88
% of cotton farmers that adopt Bt cotton[4]	77	51	11
% of cotton plots that are irrigated	24	12	2
% of respondents who have access to irrigation	42	30	27
% of respondents with access to bank credit[5]	1.12	17.54	0

Notes: The average/percentage/median statistics refer to the sample in each village in 2007-08 unless otherwise noted; [1]2001-2007; [2]2004-2005, per household; [3]These young adults include all the individuals between the ages of 15 and 26; [4]2001-08; [5]The respondent was asked to imagine he would need credit for agricultural inputs, who would he approach and how likely would he be to receive credit from this individual/organization. Multiple answers were possible. This percentage mentioned here corresponds to the respondents who said they have access to government bank or private bank credit.

[2]ICRISAT followed 300 households from six villages every three weeks during the period 1975-1985 . This dataset, known as the first generation VLS, contains detailed household and plot level data. In 2001, ICRISAT restarted the panel, revisiting 185 of the first generation VLS households and their split-offs, in addition to 261 newly added households, to make the sample representative for each village in terms of land-holding size. For an overview of the goals, methods and outcomes of the first and second generation VLS, respectively, see (i) Singh et al. (1985) and Walker and Ryan (1990) and (ii) Bantilan et al. (2006) and Rao and Charyulu (2007).

[3]Due to the nature of the experiment, we could only conduct it among the ICRISAT-VLS respondents who have farmed in the past seven years or were thinking of farming in the future.

Aurepalle, with 925 households, is the largest of the three villages. It is located in the drought-prone, poor, Telangana region of Andhra Pradesh and in terms of average income situated between the richer Kanzara and poorer Kinkhed. Kanzara and Kinkhed, with 319 and 189 households, respectively, are located in the less drought-prone Akola district of West Maharashtra. The VLS sample includes 128, 63 and 55 households in Aurepalle, Kanzara and Kinkhed, respectively. In both Akola villages, households own, on average, 5-6 acres of land. In Aurepalle, this is significantly less, about 3.2 acres. The average size of a household is between 4 and 5 members in all three villages.

The average education level of the respondent (i.e., the main decision-maker with regard to agriculture) is low, especially in Aurepalle (2.31 years), but as enrollment is relatively high (93 % up the children between 6 and 15 years of age are in school in the sample), the average education level of the next generation can be expected to be much higher. The current (average) maximum level of education in the household is 7 years in Aurepalle, and 10 years in Kanzara and Kinkhed. The enrollment rate drops sharply at the higher education level, only 21 % of the 19 to 21 year old are enrolled in an education institute. This might be partially due to credit constraints. Higher education is expensive in India, ranging from thousand to hundred thousand Rs for a degree. Very few of the farmers in Aurepalle and Kinkhed report having access to bank credit and only about 18 % in Kanzara claims to have access to bank credit.

As is typical for this region of India, cotton is the main cash crop in all three villages. Over 80 % of the households in Kanzara and Kinkhed farmed cotton in the last seven years. In Aurepalle, due to the relatively large number of land-

less families, this number is lower, 60 %. The average cotton yield in 2007-08 is around 9 Q/acre in Aurepalle, 3.5 Q/acre in Kanzara and 2 Q/acre in Kinkhed.[4] As the average cotton yield in Kanzara and Kinkhed was lower in 2007-08 compared to the previous years due to excess rainfall and flooding, it's really only the Kinkhed farmers who were presented with a distribution that (on average) significantly exceeded what they are usually getting.

Losses in cotton production in this region are mainly due to its predominant cultivation under rainfed conditions and its susceptibility to 166 different species of insects, pests and diseases. As such, the yield and profit an individual farmer obtains is very much dependent on the pest protection used and whether or not the farmer has access to a functional irrigation source. One of the new pest protection technologies available in India is the genetically modified Bt cotton. Bt cotton cultivars contain a gene sourced from the soil bacterium *Bacillus thuringiensis* (Bt) in their DNA sequence which produces a protein that is toxic to bollworms, one of the major pests affecting cotton production in India. The difference in average yield therefore largely reflects the difference in Bt cotton and irrigation uptake rates between the villages. Among the cotton farmers, 77 %, 51 % and 11 % have adopted Bt cotton since it was introduced in the region in 2002, in Aurepalle, Kanzara and Kinkhed, respectively.

While the Bt cotton seeds can be priced up to four times the non-Bt price, Bt farmers on average still do better than non-Bt farmers.[5] In Aurepalle, the average profit of Bt cotton farming stood at 7,760 Rs/acre (st. dev. 6,055 Rs/acre).

[4] 1 Quintal (Q) = 100 kg.

[5] For more details on the profitability of Bt cotton versus non Bt cotton in these villages, taking into account the use of other inputs and farmer-fixed unobservables such as soil quality, see Maertens (2010).

In Kanzara the average profit of Bt cotton farming stood at 7,760 Rs/acre (st. dev.5,299 Rs/acre) versus 1,367 Rs/acre (st. dev. 3,056 Rs/acre) for non-Bt cotton. In Kinkhed the average profit of Bt cotton farming stood at 1,176 Rs/acre (st. dev.3,842 Rs/acre) versus 153 Rs/acre (st. dev. 2,762 Rs/acre) for non-Bt cotton.

In 2007-08, 42 %, 30 % and 27 % of the respondents in Aurepalle, Kanzara and Kinkhed, respectively, reported to have access to irrigation as some point in time during the year. In Aurepalle, 24 % of the cotton plots in were irrigated, versus 12 % in Kanzara and 2 % in Kinkhed.[6] Irrigation not only reduces rainfall related risks during the rainy season but also allows for cultivation during the dry season. Both surface water (rivers, canals, ponds and basins) and groundwater (wells) are used as irrigation sources. The water is applied on the field through flood irrigation, drip irrigation or sprinkler irrigation. The cost of a well or a drip/sprinkler irrigation system is substantial, amounting to several times the (average) annual income and the investment entails some risk as not all wells successfully reach the groundwater table. Considering the benefits of irrigation in the rainy season, the average profit for an irrigated cotton plot is 6,030 Rs/acre (st. dev. 8,298 Rs/acre) versus 4,051 Rs/acre (st. dev. 5,348 Rs/acre) for an unirrigated plot.

[6]Note that cotton is a rainy season crop, and that Kanzara and Kinkhed receive - on average - more rainfall and, as such, there might be less need for supplementary irrigation during the rainy season in those villages compared to Aurepalle.

4.3 Description of the experiment and preliminary results

Typically, the experimental approach to measuring risk aversion elicits individuals' preference ranking over two gambles or their Willingness-to-Pay (WTP) for a certain gamble. Imposing Expected Utility (EU), one can use these data to derive a measure of risk aversion for each individual. Or one could extend the EU approach to include various behavioral anomalies, such as, ambiguity aversion, loss aversion and non-linear probability weighting[7] (Engle-Warnick et al. 2006, Liu 2008).

In this study, the farmers' attitudes towards risk were appraised via their evaluations of hypothetical but realistic farm alternatives involving various risky outcomes. The experiments were conducted among all ICRISAT-VLS respondents who have farmed in the past seven years or who were thinking of farming in the future. Henceforward, we will refer to this set of respondents as the "farmers".

Concretely, the risk experiment, based on Lybbert and Just (2007) and Just and Lybbert (2009), consists of four hypothetical farming seasons.[8] For each

[7] Ambiguity aversion (also known as uncertainty aversion) describes a preference for known risks over unknown risks. Loss aversion says that people are significantly more averse to losses relative to the status quo than they are attracted by gains, and more generally that people's utilities are determined by changes in wealth rather than by absolute levels. Non-linear probability weighting refers to over or under weighting certain probabilities in the distribution.

[8] The effect of using hypothetical payments versus real payments has not yet been settled in the literature. The validity depends on the nature of the experiment and elicitation method. In this case, "the subjects have no special reason to disguise their true preferences" (Kahneman and Tversky, 1979, p. 265) as their decision has no financial consequences. hence, it is often considered preferable to use high hypothetical payments over the low real payments that would have been feasible within the project's budget. Binswanger (1980) confirms that in the VLS-ICRISAT villages hypothetical and real gambles eliciting risk preferences produce comparable results. However, Holt and Laury (2002) find that respondents behave more risk loving when

"season" the farmer was asked how much he would be willing to pay for a bag of cotton seed that gives a particular yield distribution (sufficient to sow one acre of cotton). We used Fisher Price building blocks, vertically stacked, to represent cotton yield distributions (in quintal (Q) per acre). Each block represents 5%. We started with two trial distributions to learn the game and then did four experiments, in the order reflected in Table 4.2.

Table 4.2: Risk experiment design

	Distribution 1	Distribution 2	Distribution 3	Distribution 4
4 Q/acre	25	30	30	10
6 Q/acre	50	40	30	55
8 Q/acre	25	30	40	35
Average	6.0	6.0	6.2	6.5
Standard deviation	2.0	2.4	2.8	1.6

Note: 1 quintal (Q) = 100kg

The first baseline distribution, has an average yield of 6 Q/acre and a standard deviation of 2 Q/acre. The second distribution has the same average, but a higher standard deviation than the first distribution, namely 2.4 Q/acre. Thus the first distribution second-order stochastically dominates the second distribution. The third distribution has a higher average yield than the first one, but also a considerably higher standard deviation, 6.2 Q/acre and 2.76 Q/acre, respectively. The fourth distribution first-order stochastically dominates the first distribution with an average of 6.5 Q/acre and 1.55 Q/acre. Comparing these outcomes with the average yield levels in the villages in 2007-08, one can see that for the Aurepalle farmers, this distribution is at the lower end of what they are currently achieving, and for the Akola farmers this distribution it's at the higher end.

using hypothetical experiments. See also and Laury and Holt (2008) for a general discussion of this topic.

Using information on individual-level output prices and costs for inputs other than seeds, one can derive the respective profit distributions from these yield distributions.[9] The average output price is 2,086 Rs/Q (st. dev. 253 Rs/Q). There is little variation between farmers, but some variation between states (Andhra Pradesh versus Maharashtra). The average costs of inputs, other than seeds (not including the costs of the quasi fixed investments such as land, irrigation and machinery) is 5,781 Rs/acre (st. dev. 2,942 Rs/acre).

Table 4.3 presents the main descriptive statistics of the WTP for the various distributions outlined in Table 4.2.[10] Recall that the first distribution second-order stochastically dominates the second distribution. A risk averse farmer obeying standard EUT should be willing to pay less for the latter. The summary results in Table 4.4 indicate that this is often not the case. The average WTP for the first distribution is 495 Rs in Aurepalle, 643 Rs in Kanzara and 1,141 Rs in Kinkhed and the average WTP for the second distribution is 546 Rs in Aurepalle, 647 Rs in Kanzara and 1,370 Rs in Kinkhed. As indicated in Table 4.4, in Aurepalle, Kanzara and Kinkhed, 100 %, 45 % and 98 %, respectively, of the

[9]To obtain these individual level output prices and input costs, we regress the output price and input costs of 2007-08 on numbers of children, numbers of adult household members, acreage of dryland, acreage of irrigated land, soil fertility (how many plots the farmer has of self-reported good and very good quality), education of the decision maker and a village dummy variable. The input cost excluded the cost of seed but includes including the value of family labor and self-produced inputs. Using the predicted values instead of the actual values has the advantage that one has a value for all farmers, not only the ones who farmed cotton in 2007-08, and that one avoids the bias caused by unobservables which are correlated with both price and the other explanatory variables in the main regression. See also Appendix A.

[10]In the case of Kanzara and Kinkhed, the WTP is often lower than the minimum profit in any of the gambles. Informal conversations with the respondents after the experiment took place revealed that if the actual seed price would exceed what the farmer was willing to pay, the farmer would still invest his time, money and land in a different crop, cultivar or activity. For instance, the farmer could lease out his land, cultivate pulses and grain for self-consumption only or leave the land idle. This implies that, unlike in the case of small stakes experiments, where the outside option of "not accepting the bet" is the initial wealth level, in our experiments, the outside option is the next best investment plus the initial wealth level.

respondents were willing to pay more for the second distribution than the first. The differences between Aurepalle and Kanzara, on the one hand, and Kinkhed on the other, in terms of absolute numbers could be partially explained by the fact that Kinkhed farmers have a lower reference point as their average cotton yield is lower. Likewise, the fourth distribution first-order stochastically dominates the first one, yet 46 % of the respondents in Kanzara preferred the first distribution over the fourth distribution.[11] This is strange, but might be explained by the fact that may be not all respondents could recall the first distribution by the time they were asked about the fourth distribution.

Table 4.3: Descriptive statistics of Willingness-to-Pay

VILLAGE = AUREPALLE (N=95)

	Distribution1	Distribution 2	Distribution3	Distribution 4
Mean	495.21	546.21	647.63	743.11
St. Deviation	187.35	232.75	338.62	395.26
Median	465	500	600	660
Min	200	260	280	360
Max	1500	2000	3000	3500

VILLAGE = KANZARA (N=57)

	Distribution1	Distribution 2	Distribution3	Distribution 4
Mean	643.42	647.81	757.89	721.93
St. Deviation	306.21	336.88	380.32	385.22
Median	550	550	650	600
Min	200	150	200	200
Max	1800	2000	2100	2000

VILLAGE = KINKHED (N=54)

	Distribution1	Distribution 2	Distribution3	Distribution 4
Mean	1141.67	1370.37	1597.22	1481.48
St. Deviation	344.46	401.83	445.25	419.60
Median	1100	1300	1600	1500
Min	250	350	400	500
Max	2000	2200	2600	2500

[11] Note that as 94 percent of the farmers reports having access to credit for small loans, paying for seeds should not be a problem.

Table 4.4: Further characterizing Willingness-to-Pay

[in percentage]

	Aurepalle	Kanzara	Kinkhed
Pay more for distribution 2 than for distribution 1	100	45	98
Pay more for distribution 1 than for distribution 4	0	46	0

To get a better understanding of what drives these WTP, we regressed the WTP on the probabilities of the distributions, education level, income, (predicted) input costs and (predicted) output prices.[12] Note that we corrected the standard errors through bootstrapping, as the input costs and output prices are both predicted variables and, as such, may have artificially low standard errors. Table 4.5 presents the results of this analysis. Model 1 is the basic set-up. Model 2 includes also the 2007-08 cotton output as a regressor, as this may influence the results by setting a reference point in the farmer's mind. Note the number of observations used to estimate model is substantially less than model 1 as it includes only the respondents who farmed cotton in 2007-08, as opposed to all farmers. Model 3 includes individual fixed effects.

The results in Table 4.5 indicate that the WTP is largely driven by the probability of the best outcome. Increasing the probability of the best outcome by 10 % increases the WTP by, on average, 152 to 157 Rs, while increasing the probability of the worst outcome with the same magnitude decreases the WTP by, on average, 32 to 37 Rs. A standard EU model with a concave Bernoulli utility function would imply that the magnitude of the effect of an increase of the probability of the worst outcome is larger than the corresponding effect of an

[12]Note that even when one uses a simple expected utility model to explain the variation in the WTP, the WTP will depend in a non-linear manner on the characteristics of the distribution of the outside option and the distribution of the gamble presented to the respondent, the output price and input cost, and the respondent's preferences with regard to risk and time.

Table 4.5: Determinants of Willingness-to-Pay

OLS regression / fixed effects	WTP		
	OLS		Fixed effects
	Model 1	Model 2	Model 3
Probability to obtain 4 Q/acre	-390.605**	-326.576**	-366.859***
	(178.65)	(172.015)	(52.642)
Probability to obtain 8 Q/acre	1578.763***	1554.009***	1572.489***
	(231.039)	(237.819)	(80.654)
Output price, predicted (Rs/Q)	0.097	0.77*	
	(0.447)	(0.414)	
Input costs, predicted (Rs/acre)	-0.063	-0.065	
	(0.037)	(0.041)	
Education level of decision-maker (years)	3.998	1.540	
	(5.208)	(5.360)	
Wealth (land) per capita (1000 Rs)	-1.256**	-0.336	
	(0.411)	(0.306)	
Wealth (other assets) per capita (1000 Rs)	0.265	-0.150	
	(0.214)	(0.327)	
Aurepalle fixed effect	295.917	487.973**	
	(205.445)	(222.423)	
Kinkhed fixed effect	687.224***	735.044***	
	(60.524)	(69.405)	
Output produced in 2007-08 (Q/acre)		-8.017	
		(7.385)	
Constant	287.652	-1169.04	414.547***
	(904.709)	(852.547)	(20.626)

Notes: *** p<0.01; ** p<0.05; * p<0.1; bootstrapped (N=10000) standard errors are reported in parentheses below the coefficient. Number of observations model 1 = 820; number of observations model 2 = 516; number of observations model 3 = 824; Adj. Rsquare model 1 = 0.50; Adj. Rsquare model 2 = 0.59. For one farmer we did not have the education level, this observation was left out. The value of other assets was estimated using the 2006-07 ICRISAT data.

increase of the probability of the best outcome. As such, it is clear that standard EUT will not explain the variation in the data well.

Overall, the three models give very similar results. The output produced in the last season, 2007-08, has a negative but statistically insignificant effect on WTP in model 2. However, as the regressions include a village fixed effect, one cannot conclude from this result that the reference point has no impact on WTP, as the village fixed effects might be absorbing the majority of this variation.

Increasing the output price of cotton has, as expected, a positive effect on WTP, which is only significantly different from zero at the 10% level in model 2. And increasing the input costs has a negative effect on WTP, which is almost sig-

nificantly different from zero at the 10% level in model 2. We included two separate wealth controls as asset market imperfections are likely to constrain substitution across types of wealth. Increasing the value of land owned (per capita) decreases WTP. Increasing the value of all the other assets (per capita), which include livestock, residence, agricultural machinery and savings, increases WTP.

The Aurepalle and Kinkhed fixed effects are substantial in magnitude. These fixed effects could be capturing time preferences, the outside options available to the farmer if the "bet" is rejected, or reference point effects. The positive sign implies that the outside options for farmers in Kinkhed are less favorable compared to the Kanzara farmers. Education, which could influence WTP through changing the outside option of the farmer, does not appear to influence WTP.

These preliminary results seem to reject a simple model of expected utility maximization with a unique concave Bernoulli utility function. Instead, we find that farmers pay a disproportionate amount of attention to the likelihood of the best outcome.

We propose the following explanation for this phenomenon. Imagine that a farmer is credit constrained and there are various production functions available to him, the optimum which depends, among others, on his level of capital (e.g., land). In such a case, it might be optimal for a farmer to take on risky projects (as long as subsistence is guaranteed), and utilize the high returns (if realized) to make a large investment, such as installing an irrigation system or investing in higher education which would allow the farmer to move to a higher production function, and consequently, a higher income level and standard of

living in future periods. The next section presents a simple two period model, describing such a behavior.

4.4 A simple model of behavior under risk

In this section we present a simple two-period model to motivate the empirical specification. The model is inspired by the work of Lybbert and Barrett (forthcoming) and Lybbert et al. (2010). We purposely abstract from several aspects of the agricultural decision making process, such as, the pesticide and fertilizer, and other variable input decisions, with the goal of focussing on the discrete technology choice in the first period. In addition, we assume the farmer is credit constrained (he has no access to credit), has a fixed amount of land $\bar{L}$, no savings in the first period and no irrigation system set up in the first period.

Assume that the farmer, in each time period $t \in \{1, 2\}$ (denoted by a subscript), can choose between two technologies: a 'safe' technology which always yields $f(L, R)$ where R indicates whether or not the land is irrigated ($R \in \{0, 1\}$), and a 'risky' technology which yields $f(L, R) - \epsilon$ with probability $1/2$ and $f(L, R) + \epsilon$ with probability $1/2$, with $\epsilon > 0$ denoting the random component of the production function. Assume that $f(\bar{L}, 1) > f(\bar{L}, 0)$, meaning, irrigation increases the average land productivity.

The per-period Bernoulli utility function is denoted as $u(c)$, where c denotes consumption. We assume that this function is strictly increasing and strictly

concave, i.e., the farmer is risk averse. Acquiring an irrigation system requires a lump sum fixed investment $r > 0$.

In each period, the farmer has to first make a choice between the two available technologies, and then decide whether or not to invest in an irrigation system. The uncertainty of the risky technology is resolved in between these two decisions, i.e., after the farmer chooses the technology but before the investment decision is made.

In the second (and final) period, the farmer's decision problem is, where E is the expected value operator:

$$\begin{aligned} & \max_{\{,f_{t2},r_2\}} E\left[u(c_2)\right] \\ c_2 &= f_{t2}(\bar{L}, R_2) - r_2 \\ r_2 &\in \{0, r\} \\ f_{t2} &\in \left\{f(\bar{L}, R_2), f(\bar{L}, R_2) \pm \epsilon\right\} \end{aligned} \tag{4.1}$$

It is clear that, as this is the final period, $r_2^* = 0$. The farmer will compare the expected utility of the safe technology, $u(f(L, R_2))$, and the expected utility of the risky technology:

$$\frac{1}{2}u(f(\bar{L}, R_2) - \epsilon) + \frac{1}{2}u(f(\bar{L}, R_2) + \epsilon) \tag{4.2}$$

As the farmer's Bernoulli utility function is (strictly) concave:

$$u(f(\bar{L}, R_2)) > \frac{1}{2}u(f(\bar{L}, R_2) - \epsilon) + \frac{1}{2}u(f(\bar{L}, R_2) + \epsilon) \tag{4.3}$$

So the farmer will opt for the safe technology in the second period.

In the first period, the farmer will take into account his choices of the second period and will face the following decision problem:

$$\max_{\{f_{t1}, r_1\}} E\left[c(c_1) + \delta u(c_2)\right] \tag{4.4}$$

$$\begin{aligned} c_1 &= f_{t1}(\bar{L}, 0) - r_1 \\ c_2 &= f(\bar{L}, R_2) \\ R_2 &= \begin{array}{l} 0 \text{ if } r_1 = 0 \\ 1 \text{ if } r_1 = r \end{array} \\ r_1 &\in \{0, r\} \\ f_{t1} &\in \left\{f(\bar{L}, 0), f(\bar{L}, 0) \pm \epsilon\right\} \end{aligned}$$

Assume that:

$$u(f(\bar{L}, 0)) - u(f(\bar{L}, 0) - r) > \delta u(f(\bar{L}, 1) - \delta u(f(\bar{L}, 0) \equiv C(\bar{L}) \tag{4.5}$$

(4.5) states that if the farmer opts for the safe technology, he will not invest in an irrigation system. And assume that:

$$u(f(\bar{L}, 0) + \epsilon) - u(f(\bar{L}, 0) + \epsilon - c) < C(\bar{L}) \tag{4.6}$$

(4.6) states that if the farmer opts for the risky technology, and one obtains the high yield, one will invest in an irrigation system. Then, if in addition:

$$C(\bar{L}) > u(f(\bar{L}, 0)) - u(f(\bar{L}, 0) - \epsilon) + u(f(\bar{L}, 0) - u(f(\bar{L}, 0) + \epsilon - c) \tag{4.7}$$

Under conditions (4.5), (4.6) and (4.7), one can show that the farmer will opt for the risky technology in the first period and invest in the irrigation system if and only if the farmer obtains the high yield.

It is clear that whether these three conditions are satisfied depends on $\bar{L}, \epsilon, r$ and the exact shape of the production and utility function. If the farmer has irrigation installed already in the first period, he will never opt for the risky option in this model.

4.5 Results

According to the model, the willingness to accept risk will depend on the amount of dryland the farmer owns. One could imagine a similar model, substituting labor for land, and higher education for irrigation. In particular, farmers who have school going children who are over the age of 15 years might benefit substantially from a bumper yield, which would give them the option to pay out-of-pocket for one or two years of higher education. Similarly, farmers with a substantial amount of dryland might benefit from installing an irrigation system, which would allow them to cultivate also in the dry season and substantially increase (average) yield in the rainy season.

So, to test the basic implications of the model proposed we include information on the asset position of the farmer, in particular, dryland, irrigated land and school going children in a relevant age range.

In particular, we analyze the relationship between these assets and the willingness to take on risks. Recall that distribution 1 second-order dominates distribution 2; therefore any risk averse farmer should prefer distribution 1. In practise, we see that many farmers prefer distribution 2 (Table 4.4). We use the

difference in WTP between distribution 2 and distribution 1 to reflect the willingness to take on risks, i.e., the dependent variable is $WTP_2 - WTP_1$. Figure 4.1 presents the distribution of the dependent variable. The mean of this distribution is 84 Rs and the standard deviation is 127 Rs. We can see that the large majority of the farmers are situated between 0 and 100 Rs.

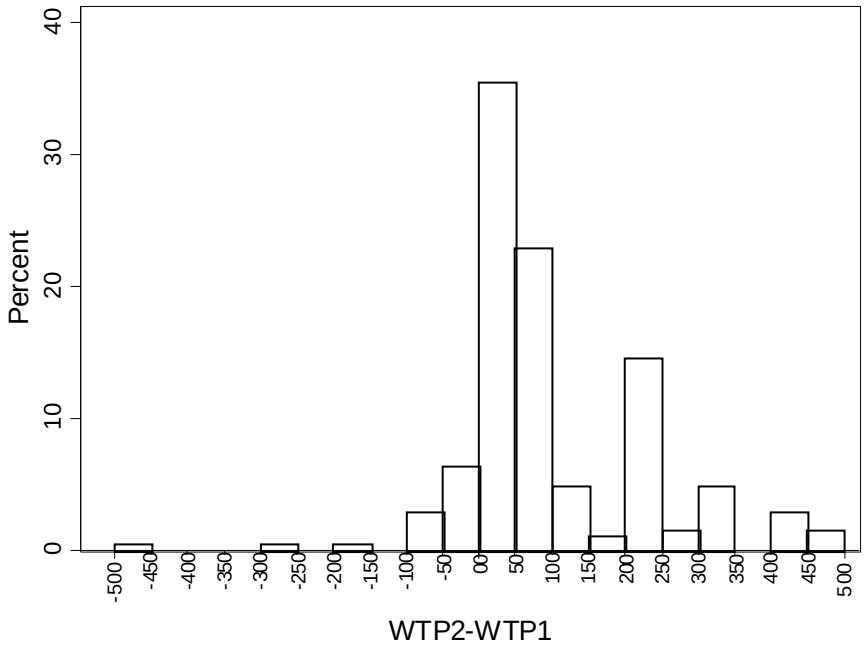

Figure 4.1: Difference in WTP

Table 4.6 presents the regression results. Note that as most farmers appeared to be credit constrained, we did not include credit constraints as a separate regressor. The first column, given for purpose of comparison, includes the standard set of variables: output price, input costs, education, income and village fixed effects. The second column (model 2) adds the number of school going

children above the age of 15 years, acreage of dryland and the acreage of irrigated land.[13]

Table 4.6: Determinants of difference in Willingness-to-Pay

OLS regression	WTP_2-WTP_1	
	Model 1	Model 2
Number of school children		17.153*
		(10.075)
Dryland (acre)		7.293*
		(4.164)
Dryland * dryland (acre)		-0.235
		(0.286)
Irrigated land (acre)		0.542
		(3.981)
Irrigated land * irrigated land (acre)		0.097
		(0.164)
Output price, predicted (Rs/Q)	-0.274	-0.525**
	(0.187)	(0.245)
Input costs, predicted (Rs/acre)	-0.015	-0.011
	(0.017)	(0.018)
Education level of decision-maker (years)	0.759	2.284
	(2.844)	(3.03)
Wealth (other assets) per capita (1000 Rs)	-0.074	-0.13
	(0.093)	(0.128)
Aurepalle fixed effect	35.685	-40.13
	(89.787)	(100.876)
Kinkhed fixed effect	192.272***	179.836***
	(27.15)	(28.923)
Constant	670.826**	1178.416**
	(369.015)	(496.394)

Notes: *** p<0.01; ** p<0.05; * p<0.1; bootstrapped (N=10000) standard errors are reported in parentheses below the coefficient. Number of observations model 1 = 205; number of observations model 2 = 205 ; Adj. Rsquare model 1 = 0.49; Adj Rsquare model 2 = 0.49. The dependent variable is the WTP for the second distribution minus the WTP for the first distribution. Note that as the majority of the farmers have either one or zero schoolgoing children, this effect captures really the dummy effect from having a schoolgoing child. See also notes of Table 5.

The results in Table 4.6 indicate that having a school going child over the age of 15 years in the household, statistically significant increases the difference in WTP by, on average, 18 Rs. Increasing the dryland owned by one acre statisti-

[13] As most farmers have zero or one school going child above the age of 15, we did not include a quadratic term for this variable.

cally significant increases the difference in WTP by, on average, 7 Rs. Increasing the acreage of irrigated land owned does not have a statistically significant impact on the difference in WTP.

Increasing the output price decreases the difference in WTP. In model 2, increasing the output price with 100 Rs, decreases the difference in WTP, on average, with 52 Rs. This number is statistically significant from zero at the 5% level. There is little within village price variation in the output price of cotton, but as the larger farmers tend to fetch a somewhat higher price (as they have better storage, and often do not need cash urgently), this effect might partially capture some residual, (uncontrolled) wealth effects.

The effects of input costs, education and income on the difference in WTP are not significantly different from zero. Again, the village fixed effects are substantial. The difference between what one pays for the higher variance gamble and the lower variance gamble is, on average, 180 to 192 Rs more for Kinkhed farmers compared to Kanzara farmers (significantly different from zero at the 1% level). This might reflect the fact that for Kinkhed farmers the distributions presented (significantly) exceed what they usually get, i.e., these distributions very much exceed their expectations.

4.6 Conclusions

The goal of this paper is to empirically examine how forward-looking investment behavior influences risk behavior in the current period. We are able to

take advantage of a unique data set, collected among farmers in India's semi-arid tropics, which contains information on farmer's assets and the results of a risk experiment.

This risk experiment consists of four hypothetical farming seasons. For each season the farmer was asked how much he would be willing to pay for a bag of cotton seed that gives a particular yield distribution. Comparing the willingness to pay for the various yield distributions, we find that 85 % of the farmers are willing to pay more for a yield distribution which is second-order stochastically dominated by the baseline distribution presented to them.

This is not consistent with what has been found to date in these contexts. While none of the methods used in these studies are directly comparable to the method we used in our study, it is useful to think through the reasons why what we find might be different. Our study was set in the same villages as Binswanger and Antle, but 30 years later. Our study elicits risk preferences using hypothetical high stakes gambles which are explicitly forward-looking, while these studies all use low to medium stakes gambles or an econometric approach using data of input choices.

Using high stake gambles implies, by definition, that the farmer may stand to gain and loose a substantial amount. The gains might be sufficient for a large fixed-cost investment, which could shoot the farmer and his family to a higher level equilibrium. For instance, one can use the returns of a bumper harvest to invest in an irrigation system and be able to cultivate one's land twice a year instead of once a year from that point onwards. So as long as the potential losses

do not bring the farmer and his family under some subsistence threshold, one can understand how, if one takes into account the asset position of the farmer, high stake gambles might be attractive for a certain set of farmers. Thinking about risk preferences from this angle, it is also clear why this forward-looking investment behavior would not be likely to impact risk preferences in case of small stakes gambles, as, indeed, the proceeds from a small stakes gamble are unlikely to be sufficient for the investment needed to move to a higher production function.

We find evidence in the data that supports this explanation. For farmers who are credit constrained, increasing the dryland owned by one acre increases the difference in WTP between a high variance distribution and a low variance distribution (with the same mean), by, on average, 4 Rs, which is 8 % of the average difference in WTP. This indicates that farmers who own more dryland are more risk loving compared to farmers who own less dryland or farmers who own only irrigated land. Also, having one more school going child over the age of 15 years in the household, increases the difference in WTP by, on average, 20 Rs, which is 20 % of the average difference in WTP. This indicates that farmers who might want to invest in higher education of their children (but are credit constrained) are more willing to take on risks compared to farmers who have no school going children in this age range.

Note that we do not test the models by Lybbert and Barrett (forthcoming) and Lybbert et al. (2010) directly. To test these model directly, one would have to derive the dynamic asset threshold from the data, calculate the distance to this asset threshold for each farmer and analyze the relationship between this

position and the willingness to take risk. As in human capital is becoming more and more important as part of the asset stock, this asset threshold would need to take into account the education level of the various family members. This, together with the fact that we do not have information on all assets in 2007-08, and are missing information on the value of some of the critical assets such as borewells and tubewells even for the earlier years, makes it difficult to pursue a direct asset threshold based approach.

Also, when interpreting the results of this study one needs to keep in mind that the experiments that we conducted were hypothetical, i.e., we did not actually pay out the farmers. The nature of the experiment, where we imitated the experience of a farmer choosing a type of seed in the input dealer's shop, would not have allowed for an actual payout. Also, as the experiment was framed as an actual purchase decision the farmers could easily relate to as it is a decision they make each years, we are less worried about any bias originating from the hypothetical method.

Binswanger (1980), who conducted monetary experiments among the same households as we did, confirms that hypothetical and real gambles eliciting risk preferences produce comparable results for low and medium stakes gambles. However, Holt and Laury (2002), who conducted experiments with students at three US universities find that respondents behave more risk loving when using hypothetical high stakes (over 100 US dollar) experiments versus real high stakes experiment. They conclude that "respondents cannot imagine how they would actually behave under high-incentive conditions". We hope that, as the

experiment we conducted was framed as a seed buying experience, our respondents had less problems imagining how they would react.

Finally note that the validity of the results also depend on the estimated individual-level output price and input costs. If these are systematically over or underestimated, and this bias is correlated with the variables of interest, such as credit constraints, acreage of dryland or even the probabilities of the experiment, the results will be incorrect.

To conclude, in this paper we provide a first empirical test of the role of forward-looking investment behavior in risk behavior. We find that, when facing (realistic) high stakes gambles, the large majority of the farmers behave as if they love risks. We find evidence of the following explanation for this risk loving behavior: in the absence of functioning credit markets, a lucky draw might enable the farmer to make a large fixed-cost investment, such as an irrigation system or higher education for one's children. This investment could allow the farmer, and his family, to move to a higher level equilibrium characterized by a higher annual income, and consequently higher standard of living in future periods.

APPENDIX

Table 4.7: Predicting individual level output price and input cost

OLS regression	output price [Rs/Q]	input cost [Rs/acre]
Number of members	-11.3	216.8
	(17.3)	(155.8)
Number of adult members	20.3	-230.3
	(22.3)	(199.8)
Dryland (acres)	2.7	-5.7
	(5.1)	(45)
Irrigated land (acres)	3.4	40.6
	(5.2)	(40.9)
Access to a plot of good soil quality	50.8	1044.9
	(81.2)	(712.5)
Number of plots of good soil quality > 1 acre	14	-24.1
	(20.6)	(182.7)
Education level of decision-maker (years)	9.6	-2.7
	(5.9)	(53.7)
Aurepalle fixed effect	-211.2***	3763.8***
	(57.8)	(522.6)
Kinkhed fixed effect	-39.8	-799.3
	(58.2)	(508.2)
Constant	2037.1***	3133.6***
	(95.1)	(836.7)

Notes: ** p<0.01; ** p<0.05; * p<0.1; number of observations 130 ; Adj Rsquare output price = 0.23 ; Adj. Rsquare input costs 0.52.

BIBLIOGRAPHY

[1] Antle, J. 1987. "Econometric Estimation of Producers' Risk Attitudes." *American Journal of Agricultural Economics*, 69:3, pp. 509-22.

[2] Bantilan, M.C.S., P. Anand Babu, G.V. Anupama, H. Deepthi, and R. Padmaja. 2006. "Dryland Agriculture: Dynamic Challenges and Priorities." Research Bulletin No. 20, GT-IMPI, ICRISAT.

[3] Binswanger, Hans P. 1980. "Attitudes toward Risk: Experimental Measurement in Rural India." *American Journal of Agricultural Economics*, 62:3, pp. 395-407.

[4] Dillon, John L. and Pasquale L. Scandizzo. 1978. "Risk Attitudes of Subsistence Farmers in Northeast Brazil: A Sampling Approach." *American Journal of Agricultural Economics*, 60:3, pp. 425-35.

[5] Engle-Warnick, Jim, Javier Escobal, and Sonia Laszlo. 2006. "Risk Preference, Ambiguity Aversion and Technology Choice: Experimental and Survey Evidence from Peru." NEUDC 2006 at Cornell University.

[6] Just, David R. and Travis J. Lybbert. 2009. "Risk Averters that Love Risk? Marginal Risk Aversion in Comparison to a Reference Gamble." *American Journal of Agricultural Economics*, 91:3, pp. 612-26.

[7] Just, David R., and Hikaru Hanawa Peterson. 2003. "Diminishing Marginal Utility of Wealth and Calibration of Risk in Agriculture." *American Journal of Agricultural Economics,* 85:5, pp. 1234 - 1241.

[8] Just, David R. and Hikaru Hanawa Peterson. 2010. "Is Expected Utility Applicable: A Revealed Preference Test." *American Journal of Agricultural Economics*, 92(1), pp. 16-27.

[9] Just, Richard E. and R.D. Pope. 2003. "Agricultural Risk Analysis: Adequacy of Models, Data, and Issues", *American Journal of Agricultural Economics* 85:5, pp. 1249-1256.

[10] Holt, Charles A. and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92:5, pp. 1644-55.

[11] Kahneman, D. and Tversky, A.1979. "Prospect Theory: An Analysis of Choice Under Risk." *Econometrica*, 47:2, pp. 263-91.

[12] Keuhberger, Anton, Michael Schulte-Mecklenbeck, and Josef Perner. 2002 "Framing decisions: Hypothetical and Real." *Organizational Behavior and Human Decision Processes*, 89:2, pp. 1162–75.

[13] Laury, Susan K. and Charles A. Holt. 2008. "Payoff Scale Effects and Risk Preference Under Real and Hypothetical Conditions," in *Handbook of Experimental Economics Results*, Charles Plott and Vernon Smith (eds), Elsevier, pp. 1047-53.

[14] Liu, Elaine. 2008. "Time to Change What we Sow: Risk Preferences and Technology Adoption Decisions of Cotton Farmers in China." Department of Economics. Princeton University.

[15] Lybbert, Travis J., Christopher B. Barrett, Solomon Desta, and D. Layne Coppock. 2004. "Stochastic Wealth Dynamics and Risk Management Among A Poor Population." *Economic Journal*, 114:498: pp. 750-77

[16] Lybbert, Travis J. and Christopher B. Barrett. 2007. "Risk Responses to Dynamic Asset Thresholds." *Review of Agricultural Economics*, 29:3, pp. 412-18.

[17] Lybbert, Travis J. and Christopher B. Barrett, forthcoming. "Risk Taking Behavior in the Presence of Nonconvex Asset Dynamics," *Economic Inquiry*.

[18] Lybbert, Travis J. and David R. Just. 2007. "Is Risk Aversion Really Correlated with Wealth? How Estimated Probabilities Introduce Spurious Correlation " *American Journal of Agricultural Economics*, 89:4, pp. 964-979.

[19] Lybbert, Travis J., David R. Just, and Christopher B. Barrett. 2010. "Estimating Risk Preferences in the Presence of Bifurcated Wealth Dynamics: Do We Misattribute Dynamic Risk Responses to Statis Risk Aversion?" Working Paper Cornell University and UC Davis.

[20] Maertens Annemie,. 2010. "Who Cares What Others Think (or Do)? Social Learning, Social Pressures and Imitation in Cotton Farming in India." Department of Applied Economics and Management. Cornell University.

[21] Moscardi, Edgardo and Alain de Janvry. 1977. "Attitudes toward Risk among Peasants: An Econometric Approach." *American Journal of Agricultural Economics*, 59:4, pp. 710-716.

[22] Pemsla, D., H. Waibela, and J. Orphalb. 2004. "A Methodology to Assess the Profitability of Bt-cotton: Case Study Results from the State of Karnataka, India " *Crop Protection*, 23:12, pp. 1249-57.

[23] Pratt, John W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica*, 32:1/2, pp. 122-136.

[24] Qaim, Matin and David Zilberman. 2003. "Yield Effects of Genetically Modified Crops in Developing Countries." *Science*, 299: 7 February 2003, pp. 900-902.

[25] Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68:5, pp. 1281-92.

[26] Rao, K.P.C. and Kumara D. Charyulu. 2007. "Changes in Agriculture and Village Economies." Research Bulletin no 21, GT-IMPI, ICRISAT.

[27] Samuelson, Paul. 1963. " Risk and Uncertainty: A Fallacy of Large Numbers." *Scientia*, 98, pp. 108-13.

[28] Singh, R.P., Hans P. Binswanger, and N.S. Jodha. 1985. "Manual of Instructions for Economic Investigators in ICRISAT's Village Level Studies (Revised)." GT-IMPI, ICRISAT.

[29] Starmer, Chris. 2000. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature*, 38:2, pp. 332-382.

[30] Townsend, Robert M. 1994. "Risk and Insurance in Village India." *Econometrica*, 62:3, pp. 539-591.

[31] Walker, Thomas S. and James G. Ryan. 1990. *Village and Household Economies in India's Semi-Arid Tropics*. Baltimore and London: John Hopkins University Press.

[32] Yesuf, Mahmud and Randall A. Bluffstone. 2009. "Poverty, Risk Aversion, and Path Dependence in Low-Income Countries: Experimental Evidence from Ethiopia" *American Journal of Agricultural Economics*, 91:4, pp. 1022–37.

APPENDIX

The villages selected for this study are Aurepalle and Dokur in the Mahbubnagar district of Andhra Pradesh, Shirapur and Kalman in the Solapur district in Maharashtra, and Kanzara and Kinkhed in the Akola district in Maharashtra.

I conducted the data collection in collaboration with the Village Level Study program (VLS) villages of the International Crop Research Institute of the Semi-Arid Tropics (ICRISAT). This program has been collecting data in these six villages since 1975. The majority of the data collection was done by the ICRISAT resident enumerators (junior investigators) who live in the villages. Some of the data collection was done by myself (through translation), and by enumerators I hired myself through the universities.

I structured the data collection in five rounds: (i) a qualitative round among selected households, teachers and village pradhans (elected village leader) to get some general insights and further narrow down the research topic, (ii) a trial round during which I tried out the (largely) quantitative questionnaires, (iii) a training round during which I trained the ICRISAT enumerators and the enumerators I hired through the universities, (iv) an actual data collection round during which the trained enumerators collected household and child level data, in addition to village, school, and progressive farmers' data, and (v) a data validation round which happened after I checked each questionnaire for inconsistencies – and where I went back to the villages to the respective households to

ask the questions again which had problems. The entire process took almost 2 years, from October 2007 to July 2009.

This appendix provides an informal narrative of selected aspects of this data collection process, focusing on the qualitative round and the trial round as these shaped the final questionnaires. The final questionnaires are available from my personal website (http://sites.google.com/site/maertensannemie/), in addition to the manuals used by the enumerators.

The qualitative round took place in October-November 2007. During this trip, I visited all six villages, starting with Shirapur and Kalman, going on to Kanzara and Kinkhed and finishing with Aurepalle and Dokur.

Schedule and general comments on the qualitative survey round

During this trip I tried to keep my main goals in mind (get a general idea of what determines the education, market participation and technology adoption decisions and decide on two specific topics for the two empirical papers needed for my dissertation). The discussions ended up being semi-structured. I had some topics I mind that I wanted to discuss with each family, but how much each topic was explored depended on their answers.

The senior ICRISAT investigators of Maharashtra and Andhra Pradesh accompanied me throughout the entire trip. In the villages we met with the junior ICRISAT investigators who reside in the villages. I conducted the interviews,

together with the senior investigators. The junior investigators observed and learned about the project.

The format of the interview – I ask a question – the senior investigator translates it – the answer comes back to me – I ask the next question, worked surprisingly well. Due to the translation delay, I had plenty of time to write down the answer and to observe the reactions of the people to my questions. In general, people were very willing and happy to talk to me. The interviews lasted between 1 and 1.5 hours per family, about 0.5 hours for the pradhan, teachers and junior ICRISAT investigators.

I started the interview by stating that participation is voluntary and the respondent could at each point in time withdraw from the interview and/or ask us questions. At the end of each interview, I asked the respondents whether they had any questions or suggestions for us. This invariably led to a second interview during which I got questioned.

During the first five days, quite a bit of time was lost due to the fact I had not established a clear daily schedule yet. One single day needed to incorporate the interviews, but also some social time with the villagers and discussions between the senior and junior ICRISAT investigators. As a result I did not collect as much data from Shirapur and Kalman as I did from the other four villages. In each village I also did one or two field visits to get to know the soils, the crops and the area.

Even though I had planned to interview non-sample households, I ended up interviewing quite a few of sample households. There were two main reasons

for this. First, ICRISAT offers the sample households each year a one or two day excursion. If I would interview non-sample households, these people would also start to ask for the trip. In addition, the non-VLS households were difficult to get a hold of and as the daily cost of this fieldtrip was very high and the trip could not be interrupted or postponed, I sometimes had to settle for VLS respondents instead. I did give priority to non-VLS respondents, but if no-one was available I switched to VLS households.

Deciding on a dissertation topic

One of the main goals of the qualitative round was to narrow down the topic of the dissertation. As such I explored different kinds of agricultural technologies, market participation behavior and various education decisions. It was clear from the beginning that the education decisions were very much influenced by identity and different social norms. It took me the remainder of the trip to decide on a second topic.

It was difficult to find a suitable agricultural technology to focus on for several reasons. First of all there were only a few new technologies: high yielding crops, mechanization (tractor, harvester and thresher) and some new types of irrigation techniques. Many farmers in Kanzara, Kinked and Aurepalle (and some in Dokur) experiment a lot with the different varieties. Almost every year or two years they try out a new variety for their crops. However for both the mechanization and the new irrigation techniques, I could not find any evidence of the presence of identity and social network effects. In addition, only few

people in each village adopted these mechanization and irrigation techniques. Moreover, in several villages, like Kalman and Dokur, agricultural techniques have not changed much in the last ten years. In Kalman, where mainly dryland farming is practiced, farmers said that there is little scope for improvement. In Dokur, where mainly paddy is cultivated, the technique and varieties used have not changed much either. Some new crops appeared in the villages, like fruit orchards (mangos and citrus), grapes and soybeans. In addition, both pesticide use and fertilizer use increased in all of the villages.

I did explore the option of looking at water management. This region of India seems to have been severely hit by climate change. Water is one of the main production constraints. Both surface water (rivers, canals, ponds and basins) and groundwater (wells) are used as sources. The water is applied on the field through flood, drip or sprinkler irrigation. In general, surface water systems are provided by the government. There is a main canal, which a bunch of sub-canals. The entire system is build by the government and managed by a 'water committee'. That is, there is supposed to be a committee. If the committee is not there or does not function, problems arise such as: who cleans the canals?, who cleans which canal?, what if I am farmer at the end of a canal and I want water, am I supposed to clean also the beginning of the canal? The farmers pay a fixed price per season by acre for a fixed number of irrigations. For additional irrigation they need to pay an additional sum. Sometimes there is not enough water, and during that year, no irrigation will be provided. In addition, the land tax on this 'irrigable' land is higher than the land tax on the 'non-irrigable land'

(independent of the use). Also, several farmers complained about the quality of this surface water.

In case of groundwater, the investment is private. The farmers dig a well on their own fields. These investments are very costly (equivalent to several years of income) and tend to be a risky investment, depending on the village, 20% to 80% of the wells were dry or became dry after a few years. However, I could not find any significant identity and social network effects in this private investment decision. There are several government rules however related to where to build a well, and how to finance the well etc, and these government rules differ by state. Also, they locate the well using interesting tools like sticks from some tree and lemons. As such, I concluded that even though this topic of water management is a fascinating one, I would not embark on this research now for my dissertation.

In the end I decided on Bt cotton for the technology adoption part of my dissertation. Cotton is a major crop in three out of six villages. In the last three to five years Bt cotton has gradually been adopted in those three villages. As such, as I have panel data of the last seven years of the villages, with a lot of details on their farming activities, this decision seemed a suitable one to study.

I also considered the topic of market participation. Most farmers decide simultaneously on their cropping pattern and market participation (as one cannot consume cotton, castor or an entire orchard of mangos). There are four ways to market the output.

The first, and most common one, is the government regulated market in the nearby towns. These government regulated market have a 'market committee'. This committee (in theory) ensures that the government minimum price is respected, that no cheating (in weights) occurs and that disputes between farmers and traders are resolved. In this case the farmer sells his produce to the commission agent (he is often in the village) or the trader (in the market).

The relationship between farmer and commission agent/trader is a complex one. In the case of a commission agent, he often also provides credit for inputs. The farmer gives his harvest to the agent and the agent goes to the regulated market with the harvest to participate in an open auction. The farmer often sits next to the agent when the auction takes place (some expect cheating when they do not go along to the market). While the farmer has no influence over the price, he can always refuse to sell at the price they offer him. In this case the farmer will store the produce for a few days (at the market) and sell later.

If the farmer does not use a commission agent he will go to market himself and meet the trader (with or without harvest). In this case the trader immediately announces his price of the day (also announced on a board next to the shop). Again, the farmers said they had no influence on the price (some of them recalled that things used to be different in the past and farmers then did have some group influence, but today, they complained, farmers were not a 'unit' any more). The farmers had the impression that traders formed a 'block' and seemed to agree in the morning on the price. This is why, according to them, the price was the same in the market on a given day. As such, in order to get a better price, farmers sometimes wait to sell their produce and visit or call the

trader to get to know the price. How long they can wait depends on the storability of the produce and their financial situation (some need to pay back debts immediately). In many cases, the trader provides credit and inputs. In this case the farmer is often required to sell his output to the same trader.[14] While switching traders does not happen frequently, unwillingness to provide credit is often cited as a reason to switch traders if it does occur.

The farmers have some idea of the prices of the crops and the expected output price enters their cropping pattern decision. Only a few farmers told me that they did not think about the price at all when making the cropping pattern decision. When the crop is growing, the farmers often sit in the field together and chat about what they think the price will be this year. There was a general sentiment that for the crops for which there is no government minimum price, price fluctuations are on the increase. Also, in the past, farmers were better able to predict the price. Now, (in their view) due to inter-state trade and imports, the price is less predictable. When the crop is harvested, the farmers in general will inform themselves about the market price in the regulated market. When the farmer sells his produce to the commission agent, he will only get to know the exact price at which he sold after the agent is back from the market. Similarly when the farmer sells his produce to a trader, he will only know the price once he meets the trader in the market.

[14]When asked about what would happen if they do not do this, farmers told me that they are allowed to pay back the loan and then sell the output to someone else, but because the daily price on the market is the same for all traders, no one has done this, as this would spoil the relationship.

The second way to sell the output is to sell the crop within the village. This is not common. It used to happen more often, but now is done only by small, marginal farmers.

The third way is to sell the produce to traders who come to the village. This is only done by larger farmers, who can sell large amounts at once. These farmers compare the traders' offers and bargain. Once they agree on a price, the trader comes and picks up the produce with his own tractor.

The last way is to sell the produce immediately to a factory. This is the case for sugarcane in Shirapur and paddy in Dokur. In Shirapur, the relationship between the farmer and the sugarcane factory owners is complex. The factories often provide credit and inputs to the farmers. Sugarcane factory owners are often also local politicians. The farmers complained about the low price (it was unclear whether this was a fixed price – even though the government. has a minimum price on this crop). In addition, a few farmers told me that they would like to switch crops, but in that case they would be unable to get credit for inputs (apparently the sugarcane factory owners influence could reach the moneylenders etc.).

Clearly, the topic of markets is very interesting to investigate. But I decided not to dedicate another paper to this topic for a couple of reasons. First, it was unclear to me which decision I would focus on. All produce is sold, the only main decision left to the farmer is when to sell, and this decision is determined by mainly non-identity and non-social network reasons. For the farmers who did not buy inputs on credit, they might still have to decide where to sell their

crop. But most farmers have sold to the same person for the last 5 to 10 years. With so little variation over time, panel data analysis would probably not be very interesting. Second, I did not have the funds to interview the traders. The power relations between trader and farmer appear too complex for me to embark on this topic as a side study. As such I decided to consider only the input and output markets of one single crop, cotton.

Difficult questions

The first kind of questions that were sometimes difficult to answer are the hypothetical questions which are forward looking in nature, like for instance, 'if you a friend of your daughter would be absent from school for a long time, would you go up to her parents and say something?' Often such a question was answered with, 'that would never happen'. In such a case, explaining that we just are interested in what the respondent would do in such a situation rectifies the problem.

A second kind of question that was difficult to answer concerns (detailed) beliefs of the past, like for instance, 'five years ago what did you think the cost of one year in a private school in town X would be?' I found that people recall their general thoughts and ideas about the issues I was interested in (education, technologies and markets), like 'fifteen years ago I did not realize the importance of education, now I do, and I regret my decisions' or 'when my son was young, I though he could become a doctor'. But recalling numbers/amounts of

what they thought in the past about the future at the time is not possible (one tends to project current beliefs on the past).

Recalling actual facts was not very difficult, and I found that, depending on their education and intellectual abilities, people could recall up to 3 to 5 years of cropping patterns, input use, market prices, wages and educational costs. These kinds of questions were fun for the respondents as I asked 'do you remember the price you received for your cotton one year ago, yes? and now two years ago?, yes, and now three years ago? etc'.

Eliciting quantitative beliefs about yields and wages went relatively well. I also tried to elicit expected wages conditional on education level. It took us a while to get the formulation of the question right as when you ask 'what do you think that your daughter would earn if she finished grade 10', a common answer was 'God will decide that', as such I needed to reformulate the question, thereby emphasizing that even though God is influential, I was interested in her personal expectations. For the parents with very small children, the question could not be asked in terms of 'their daughter' but had to be asked in a general way, otherwise I invariably got the answer that 'it was too soon to be thinking about these things'. So I rephrased the question and asked for a 'if a child similar to their daughter in ability, caste etc would reach 8th standard'. I also tried to see whether people had some idea of the likelihood of obtaining such a job, and they were able to answer these questions as well. I interviewed too few people to say anything at this point, but I had the impression that the wage expectations and the likelihoods of getting a formal job differed by caste.

Eliciting beliefs about agricultural prices was more difficult. A few farmers told me that they never thought about output prices, they only thought about input prices and then maximizing yields. When asking the question how they predict the price of next year, some farmers said that they took the price of last year, other mentioned looking at the prices of the last few years, while other has a more complex model in mind thinking about the expected production of that year.

I did not find a suitable way to ask whether 'discount rates' were different among gender/caste. It seems to me that the standard experimental economics games are not relevant for the decisions I am considering. For education decisions, time preference would matter and indeed more than once farmers referred to certain communities and castes (like Muslims) as 'impatient' in that regard. Related to the Bt cotton decision, which is a yearly decision, it is not clear how a different discount rate might matter as switching to Bt cotton does not necessarily mean trading future consumption in for present consumption as the seed are often bought on credit (while the normal cotton seeds are not).

I did not try to elicit risk aversion either using standard experimental techniques. I did, however, have a several discussions with the farmers where I let them talk freely about how they felt about risk in farming and the risk in the education decision. Most farmers thought that being a farmer is by nature a risky job. As such many of them had an extensive set of coping mechanisms to minimize risk and to deal with risk once it occurs. The riskiness of the job was also the main reason why farmers wanted their brightest sons to move out of agriculture. Only the large (almost industrial) farms saw for their sons a future

in farming. Related to riskiness of education investment, most people felt that there was little they could do about it, and as such they could only educate their daughter and son up to the ideal level and then they would 'leave things up to God'.

Caste and gender

With regard to caste, the neighborhoods in the villages are without exception structured along caste lines. The previous untouchables, the Scheduled Castes (SC), often live at the edge of the village. As people tend to interact with their neighbors mostly, a high correlation between social networks and caste-identity can be expected. Quite a few of the SCs live in concrete houses build by the Indira Ghandi Housing Scheme. In general, government programs are structured along caste and gender lines.

Social norms related to marriage, education and labor market participation differ between men and women. In addition, social networks and social connections are clearly structured along gender lines.

Social connections

The general questions related to how social life in the village has changed (this question asked to all pradhans, several school teachers and some household members), invariably let to the same kinds of answers. People complain about

the fact that the respect for elderly has declined in the last ten years, that people tend to be less willing to help each other (provide labor, credit etc), that the norms of reciprocity have changed[15]. Bad habits (like drinking wine) are on the increase,[16] and that the joint family system is being replaced by the nuclear family system. It was easy to elicit the several groups active in the village from the pradhan, even though sometimes they forgot to mention a few. In most villages there were several self-help groups (SHG), chit funds and caste-based organizations. Some villages had a school committee and a water committee. It was difficult to ask the pradhan questions related to who-interacts-with-whom. I often got the answer that 'everyone in the village interacts with everyone else' or 'everyone likes everyone'.

To get an idea of the social connections of the villagers, I had several more general discussions with the senior and junior investigators on the topic of who-knows-whom?, who-interacts-with-whom?, and are many of the contacts outside of the village?. I was unable to have many of these general discussions with the villagers themselves. With them I therefore asked very specific questions, like, who would you go to if you need some flower immediately?

In general, the villagers seem to know more people outside of the village than ten years ago. The structure of the marriage networks however has not changed much in the last ten years. In Maharashtra, the marriage networks are generally within-caste, but outside the village, and sometimes even outside the

[15]Many villagers cited the example of financial help. If help is needed before help used to come unconditional and from many people, now only few people offer to help, and many people keep mental accounts or explicitly impose conditions.

[16]Related to those bad habits, in Maharashtra these are generally frowned upon while in Andhra Pradesh these are sometimes also considered as a sign of wealth.

district. In Andhra Pradesh, the marriages often take place within the family (for instance, marrying your brother's or sister's child is very common). The majority of the women in the villages interact mostly with people in the village. The men often have several contacts outside the village, mostly market-related contacts. These market-contacts outside of the village are a relatively new phenomenon in the villages (last ten years).

The social networks differ by the use of the network: different sets of people are contacted and relied upon for information related to agriculture, for information related to education, and for credit. Even within the agricultural information networks, people often rely on different people depending on the crop they need information on. Similarly, if credit is needed, who is contacted depends of what the credit is needed for, the amount and when the farmer expects to be able to pay back. These different credit networks are function according to different rules. For instance, if credit is needed for agricultural inputs either the village money lender or input dealer (in the city) will be contacted. In this case a relatively high interest rate will have to be paid (30-40 % per year). If credit is needed for the marriage of a daughter, family members will be contacted. The terms of repayment are much more flexible in this case. Repayment can be spread over several years, often interest free, and sometimes the money will be transferred as a gift. If a small amount of money is needed for immediate consumption, friends and neighbors in the village (often cast fellows) will be contacted, the money will be repaid immediately with very little to no interest.

Schedule and general comments on the trial survey round

During the trial survey round, which took place in December 2007, I tried out two draft questionnaires, one on Bt cotton and one on education, through 16 interviews: 4 in Kinkhed, 4 in Kanzara, 4 in Aurepalle and 4 in Dokur. I did not do all interview questions with each household, but worked through selected sections depending on how well the sections went. It was not an option to work through all sections as trying out new questions tends to take more time than asking a routine question: looking at their reaction, reformulate things, try them again, and all this through translation. The respondents were told that we were doing a trial round, and were asked to be critical of the questions. Most of them took up their job as 'judge' very willingly, and suggested changes in the questionnaire. If I saw that two or three households had no problems with a certain section, I moved on to another section. I had to rewrite several sections multiple times, notably the identity section, the general education section and the learning and norms sections.

As each questionnaire was about 60 pages long, I decided to limit make three village 'education only' and the three others 'Bt cotton only'. Aurepalle, Kinkhed and Kanzara, a total of 248 households, would receive the Bt cotton questionnaire. The remaining 352 households in Dokur, Shirapur and Kalman would receive the education questionnaire.

Rewriting the perceived identity section

I rewrote the identity section (in Section II of both questionnaires) almost entirely. The caste/religion section needed to incorporate the fact that people can belong to two castes at the same time: a personal one and an official one. There are still very few to no inter-caste marriages. In the case of an inter-caste marriage the wife sometimes takes over the caste of her husband. Sometimes they both keep their own caste and the children end up without a caste. To incorporate these cases I have asked for the caste/religion at birth, the current (personal) caste/religion and current (official) caste/religion. In addition, I have moved this section towards the end of the identity section, as to begin the questionnaire with such sensitive questions did not seem like a good idea.

I also rewrote the next set of questions (on perceived ability, risk aversion, impatience and degree of fatalism). The goal of this section was threefold: (i) understand how the respondent views himself with respect to his relevant information network in terms of qualifications related to farming and education, (ii) understand how risk averse the respondent views himself with respect to his relevant farming information network, (iii) get a measure for time preferences. For the latter, I was inspired by the recent work of Loewenstein et al. (2001). They distinguish (i) impulsivity, (ii) planning/compulsivity, and (iii) behavioral/emotional inhibition. Ideally, these elements should be measured on an absolute scale to allow comparison between villages, but it seems that it is easier to get variation if one measures these on a relative scale (compared to others).

It appeared impossible to ask these perceived identity questions without a benchmark. For instance, if one asks 'how often do you have doubts related to agricultural practices?', all respondents chose 'sometimes', or if one asks the degree to which one agrees with a statement, the respondents invariably respond 'somewhat'. As such, to get after perceived ability, I decided to include in the sentence 'with respect to others'. This made a huge difference in the way these questions were answered, and I managed to get a decent amount of variation in the answers. Thinking about this inclusion of 'with respect to others' in the question, I initially thought that this was not a good idea, as in this way, I would not be able to compare the farmers across villages as 'the others' tend to refer to the other villagers. On the other hand, even when the question is asked in a non-comparative way, people tend to think in a comparative way in their mind, and as such, it is probably a good idea to make this comparison explicit. It appeared not possible to get after who 'the others' were in these questions. I then tried to answer the question myself, and indeed the 'others' were difficult to define. From that moment, I started to use the following ground rule: if I can't answer the question, I won't ask the farmer, and as such, I did not bother asking about 'the others' any more.

With regard to the time preference questions specifically, Loewenstein et al. (2001) suggest two ways: one is to ask questions related to the individual's character, the other way is to ask questions related to the individual's behavior. The problem with the behavior questions proposed by them is that none translates well in the Indian context. The impulsivity and inhibition character questions did work, and were a source of enjoyment and laughter for the respondents.

However, one should be careful in the interpretation of the impulsivity question. People sometimes told me that they often 'had' to make decisions quickly (even though they do not always like to make decisions quickly). I thought of changing the question to 'compared to others do you enjoy making decisions quickly', but even though I never tried this version out (as I only thought of this after I came back), I fear that this formulation will make the question too complex. The planning question was problematic. All people say that they 'somewhat' plan ahead and using the agree – disagree scale, the majority ended up in the middle. Similarly, all respondents told me that 'if they make a plan, they stick to it'. When I then decided to ask planning ahead towards specific things, like a lump-sum investment for agriculture, wedding daughters, saving for education etc, not all people had experienced these and in addition, many people answered these questions by referring to their financial capacities (so instead of capturing the 'intentions', we end up capturing the 'constraints'). I then finally decided not to ask questions related to certain behaviors, as not all respondents could relate to these, but to stick to the general attitude question related to planning, but to add a 'with respect to others' as well.

Rewriting the Bt cotton survey

Related to the landholding section, the ownership status was more complex than I thought. The share-in/share-out plots don't only share outputs but often also inputs. If the landlord shares the input he will get 1/2, if he shares only output he will get only 1/3.

I decided to add functionality of the irrigation source in the recall section rather than in the landholding section. The reason is that while irrigation source is mentioned in the VLS data of the last 7 years, the functionality of these sources is not mentioned. It is not uncommon that irrigation is present but the amount of water available is insufficient. For instance, in the case of canal irrigation, the canals are often dry, and in the case of well irrigation, the groundwater percolation into the well from the aquifer might be too slow. People tend to recall whether the irrigation source they had was functional.

I dropped the question in the random-matching-within-sample excercise on whether people were willing to talk to someone, as everyone always answered yes to this. I tried to go back with several of the questions in time (seven years back) to see whether the relationship and/or position field changed, but I always got a negative answer on this. So in the interest of time and space, I decided to drop the recall questions in this module. I had to include the option that the other farmer belongs to a household, and sometimes even though the two farmers don't talk, their wives do. Surprisingly, the farmers tend to know a lot of details about each other's plot, like soil, and irrigation facilities; even recall questions worked fine here (see Hogset and Barrett 2010 on the accuracy of self-reported peer behavior). I decided to drop the recall questions though, also in the interest of time and space.

In the input/output schedule, I decided to ask for less details related to the dates of the operations as people do not tend to remember these. They tend to remember the week and month of sowing, and the week and month of harvest. The other field operations, people tend to remember in broad intervals ('about

half a month after I did operation X, I did operation Y'). Also, land preparations are often done in stages, and people do not tend to remember when they did these (other than the months).

In the section on additional recall details on cotton production, I dropped the question on the level of bollworm attack, as how the farmers perceived this depends on whether or not he used Bt cotton and pesticides. I also dropped the question on the producer of the seed as the farmer does not know the answer to this in general. In addition, the farmers in general do know whether their seed is legal or illegal; as such I dropped that question as well. In Aurepalle, the farmers also explained me that they often do not buy pesticides together with the seed. Many of them only buy pesticides at a later stage, when they need it. They told me that it is dangerous to keep pesticide, which is essentially a poison, in the house with small children around. I also dropped the question on whether the yield of cotton was more or less than expected during the last seven years (in the section on beliefs, the last section). The reason is that we would need the expectation with respect to what they thought at the beginning of the season. People do not remember exactly what they thought at the beginning of the season and answer these questions in a back-ward looking fashion: judging now whether that was a good or bad year. It is doubtful that this will as such generate any useful information.

After the training round, I decided to give the questionnaire one more revision. I rewrote the asset schedule on machinery based on the World Bank's Living Standard Survey questionnaires (Gross and Glewwe 2002). Then, I also decided to split up the machinery into: big investments and small investments.

I did this because almost all farmers have smaller machinery and it won't be possible for them to recall the details, as they bought these items several years ago.

In the random-matching-within-sample excercise, I changed the formulation of the question to incorporate the possibility that it's not always the household heads that talk to each other, but also their wives, etc.

Finally, I changed the elicitation of the yield to incorporate the case that the respondent does not have any land and is not farming. In this case a hypothetical plot will be used of average soil quality.

Rewriting the education survey

I realized on the field that the concept of household for the education questionnaire was not trivial. In this survey, we go back 25 years in time, during which people have moved in and out, people have died, and new members are born. Imagine a joint family as in Figure A1. In this family any one of the males could be the respondent, and, as such by referring to the sons and daughters of the respondent, one might not capture the relevant children.

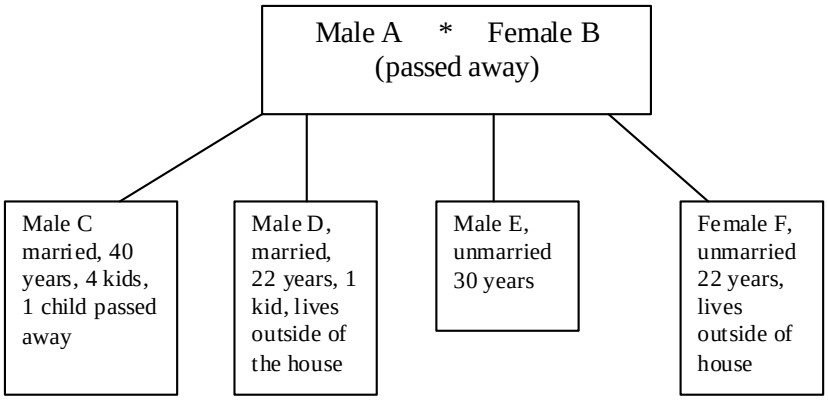

Figure 4.2: Complex family structure

I had to carefully think about who we need to include for what purpose. There are four sections of the questionnaire involved where the structure of the household matters: the identity and household composition section (section II), the wealth section (IV), the education of the older members (V) and the education of the younger members (VI). I first decided on the following structure. Section II (identity) includes all members included in Section V and Section VI. Section IV (wealth) includes all members included in Section V and Section VI and the members who used to live in the house and passed away. Section V (education of older members) includes all members older than 25 who currently belong to the household or who used to belong to the household and also all daughters in law. Section VI (education of younger members) includes all members younger than 25 years who currently belong to the household or used to belong to the household.

The wealth section (section IV) includes recall questions on assets on wealth. This is needed to reconstruct income as the VLS only has income data from 2001-

2007, while my education panel data structure goes back to 1985 for the oldest kids. I decided upon the following wealth variables: irrigable land, dryland, the number of working male/female adults/children. This recall section follows the household of the respondent, assuming this is the household in which the children of the household (of section VII) are growing/grew up.

For the education variable, we need to explain the education level of all children of the household under the age of 25. This includes the young adults who have moved out by now, but who used to live in the household. For instance, this includes female F and male D. In theory, this includes also the children who would have been between 0 and 25 now but who passed away. But talking about children who passed away is difficult, and as such, we should keep these questions to an absolute minimum. So I decided to transfer all the questions related to the family members who passed away to a small section in the wealth section (section IV).

During the training round, the enumerators found it difficult to work with this structure. As such, I decided to focus section II only on current household members and past household members under the age of 25. As such, we are sure to catch all the kids in the relevant age group. The basic information on the current household members older than 25 is still asked in section V, but the past family members are dropped – instead this information is directly included in the social learning section. Section VI remains the same as does section IV.

Section V of the questionnaire includes the education information of the current and past family members older than 25 and the daughters/sons in law. I do

not explain these educational choices, but need to collect sufficient information on these members to serve as control variables. I tried out several additional questions here on details of education. I was surprised by the details the respondents could remember of their own education 30 to 40 years ago. They remembered the schools they had been to, how they commuted, how well they did in school, and whether they attended frequently. I did not bother asking them about the cost, as it were their parents who paid the costs, I doubt that reliable information could have been obtained. A selection of all these possible questions had to be made, keeping in mind the goal: control for unobservable genetic (ability) factors and unobservable parental inputs in the child's education. I could do the Raven's matrices test which captures innate ability (formal schooling is supposed not to affect the outcomes) or even a cognitive achievement test of the respondent. However, these tests are bound to take a lot of time, and due to practical reasons we will only be able to do one of the parents. As such, I decided against this. An imperfect control of the parent's ability is their education level. This education level can also serve as an imperfect control of the parent's inputs in the child's education. I added a question on employment. As the immediate family is the first information neighborhood, the answers to these questions can be used to construct the social learning term.

Section VI is an extensive section on education of the current and past household members between 0 and 25 years (excluding family in law). The majority of these questions worked well on the field.

Regarding social norms, I tried several questions in this section: 'how did your environment react when you enrolled the child in this school', if the stu-

dent attended irregular 'did anyone come and say something about this to you"). These questions all worked fine. Past hypothetical questions such as 'your child attends frequently, would someone have come and said something if he did not?' does not work. In general, forward looking hypothetical questions work, as one can still (with some effort) imagine such a thing to happen, but to imagine what would have happened if the past was different doesn't work. I decided to drop the question for children who are currently enrolled: 'Did you ever think of stopping this child's education'. Even though this question gave interesting responses, I will just capture a few interesting stories, but bother all of the respondents with a longer questionnaire.

Even though I know that discrimination in school of lower castes and girls used to be very much an issue, I decided to drop these types of discrimination questions. I asked these questions during the trial round and most parents say things were quite all right in this respect. But, I had doubts related to the translation. The enumerators themselves did not believe there could be discrimination in the schools, and as such, their view might have (unintentionally) influenced the average response we got.

I decided to add a section VII with all the remainder questions that have to be asked for each individual member between the age of 0 and 25. These questions are not conceptually related, but still I've put them together as this section will need to be asked to each member, so from a practical point of view, this was more convenient. These include additional questions on exam results, and dowry and marriage. The marriage questions of the trial questionnaire were fine, but the dowry questions had to be adapted.

The parents do not think in terms of gifts to the couple and gifts to the family, but rather in terms of gifts to the child and gifts to the family. In addition, it seems impossible to distinguish gifts from dowry. Dowry is often described as a large amount, and an amount that has been agreed upon, while gifts are typically of a smaller value and were not discussed before hand. My estimate would be that for the poorer families the majority of the gift and dowry expenses are in the form of a dowry, while for the richer families, there are mainly gifts. Dowry can be paid in cash, but this is not necessary. Gifts are mainly non-cash gifts. I decided to drop the question on 'on which do you decide first dowry or education, as it's seems to be an alternating decision-making process which can not be discovered by asking such a relatively simple question' Naturally, the dowry decision comes later, physically speaking, but people keep the dowry in mind when deciding on the education level.

The last part of section VII included a health question and child-dependent expectations. I tried out several different versions of the stone game on the field. The first way was 'if this child finished 12th what are the chances of finding a job?'. The respondents interpreted this question as 'what are the chances of finding a salaried job?' Indeed, in the end, everyone works, as one cannot afford not to work. The respondents had no problems with 'double' conditional questions such as 'if this child finishes 12th standard and the child finds this (salaried) job, what is the minimum that you would expect her/him to earn?' I also tried questions for specific 'majors' of study, such as 'if this child gets an engineering degree, what are the chances of finding a job as an engineer?' and the double conditional one, 'if he or she finds that job, what is the minimum/maximum

earnings to be expected'. The question 'if he/she gets an engineering degree what is the minimum/maximum earnings to be expected', works as well. I was only able to go up to three categories in the stone game. It seems that parents think less of the risky-ness of the educational investment than they do about the risky-ness of their agricultural activities. More than three boxes resulted in confusion, with uniform distributions, stones left over or all stones in one box.

I thought quite a bit about the issue of power. Intuitively, it seemed to me that the concept is related to the concept of social networks. Two kinds of power influence are of interest to me: power to influence the education admission decision, and power to influence the job-hiring/earnings. The source of this power is the position in a social network and/or interlinked markets. To understand this power of a particular individual, one would have to look at the web of networks surrounding this individual. This is clearly not feasible in this study. I did include however a question on whether 'influence' was used to enroll the child in a particular school in section VI. It seemed not possible to include questions on whether power was used to obtain particular jobs, as the number of jobs that the family members could have had over the last 25 years is just too high. I also decide to drop the power question in the open question on social learning ('who would you go to') as whether or not the influence of those people is relevant very much depends on the kind of job they are applying for.

Section VIII aggregates all remaining household level questions on social learning, child ability, and contribution of the children during old age.

Related to social learning, one of my main underlying assumptions appeared incorrect: the population of the information networks related to costs and benefits of education could not be approximated by the village. I found this through questions such as 'who would you go to if you need information related to education?' and 'how many people do you know (by first name) that have a BA degree, and how many of these live in the village?' As such, the random matching within sampling approach is not the right approach any more to get after the information networks. I therefore tried out several alternative approaches.

First, I wanted to get some control variable for the information coming from institutional sources. It is of course impossible to ask for every event when they received information from media, etc. I asked the question during the trial round in an open manner, as people answered this question telling me they 'never' or 'sometimes' received information from the source I mentioned, I decided to frame the question as such using these options.

Second, one can ask the respondent approximately how many people he knows who finished Xth standard, XIIth standard, diploma, bachelor's, master's etc. The possible answers are none, one, two, three, four, five, between five and ten, between ten and twenty, between twenty and fifty, between fifty and one hundred, etc. Ask the respondent what the approximate share is relatives and people in the village. As the number of people that the respondent tends to answer is too high, asking detailed questions about the properties of these people is not possible.

Third, ask the respondent more position-based questions, like 'How many people do you know with job X?' By phrasing the question in a more specific way, people are better able to recall the information. In addition, these questions might help in distinguishing learning within the several higher education careers.

Fourth, I added an open question (hypothetical question) on who they would go to if they need information related to education and jobs. This open question is mainly relevant for the future decisions and in addition, will give some idea of the structure of the strongest links in the information network. In this section, I dropped the question on 'would you have gone up to the person 5 to 10 years ago etc?' as the answer to this question was always yes if they knew the person at that time, and no otherwise (during the trial round). I did not distinguish the different education levels here, as everyone interviewed during the trial round here points at the same set of people. In this section, dropped the question on learning about job opportunities (actual job openings) as the occasions are too numerous to count here, a different kind of questionnaire and focus would be needed to justify these questions (in addition in this case, people do hear from different people depending on the job). Naturally, questions on the actual use of job-information networks of the currently employed members would lead us too far once more, so these questions were not included either.

BIBLIOGRAPHY

[1] Grosh, Margaret and Paul Glewwe (eds). 2000. *Designing Household Survey Questionnaires for Developing Countries, Lessons from 15 Years of the Living Standard Measurement Study*. The World Bank, Washington DC.

[2] Hogset, Heidi and Christopher B. Barrett. 2010. "Social Learning, Social Influence, and Projection Bias: A Caution on Inferences Based on Proxy Reporting of Peer Behavior." *Economic Develoment and Cultural Change*, 58:3, pp. 563-589.

[3] Loewenstein, George, Roberto Weber, Janine Flory, Stephen Manuck, and Matthew Muldoon. 2001. "Dimensions of Time Discounting." Presented at conference on Survey Research on Household Expectations and Preferences, Ann Arbor, MI, Nov. 2-3, 2001.