GRAIN BANKS: AN INSTITUTIONAL AND IMPACT EVALUATION

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

Ruchira Bhattamishra

January 2008

GRAIN BANKS: AN INSTITUTIONAL AND IMPACT EVALUATION

Ruchira Bhattamishra, Ph.D.

Cornell University 2008

This dissertation consists of an institutional and impact evaluation of grain banks in southwest Orissa, India. Grain banks are village-level institutions which provide loans in the form of grains to help households smooth food consumption over the agricultural cycle.

In the first chapter, I discuss grain banks in the context of the food insecurity situation in Orissa and compare them to similar interventions in India and elsewhere.

In the second chapter, I examine the key institutional and socioeconomic determinants of the likelihood and duration of grain bank survival. This study is motivated by the fact that the majority of grain banks have ceased to function over time. I find that a number of village-level factors have a significant impact on the likelihood and duration of survival, indicating the importance of the socioeconomic environment where grain banks are implemented and the need for careful geographic targeting for improving sustainability.

In the third chapter, I examine the impact of grain bank participation by households on young children's health outcomes using propensity score matching. This study is motivated by the gap in knowledge on the impact of grain banks on household and individual food insecurity. I also examine whether this impact varies by the lifespan of the grain bank to test the hypothesis that children in participating households in longer-lived grain bank villages benefit more due to intergenerational effects. I find that grain banks do not have a statistically significant impact on various anthropometric measures examined, indicating that they may not be effective in improving children's health status.

In the fourth and final chapter, I examine the impact of grain bank participation by households on the incidence of borrowing from private, informal moneylenders, again using propensity score matching. This study is motivated by anecdotal evidence that grain banks displace borrowing from moneylenders. I also examine how this impact varies by the lifespan of the grain bank. I find that grain banks have a large, statistically significant displacement effect on moneylenders, and this effect is even larger in longer-lived grain bank villages.

BIOGRAPHICAL SKETCH

Ruchira Bhattamishra was born and grew up in India. She received an A.B. degree in economics from Bryn Mawr College, USA in 1998. She also received a M.Sc. degree in Politics of the World Economy from the London School of Economics and Political Science (LSE), UK in 1999, and a M.A. degree in economics from Cornell University, USA in 2006.

Dedicated to the memory of Dada, Jiji, Aja and Aai

ACKNOWLEDGMENTS

I would like to thank my Special Committee members, Ravi Kanbur, Christopher B. Barrett and George Jakubson, for their guidance and encouragement. I would also like to thank Agragamee, in particular Achyut Das and Vidhya Das, for their support and assistance towards the field research, as well as the inhabitants of the survey villages in Kashipur and Dasmantapur who generously shared their time and information with the survey team. Funding for the data collection was provided by the National Science Foundation; the Mario Einaudi Center for International Studies, Cornell University; the Department of Applied Economics and Management, Cornell University; and the Graduate School, Cornell University. I would like to gratefully acknowledge all the above sources of funding. Finally, I would like to express immense gratitude to my parents and husband for their constant support and encouragement. Responsibility for all findings and any errors are my own.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

CONTEXTUALIZING GRAIN BANKS

1.1 Introduction

This study consists of an institutional and impact evaluation of community grain banks in the southwest part of Orissa, a state in the eastern part of India. This region, which has a high proportion of Orissa's tribal population, is one of the most economically depressed and food-insecure parts of the state. In recent years, community grain banks have become an important component of the food security strategy of the government as well as rural development NGOs active in this region. Described simply, a grain bank is a village-level institution which allows households to borrow grains during the lean season which are returned with interest (also in the form of grains) at the conclusion of the following harvest season. By extending consumption credit, grain banks provide households with the opportunity to smooth food consumption over the agricultural cycle. Grain banks also provide an alternative source of credit to households in a region with limited credit market competition. Anecdotal evidence collected initially suggests that grain banks have had a displacement effect on borrowing from local informal moneylenders, who have traditionally offered credit to households at unfavorable terms.

The main aims of this project are to examine (1) the institutional and environmental factors behind the continuing operation of community grain banks, given that a large proportion of grain banks have ceased to function following their establishment; (2) the impact of household participation in grain banks on young children's health status using anthropometric measurements; and (3) the impact of

household participation in grain banks on the incidence of borrowing from local informal moneylenders.

The remaining sections of the chapter are structured as follows. In Section 1.2, we discuss the food security situation in India as well as the performance of existing food security programs, followed by a reexamination of the same issues with respect to the state of Orissa, and tribal Orissa in particular. In Section 1.3, we describe the grain banks in Orissa and compare them to similar interventions elsewhere. In Section 1.4, we provide an overview of the data used for the study. Finally, in section 1.5, we present a brief summary of the questions posed in the research project, the motivations for the questions, the empirical methods applied to address them and the main findings.

1.2 An overview of the food security situation

Food security can be broadly defined as the condition of having access to adequate nutrients in order to lead a healthy life. The concept of food security has evolved over time, with earlier definitions looking at the availability of food at the aggregate level to recent, more disaggregated definitions, both static and dynamic, that take household and individual-level access to food into account (Barrett 2002).

In this section, we first present an overview of the food security situation in India, and discuss briefly the issues of both availability and access to food. We then discuss the implementation of the main government programs that address food insecurity in India generally, and in Orissa specifically.

1.2.1 Food security in India

India has been self-sufficient in food grain production since the 1970s.[1] In 2002, government of India buffer stocks stood at 60 million tons, more than three times the minimum buffer stock norm of 18 million tons (Ministry of Finance 2003). However, regional production patterns vary widely. For example, the northern states of Punjab, Haryana and west Uttar Pradesh have much higher levels of production than the rainfed central and eastern states which include Orissa. Regardless, adequate food availability at the national level does not necessarily imply adequate food access for all, as evident from recent household surveys on nutrition and health. For example, the National Family Health Survey (NFHS 1998-99) finds that 47 percent of children below 3 years are underweight and that 40 and 37 percent of adult females and males are chronically energy-deficient, respectively.

Insofar as income is seen as an important determinant of access to adequate food, it is useful to look at recent income poverty trends in India. Adjusted estimates by Deaton and Dreze (2002) based on National Sample Survey (NSS) data indicate that there has been a decline in rural and urban poverty from 33 to 26 percent and 18 to 12 percent during the 1990s, respectively. Recent poverty estimates have however been mired in several controversies; consequently, the reliability of the poverty figures have been called into question.[2] Notwithstanding, income poverty estimates as well as the extent of contribution to the overall decline in poverty vary significantly across regions and social groups with eastern and central states, which have a higher proportion of scheduled tribes (ST), showing smaller declines in poverty (ibid.).

---

[1] We look at food grain production and access as indicators of food security, since grains constitute the largest component of caloric intake, especially in poor households. This is discussed later in this section.

[2] The main issues of contention appear to be the use of the established 'caloric norm' basis for poverty measurement as well as the use of different recall periods in consumption expenditure surveys used in different rounds of the NSS. There is also some concern that the poverty calculations may have been flawed. See Deaton and Kozel (2005) for more on these issues.

Even if it is agreed that poverty has fallen overall, the question that naturally arises is: Has the decline in poverty translated into improved access to food and greater food security?[3] Using anthropometric measures, estimates from the NFHS show that the percentage of underweight children fell from 53.4 percent in 1993-94 to 47 percent in 1998-99.[4] Similarly, BMI estimates from the National Nutrition Monitoring Bureau (NNMB) show that the proportion of chronically energy-deficient (CED) adult females declined from 45.8 in 1999 to 39.4 percent in 2001. While these declines in the incidence of malnourishment are consistent with declines in income poverty, the absolute proportion of underweight children is much higher than the official and unofficial income poverty estimates for that year. The finding is effectively the same if we take the proportion of CED adult males or females or stunted children.

The food security picture is even bleaker if we consider micronutrient adequacy. For example, estimates from NFHS 1998-99 reveal that 74.3 percent of children between 6-35 months and 51.8 percent of women between 15-49 years suffer from anemia. These figures are much higher than the income poverty figures for the same period, which again points to the fact that adequate food production and access to calories are not sufficient for ensuring adequate micronutrient intake and absorption. As mentioned before, dietary diversity as well as intervening environmental factors such as hygiene and sanitation, safe drinking water, primary health care, and hygiene awareness play an important role in achieving food security.

---

[3] The NSS collects data on self-reported food adequacy, by asking a question to the head of the household regarding adequacy of food for all household members in the past year. According to NSS data, households reporting that they have not had adequate food for some part or all of the past year have also declined systematically between the 1980s and 1990s, from 18.6 percent in 1983 to 5.1 percent in 1993-94 and 3.4 percent in 1999-2000 in rural areas. The corresponding numbers for urban areas are 6.4 percent, 1.6 percent and 0.9 percent respectively. However, given that these figures differ hugely from more objective data on malnourished children and adults (using anthropometric measures, such as in the NNMB, NFHS, etc.), the NSS food adequacy data cannot be considered credible.

[4] The 1993-94 NFHS figures are for children below 4 years, while the 1998-99 figures are for children below 3 years.

Household surveys of the type conducted by the various major data collection agencies typically do a poor job of capturing the problem of seasonal food insecurity. As discussed in Zeller et al. (1997) and Sahn (1989), peasant societies that are characterized by seasonal variability in agricultural output are also often highly vulnerable to seasonal variability in consumption, which has adverse impacts on health and asset holdings both in the short- and long-term. For example, using ICRISAT data for southern rural India, Behrman and Deolalikar (1989) find that reduced food intake during the lean season has a negative impact on future agricultural productivity. Similarly, Lawrence et al. (1989) find for the Gambia that seasonal food shortages lead to losses in female body weight as well as low birth weights. Clearly, seasonal food shortages can negatively impact health and productivity and can create a perpetual cycle of low productivity, low output and poor health. Zeller et al. (1997) discuss the importance of extending consumption credit in order to enable agricultural households to smooth consumption and break the cycle of low productivity, depleted asset holdings, low output and food insecurity. They cite empirical evidence from rural Nigeria, China, Pakistan, the Gambia and Madagascar which suggests the importance of credit for consumption purposes, particularly among poor households.

1.2.2 Government interventions to promote food security

The government of India has established a host of programs designed to improve access to food, including three of the world's largest food security programs in the form of the Public (Food) Distribution System (PDS), the Integrated Child Development Scheme (ICDS) and Mid-Day Meals Scheme (MDM). However, none of these programs were designed with seasonal food insecurity in mind; if anything, the intended objective of these programs was to address perennial food insecurity.

Below, we provide a brief overview of these three programs, describing their goals, important design features, and implementation issues and performance.

The PDS, which comprised around 455,000 fair price shops (ration outlets) in 2000 (Ministry of Food and Consumer Affairs 2000, as quoted in Umali-Deininger and Deininger 2001), distributes mainly subsidized rice and wheat worth over Rs. 150 billion to around 180 million households (Asthana and Medrano 2001). In 1997, the PDS was transformed from a universal to a targeted program; this was done largely in an attempt to stem leakages and to reduce the fiscal burden of food subsidies. Under the targeted PDS, households were classified as either Above Poverty Line (APL) or Below Poverty Line (BPL) with BPL households entitled to 35 kilograms of grains per month at specially subsidized prices. The reforms also included changing PDS rations from a per-adult equivalent basis to a per-family basis. As discussed in Swaminathan (2000), the monthly entitlement for a BPL household with 5 persons translates to 35 percent of the caloric intake norm recommended by the Indian Council of Medical Research.

Due to its high procurement, storage and distribution costs, the PDS cannot function without a hefty central government subsidy, which has nearly tripled in real terms over the last two decades, from Rs. 29.4 billion in 1980/81 to Rs. 90 billion in 1998/99 (Umali-Deininger and Deininger 2001). Imperfect targeting, poor coverage, the misappropriation of food grains, corruption, the irregular supply in fair price shops and the low quality of food grains are some of the major problems that prevent the PDS from achieving its objective of providing food security for the poor.[5] The diversion of grain from the PDS to the open market is estimated to be of the order of 36 and 31 percent in the case of wheat and rice, respectively (Asthana and Medrano

[5] See, e.g., Dreze 2001, Surayanarayana 2001, and Kriesel and Zaidi 1999 (as cited in Umali-Deininger and Deininger 2001).

2001). Food stock deterioration is also a serious problem for the Food Corporation of India (FCI) (the parastatal body in charge of running the PDS), with about half of its stocks being more than 2 years old in 1997 (Sinha 1997, as cited in Umali-Deininger and Deininger 2001). Moreover, as FCI stocks also feed into other government programs such as employment schemes, ICDS and MDM programs, the impact of poor quality FCI grains is not confined to the PDS alone.

The ICDS and MDM schemes also aim to address the nutritional needs of members of poor households. Initiated in 1975, the ICDS has universal coverage (at least in principle) and provides free services with the intent of addressing the nutritional, health care and educational needs of children under the age of 6 as well as the needs of pregnant or nursing mothers and young girls (Radhakrishna et al. 1997, Gupta et al. 1998). Under the ICDS program, an *anganwadi* center, literally a courtyard, is made available in each settlement of 1,000 people (in the case of tribal areas or rough terrains, in settlements of 700 people). The *anganwadi* is the focal point of delivery of ICDS services, which, among others, include a cooked mid-day meal that provides 300 calories and 10-12 grams of protein to children and 500 calories and 20-25 grams of protein to mothers. By locating *anganwadis* in poor localities, typically in villages or urban slum areas, the ICDS employs self-selection as its targeting mechanism. However, as discussed by Gragnolati et al. (2006), states which are poorer and with the highest rates of undernutrition, have the poorest coverage and the smallest government budgetary allocations per malnourished child. In addition, Das Gupta et al. (2005) find that a larger proportion of wealthier villages have the ICDS program than do poorer villages, indicating that the ICDS may not be reaching all its intended beneficiaries. Among other things, the ICDS is also characterized by poor quality services, poor quality of equipment for weighing and

growth promotion, irregular food availability and overburdened *anganwadi* workers (Graganolati et al. 2006).

Although the MDM program was formally initiated in 1995, it was not implemented in the majority of states until 2001 following a Supreme Court order. At present the program covers about 50 million children in primary schools. However, it is still to be introduced in some states such as Bihar, Uttar Pradesh and Jharkhand (Dreze and Goyal 2003). Under the MDM program, a nutritious cooked mid-day meal with a minimum content of 300 calories and 8-12 grams of protein is to be provided each day of school for a minimum of 200 days per year in all government and government-assisted primary schools. PDS grains are provided free of cost by the central government. Like the ICDS, the MDM program also uses self-targeting as the mechanism to reach children in poor households. Unfortunately, both the ICDS and the MDM program are faced with a host of problems, including leakages to non-priority groups, poor infrastructure, lack of universal coverage (due to incomplete reach), irregular food availability, and large differences in the quality of meals across and within states.[6]

1.2.3 Reduction of cereal consumption in the 1990s

Given that the provision of cereals (such as rice and wheat) is the focus of food security programs in India, we now turn to a brief discussion of the importance of cereals in the diets of low-income households in India. As noted in Suryanarayana (2000), cereals account for more than 85 percent of caloric intake in poorer households, especially in rural areas. In addition, in terms of food expenditure, cereals constitute about 60 percent of total expenditure for these households. Although estimates based on NSS data indicate a decrease in average cereal consumption over

[6] See, e.g., Dreze and Goyal 2003, Graganolati et al. 2006

the 1990s, Deaton and Dreze (2002) contend that this trend is largely driven by a decrease in cereal consumption and substitution towards 'superior food items' (such as vegetables, milk, fruit, fish and meat) among higher income households.[7] Hanchate and Dyson (2000), as cited in Deaton and Dreze (2002), compare rural food consumption patterns in 1973-74 and 1993-94 to show that while average per capita cereal consumption declined in this period, it actually rose amongst the poorest households. Thus, the overall decline can be attributable to a reduction in cereal consumption among higher expenditure groups. Deaton and Dreze (2002) also find a similar relationship at the state level, with poorer states such as Orissa and Bihar having higher cereal consumption than richer states such as Punjab or Haryana. We can thus conclude that cereals account for the largest share of calories consumed by low-income households and that this share appears to be holding steady among low-income households. In light of these facts, cereals (should) occupy an important place in any food security intervention for the poor.

1.2.4 Food security in Orissa

In this subsection, we provide an overview of existing food security programs in the context of the food insecurity problem in tribal Orissa. According to estimates based on the 50th round of the National Sample Survey (NSS), 15.4 percent of the population of Orissa reported not having two square meals a day for part or whole of the year in the year preceding 1993-94 (NSSO 1997). More objective indicators of nutritional status, such as anthropometric measures, indicate an even more serious problem. For example, National Family Health Survey (NFHS-2) estimates indicate that the infant mortality rate in Orissa between 1994-98 was 81.0, compared to the national average

[7] As discussed in Deaton and Dreze (2002) average cereal consumption per capita fell from 13.5 kg per month to 12.7 kg per month in rural areas and from 10.6 kg to 10.4 kg in urban areas between 1993-94 and 1999-2000.

of 67.6.[8] Similarly, the proportion of underweight children in Orissa in 1998-99 was found to be 54 percent, compared to the national average of 47 percent. To cite another example, 48 percent of women between 15-49 years of age were found to be chronically energy deficient in 1998-99, compared to the national average of 35.8 percent.

While Orissa fares poorly on a range of social and economic development indicators compared to the rest of the country, the state of social and economic development of the tribal population within Orissa is considerably worse even compared to the Orissa average. Recent data on adequacy of food intake collected by the NSS reveal that while 8 percent of the rural population in Orissa reports not having two square meals a day for part or whole of the year in the year preceding 1999-2000, the corresponding number for the tribal population is 10.1 percent (NSSO 2001). Table 1.1 presents a range of health-related indicators by socioeconomic groups, all of which consistently show that the tribal population in Orissa fares worse compared to all other socioeconomic groups in the state. While the infant mortality rate (or the probability of dying before the first birthday) for Orissa is 89.5 per 1,000 live births, the corresponding figure for the tribal population is 98.7.[9] Similarly, the child mortality figure (or the probability of dying between the first and the fifth birthdays) in Orissa is 28.8 per 1,000 live births, while the corresponding figure for the tribal population is considerably higher, at 44 per 1,000 live births. As discussed by IIPS and ORC-Macro (2000b), one of the main objectives of the child health-care system in India is to achieve universal immunization coverage against six preventable diseases (namely, tuberculosis, measles, polio, diphtheria, whooping cough and tetanus) under the Universal Immunization Programme (UIP). NFHS-2 data from Orissa, however,

[8] Infant mortality rate calculated per 1,000 live births.
[9] Infant and child mortality rates calculated for the 10-year period preceding the NFHS-2 (1998-99) survey.

reveal that by the age of 23 months, only 43.7 percent of children had received all childhood vaccinations, with the corresponding figure for tribal infants even lower, at 26.4 percent. A related objective of the UIP was to record all vaccinations received on a card. However, according to NFHS-2 figures, less than half of children in Orissa below 23 months had a vaccination card at the time of the survey, while the corresponding figure for tribal infants was below 30 percent. Similarly, the tribal population consistently performs poorly, compared to the all-Orissa average, on other indicators of health (such as vitamin A supplementation), health-related knowledge (such as knowledge of diarrhea treatment using oral re-hydration) as well as taking up treatment by going to a health facility/provider.

Table 1.1: Selected health indicators in Orissa, by socioeconomic group

Health indicator	Scheduled Tribe	Scheduled Caste	Other Backward Castes	Other	All-Orissa
Infant mortality [a]	98.7	83.9	95.6	79.1	89.5
Child mortality [b]	44.0	42.4	20.1	15.0	28.8
Percent of children age 12-23 months who received all childhood vaccinations	26.4	44.5	48.5	49.3	43.7
Percent of children age 12-23 months showing vaccination card	29.7	44.7	49.5	56.1	46.2
Percent of children age 12-35 months who received at least 1 dose of vitamin A	30.5	42.1	45.8	46.3	42.0
Percent of mothers who know about ORS packets for treating diarrhea	53.7	76.5	76.8	82.1	72.9
Percent of children under age 3 who had diarrhea in 2 weeks preceding survey and were taken to health facility/provider	21.2	48.1	55.1	53.8	46.9

Source: NFHS-2.
Notes: [a] Probability of dying before first birthday, calculated on a per-1,000 basis, for the 10-year period preceding the survey. [b] Probability of dying between first and fifth birthday, calculated on a per-1,000 basis, for the 10-year period preceding the survey.

We now turn to a brief discussion of Orissa's tribal population, followed by an overview of the food insecurity problem in tribal Orissa.

Overview of Orissa's tribal population and food insecurity problem

Of the 533 communities currently recognized by the Indian government as Scheduled Tribes (ST) (Ministry of Tribal Affairs 2000-01), the largest number, 62 tribes, reside in the state of Orissa. With over 7.6 million tribal people in rural areas, Orissa also has one of the highest concentrations of tribal populations among Indian states.[10] While the ST population comprises about 8 percent of the national population, the tribal population comprises about a quarter of the total rural population of Orissa (Census 2001). More significantly, in many of the less economically developed districts in Orissa, in particular, in the southwest, the tribal population comprises over half the total population.

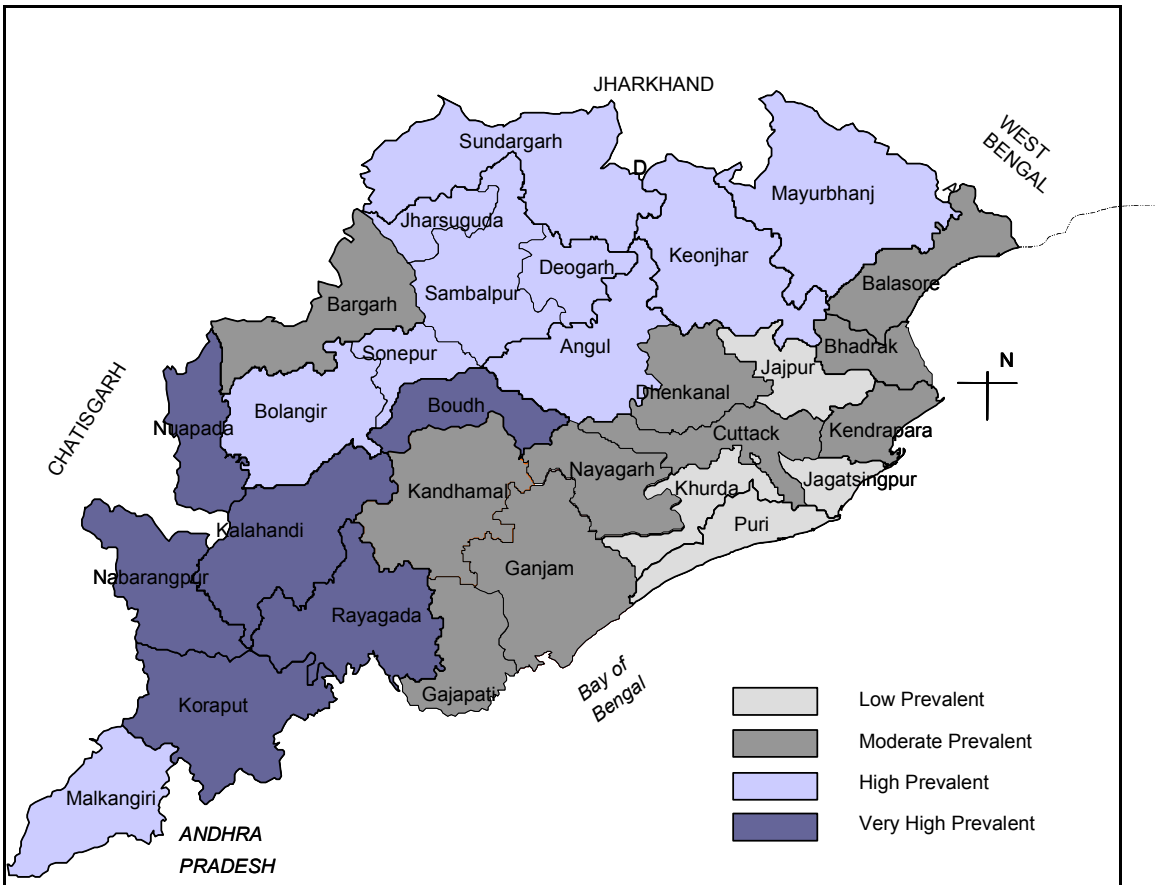
Until 1992, Orissa was divided into 13 districts. Of these, the KBK districts – Kalahandi, Bolangir and Koraput – in the southwest region comprised the tribal belt of Orissa.[11] After 1992, the 13 existing districts were subdivided into 30 smaller districts. The KBK region currently includes the 8 districts of Nuapada, Kalahandi, Bolangir, Sonapur, Rayagada, Nabarangbur, Koraput and Malkangiri.[12] The region accounts for 30.6 percent of the geographical area of the state but only 20 percent of the population of Orissa, indicating the low population density relative to other parts of the state. It is also one of the poorest parts in Orissa, as well as in India. This study will use household, village and grain bank data from two of the new districts –

[10]The tribal population in India is classified under the Scheduled Tribe (ST) social group as defined by the Census of India. These numbers therefore reflect the population classified as ST.

[11] The ST population also forms a majority of the population of Mayurbhanj district in the northeast and Sundargarh district in the northwest of Orissa.

[12] Erstwhile Kalahandi was divided into Nuapada and Kalahandi, erstwhile Bolangir into Sonapur and Bolangir, and erstwhile Koraput into Rayagada, Nabarangbur, Koraput and Malkangiri.

Koraput and Rayagada – carved out of erstwhile Koraput. Figure 1.1 shows a district map of Orissa denoting malnutrition levels. The map shows that the malnutrition rate is much higher in the southern KBK districts compared to the rest of the state.



Notes: Data from DWCD (2002-03) for age-group 0-3 years, based on data for Grades II, III, and IV malnutrition. The Districts are classified according to the prevalence of malnutrition. Districts which have a percentage of malnutrition lying 2-3 standard deviations (SD) below the overall mean are classified as 'Low Prevalent'; 0-2 SD below the mean as 'Moderate Prevalent'; 0-1 SD above the mean as 'High Prevalent' and 2-3 SD above the mean as 'Very High Prevalent' districts.

Figure 1.1: Prevalence of malnutrition in Orissa, by district

Table 1.2 indicates the extent to which the ST population of Orissa is concentrated in Rayagada, Nabarangpur, Koraput and Malkangiri districts, carved out

of erstwhile Koraput. Firstly, these districts together are home to almost a quarter of the total ST population in Orissa. The largest tribes include the Kandha, Paraja, Sabara, Gadaba, Bhumia, Koya, Durua, Matia, Bonda, Halwa and Didayee. In addition, while the ST population comprises 22 percent of the total population of Orissa, the ST population of these 4 districts comprises 54 percent of their total population (Census 2001).

Table 1.2: Orissa, KBK region and districts of erstwhile Koraput

	Total population	Share of population that is Scheduled Tribe (percent)	Literacy rate (percent)	Share of total workers who are cultivators or agricultural laborers (percent)
Orissa	36,804,660	22	54	65
KBK region	7,286,923	38	36	77
Koraput	1,180,637	50	30	73
Rayagada	831,109	56	30	75
Nabarangpur	1,025,766	55	28	83
Malkangiri	504,198	57	25	83

Source: Census of India 2001.

The population of erstwhile Koraput is predominately rural and characterized by low levels of development. In Table 1.2, we see that the literacy rates in the selected districts, and the KBK region as a whole, are much lower compared to the Orissa average. In addition, these districts also have a higher share of workers engaged as cultivators or agricultural laborers relative to the Orissa average, indicating the high degree of dependence on agriculture as a means of livelihood in the KBK region.

The KBK region spans the Southwestern Plateau (part of the Deccan Plateau) and a section of the hills of the Eastern Ghats (Meher 2001). The rugged terrain, together with the poor physical infrastructure in the region, makes access and communication a problem for many villages in this region. Electricity, telephone

cables and satellite coverage are virtually nonexistent in the rural areas. Many settlements in rural areas, which are characterized by extremely low levels of population density, are located away from motorable roads, limiting access to public and private social services and markets. During the monsoon season, some villages in this region are inaccessible as they are not connected by an all-weather passable road, indicating the extreme level of physical isolation of tribal settlements in this region.

In addition to the hilly terrain, the KBK region is also characterized by poor soil and low-productivity agriculture (Meher 2001). However, as evident from Table 1.2, there is a high level of dependence on agriculture, which is characterized by primitive agricultural tools and a lack of mechanization. *Podu* cultivation, or cultivation on hill slopes cleared by burning forest cover, is a common feature in this area, and has contributed to the increasing level (and rate) of deforestation in erstwhile Koraput. This has adversely affected the collection of forest products (such as *tendu* leaf, *sal* leaf, honey), which have traditionally formed a major component of the tribal economy along with shifting as well as settled agriculture. Agriculture in the KBK region is mainly rain-fed, although traditional irrigation systems such as *munda* (check dams) and terracing are also used. Few government-sponsored irrigation schemes are present, and recurring droughts in the region have adversely affected the economic circumstances of its people.

Poverty levels among the tribal population are higher than poverty levels for the general population. According to a survey conducted by the Panchayati Raj Department, Government of Orissa (as cited in Meher 2001), rural poverty in the KBK region is higher than in rural Orissa overall. For example, 91.6 percent of rural households in erstwhile Koraput were found to be Below Poverty Line (BPL)

compared to 79.1 percent of rural households in Orissa.[13] More significantly, more than half of these BPL households belonged to the tribal population. One of the most pressing problems faced by these households is the lack of food security.

Unfortunately, the performance of government schemes in reducing the prevalence and acuteness of food insecurity in Orissa in general, and tribal Orissa in particular, is abysmally poor. According to Shariff and Mallick (1999), only 4-5 percent of poor and ST/SC households used the PDS in 1990. As in the rest of the country, leakages, poor grain quality, inadequate ration quantity, as well as poor coverage are some of the main problems that plague the PDS in Orissa. Even after targeting reforms, there is a very small difference between purchase of grains from the PDS between eligible and ineligible households. Using data from the 55th round of the NSS (1999-2000), Dreze and Khera (2003) find that the average grain consumption from the PDS was 1.74 kilograms per person per month for the poor and 1.44 kilograms per person per month for the non-poor in rural Orissa. Many eligible households in rural Orissa, and the KBK region in particular, cannot participate in the program as they have not received the BPL cards required to avail themselves of low-priced foodgrains. For example, about a third of the households sampled in the household survey serving as the source of data for this study have not received the BPL card, in spite of being eligible. In addition, since the PDS monthly quota has to be purchased in a single installment, the poorest households are not able to avail themselves of their full quota as they lack the ability to pay a lump-sum amount. Again, using data collected for this study, we find that approximately three-quarters of the sampled households report not being able to afford their PDS monthly quota.

[13] A BPL household is defined as a family with an annual income below Rs. 11,000 which is roughly equivalent to a dollar a day (1992 rupees).

Although the coverage of the ICDS and MDM programs is higher than that of the PDS, the quality of services delivered by these programs is poor and susceptible to irregular grain supply and corruption (Misra and Behera 2001). For example, due to the poor physical infrastructure, hilly terrain and low population density in the KBK region, many *anganwadi* centers are not attended regularly by the workers. In addition, caste differences between the *anganwadi* workers (who are typically higher caste) and ICDS beneficiaries – tribal mothers and children – also result in non-delivery of important ICDS services, such as taking weight measurements of children to monitor health and development status. For example, in more than half of the villages sampled for the household survey in this study, ICDS beneficiaries reported not feeling 'comfortable' with the *anganwadi* worker.

In addition to the PDS and ICDS programs, a set of programs that indirectly promote food security includes wage employment and self-employment generation schemes. These include the Jawahar Rozgar Yojana (JRY) and the Employment Assurance Scheme (EAS). The latter has the explicit goal of providing 100 days of employment for constructing public works during the agricultural lean season. However, between 1992-98, these and other anti-poverty employment programs together covered less than 16 percent of BPL households in the KBK region, indicating that many poor households do not receive any assistance from these programs (Meher 2001). Even households that do receive assistance often do not receive 100 days of employment or do not receive wages for days worked. In addition, while public works schemes provide temporary employment during the summer months, they are not implemented during the monsoon months of the agricultural lean season.

For generations, the principal strategy used by tribal households confronted with food shortages in the lean season has been to borrow grains (or cash in order to

buy grains) from the local moneylender at exorbitantly high interest rates or to participate in bonded labor markets under exploitative conditions and terms. Other coping strategies commonly adopted in the lean season include distress migration, distress sales of livestock and other productive assets, and the consumption of undesirable forest food products such as wild roots and leaves.[14]

Unquestionably, there is a genuine and urgent need for superior interventions that smooth food consumption across the agricultural cycle for poor households. In this context, grain banks have been touted as a highly successful food security intervention, particularly with respect to addressing seasonal food insecurity. Over the past two decades, NGOs in Orissa have established grain banks in order to directly confront cyclical episodes of hunger and food insecurity. They have also been adopted by the government as an integral component of the food security strategy for tribal communities.

1.3 Grain banks: Concept and rationale

What are grain banks? Grain banks in tribal Orissa are a descendent of the traditional system of grain *golas* in tribal villages, where surplus grains post-harvest were collected into a common pool which was controlled by the village head and from which disbursements were largely discretionary. This system has long since disappeared. The version found on the ground now has appeared only in more recent times.

[14] The practice of turning to forest product collection from agriculture during the lean season is a common coping strategy in traditional societies that are not integrated into the modern economy. Scudder (1962) and Newman (1970) record similar customs among the Gwembe Tonga of Zambia and the Sandawe of Tanzania, respectively. Reardon and Matlon (1989) document the consumption of leaves gathered from common property resources as a famine food strategy of farmers in the Sahel region of Burkina Faso. However, this strategy may not continue to be feasible given ecological changes, deforestation and environmental degradation.

The current grain bank is initiated by a one-time grant from an external agency – an NGO or the government – with or without the requirement of contributions by participating (member) households. Once established, the grain bank is managed by the member households themselves. The grain bank provides loans in the form of grain to member households at times of food scarcity, typically during the lean season. These loans are returned with interest (also in the form of grain) after the following harvest season.

1.3.1 Potential advantages of grain banks

What are the advantages of grain banks? To begin with, grain banks are unique across the gamut of food security interventions in that they provide consumption credit. While production credit and income generation programs can assist households with asset generation and growth and hence increase food security in the long-term, they are less able to address in a timely manner food shortage problems that arise in the short-term. By channeling savings from the harvest period to the lean period, grain banks can enable households to self-insure and smooth consumption (Ministry of Tribal Affairs 2002).

Second, grain banks can respond to seasonal food shortages in a timely manner since they are situated in and managed by the beneficiary community (ibid.). Unlike existing government food security and nutrition programs, there are no transportation or distribution costs, as grains are stored and disbursed locally. This aspect is even more important considering that tribal villages tend to be remote and have poor transport and communication facilities. In addition, grain banks appear to be highly cost-effective as their operating costs are minimal – the only major cost is the one-time expenditure related to the construction of the grain bank storage facility which is present in few villages. This compares favorably with the PDS, which is an extremely

expensive program and, barring radical reforms, can only be maintained with a hefty subsidy from the central government (Umali-Deininger and Deininger 2001). Since decisions are made and executed by the beneficiary community, there are fewer bureaucratic lags or barriers and a higher level of operational flexibility, compared to the PDS.

Third, community oversight of grain banks can also enable high levels of accountability and transparency. Since small village communities typically have intimate knowledge regarding the circumstances and needs of member households, grain banks have a comparative advantage in targeting benefits to those households in need compared to external interventions. As evident from a survey of social safety net programs in different countries, community participation in program design, implementation, and monitoring can lead to better targeting (Subbarao et al. 1997). Another example of the effectiveness of community participation in achieving better targeting outcomes is provided by Alderman (2002). Using data from Albania, he finds that its social assistance system, which allows for community discretion in determining distribution, is better targeted to the poor relative to safety net programs in other countries which do not allow for community involvement. He also finds that the poverty targeting in the social system in Albania achieves better outcomes than could be expected based on proxy indicators of targeting using household survey data, and concludes that community-level discretion in determining distribution permits the use of local information that is unlikely to be obtained from survey instruments. Conning and Kevane (2002) also discuss the potential benefits of community-based targeting for the implementation of social safety nets. More recently, in a review of targeted anti-poverty programs in 48 countries, Coady et al. (2004) show that, on average, community-based selection performs well. In the context of grain banks, members possess intimate knowledge of the output, asset holding and other resources

of fellow members, which can improve targeting outcomes. However, like all other community-based institutions, grain banks are also susceptible to leakages, misappropriation and capture by local elites.[15]

1.3.2 Relationship to cereal banks in Africa

The grain banks of India have a distant cousin in the form of "cereal banks" in different parts of Africa, especially in the Sahel. As described in the gray literature on cereal banks, these institutions largely function as village cooperatives that buy, store and sell food grains. In the quintessential model, villagers receive a start-up grant or loan from an external agency (usually an NGO) to purchase grains after the harvest, when prices are low (CRS 1998). During the lean season when prices are high, the cereal bank sells its stock in the village above the original purchase price but below the prevailing market price, using the revenues generated as a revolving fund to refinance its operation in the following year (ibid.). Apart from providing the start-up grant or loan, the external agency also finances the construction of a storage facility. The cereal bank also assists producers to market their grains in urban markets where consumer prices are higher. Thus, the main objectives of cereal banks appear to be the provision of better marketing services for grain producers and consumers at the village level, the reduction of post-harvest losses, the creation of local emergency stocks as well as building organizational capacity at the village-level (ibid.).

Given their storage and trading roles, African cereal banks are probably best described as mechanisms for commodity-price stabilization. There is a well-developed literature on commodity-price stabilization (e.g., Newbery and Stiglitz 1981, Williams and Wright 1991, Deaton and Laroque 1996) which discusses the role

[15] See Mansuri and Rao (2004) for a summary discussion of imperfections in community-based targeting due to capture by local elites.

of storage by forward-looking agents in order to decrease variability in prices. In the absence of storage or futures markets, government intervention in the form of price stabilization schemes can improve the welfare of consumers and producers. As discussed in Newbery and Stiglitz (1981), the main objectives of such price stabilization schemes include, inter alia, raising producer prices and incomes and reducing the price-related risks faced by both producers and consumers.

Commodity price stabilization via storage and improvements in infrastructure can also improve welfare in the presence of interseasonal flow reversals, that is, the post-harvest flow of grains at low prices from rural areas to urban areas followed by the reverse flow of higher priced grains during the lean season as rural grain stocks become depleted . Providing evidence of price reversals in Madagascar, Barrett (1996) shows how this phenomenon can adversely affect peasant households that are net grain buyers in the lean season. He concludes that interventions that diminish the spatial concentration of food grains storage help to improve the welfare of peasant households. Are the grain banks found in Orissa motivated by this concern? This is probably unlikely, since markets do not appear to be well-integrated in southwest Orissa and there is scant evidence of inter-seasonal flow reversals in this area. In addition, the majority of rural households are semi-subsistence producers, engaging in market trade only to a limited degree, usually in local rural markets.

Clearly, there exist substantive differences between the grain banks in tribal Orissa and the cereal banks in Africa. We draw attention to two important distinguishing elements. First, they do not engage in arbitrage either spatially (taking advantage of the difference between high urban and low rural post-harvest prices) or temporally (taking advantage of the different between low postharvest prices and high lean season prices). Second, and more importantly, the raison d'etre of grain banks in Orissa is to provide consumption credit to member households for the purpose of

smoothing food consumption across the agricultural cycle rather than to serve as a mechanism for stabilizing grain prices across the agricultural cycle via storage. Notwithstanding these differences, grain banks in Orissa are similar to their African counterparts in that they are both managed by the beneficiary community.

To some extent, the source of inspiration for grain banks is most likely found in the homegrown Self-Help Group (SHG) movement in the informal banking sector in India. The SHG movement was started by social development NGOs in the 1980s to promote group savings and lending associations, primarily among women (Morduch and Rutherford 2003). SHGs mobilized women to pool their savings into loans, often for short-term consumption. Features that are common across grain banks and SHGs include the provision of consumption credit as well as member ownership, though grain banks lack the feature of joint liability which typifies many SHG credit schemes.

Government of India's Village Grain Bank Scheme

Under the village grain banks scheme initiated by the Ministry of Tribal Affairs in 1996-97, community grain banks are established in villages having a majority ST population (Ministry of Tribal Affairs 2002). Member households can borrow grains from these grain banks during the lean season as well as during natural disasters such as droughts or cyclones in order to combat food insecurity and starvation. In addition, they can also borrow grains to cope with idiosyncratic shocks such as illness or death of a family member (ibid.). The stated objective of the scheme is to provide a safety net to protect households, especially for young children and pregnant and lactating mothers, from a drop in nutritional standards in the face of covariate as well as idiosyncratic shocks.

According to the Ministry of Tribal Affairs, there are plans to expand the scheme to cover all endemically drought- and migration-prone areas and tribal villages across the country (Press Information Bureau 2002, 2004).[16] Apart from Orissa, the scheme is currently operational in 12 other states including Andhra Pradesh, West Bengal, Bihar, Gujarat, Madhya Pradesh, Tripura, Rajasthan, Tamil Nadu, Kerala, Maharashtra, Uttar Pradesh and Manipur. Out of the 1483 grain banks established between 1996-97 to 2003-04 under this scheme, the largest number – 530, or about 37 percent – are in Orissa. The scheme is operational in 69 blocks in the undivided KBK districts. However, to date, these have been no rigorous evaluations of the efficiency and impact of grain banks, and their reputation is based largely on anecdotal evidence and case studies of 'model' grain banks. The aim of this project is to fill this gap in knowledge.

1.4 Data

Given the unavailability of relevant data required for performing either the impact or institutional analysis, a small-scale panel village and household survey was conducted in the predominantly tribal districts of Rayagada and Koraput where a high concentration of grain banks have been set up by Agragamee, one of the more prominent rural development NGOs currently active in Orissa and the largest player in terms of the number of grain banks instituted in the state. Using these data, a number of questions on the impact and survival of grain banks are investigated.

[16] Recently, the grain bank scheme has been revised and transferred to the Ministry of Consumer Affairs, Food and Public Distribution. Under the revised scheme, there is an increased focus on targeting and community participation (especially of women) and enforcing repayment, by linking the grain banks scheme to the TPDS entitlements of members. BPL/*Antyodaya Anna Yojana* families in villages identified by the government as chronically food deficit areas are targeted for this scheme.

Survey data collection comprised of three phases, and involved the use of village, grain bank, and household survey instruments. Details on the survey methodology and sampling issues are provided in a separate appendix.

The first wave of the household survey was conducted in the post-harvest season (January-March 2005) in Kashipur block in Rayagada district. The total sample size was 28 villages, which included 14 villages with functioning grain banks and 14 villages with failed grain banks or villages where grain banks were never set up were sampled. Within each selected village, 20 households were sampled randomly. The total usable sample size was 544 households. Out of these households, 269 lived in villages where grain banks were active, and 275 in villages where grain banks were absent.

A village and grain bank survey was also implemented in these 28 villages to collect village and grain bank information. In villages where grain banks failed, only retrospective information on grain bank design and functioning was gathered, whereas in villages with surviving grain banks, both contemporaneous and retrospective information on grain bank design parameters and functioning was gathered.

The first wave of the household survey was followed shortly by a village and grain bank survey (April-May 2005) in the adjoining block of Dasmantpur, Koraput district, which served as the data for the institutional analysis. The survey was fielded in a balanced sample of 40 villages where grain banks were operational and 40 villages where grain banks were established but were no longer functional as of the time of the survey.[17]

The second wave of the household survey (accompanied by a pared down version of the original village and grain bank survey) was conducted in Kashipur in

[17] Some of the villages have no hamlets, i.e., all households live in one cluster. This is true for most of the smaller villages. However, some villages have more than one hamlet, and grain banks were established separately in each hamlet. In this case, data was collected at the level of the hamlet.

the lean season (August-early October in the same year), when food stocks tend to be at their lowest. The timing of the second wave coincided with the period when loans are provided to member households, and was completed before the minor crops (small millets) are harvested. The survey was conducted in 26 of the 28 villages selected for the first wave. A rotating panel sample design was adopted for the selection of households. Within each sample village, attempts were made to contact 15 out of the 20 sample households selected in the first wave and to contact 5 new households based on the original household listings. Contact was reestablished with 400 households from the first wave, and 99 new households were added. This brought the total usable sample to 499 households; 250 households were in grain bank villages, and 249 were in non-grain bank villages.

1.5 Objectives and methodology

The objective of this study is to implement an institutional and impact analysis of grain banks. The remaining chapters are organized as follows. In Chapter 2, we examine what factors promote or impair grain bank survival. The study is motivated by the fact that a large share of grain banks have ceased to function over time. It is also motivated by the fact that, despite this, the government and rural NGOs are increasingly scaling up their grain bank activities. In light of this, the insights provided here could help inform decision-making on design and implementation. The factors that we examine comprise of village, grain bank member, and grain bank design and operational characteristics. We examine the determinants of the probability of grain bank survival as well as the duration of grain bank survival using a binomial probit model and Cox proportional hazards model respectively. We find that village characteristics appear to matter for both outcomes, whereas grain bank member characteristics and grain bank institutional features, given the variation found in the

data, mainly do not appear to matter. In sum, the findings suggest the particular importance of program placement (i.e., site characteristics) for grain bank sustainability.

In Chapter 3, we examine the impact of community grain banks on young children's health outcomes. The study is motivated by the Indian government's proposal of expanding the grain bank initiative with the objective of improving nutritional outcomes in spite of the lack of quantitative evidence of the impact of existing grain banks. In particular, we examine children's height-for-age and weight-for-height standardized scores as well as the change in children's height, which are standard, objective indicators of children's nutritional status. The study contributes evidence on the impact of grain banks on children's health outcomes in order to inform future food security policy interventions. In order to arrive at an appropriate comparison group, we use propensity score matching estimators. We also examine if the effect is stronger in villages with older surviving grain banks to test the hypothesis that children in participating households in longer-lived grain bank villages benefit more due to intergenerational effects (through, for example, the transmission of health benefits to children via mothers). We find that grain banks do not have a statistically significant impact on any of the outcomes we examine, regardless of the duration of their lifespan. This evidence suggests that grain banks, given their current design, may not be effective in meeting the objective of improved health outcomes for young children.

In Chapter 4, we examine whether community grain banks have displaced informal moneylenders, traditionally the main source of credit for households. The study is motivated by widespread anecdotal evidence of such an effect. In addition, it is motivated by the possibility of welfare gains for grain bank beneficiary households given that the institution offers credit at more favorable terms. The study contributes

evidence on the effects of grain banks, at a time when the government and rural NGOs are increasingly scaling up their grain bank activities, as well as contributes to the presently thin literature on the effects of increased competition in credit markets. We estimate the effect of grain bank participation by households on the incidence of borrowing from moneylenders via propensity score matching. We find that participation in grain banks has a large negative effect on the incidence of borrowing from moneylenders. Further, we find that this effect is even larger when we examine participation in villages with older surviving grain banks. This evidence combined with some less rigorous evidence of smaller loans when grain bank beneficiary households do borrow from moneylenders suggests welfare gains for them.

While the findings of this study are not meant to be generalizable to grain banks in other parts of Orissa (such as the coastal districts) or India, the survey site represents one of the poorest and most food-insecure regions of the country. Thus, an understanding of the functioning and effectiveness of a promising food security intervention in this region can act as a useful resource in enabling similarly vulnerable tribal communities to achieve an improvement in their food security status, and informing the grain bank movement generally.

REFERENCES

Alderman, Harold. 2002. Do Local Officials Know Something We Don't? Decentralization of Targeted Transfers in Albania. *Journal of Public Economics*. v83 (n3):375-404.

Asthana, M.D. and Pedro Medrano. 2001 (eds.) *Towards Hunger Free India: Agenda and Imperatives.* New Delhi: Manohar.

Barrett, Christopher B. 1996. Urban Bias in Price Risk: The Geography of Food Price Distributions in Low-Income Economies. *Journal of Development Studies*. v32 (n6):830-49.

Barrett, Christopher B. 2002. Food Security and Food Assistance Programs. In Boris Gardner and G. Rausser, ed., *Handbook of Agricultural Economics, Volume 2*. Elsevier Science B.V.

Behrman, Jere R. and Anil B. Deolalikar. 1989. Agricultural Wages in India: The Role of Health, Nutrition and Seasonality. In David E. Sahn, ed., *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Catholic Relief Services (CRS). 1998. Notes from the Workshop on Community-level Grain Storage Projects (Cereal Banks): Why do they rarely work and what are the alternatives? January 19-22. http://www.foodaid.org/pdfdocs/cmgmt/grainstorageworkshop.pdf (accessed on December 10, 2004).

Census of India. Various years. Government of India.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. Targeting Outcomes Redux. *World Bank Research Observer*. v19 (n1):61-85.

Conning, Jonathan and Michael Kevane. 2002. Community-based Targeting Mechanisms for Social Safety Nets: A Critical Review. *World Development*. v30 (n3):375-94.

Das Gupta, Monica, Michael Lokshin and Oleksiy Ivaschenko. 2005. Improving Child Nutrition: The Integrated Child Development Services in India. *Development and Change*. v36 (n4):613-640.

Deaton, Angus, and Valerie Kozel. 2005. Data and Dogma: The Great Indian Poverty Debate. *World Bank Research Observer*. v20 (n2):177-199.

Deaton, Angus, and Jean Drèze. 2002. Poverty and Inequality in India: A Reexamination. *Economic and Political Weekly*, 7 September.

Deaton, Angus, and Laroque, Guy. 1996. Competitive Storage and Commodity Price Dynamics. *Journal of Political Economy*. n104 (n5):896-923.

Department of Woman and Child Development (DWCD). 2002-03. *Annual Activity Report 2002-03.* Government of Orissa.

Drèze, Jean and Reetika Khera. 2003. Handout prepared for workshop on Reform of the PDS. New Delhi: Centre for Development Economics (Delhi School of Economics), 9 January.

Drèze, Jean and Aparajita Goyal. 2003. The Future of Mid-Day Meals. *Economic and Political Weekly* (November 1): 4673-83.

Drèze, Jean. 2001. Starving the Poor. *The Hindu*. February 26.

Gragnolati, Michele, Caryn Bredenkamp, Monica Das Gupta, Yi-Kyoung Lee and Meera Shekar. 2006. ICDS and Persistent Undernutrition: Strategies to Enhance the Impact. *Economic and Political Weekly* (March 25).

Gupta, Devendra B., Anil Gumber and Anuradha Banerjee. 1998. Strengthening Integrated Child Development Services Programme: Key Findings From a Pilot Study. *Margin*. v30 (n2).

International Institute for Population Sciences (IIPS) and ORC Macro. 2000a. *National Family Health Survey (NFHS-2), 1998-99: India*. Mumbai: IIPS.

International Institute for Population Sciences (IIPS) and ORC Macro. 2000b. *National Family Health Survey (NFHS-2), 1998-99: Orissa*. Mumbai: IIPS.

Lawrence, Mark, F. Lawrence, T. J. Cole, W. A. Coward, J. Singh and R. G. Whitehead. 1989. Seasonal Pattern of Activity and its Nutritional Consequence in Gambia. In David E. Sahn, ed., *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Mansuri, Ghazala and Vijayendra Rao. 2004. Community-Based and –Driven Development: A Critical Review. *World Bank Research Observer*. v19 (n1):1-39.

Meher, R. 2001. Socio-Economic Profile of KBK Districts, in *Orissa Development Report*. Planning Commission, Government of India and Government of Orissa. 2001. Bhubaneswar: Nabakrushna Choudhury Centre for Development Studies.

Ministry of Finance. 2003. *Economic Survey 2002-03.* New Delhi: Ministry of Finance, Government of India.

Ministry of Tribal Affairs. 2000-01. Annual Report. New Delhi: Ministry of Tribal Affairs, Government of India.

Ministry of Tribal Affairs. 2002. Grain banks. Ministry of Tribal Affairs, Government of India website, 2002, http://tribal.nic.in/grain bank1.html. Accessed 5th April 2004.

Ministry of Tribal Affairs. 2003-04. Annual Report. New Delhi: Ministry of Tribal Affairs, Government of India.

Misra, S.N and M. Behera. 2001. Poverty and Living Conditions, in *Orissa Development Report*. Planning Commission, Government of India and Government of Orissa. 2001. Bhubaneswar: Nabakrushna Choudhury Centre for Development Studies.

Morduch, Jonathan and Stuart Rutherford. 2003. Microfinance: Analytical Issues for India. In Priya Basu, ed., *India's Financial Sector: Issues, Challenges and Policy Options*. Oxford University Press, forthcoming.

National Sample Survey Organization (NSSO). 1998. *Reported Adequacy of Food Intake,1993-1994.* New Delhi: NSSO, Ministry of Statistics and Programme Implementation, Government of India.

National Sample Survey Organization (NSSO). 2001. *Reported Adequacy of Food Intake in India, 1999-2000.* NSS Report No. 466. New Delhi: NSSO, Ministry of Statistics and Programme Implementation, Government of India.

Newbery, David M. G., and Joseph E. Stiglitz. 1981. *The Theory of Commodity Price Stabilization: A Study in the Economics of Risk.* Oxford University Press.

Newman, James L. 1970. *The Ecological Basis of Subsistence Change among the Sandawe of Tanzania*. Washington, D.C.: National Academy of Sciences.

Ota, A.B. 2001. An Overview of the Scheduled Tribes in Orissa, in *Orissa Development Report*. Planning Commission, Government of India and Government of Orissa. 2001. Bhubaneswar: Nabakrushna Choudhury Centre for Development Studies.

Perspective Planning Division. 1979. *Report of the Task Force on Projections of Minimum Needs and Effective Consumption Demand*. Planning Commission, New Delhi.

Planning Commission, Government of India and Government of Orissa. 2001. *Orissa Development Report*. Bhubaneswar: Nabakrushna Choudhury Centre for Development Studies.

Planning Commission. 2001. Public distribution system and food security, Report of the Working Group for the Tenth Five Year Plan. November. Government of India. http://planningcommission.nic.in/aboutus/committee/wrkgrp/wg_pds.pdf . Accessed 5th April 2004.

Press Information Bureau. 2002. Towards a Hunger Free Tribal India. *Features*, January 2002. Press Information Bureau, Government of India website, 2002, http://pib.nic.in/feature/feyr2002/fjan2002/f210120022.html. Accessed 5th April 2004.

Press Information Bureau. 2004. Ministry of Tribal Affairs: Year End Review. Press Information Bureau Release. Press Information Bureau, Government of India website, December 31, 2004, http://pib.nic.in/release/release.asp?relid=6173. Accessed 10th June 2006.

Radhakrishna, R. and Kalanidhi Subbarao with S. Chandrakant and C. Ravi. 1997. *India's Public Distribution System: A National and International Perspective*. World Bank Discussion Paper 380, World Bank, Washington D.C.

Reardon, Thomas and Peter Matlon. 1989. Food Insecurity in Burkina Faso. In David E. Sahn, ed., *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Sahn, David E. 1989 (ed.) *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Scudder, Thayer. 1962. *The Ecology of the Gwenbe Tonga*. Manchester: Manchester University Press.

Shariff, Abusaleh and A. C. Mallick. 1999. Dynamics of Food Intake and Nutrition by Expenditure Class in India. *Economic and Political Weekly* (3 July): 1790-1800.

Subbarao, Kalanidhi, Aniruddha Bonnerjee, Jeanine Braithwaite, Soniya Carvalho, Carol Graham and Alan Thompson (1997). *Safety net programs and poverty reduction: Lessons from cross-country experience*. Washington, DC: World Bank.

Suryanarayana, M.H. 2000. Food Security and Caloric Adequacy across States: Implications for Reform. In Krishnaji N. and T.N.Krishna, ed., *Public Support for Food Security: The Public Distribution System in India*. Strategies for Human Development in India Vol. 1. New Delhi: Sage Publications.

Suryanarayana, M. H. 2001. Economic Reform versus Food Security: Kerala's Gordian Knot. *Journal of International Development*. v13 (n2):239-53.

Swaminathan, Madhura. 2000. *Weakening Welfare: The Public Distribution of Food in India.* LeftWord Books, New Delhi.

Umali-Deininger, Dina and Klaus W. Deininger. 2001. Towards Greater Food Security for India's Poor: Balancing Government Intervention and Private Competition. *Agricultural Economics*. v25 (n2-3):321-35.

Williams, Jeffrey C., and Brian D. Wright. 1991. *Storage and Commodity Markets*. Cambridge; New York and Melbourne: Cambridge University Press.

Zeller, Manfred, Gertrud Schrieder, Joachim von Braun and Franz Heidhues. 1997. *Rural Finance for Food Security for the Poor: Implications for Research and Policy*. International Food Policy Research Institute, Washington DC.

CHAPTER 2

DETERMINANTS OF GRAIN BANK SURVIVAL AND LONGEVITY

2.1 Introduction

In this chapter, using data from a small-scale village and grain bank survey in Koraput, Orissa, we investigate the key institutional and socioeconomic factors that affect the likelihood and the duration of survival of community grain banks. To date, there has been no rigorous quantitative analysis of grain bank sustainability despite the fact that the majority of grain banks that were established in this region have ceased to function. Such an analysis is important in light of the fact that rural development NGOs as well as the government of India have adopted grain banks as an integral part of their tribal food security strategy, although not much is known about the factors that affect their sustainability and performance. In addition, grain banks have some unique features, discussed in Chapter 1, that make them a promising complement to existing food security programs, such as the provision of consumption credit, membership-based management and ownership and the ability to respond to food shortages in a more timely manner, given that the institution is situated locally within the community of intended beneficiaries.

The remaining sections of the chapter are organized as follows. In Section 2.2, we provide an overview of the institutional features of grain banks in the region, including a discussion of whether and how grain banks help households to manage risk and their role in promoting household savings. In Section 2.3, we describe the data used and provide an overview of the socioeconomic characteristics of the sample villages. We also discuss the salient design features of the sample grain banks. In Section 2.4, we lay out the methodological framework for the empirical analysis of the determinants of grain bank survival and duration. In Section 2.5, we present our

empirical results and discuss them. We also present how our results fare when subjected to a series of robustness checks. Section 2.6 summarizes the main findings and concludes.

2.2 Institutional features

Grain banks fall under the rubric of membership-based voluntary community-based organizations (CBOs) that play a prominent role in the provision of credit, insurance and other financial services, especially in rural areas in developing countries where formal financial institutions are largely absent. They are membership-based, in that they are owned and managed completely by the beneficiaries themselves. Grain banks are established in each village community, with participation limited to community members, as a result of which individuals are likely to possess a rich set of information on their fellow members. In addition, participation is voluntary, as no one is forced to become a member of a particular group, unlike stricter kinship-based systems (see Dercon et al. 2005 for further discussion).

Why did grain banks originate in their current form? Given issues of poor coverage and irregular supply endemic to the major government food security and nutrition programs especially in rural and remote parts of Orissa as well as the recurring occurrence of seasonal food shortages and hunger in these parts, grain banks were introduced by social activists in Orissa as a community level solution that was not susceptible to the vagaries of these government programs. By lending and collecting loans in kind, grain banks fit into the barter system that was familiar to the tribal population. In the absence of formal credit markets, they provided an alternative source of (consumption) credit at interest rates lower than those charged by local moneylenders, without the threat of the loss of collateral. Grain banks were swiftly and enthusiastically embraced by the NGO sector engaged in promoting food security

and reproduced widely in different parts of Orissa and elsewhere. Lately, the government of India has started to establish its own grain banks in tribal villages across Orissa and other states as part of its tribal development program.

Insight into the main institutional features of grain banks can be gained by comparing them to the cereal banks of the Sahel region in Africa.[18] Prima facie, the institutions appear to be very similar. Both are community-based food security interventions. They are initiated by an external agent (usually a NGO) and subsequently managed by members. However, the main objective of cereal banks is to enable commodity price stabilization across the agricultural cycle through the provision of a physical storage facility. On the other hand, the main objective of grain banks is to enable consumption smoothing across the agricultural cycle through the provision of consumption credit, and it does not depend on the provision of a dedicated physical storage facility. As discussed later in Section 2.3.3, such a facility was constructed in very few instances in the case under study. Thus, unlike cereal banks, the grain bank intervention provides mainly institutional infrastructure, not physical infrastructure.

To some extent, grain banks enable member households to cope with idiosyncratic risk. However, as discussed below, their main function appears to be the provision of a commitment savings product, possibly for addressing issues of self-control, problems of intra-household allocation or coordination problems associated with social claims on liquid assets.

[18] For a more detailed comparison of the main features of African cereal banks and Indian grain banks, see Chapter 1.

Grain banks for managing household risk

Due to the fact that grain bank membership is confined to members of the same village, community grain banks cannot help member households cope with covariate risk such as climatic shocks. Carter (1997), using data from semiarid Western Africa, shows that households are able to insure idiosyncratic risk locally, but not covariate risk. To the extent that covariate risk is important in this region, grain banks are not effective arrangements for the provision of mutual insurance, as all member households default simultaneously in the event of drought and associated crop failure.

However, studies such as Townsend (1994) have shown that idiosyncratic risk dominates covariate risk for poor rural households. This is also shown by Lybbert et al. (2004) among pastoralists in southern Ethiopia, a population that is extremely vulnerable to climatic, epidemiological and other covariate risks. To the extent that grain banks provide consumption loans to cope with illness shocks or household-level crop failure, they can indeed help households manage idiosyncratic risk. These are examples of grain banks insuring risk via mutual insurance. In addition, as discussed later in Section 2.3.2, grain bank participants that default on their loan repayments typically do not immediately lose their membership, but instead given a grace period and allowed to return their loans with interest in the following harvest season. Thus, the grain bank credit contract implicitly includes insurance.

Grain banks for enabling savings

Open-ended conversations with survey respondents as well as Agragamee field staff revealed a two-fold strategy households use to cope with food shortages in the agricultural lean season: a decrease in consumption and consumption credit from the local, private moneylender. Typically, the moneylender was a non-tribal individual who was a trader or shopkeeper. He did not require collateral for giving the loan, but

charged extremely high interest rates. Repayment of loans to the moneylender constrained the ability of villagers to save during the harvest season, resulting in a cycle of borrowing and debt. Behavioral factors – the lack of a savings habit and long-term planning – were also cited as a reason to explain the absence of savings prior to the establishment of grain banks. According to respondents, by providing a one-time external grant, grain banks could break the cycle of debt. Agragamee field staff reported that grain banks also provided villagers an opportunity to cultivate a savings habit that did not exist prior to the intervention.

Using the terminology of behavioral economics, we posit that evidence of post-harvest festivals marked by high food and alcohol consumption in tribal Orissa, in contrast with severe food shortages in the lean season, may indicate time-inconsistent or hyperbolic preferences and help explain the lack of a savings habit prior to grain banks. According to standard economic theory, individuals are assumed to have time-consistent preferences, which imply exponential discount rates. However, a large body of literature has recorded evidence from both laboratory and field experiments that individuals suffer from time-inconsistency problems and do not discount the future exponentially (see, e.g., Laibson 1997 and O'Donoghue and Rabin 1999). As a result, they exhibit more impatience for near-term trade-offs compared to future trade-offs. Since savings requires a delay in immediate rewards for future (greater) rewards, individuals with time-inconsistent preferences have a preference for short-term "over-consumption". In the context of food consumption in tribal Orissa, time-inconsistent preferences may result in over-consumption of food stocks after the harvest, even if that implies adverse consequences on consumption in the agricultural lean season.

The evidence of cycles of reduced food intake and ritual feasting is not limited to the case of the tribal population in Orissa. Fortes and Fortes (1936), as cited in

Messer (1989), interpret consumption practices among the Tallensi of Ghana as a problem of "misallocation of resources across seasons". The study documents "over-consumption" in the post-harvest period in the form of liberal use of grains in beer brewing in contrast to the inadequacy of grains stored for the lean season, when energy requirements are high due to planting activities.

More generally, there is evidence from different parts of the world that individuals demand appropriately-designed savings products or mechanisms that can reduce commitment problems and enable savings.[19] For example, using data from Kenya, Gugerty (2001) shows that participants explicitly design their rotating savings and credit associations (ROSCAs) to enable exercise of self-control. Similarly, Rutherford (1999) discusses numerous commitment devices used by individuals in rural East Africa to stick to savings plans, including buying a lock box and throwing away the key and the use of "money guards" in which individuals entrust their savings to someone else so that they cannot spend it. Ashraf et al. (2006) provide evidence from an experiment in rural Philippines that individuals with time-inconsistent preferences have a higher demand for a commitment savings product, and that by use of this commitment savings product, they are able to increase both short-term and longer-term savings. Duflo et al. (2005) use experimental data from rural western Kenya to show that commitment problems are a key factor determining the lack of use of fertilizer, in spite of the fact that it increases productivity and profitability. They show that programs that help farmers to commit to buying fertilizers at the time when they have money result in the use of fertilizer in the future, demonstrating that farmers exhibit time-inconsistent preferences but are aware of their self-control problems and therefore have a demand for savings and commitment devices.

[19] See Ashraf et al. (2003) for a review of commitment savings products

These studies imply that in the presence of time-inconsistency and self-control problems, products that place restrictions on present consumption can alter savings habits. Similarly in tribal Orissa, post-harvest "over-consumption" and the absence of a savings habit prior to the establishment of grain banks can be interpreted as a result of time-inconsistent preferences. In this context, grain banks can enable savings for members with self-control problems.

Alternately, grain banks can potentially enable savings by addressing problems of intra-household allocation. For example, Anderson and Baland (2002) show that ROSCAs can provide "spousal control" as they allow individuals to hide money from their spouse. Informal interviews in tribal Orissa indicate that the unitary model of household decision-making may not be a correct representation of tribal households in these areas. One piece of evidence is the presence of anti-liquor campaigns by women against liquor consumption by men in their households, which suggests differing preferences over the allocation of income to expenditures. Since the grain bank committee, which oversees management of grain banks, is typically dominated by women, they may provide women with greater "spousal control", in the case that women have a greater interest in ensuring household food security than men.

Grain banks can also enable savings by reducing the rate of social taxation, or, in other words, resolving the coordination problem associated with social claims on liquid assets. As discussed by Armendáriz and Morduch (2004), there is evidence that poor households have a desire to save, but due to the lack of access to formal savings facilities, are often compelled to save through imperfect informal means, such as leaving money with friends, neighbors or a deposit taker, hiding it in the house or joining ROSCAs. These means are not efficient and vulnerable to losses, often due to constant requests for aid from friends, relatives and spouses. In the case of grain banks, saving by oneself may not be efficient as it may be vulnerable to high rates of

social taxation due to requests from relatives and neighbors during times of need. However, saving in a community-level institution where decisions on disbursement are made only a few times a year and require the sanction of the management committee can lower the rate of social taxation, thereby making the savings decision attractive.

Mechanisms for contract enforcement

In chapter 1, we mentioned that group lending and joint liability, mechanisms used by local credit groups to overcome informational asymmetries, are not present in grain banks.[20] How, then, are grain bank contracts enforced?

Tribal agrarian societies are bound by close ties of clan and kinship, and individuals typically possess a rich set of information regarding fellow members. Thus, peer monitoring can be an effective and low-cost instrument for attenuating moral hazard problems. In addition, the cost of social sanctions is high. Therefore, the threat of social sanctions can maintain high repayment rates and overcome free rider problems in activities with a public good character, such as peer monitoring and auditing. This has been shown theoretically by Besley et al. (1993) and Besley and Coate (1995), and empirically by Miguel and Gugerty (2004) in the provision of public goods in Kenya.[21]

Grain banks share the feature of dynamic incentives with prominent microcredit institutions such as Grameen Bank in Bangladesh, Banco-Sol in Bolivia, Bank Rakyat in Indonesia and the Foundation for International Community Assistance

[20] In any case, joint liability is not a panacea for overcoming information problems and enforcing contracts. Besley and Coate (1995) construct a model in which group lending generates both positive and negative effects on repayment incentives.

[21] However, Mude (2006) shows how the informational advantages provided by close kinship ties in small communities are diminished if incentives for patronage and favor-peddling are present, providing a cautionary note on the limitations of peer monitoring in small, traditional communities.

(FINCA) village banks. As discussed by Morduch (1999), dynamic incentives in repeated games, whereby loan sizes are increased over time, can also result in low default rates, even in the absence of group lending. Dynamic incentives are more effective in areas with relatively low mobility, since a threat to withhold future (larger) loans is not credible if the lending game is finite. In tribal regions of Orissa (and elsewhere in India), seasonal food shortages can therefore be interpreted as a repeated game in which there are few (and no better) alternatives than grain banks.

2.3 Data and sample

Background

The data used in this study come from Dasmantpur block, Koraput district in the tribal belt of southwest Orissa, one of the poorest regions in the state (see Figure 2.1 for a district map of Orissa with the tribal belt shaded in gray).

Figure 2.1: District map of Orissa

According to a population census conducted by the government of Orissa in 1997, 89 percent of households in Dasmantpur were classified as poor or Below Poverty Line (BPL).[22] In terms of occupational distribution, 47.4 percent of all workers are self-employed farmers, followed by 40.6 percent as agricultural laborers (Census 2001). The overall adult literacy rate is 23.3 percent, with female literacy abysmally low at 11.8 percent. Roughly 61 percent of Dasmantpur's population is tribal, with the Paraja and the Kandha constituting the two main tribes.

Due to the lack of storage mechanisms and poor transport and communication facilities in this region, markets tend to be poorly integrated over space and time. As a result, food consumption is tied closely to the agricultural calendar and the composition, quality and quantity of food consumption changes with the seasons. Between October-February, when produce is harvested, the tribal population consumes staples such as rice, millet and maize.[23] Between March-May, consumption depends on food stocks as well as purchases using earnings from daily wage labor (mainly in labor-intensive public works programs). Food shortages are experienced in the 'hungry' season (June-September), during which time food stocks tend to be low and daily-wage labor employment is unavailable due to the monsoon rains. During this period, meals are limited to gruel made from millet flour or flour from dried seeds (tamarind and mango), food consumption levels generally fall, and a decrease in the average body mass is observed.[24] Existing public assistance programs have not been able to eliminate these seasonal food shortages. In recent years, grain banks have been

[22] A BPL household is defined as a family with an annual income below Rs. 11,000 which is roughly equivalent to a dollar a day (1992 rupees). The percentage of households classified as BPL in Koraput district by the same survey is 84 percent.

[23] Cereals form the most important part of tribal diets. Pulses (mainly *kandula*, which is grown locally) are also consumed. However, few fruits and vegetables are consumed, as these are grown mostly for marketing purposes. Meat is only consumed during festivals or when there are guests. Milk and eggs are generally not consumed.

[24] Individual-level data on body weights was not collected for the village survey. However, as mentioned in chapter 3, data collected during the post-harvest and lean seasons as part of the household survey in Kashipur indicate a decrease in average body weight for adult women across the seasons.

seen as a welcome alternative for coping with seasonal food shortages and addressing credit constraints. We discuss below the grain banks implemented by Agragamee, a rural development NGO that has been active in the tribal region of Orissa for over two decades.

Community grain banks constitute one of the core components of the NGO food security strategy in tribal Orissa. Agragamee, one of the pioneers of the grain bank movement in this region, began its grain bank initiative in Kashipur, Rayagada in 1981 to combat starvation deaths and hunger there. Initially, it established grain banks in a few villages that were in the vicinity of its headquarters in the town of Kashipur. Following heavy rains in Rayagada, Koraput, Kalahandi and Bolangir districts during the monsoon of 1992, press reports of starvation deaths in the region prompted action by the government of Orissa. Together with UNICEF, the state government sponsored a food security project that was implemented by two NGOs: Agragamee in Rayagada and Koraput, and by Friends Association for Rural Reconstruction (FARR) in Kalahandi and Bolangir. For this project, within each district, one block where food insecurity was a particularly acute and chronic problem was selected – Kashipur block in Rayagada, Dasmantpur in Koraput, Tureikela in Bolangir and Lanjigarh in Kalahandi. Agragamee and FARR were chosen to administer the project on the basis of their long experience working with the tribal population in the region.

Thus, in 1993, supported by funds from the government of Orissa and UNICEF under the Orissa Household Food Security Project (OHFSP), Agragamee expanded its grain bank initiative to all villages in Kashipur block (Rayagada district) and adjoining Dasmantpur block (Koraput district). The interventions introduced as part of this project included the promotion of grain savings through the establishment of grain banks, increase in food availability by strengthening the public distribution

system, income generation through participation in self-help groups and the establishment of village-level committees to oversee the project.

Survey sample

The data used in this study come from a small-scale survey of grain banks conducted in Koraput and Rayagada districts in southwest Orissa, an area with a high concentration of grain banks.[25] A total of 232 grain banks were established in Dasmantpur block, Koraput district between 1993-98 by Agragamee. In April-May 2005, a village and grain bank survey was fielded in 40 villages where grain banks were operational (surviving grain bank villages) and 40 villages where grain banks were once present but were no longer operational at the time of the survey (failed grain bank villages).[26]

The sample numbers of surviving and failed grain bank villages are balanced and not proportionate to the population numbers of surviving and failed grain bank villages. According to the most recent records available from Agragamee, out of the 232 grain banks initially established, 71 were still operational. Thus, the population proportion of operational grain banks is 30.6 percent. Proportional sampling would have resulted in a lower number of observations of surviving grain banks, and given the objectives of the study to study the factors that influence grain bank sustainability, we intentionally oversampled operational grain bank villages. We however check the sensitivity of our empirical findings to this sampling decision and find that they are robust to correcting for the reported population proportion of operational grain banks

[25] Data for the impact analysis implemented in Chapters 3 and 4 is from a household and grain bank survey conducted in a separate set of villages in adjoining Kashipur block, Rayagada district.
[26] Details of the survey and sampling methodology are provided in a separate appendix.

as well as proportions bracketing the reported population proportion, given possible inaccuracies in Agragamee's records.[27]

The survey questionnaire was addressed jointly to village elders and as well as villagers who were likely to have a detailed knowledge of grain bank functioning, such as former or existing committee members. The survey was administered by trained local enumerators. To reduce response bias, each survey team was formed by enumerators from Dasmantpur block together with enumerators from Kashipur block, who would not have any close social connections with the respondents.

2.3.1 Overview of sample villages

Table 2.1 presents summary statistics for selected variables for the survey sample. On average, 65 percent of the households in sample villages faced food inadequacy for at least one month of the year.

Table 2.1: Descriptive statistics of selected village characteristics

Variable	Mean	Std. Dev.	Min.	Max.
Distance from closest weekly market (km)[1]	6.36	4.29	1.00	35.00
Distance from main road (km)	4.03	6.03	0.00	50.00
Distance from block headquarters (km)	25.98	17.34	1.00	62.00
Distance from closest Agragamee field office (km)	6.81	3.69	0.00	21.00
Total number of households	53.46	35.04	5.00	179.00
Share of households without land[2]	0.29	0.24	0.00	0.92
Share of households that reported food insufficiency	0.65	0.29	0.03	1.00
Share of ST households	0.77	0.32	0.00	1.00
Share of SC households	0.09	0.17	0.00	0.90

Notes: $N = 80$. [1] Data for all distance variables was collected as part of the village survey and reflect reports by a group of village leaders and elders. They should be interpreted as the approximate distance by foot, as this was the most common mode of transport. [2] Without land here actually means without formally titled land.

[27] Specifically, apart from estimating our regression model taking account of the actual population proportion of surviving grain banks, namely 0.31, we also estimate the model using a lower population proportion of 0.25 (1 in 4 surviving) and a slightly higher population proportion of 0.33 (1 in 3 surviving). These additional results are presented in an appendix of tables at the end of this chapter.

The sample villages and hamlets vary in size from 5 households to 179 households, with more than half the villages composed of 21-50 households (see Table 2.2). As a result of the small village size and strong kinship ties, members of the village community are likely to possess a rich set of information regarding fellow members.

Table 2.2: Distribution of sample villages by number of households

Number of households (h)	Frequency	Share (%)
$h \leq 20$	6	7.5
$20 < h \leq 50$	45	56.3
$50 < h \leq 80$	13	16.3
$80 < h$	16	20.0
	80	100.0

Ethnic composition

Seventy-six of the sample villages have some tribal population, with the Kandha and Paraja tribes constituting the two main tribes. Twenty-nine of these villages have exclusively ST population, while the other villages have a mixed population (Scheduled Castes (SC), Other Backward Castes (OBC) and General caste). The high proportion of tribal population indicates the high level of ethnic homogeneity in this region.

Transport and communication infrastructure

The transport and communication facilities of villages in this area are very poor. The mean distance from the sample villages to the block headquarters, which is the seat of local government offices and public agencies, was reported to be about 26 kilometers (roughly about 16 miles). Note that this distance has to be interpreted in the context of the extremely poor road and vehicle conditions, where a stretch of 25 kilometers (about 15.5 miles) can take more than an hour by road, and in light of the fact that the main mode of transport for villagers in this region is walking. Table 2.3 presents the

distribution of villages by distance from block headquarters. This information was collected from reports by villagers and reflects, in most cases, the distance of the path taken by foot (and not the distance on motorable road), as this was the most common mode of transport.

Table 2.3: Distribution of sample villages by distance to block headquarters

Distance from block headquarters (km)	Frequency	Share (%)
$0 \leq d < 5$	13	16.3
$5 \leq d \leq 25$	25	31.3
$25 < d$	42	52.5
	80	100.0

In order to travel to the block headquarters, villagers frequently reported that they had to traverse a combination of mud paths, untarred (*kachcha*) and tarred (*pucca*) roads. In more than half the sample villages, walking (either by itself or in combination with another mode of transport) was reported to be the most common mode of transport, followed by "private vehicle" run by private operators, sometimes on a weekly basis (see Table 2.4). Thus, travel to the block headquarters in order to avail of medical facilities or to obtain identification cards (such as a BPL card) could be an entire day's affair, starting with a long walk from the village to the main connecting road on which buses and other vehicles operate, usually on an infrequent and irregular schedule.

Table 2.4: Distribution of sample villages by main mode of transport to block headquarters

Main mode of transport	Frequency	Share (%)
Walking	33	41.3
Private vehicle	19	23.7
Public bus	16	20.0
Other (combination of walking & private vehicle; cycle)	12	15.0
	80	100.0

Village physical infrastructure

Physical infrastructure in these villages is generally nonexistent. When present, it is of extremely low quality. The main source of lighting in the sample villages is firewood. Only nine villages (11 percent) reported being connected to the electrical grid. In these villages, the average share of households with an electrical connection is less than 17 percent. Fifty-five sample villages (70 percent) reported that their access and internal roads are wholly *kachcha* or untarred, while the rest reported these roads as being partially or fully tarred. As a result, in the monsoon season, these roads become muddy (and are not usable in some cases). The roads are also difficult to keep from becoming contaminated and may serve as an important source of parasitic infections, as most villagers walk barefoot.[28]

Tribal settlements tended to form close to groundwater sources, typically near natural springs. Due to ecological changes in the area, mainly precipitated by collective human actions, drying up of natural water sources has made the tribal population dependent on tube wells established by the government for drinking water supply purposes. Although virtually all of the sample villages obtain drinking water from tube wells installed within the village (one village reported having access to piped water), inadequate water supply was commonly reported as a problem. Table 2.5 presents the distribution of tube wells across villages by village size categories. On average, across the 79 villages, one tube well was available for every 33 households.[29]

[28] Dirt floors in homes have also been shown to increase the likelihood of parasitic infections, diarrhea and anemia in children, which can in turn reduce cognitive ability (see, e.g., Anderson and May 1991 and Behrman 1996). Thus, poor physical infrastructure such as dirt floors and untarred roads can have serious deleterious impacts on current and future health and labor productivity.

[29] One village had no tube wells but had access to piped water.

Table 2.5: Distribution of sample villages of different sizes by number of tube wells

	Number of villages by size categories				
No. of tube wells	20 households or less	21-50 households	51-80 households	More than 80 households	Total
1	4	27	7	1	39
2	2	12	1	6	21
3	0	4	4	4	12
4	0	1	1	2	4
5	0	0	0	2	2
6	0	0	0	1	1

Notes: The remaining village had piped water but no tube well.

Economy

Agriculture was the main source of employment and livelihood in the sample villages, followed by seasonal daily wage labor. The latter was available mainly in the summer season in connection with large public works projects. The main crops cultivated in this area are cereals including paddy and different varieties of millet. Other crops include corn, oilseeds, pulses and vegetables (such as eggplant, potato, chillies and tomato). Items not produced by the villagers (such as salt and clothing) are purchased using cash or by bartering at the local weekly markets in each of the *gram panchayats*.

Food insufficiency

In all sample villages, villagers indicated that households faced food insufficiency during at least one month of the year. Thirty-seven villages reported that they faced food insufficiency for five to six months out of the year, while the rest reported facing food insufficiency for three to four months out of the year. While the months most commonly cited as food shortage months were the late summer and monsoon months of *Landi*, *Aashad* and *Shravana* (mid-May to mid-August), some villages reported

having households that faced food shortage from *Baishak* to *Kartik* (from mid-April to mid-November).

Development programs

Within the sample villages, the most commonly-found government programs include labor-intensive public works programs, especially during the summer months, and watershed management programs. Roughly two-thirds of the villages reported having had a food-for-work program within the 10 years preceding the survey, while roughly two in five villages reported having had a watershed management program within the same time period. At the time of the survey, three in four villages had one or more operational microcredit groups, called self-help groups, for organizing income-generating activities. These self-help groups generally obtain credit from state-run banks for the purpose of retail business and marketing, including making and marketing *jhadus* (brooms), dry fish, disposable leaf plates, etc. The credit is often used to buy items at a wholesale rate and then sell at a higher price, with profits shared equally among members. In villages with operational self-help groups, the share of participating households varied from a fifth to all households, with mean participation in these villages at around three-fifths of households. The size of these self-help groups varied, with larger villages often having more than one group, and groups ranging in size between 10 and 100 members, with the average membership size being about 32 members.

Health and education

At the time of the survey, 63 of the sample villages (about 78 percent) reported having functional primary schools. However, school infrastructure is of extremely poor quality, at best comprising of a single roofed room often without a blackboard. Few

resources, if any, are available to the teacher, who single-handedly manages young children of varying ages.

Sixteen villages (or 20 percent) reported not having a functional *anganwadi* center (the point of delivery of ICDS services), while the rest have a functional center or sub-center in the village. However, of the 66 villages with a functional *anganwadi* center, only 19 reported regular (daily or weekly) visits by the *anganwadi* worker. The frequency of attendance by the *anganwadi* worker is shown in Table 2.6. Only 25 of the sample villages (about 31 percent) reported having a mass immunization campaign in the five years preceding the survey.

Table 2.6: Distribution of sample villages by frequency of attendance by *anganwadi* worker

Frequency of attendance	Frequency	Share (%)
Daily	8	10.0
Weekly	11	13.8
Monthly	37	46.2
Less frequently or never	24	30.0
	80	100.0

Notes: 16 villages had no functional *anganwadi* center or sub-center.

Shocks

Roughly three in four sample villages reported experiencing three or more disasters (such as a landslide, drought, crop failure due to insect blight, flood, conflict, land seizures and forest fires) in the five years preceding the survey. Crop failure due to droughts or insect blight was a commonly occurring problem in this area. This raises the question on whether grain banks are sustainable in this area, as their sustainability depends on sufficient agricultural output. In fact, a large number of failed grain banks occurred in the years following a severe drought, which is illustrated in a later section.

Governance

Table 2.7 shows the distribution of villages by the share of villagers that attended village-level meetings to discuss issues that affected the entire village, from social functions and festivals to the progress of the village school, self-help group, etc. Apart from one village, the remaining villages reported holding village-level meetings which were attended by all or a majority of households. Roughly half of the villages reported holding meetings whenever the need arose. A little over a quarter of villages reported holding meetings on a monthly basis, while the remainder reported meeting on a weekly or fortnightly basis.

Table 2.7: Distribution of villages by the share of households that attend village meetings

Share of households	Number	Share (%)
All	56	70.0
More than half	21	26.3
Less than half	2	2.5
	79	100.0

Notes: One village reported not having village-level meetings.

Surviving grain bank villages versus failed grain bank villages

Tables 2.8 and 2.9 summarize the means of important variables for the sample of surviving grain bank villages and failed grain bank villages.

In Table 2.8, among other things, we see that the share of households reporting food inadequacy for at least one month in the year preceding the survey was significantly higher in villages with surviving grain banks than those with failed grain banks.

Table 2.8: Means of continuous variables, by village grain bank status

Variable	SGBV (1)	FGBV (2)	SGBV-FGBV (1)-(2)
Village-related variables			
Distance from main road (km)	3.74	4.33	-0.59
Distance from block headquarters (km)	23.95	28.01	-4.06
Time taken to travel to block headquarters by main mode of transport (minutes)	99.00	123.00	-24.00
Distance from closest Agragamee field office (km)	6.88	6.74	0.14
Distance from the closest weekly market (km)	7.00	5.71	1.29
Distance from the closest post office (km)	5.26	4.23	1.03 *
Total number of households	55.01	51.9	3.30
Share of landless households	0.28	0.29	-0.01
Share of ST households	0.76	0.78	-0.02
Share of households that have reported food inadequacy for at least 1 month in past year	0.73	0.57	0.16 **
Grain bank-related variables			
Grain bank membership size (at inception)	41.35	36.28	5.10
Amount of grain bank contributed by villages (at inception)	442.8	393.50	49.30
Share of ST members in grain bank	0.78	0.82	-0.03
Number of committee members	6.43	6.95	-0.52
Share of committee members who are female	0.55	0.51	0.03

Notes: SGBV refers to surviving grain bank villages; FGBV refers to failed grain bank villages.
* Statistically significant at the 10 percent level; ** at the 5 percent level; and *** at the 1 percent level.

In Table 2.9, we see that the proportion of villages with functional self-help groups, meeting on a needs-basis and also having an elected government representative in the five years preceding the survey ("ward member") is significantly higher in the surviving grain bank village sample than in the failed grain bank village sample. This may indicate that grain banks survive in villages with a higher degree of social interaction and leadership. The results discussed here are unconditional, and serve as a precursor to the conditional analysis.

Table 2.9: Means of dichotomous variables, by village grain bank status

Variable	SGBV (1)	FGBV (2)	SGBV-FGBV (1)-(2)
Share of villages (in percent) having			
Functional primary school	80.0	77.5	2.5
Midday meals at school	72.5	77.5	5.0
Anganwadi center or sub-center	77.5	82.5	-5.0
Regular (daily/weekly) visits by *anganwadi* worker	22.5	25.0	-2.5
Vaccination drive in last 5 years	30.0	32.5	-2.5
Electrical connection	7.5	15.0	-7.5
Tarred village road	22.5	40.0	-17.5 *
New road in last 5 years	75.0	77.5	-2.5
Walking as main mode of transport to block headquarters	37.5	45.0	-7.5
Self-help groups	85.0	65.0	20.0 **
Village-level meetings on a need-basis	75.0	30.8	44.2 ***
Ward member in last 5 years	77.5	55.0	22.5 **
Watershed management program in last 10 years	40.0	45.0	-5.0
Food-for-work program in last 10 years	75.0	57.5	17.5 *
Drought in last 5 years	80.0	77.5	2.5
Landslide in last 5 years	87.5	77.5	10.0
Forest fire in last 5 years	55.0	37.5	17.5

Notes: SGBV refers to surviving grain bank villages; FGBV refers to failed grain bank villages.
* Statistically significant at the 10 percent level; ** at the 5 percent level; and *** at the 1 percent level.

2.3.2 Overview of sample grain banks

In this section, we provide a brief overview of the 80 sample grain banks, which include 40 surviving and 40 failed grain banks.[30] Grain banks were established between 1993-1998. The distribution of sample grain banks by year of establishment is presented in Table 2.10 to see if there is any evidence of a difference in the failure pattern based on year of inception. For example, are the earliest grain banks more likely to fail? In other words, is there a demonstration effect or learning by doing? However, from the table below we find no statistically significant difference in the

[30] One shortcoming of the sampling frame is that the sample did not take into account the year of inception. Thus, from the current sample, we cannot make any inferences about a chronology of grain bank failure in the population. To the extent possible, we try to address this shortcoming in the data by including a variable on the year of inception in the conditional analysis.

proportion of surviving and failed grain banks based on the year in which they were established.[31]

Table 2.10: Distribution of grain banks by year of establishment

Year	Surviving grain banks		Failed grain banks		All	
	Number	Percent	Number	Percent	Number	Percent
1993	8	20.0	9	22.5	17	21.3
1994	8	20.0	9	22.5	17	21.3
1995	21	52.5	18	45.0	39	48.8
1996	2	5.0	3	7.5	5	6.3
1997	1	2.5	0	0.0	1	1.3
1998	0	0.0	1	2.5	1	1.3
Total	40	100.0	40	100.0	80	100.0

Chi-squared test of independence: Pearson chi-squared statistic = 2.55; *p*-value = 0.769.

Grain bank membership is voluntary, and in principle is open to the entire village community. In practice however, the entire village community became grain bank members in only 10 villages (or in 12.5 percent) of the sample. In villages with less than 80 households, on average about four out of every five households became grain bank members. In larger villages, on average about three out of five households became grain bank members. Two types of households did not become grain bank members – households that were food-secure and did not require borrowing grains from grain banks to smooth consumption, and households that were not accepted into grain banks because of liability risks or caste differences.

The grain banks were established in the monsoon season through a grant in the form of grains provided by Agragamee. In over three-quarters of the sample grain banks, villagers also made a matching contribution in order to double grain bank stocks. This was actively promoted by Agragamee field staff, who felt that this

[31] Note that this result is based on village and grain bank data from Dasmantpur. In Kashipur, the site of the household survey, grain banks established prior to the OHFSP differed from later grain banks in the level of involvement on the part of Agragamee.

contribution increased the community's sense of ownership of the grain bank, thereby potentially positively affecting the likelihood of grain bank sustainability.

Grain bank operations were overseen by a committee comprising of, on average, seven elected representatives, who were also grain bank members. On average, slightly over half of the committee members were female. Female representation was considered important for grain bank sustainability since the staff implementing the program felt that women would value the grain bank more. This was rooted in the implicit belief that women are more concerned about the nutritional and health status of their children than the male members of the household.[32] This design feature of Agragamee's grain banks is not unique. Transfer programs instituted by developing countries are increasingly delivering the benefits to the mother directly. For example, in Bangladesh, Mexico and elsewhere, conditional cash and in-kind transfer programs aimed at improving children's health and educational outcomes provide the benefit directly to mothers (Rawlings and Rubio 2005).

At the time of inception, interest rates were set at 100 percent for the standard duration of the loan.[33] This was done in order to help generate surplus reserves in the grain banks. In subsequent years, when sufficient reserves were achieved, the interest rate was lowered to 20-25 percent by the majority of sample grain banks.

The member households meet twice a year, once after the harvest when loan repayments are made and the cases of defaulters assessed, and once at the start of the lean season when loans are disbursed. Decisions on the loan amounts to be disbursed as well as the interest rate to be charged were usually made collectively by the entire

[32] For evidence on women's bias in channeling own-income towards expenditures on health and education for their children, see Thomas (1990) and Duflo (2003).
[33] This interest rate, although high, is lower than the interest rates set by moneylenders, which often include non-transparent and implicit rates, such as in tied credit-labor contracts.

member community, although the grain bank committee was responsible for collections, disbursements and bookkeeping.

Enforcement of repayment takes place through peer monitoring and the threat of social sanctions. Across the sample of surviving grain banks, on average, about two-fifths of households had not returned their loans from the previous year at the time of the survey. However, defaulters did not immediately lose grain bank membership. Instead, they were given a grace period and were asked to return their loans with interest in the following harvest season. Thus, the grain bank credit contract implicitly includes insurance, as in the credit contracts in northern rural Nigeria, examined by Udry (1990), in which repayments depend on realizations of random shocks by both borrowers and lenders. This practice of state-dependent repayments is also similar to other community-based rural credit cooperatives, such as the German credit cooperatives discussed by Ghatak and Guinnane (1999). These cooperatives granted an extension to borrowers who could not repay, based on their ability to see the extenuating circumstances of the borrower, such as illness or crop failure. In the case of our grain banks, the most common reasons for default included crop failure or insufficient agricultural output, followed by marriage expenses, events which are readily and highly observable to the entire community.

2.3.3 Overview of surviving sample grain banks

This section provides an overview of surviving grain banks in the survey sample. As mentioned before, 40 surviving grain banks were surveyed, all of which were established in the monsoon season, between 1993-97. However, more than half were set up in 1995 (see Table 2.10). At the time of inception, Agragamee provided a grant in the form of grains. Unlike the villages in Kashipur where Agragamee had established grain banks prior to the OHFSP, villagers were not obliged to contribute a

matching amount to the grain bank when the grain bank was established. The average number of member households at the time of inception ranged between 10 and 125 households, with the mean number of households around 41. Between the time that grain banks were established and the time of the survey, the average membership size had increased by about two households, indicating little change in membership size from inception.

Twelve of these surviving grain bank villages had only ST population, while the rest had a mixed population comprising of ST, SC, OBC or General caste members. Only three villages had no ST population. At the time of inception, in the remaining 37 villages, the average share of ST households that were grain bank members was about 85 percent.

Almost all villages cited shortage of food grains as one of the main reasons for setting up grain banks. In addition, 29 villages cited advice from Agragamee, while 18 villages cited the desire to gain respite from local moneylenders as being another important reason. The average amount of grains given by Agragamee at start-up was close to 19 kilograms per household. In contrast, the government of India granted 100 kilograms per household in its grain bank scheme (Government of India 2003-04).

The management of grain banks was undertaken by a village committee, which comprised both men and women. These committees ranged in size from three to eight members, with the average committee size equal to about six members. On average, females formed the majority of committee members, and the share of females in the committee ranged from about 38 to 100 percent. While the grain bank committee was officially in charge of loan disbursement and collection, in several villages, decision-making authority was shared more widely and involved all grain bank members. Specifically, 22 villages reported that in addition to having a grain bank committee, all grain bank members participated in decision-making regarding important matters.

The share of borrowers in the year preceding the survey ranged from 13 to 100 percent of the membership base, with the average share of borrowers across villages at 84 percent, indicating that a large proportion of grain bank members used the borrowing facility.

In terms of the reported benefits of grain banks, 39 villages indicated that the intervention had reduced their dependence on the local moneylender. Thirty-six villages reported that the availability of grains in a timely manner was another important benefit. Twenty-five villages also reported that borrowing from the grain bank allowed them to pay a lower interest rate than borrowing from other sources.

At the time of inception, the interest rate that was charged on grain bank loans was 100 percent in 32 villages, while it varied between 20-50 percent in the remaining eight villages. The reason cited for the high initial interest rate was the desire on the part of the grain bank members to increase the grain stock. In subsequent years, as the grain bank stocks increased, 34 out of 40 villages lowered the interest rate. The majority of grain banks (30 villages) charged between 20-25 percent, while the remaining 10 villages charged 50 percent. In examining the distribution of interest rates in effect at the time of the survey, we find that the median interest rate is 25 percent with an interquartile range of 12.5 percentage points.

Table 2.11 presents summary statistics on the stocks of different types of grains stored in the grain banks. Millet, the traditional staple in this region, is the most commonly stored grain, followed by paddy, and in a few cases, rice. Informal conversations with villagers and Agragamee staff indicated that millet may be more suitable for storage purposes, as it is more resistant to pests and the type of climate in the region.

Table 2.11: Summary statistics of grain bank stocks by membership size in the year preceding the survey (in kilograms)

Grain	*N*	Mean	Std. Dev.	Max.	Min.
Paddy	32	20.7	18.7	81.8	3.0
Millet	34	19.7	15.7	64.0	2.8
Small millet	29	11.0	9.0	32.0	1.8

Notes: Statistics reported only for surviving grain banks. *N* denotes the number of villages that reported positive holdings for each of the grains types reported above. Statistics not reported for rice, as only 5 villages reported positive holdings of rice.

Grains were stored in large bamboo baskets known as *dudis* or in polythene sacks. These baskets were kept in a member's house, as a dedicated storehouse for the grain bank was reported in only one of the 80 sample villages.[34] This provides an interesting and important insight into the defining characteristics of grain banks, as it indicates that the innovation is purely institutional rather than for storage purposes. It also presents an important distinction from the cereal banks in the Sahel region, one of whose main objectives is to provide an improved physical facility for storage.

Table 2.12 presents summary statistics on the share of paddy and millet holdings lost due to pests in an average year. These figures represent approximate amounts as reported by one or more committee members that possessed information on grain bank holdings and losses. The average share of paddy lost is higher than the average share of millet.[35] The lack of proper storage facilities was cited as one of the most common problems, and 39 villages reported that constructing storage facilities would improve grain bank functioning. Seven villages reported that increasing the grant amount or providing a seed grant would also help to improve grain bank functioning.

[34] We do not have household-level data on storage losses that would allow us to compare the rate of personal storage losses for households that host the grain banks' stocks with the rate of storage losses of grain banks. Therefore, we cannot test if there is a moral hazard problem whereby the host may take better care of his or her own stocks relative to grain bank stocks.

[35] This corroborates the claim by NGO field staff and management that millets are better suited for storage purposes, given the local climate and storage facilities.

Table 2.12: Summary statistics of share of grain bank losses due to pests by type of grain in the year preceding the survey

Grain	*N*	Mean	Std. Dev.	Max.	Min.
Paddy	24	0.05	0.04	0.20	0.0025
Millet	25	0.03	0.02	0.08	0.0040

Notes: Statistics reported only for surviving grain banks. In most cases, amounts reported are approximations, as reported by one or more grain bank committee member. *N* denotes the number of villages that reported positive holdings for each of the grains types reported above. Summary statistics not reported for small millet and rice as only 0 and 2 villages reported losses for these grains, respectively.

2.3.4 Overview of failed sample grain banks

Forty failed grain banks were surveyed, all of which were established in the monsoon season, between 1993-98. The average number of member households at the time of inception ranged between 5 and 139 households, with the mean number of households at 36. Seventeen of these villages had only ST population, while the rest had a mixed population comprising of ST, SC, OBC or General caste members. Only one village had no ST population. In the remaining 39 villages, at the time of inception, the average share of ST households in the grain bank membership was about 84 percent, almost the same as in the sample of villages with surviving grain banks.

Similar to the surviving grain bank villages, almost all cited shortage of food grains as one of the main reasons for introducing grain banks. In addition, 24 of these villages cited advice from Agragamee; nine villages cited respite from the moneylender as being an other important reason. The average amount of grains given by Agragamee at start-up was close to 21 kilograms per household.

As with the surviving grain bank villages, the grain bank management committee comprised both male and female members, with the average committee size equal to about seven members. Females formed the majority of committee members, while the share of females in the committee ranged roughly between 30 and 70 percent.

In addition to the grain bank committee, all grain bank members participated in the decision-making process in only 10 sample villages. This was significantly different from the surviving grain bank villages, where a larger number reported the involvement of the entire membership base.

Figure 2.2 plots the distribution of sample grain banks by year of collapse. It shows that more than half of sample grain banks collapsed between 1997-2001. These years also coincided with years of very low rainfall in Koraput district, indicating that insufficient agricultural output in those years probably played a large role in the closure of the grain banks. This may indicate that covariate shocks, rather than slow decline, drives grain bank failure.[36]

The proximate reasons for grain bank collapse, as reported by the villagers, include poor management or misappropriation by committee members (27 villages).

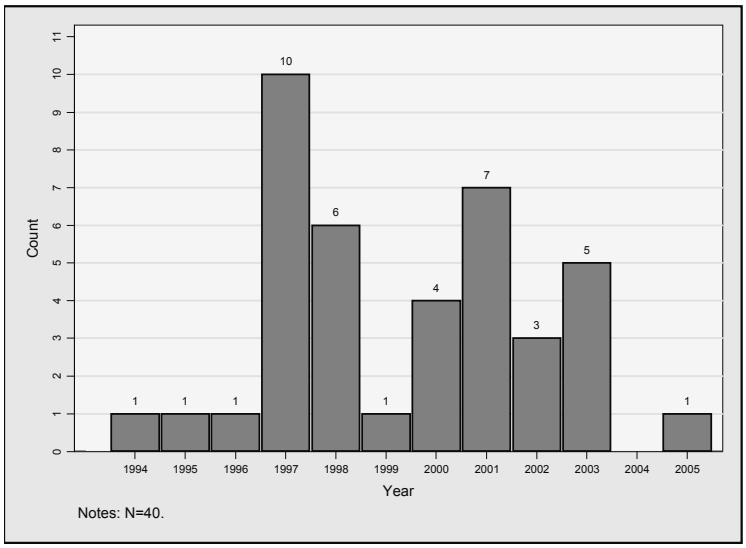


Figure 2.2: Frequency distribution of failed grain banks by year of failure

[36] We cannot perform a conditional analysis of whether rainfall shocks are a significant determinant of grain bank failure or success, since the dataset is from a small geographical region that faces the same weather shocks at any given time. Due to the lack of variation in the weather data, we are not able to include it as one of the explanatory variables, though we have reason to suspect that it may be an important one.

This indicates that there was a lack of managerial skills on the part of the grain bank committee, and that, in order to be successful, future grain bank schemes should include the necessary training to enable the grain bank committee members to perform their responsibilities in a satisfactory manner.

The second-most reported reason for grain bank failure, which corroborates the claim that poor rainfall plays an important role in grain bank failure, was crop failure or insufficient output (14 villages).

Figure 2.3 shows the distribution of the life span of failed grain banks, which ranges from 1 to 10 years. The median lifespan of failed grain banks was five years, with an interquartile range of 3.5 years. Interestingly, there does not appear to be any discernible pattern in the relationship between grain bank life span and the incidence of collapse. Thus, we do not find any evidence of sorting (i.e., "low-quality" grain banks collapse within a short time after establishment and have a short life span, while the "higher-quality" grain banks survive).

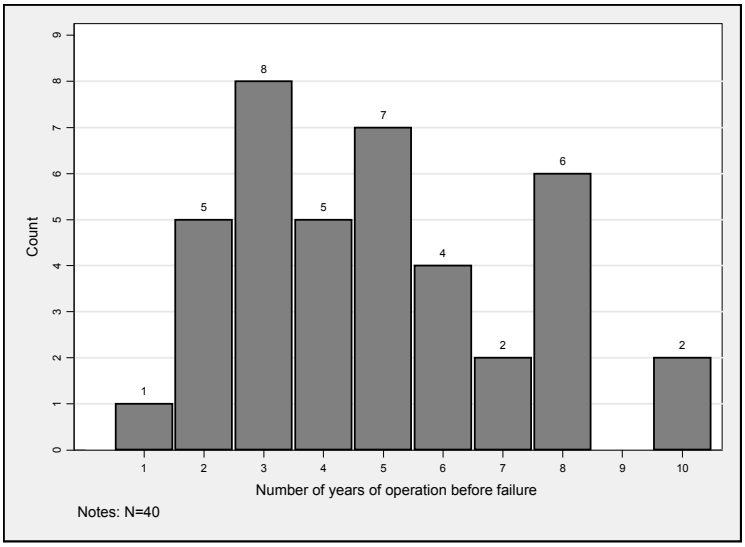


Figure 2.3: Frequency histogram of years of operation before failure

2.4 Empirical framework

2.4.1 The binomial probit model

In order to examine the determinants of a successfully functioning grain bank, where success is defined by whether the grain bank is surviving or not, we estimate a discrete choice process using a latent variable regression model. The model assumes that there is an underlying latent response variable y^* defined by the regression relationship

(1) $$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

where y^* is the net benefit obtained from a grain bank, $\mathbf{x}$ is a vector of village and grain bank characteristics, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and ε is a stochastic error term. The latent response variable y^* is not observed. Instead, a dummy variable y is observed, such that

(2) $$\begin{aligned} y &= 1 \text{ if } y^* > 0 \\ y &= 0 \text{ if } y^* \leq 0 \end{aligned}$$

From (1) and (2), we have

(3) $$\Pr(y = 1) = \Pr(\varepsilon > -\mathbf{x}'\boldsymbol{\beta}) = 1 - F(-\mathbf{x}'\boldsymbol{\beta})$$

where F is the underlying cumulative distribution function for the error term ε.

In the analysis of discrete choice models, two commonly used distributions for the error term include the normal and logistic distributions, which give rise to the probit and logit model, respectively. The logistic distribution is similar to the normal distribution except in the tails, where it is heavier. Both models are estimable via maximum likelihood estimation (MLE). For the normal distribution, the log-likelihood is

(4) $$\ln L = \sum_{y_i=0} \ln[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})] + \sum_{y_i=1} \ln[\Phi(\mathbf{x}_i'\boldsymbol{\beta})]$$

While there is no conclusive resolution on which distribution is more appropriate, Greene (2002) discusses that predictions from both models tend to be similar, especially in a balanced sample (equal number of responses, i.e., $y = 1$, and non-

responses, i.e., $y = 0$), which is true by sample construction in our case. Therefore, we only estimate the probit model.

As the estimated parameters are difficult to interpret given that they are related to the underlying latent structure, we compute the associated marginal effects, which we present in the appendix. In large samples, marginal effects evaluated at the sample means give the same result as computing the average of the marginal effects at every observation (see Greene 2002). However, the problem with the former is that if the $\mathbf{x}$ vector contains dummy variables (which is the case in this study), then we may be evaluating the marginal effect at a nonexistent or nonsensical value. In addition, the sample used in this study is small, which increases the likelihood that the marginal effects at the sample means are not the same as the average marginal effects. For this reason, we also compute the average of marginal effects computed at each observation, which is also the favored practice currently (see Greene 2002). These are also presented in the appendix.

A second set of results that take the sampling design into account is also presented. In order to have a balanced sample, data was collected from an equal number of surviving and failed grain banks villages. This was done because the number of surviving grain banks is much fewer than failed grain banks, and given that the total sample is small, proportional sampling would have resulted in very few observations on surviving grain banks. Out of 232 grain banks initially established, only 71 were still operational by the latest available Agragamee records, which implies that at the time of the survey, about 31 percent of grain banks that were initially established were still surviving. Therefore, in our survey, surviving grain banks were oversampled. Oversampling results in a choice-based sample or an endogenous-stratified sample. Maximizing a random-sample likelihood function as in (4) with choice-based sampling will yield inconsistent parameter estimates . A

relatively simple estimator which yields consistent estimates under choice-based sampling is the Weighted Endogenous Sampling Maximum Likelihood Estimator (WESMLE) developed by Manski and Lerman (1977), as cited in Greene (2002). In order to arrive at population parameter estimates, survey weights have to be used. This requires that the true population proportions of ones and zeros, ω_1 and ω_0, be known, which is the case in our study. Then, if p_1 and p_0 are the sample proportions of ones and zeros, the estimator is obtained by maximizing the weighted log likelihood function

(5) $$\ln L = \sum_{y_i=0} w_i \ln[1-\Phi(\mathbf{x}_i'\boldsymbol{\beta})] + \sum_{y_i=1} w_i \ln[\Phi(\mathbf{x}_i'\boldsymbol{\beta})],$$

where the weight is given by $w_i = y_i(\omega_1 / p_1) + (1-y_i)(\omega_0 / p_0)$.

2.4.2 The Cox proportional hazards model

As a complementary exercise, we also implement an analysis of the determinants of duration of grain bank survival, using the semiparametric Cox proportional hazards model. Survival models answer a different question than the probit model discussed previously. While the probit regression model estimates the probability of grain bank survival, the proportional hazards models estimate the likelihood of grain bank failure, conditional on having survived until the previous period. The normality assumption for time to an event conditional on the covariates is unreasonable in many cases (Keifer 1988), and models such as the probit model or linear regression model which make the normality assumption may be inappropriate to model time until failure. However, in such a case, survival models make more reasonable assumptions on the distribution of the error term. Which distribution is more appropriate depends on the assumption on the shape of the baseline hazard function. For example, if there is reason to assume that the baseline hazard function is constant over time, then the exponential model may be most appropriate. This implies that the failure rate is

independent of time. In this case, the failure process is said to lack memory. If the hazard function is assumed to be monotonically increasing or decreasing over time, then the Weibull model may be more appropriate. If, on the other hand, the hazard function is believed to be non-monotonic (i.e., increasing and then decreasing), then the log-normal accelerated failure-time model is more appropriate. A correctly-specified parametric hazards model provides efficient estimates.

However, if we do not have any reason to believe a priori that the hazard function follows a particular shape, the Cox proportional hazards model may be more appropriate as it does not assume any functional form for the baseline hazard (Cox 1972). This gives the Cox model a particular advantage over parametric models that may be using an incorrect specification for the distribution of the error term, thus biasing the parameter estimates. This gain however comes with a potential loss in terms of efficiency relative to maximum likelihood estimation of a correctly specified parametric survival model. The Cox model uses partial maximum likelihood to estimate how the covariates $\mathbf{x_j}$ shift the hazard function $h(t \mid \mathbf{x_j})$.

As discussed in Kiefer (1988), in the proportional hazards specification, the effect of the explanatory variables is to multiply the hazard function by a scale factor. In contrast, in the accelerated failure-time model, the effect of the explanatory variables is to rescale the time axis. The exponential and Weibull parametric models can be written in both the accelerated time metric as well as the proportional hazard metric. Although the results are equivalent, the advantage of choosing the hazard metric is that the estimates are more easily comparable to the Cox estimates.[37] Using the proportional hazard metric, we can write the hazard function for village *j* as follows:

[37] The log-normal accelerated failure time model has no proportional hazards interpretation. However, we find it attractive for conducting robustness checks, because of the flexible assumption on the shape of the underlying hazard function – first increasing and then decreasing.

(6) $$h(t \mid \mathbf{x_j}) = h_0(t)\exp(\mathbf{x_j}\boldsymbol{\beta}_\mathbf{x}),$$

where the vector of regression coefficient $\boldsymbol{\beta}_\mathbf{x}$ are to be estimated from the data. In our data, $\mathbf{x_j}$ is a vector of time-invariant village and grain bank covariates for village *j*. The parameter $\boldsymbol{\beta}_\mathbf{x}$ measures the semi-elasticity of the hazard with respect to $\mathbf{x_j}$. The baseline hazard, $h_0(t)$, is assumed to be the same for all observations in the data. As a result, the ratio of the hazards for the *j*th village and the *k*th village, which is given by

(7) $$\frac{h(t \mid \mathbf{x_j})}{h(t \mid \mathbf{x_k})} = \frac{\exp(\mathbf{x_j}\boldsymbol{\beta}_\mathbf{x})}{\exp(\mathbf{x_k}\boldsymbol{\beta}_\mathbf{x})}$$

is constant, giving these models their name ("proportional" hazards). As mentioned earlier, in the Cox model, $h_0(t)$ is not assumed to have any functional form.[38]

The second problem addressed by models of duration analysis, including the survival models used in this study, is censoring (see Keifer 1988). In our data, the observations on surviving grain banks are right censored, as they may fail in the time periods following the survey. This is not taken into account by the probit model used in the study, which puts all surviving grain banks at the time of the survey into one category. This problem is the lesser of the two and can be dealt with easily with the use of censored regression models such as the Tobit model, but survival models have the advantage of addressing censoring without assuming that the error term is distributed normally. It is important to note here that the right censoring resulting from when the survey was implemented is, conditional on the covariates and the grain bank's survival to a particular time, independent of the future value of the hazard for the grain bank.

---

[38] For the exponential model, we assume that $h_0(t) = \exp(a)$, where a is an extra parameter that has to be estimated from the data. In this model, the baseline hazard function is constant over time. For the Weibull model, we assume that $h_0(t) = p^{t^{p-1}}\exp(a)$, where a and p are extra parameters that have to be estimated from the data. In this model, the hazard function is monotonic. In the case of accelerated failure-time models, we have $\tau_j = \exp(-\mathbf{x_j}\boldsymbol{\beta}_\mathbf{x})t_j$. For the log-normal regression model, τ_j is distributed as log-normal with parameters (β_0, σ) which are estimated from the data.

To test the robustness of the results of the survival analysis, estimates from the different models are compared. In addition, in order to ascertain if the semi-parametric Cox model can provide potentially credible results, tests of the underlying assumption of proportional hazards are also implemented.

2.5 Empirical results and discussion

Informal conversations with Agragamee field staff indicated a belief that membership size and the degree of social heterogeneity (i.e., proportion of tribal and non-tribal households) increase loan default rates. In addition, they posited that member contributions at the time of inception of the grain bank seem to decrease loan default rates. These and other hypotheses are tested in this analysis.

The explanatory variables in the various regressions can be roughly grouped as follows:

1. village-level characteristics (e.g., village size, degree of remoteness, infrastructure, share of landless households, community cohesion, presence of other community-level groups);
2. socio-economic characteristics of grain bank members (e.g., ethnic diversity, share of females in grain bank committee); and
3. grain bank characteristics (e.g., establishment period, membership size, source of start-up grains).

Descriptive statistics for these variables are presented in Table 2.13.

Table 2.13: Descriptive statistics for estimation sample

	Mean	Std. Dev.	Min.	Max.
Village level variables				
Total households	53.46	35.04	5.00	179.00
Distance from block headquarters (km)	25.98	17.34	1.00	62.00
Share of households with no *pata* (deeded) land	0.29	0.24	0.00	0.92
Quality of village road (1 = untarred)	0.69	0.47	0.00	1.00
Meeting on needs basis (1 = yes)	0.53	0.50	0.00	1.00
Presence of self-help group (1 = present)	0.75	0.44	0.00	1.00
Grain bank level variables				
No. of member households in grain bank	38.81	25.67	5.00	139.00
Share of female members in grain bank committee	0.53	0.14	0.00	1.00
Share of grains contributed by villagers at grain bank inception	0.37	0.23	0.00	1.00
Ethnic diversity index	0.14	0.19	0.00	0.66
Established in 1993-94 (1 = yes)	0.43	0.50	0.00	1.00

Notes: $N = 80$. The summary statistics for the first three variables are reported in Table 2.1, but are presented in this table along with those of other variables that are included in the conditional analysis.

Discussion of explanatory variables

Village characteristics

The distance from block headquarters is used as a proxy for the degree of remoteness. Our hypothesis is that the more distant the village is from block headquarters, the less likely it is to access government services available there, and the greater its dependence on a village-level institution such as a grain bank. This may increase the probability of grain bank survival, as villagers with fewer alternatives may strive harder to sustain grain bank operations. In addition, given that Agragamee's head office is located in the block headquarters, the closer a village is to the block headquarters, the more likely it is to have operational support from the NGO. For this reason too, the distance from block headquarters may increase the probability of grain bank survival.

The quality of the village road – whether *pucca* (tarred) or *kachcha* (untarred) – is used as an indicator of the quality of physical infrastructure and the overall level of development of the village, since no other indicators of infrastructure with sufficient variation are available. We include a dummy variable for the quality of the village road, where 1 denotes whether the road is wholly or partially *pucca* and 0 if it is wholly *kachcha*. We hypothesize that less-developed villages obtain a higher level of utility from a functional grain bank as they face higher food insecurity risks (possibly attenuated by functional grain banks) and have fewer alternatives (possibly making functional grain banks even more valuable). Thus, we hypothesize that the probability of grain bank survival is likely to be higher in villages which are less developed (as indicated by road quality).

We include the share of landless households as another determinant of grain bank survival and duration of survival. Agriculture is the primary source of employment and livelihood of all households in the sample villages and almost all households practice *podu* cultivation on land cleared of forest cover.[39] However, they do not have formal titles to these lands, which are called *anabadi* (or encroached) land. Due to the lack of formal property rights, we define all households that do not have access to titled land as landless (even if they cultivate *anabadi* land). Using data from Pakistan, Heltberg and Del Ninno (2006) show that there is a positive correlation between landlessness and economic vulnerability. In our study, we hypothesize that villages having a higher share of landless households derive higher utility from a functional grain bank, since they are more vulnerable and face higher food insecurity.

A dummy variable for the presence of a self-help group in the village is included in the specifications as a proxy for social cohesion. We posit that the presence of a self-help group can improve grain bank functioning by increasing the

[39] See chapter 1 for more details on *Podu* cultivation.

level of social cohesion. A dummy variable indicating the frequency of village-level meetings is included as another proxy of social cohesion and community unity. In villages that reported organizing village-level meetings whenever the need arose, we hypothesize that grain banks are more likely to survive, as villagers may be more proactive and unified.

We also include the village size (as indicated by the total number of households) as a village-level variable that may impact the successful functioning of a grain bank. Anecdotes gathered during the survey from Agragamee field staff indicate that grain banks are more likely to fail in larger villages, since it could increase the costs of monitoring and coordination. This hypothesis is reinforced by a large body of literature, starting with Olson (1965), which demonstrates the negative impact of group size on collective action and the provision of public goods, as free rider problems are more likely to occur as group size increases.

Grain bank characteristics

Grain bank characteristics included as explanatory variables are the membership size, ethnic diversity of the membership, share of female committee members, share of grains contributed by grain bank members at the time of inception and establishment period.

The same argument applies for grain bank membership size as for village size in determining grain bank survival and longevity, namely, as group size increases, there may be an increase in coordination and free rider problems for the provision of activities that have features of a public good, such as peer monitoring.[40] A quadratic specification for membership size is also included to examine potential non-linearities

[40] See, for example, Ghatak and Guinnane (1999) for a discussion of group size and microcredit institutions.

in the conditional relationship between membership size and grain bank survival. Specifically, the quadratic term will help to capture if, for example, there is a concave, non-monotone effect of membership size on institutional survival – the likelihood that survival increases until a certain membership size (due to, for example, some amount of risk sharing) but then falls as it increases further (due to, for example, increasing informational costs).

Following the literature on ethnic diversity and public goods provision, we include an index of ethnic diversity as a quantitative indicator of ethnic diversity. Similar to Alesina et al. (1999), Easterly and Levine (1997) and Miguel and Gugerty (2004), we construct an ethnic diversity index ED_j for community j which measures the probability that two randomly chosen people from the population are from distinct ethnic groups. This is related to a Herfindahl index, and is calculated as

$$ED_j = 1 - \sum_e \left(\frac{n_{ej}}{N_j} \right)^2 ,$$

where n_{ej} is the number of people in ethnic group e in community j, N_j is the total population in community j, and e ={Scheduled Tribe, Scheduled Caste, Other Backward Castes, General Caste}.

The share of women in the grain bank committee is included to test the hypothesis that a larger share of women results in a higher likelihood of grain bank sustainability. Many microcredit and conditional cash transfer programs are targeted at women. For example, out of a membership of over 2.4 million, nearly 95 percent of Grameen Bank clients are female (Morduch 1999). Similarly, nearly 95 percent of the members of FINCA are women. One reason behind this is that women have been found to be less likely to default on their loans compared to men, perhaps due to fewer outside alternatives. Thus, women are better potential clients, and their presence on the committee may improve sustainability. Programs that target credit at women have

also been shown to increase consumption and children's health more than credit targeted at men (see, e.g., Pitt and Khandker 1998, Pitt et al. 2001). More generally, a study of women's participation in local government bodies, called *panchayats*, from Rajasthan and West Bengal has shown that increased female participation in decision-making bodies results in policy decisions that are closer to the preferences of women (Chattopadhyay and Duflo 2004). Similarly, we hypothesize that a higher share of women in the grain bank management committee can increase the probability of grain bank survival and duration, as grain bank survival contributes to increased consumption and household food security and better children's health outcomes, assuming these are "women's goods".

A variable indicating the amount of grains contributed by villagers as a share of the total grains collected at grain bank inception. This was included to test the hypothesis posited by Agragamee field staff that a higher level of "ownership" increased the likelihood of grain bank survival. The ostensible claim for this was that a higher level of contribution on the part of villagers increased the level of "vested interest" in it; i.e., if villagers contributed more, they had more to lose if the grain bank collapsed. While this may be true, another explanation for why a higher contribution share at grain bank inception may increase grain bank survival is that it may be indicative of a higher savings rate which enables a lower loan default rate, which, in turn, contributes to grain bank survival. However, while we can test if the contribution level is a significant determinant of grain bank survival and longevity, it is not possible to distinguish between competing explanations behind why it may be so.

We also include a dummy variable indicating whether the grain bank was established in an earlier or later period. This variable equals one if the grain bank was instituted in 1993-94 (the "early" years), and zero otherwise. This variable can

provide some evidence on whether chronology of grain bank establishment matters for survival. For example, if there is learning by doing, then we should find that the earliest established grain banks are more likely to fail.

Main findings

Here, we first present the results from estimating a binomial probit model of the determinants of grain bank survival. We then present the results from estimating a Cox proportional hazards model of the determinants of grain bank duration.

For each regression model, four sets of results are presented. We refer to these as the results for specification 1 (village variables only), specification 2 (grain bank variables only), specification 3 (both village and grain bank variables, but no interaction terms), and specification 4 (both village and grain bank variables, including interaction terms). Specifications 3 and 4 include all village and grain bank explanatory variables included in specifications 1 and 2, except village size (since it is highly correlated with grain bank membership size included in specifications 3 and 4, which we feel is a better variable to capture the impact of group size on institutional survival and longevity). These two specifications include the same explanatory variables, except for the interaction between the grain bank ethnic diversity index and membership size, which we include in specification 4 only.[41] We refer to specification 4 as the "full" specification, as it has the most complete set of explanatory variables.

A. Binomial probit estimates

Estimated coefficients from fitting a binomial probit model to the data via MLE are reported in Table 2.14. For ease of interpretation, the corresponding marginal effects

[41] Interactions between ethnic diversity and incidence of needs-based meetings /share contributed by villagers at inception/ share of female committee members were found to be statistically insignificant in both the probit and survival models, and were not included in the final specification.

at the sample means (MEMs) as well as average marginal effects (AMEs) are presented in Tables 2.A1 and 2.A2, respectively in the appendix. As expected, given the small sample size, there are significant differences between the MEMs and the AMEs.

As presented in Table 2.14, from specifications 1, 3 and 4, it appears that grain banks in villages at lower levels of development, as indicated by the quality of village road dummy, are more likely to survive. This result is highly statistically significant across the alternative specifications. We interpret the magnitudes of the estimates from the full specification by referring to the AMEs (see Table 2.A2). We find, ceteris paribus, that a village having *pucca* (tarred) roads is, on average, about 37 percentage points less likely to have a surviving grain bank. The presence of a self-help group is found to have a statistically significant and positive effect on the likelihood of grain bank survival. Specifically, the presence of a self-help group increases the probability of survival by, on average, 38 percentage points. This may indicate that conditions that support self-help group formation also support grain bank survival, or that grain bank survival may depend on the implementation of parallel schemes such as self-help groups. In the latter case, we can infer that the presence of another jointly-owned village institution such as a self-help group may increase the level of interaction between member households and the level of social cohesion, which in turn promotes successful operation of a grain bank. Finally, villages where meetings are arranged according to demand or whenever the need arises are also found to be, on average, 59 percentage points more likely to have surviving grain banks. These two findings are also highly statistically significant and robust across alternative specifications.

We now turn to the estimated coefficients associated with the grain bank level variables. In specification 2, we find that the only variable that is statistically

significant is the amount of grains contributed by villagers as a fraction of total grains at grain bank inception. However, the variable loses its significance in specifications 3 and 4, once village-level variables are controlled for. Thus, this finding is not robust to the inclusion of village-level covariates.

We find that membership size is not statistically significant in any of the specifications. The quadratic term for membership size is significant at the 10 percent level in the full specification only. Moreover, it is insignificant in practical terms, as the estimated coefficient is very small. The estimated coefficient on ethnic diversity is also statistically significant at the 10 percent level in the full specification only. Moreover, the sign of the estimated coefficient is not robust across alternative specifications. We also find that the interaction between membership size and ethnic diversity has a statistically significant effect on grain bank survival at the 10 percent level.

Table 2.15 presents the WESMLE probit estimates, using the reported population proportion. As mentioned earlier, the proportion of surviving grain banks in the population is about 0.31. The results are similar to those from the MLE probit estimates, both in terms of magnitudes as well as signs. Quality of village roads, the presence of a functional self-help group and the incidence of needs-based meetings, all continue to be highly statistically significant. Like the MLE probit estimates, the share contributed by villagers is statistically significant in specification 2, but this variable is not robust to the inclusion of village-level variables. Similarly, like the MLE probit estimates, the interaction between ethnic diversity and membership size, in specification 4, is statistically significant at the 10 percent level.

Table 2.14: Determinants of grain bank survival: Estimated coefficients
MLE binomial probit regression estimates

Dependent variable: Surviving grain bank (1 = yes)

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.007			
	(0.006)			
Distance from block headquarters	0.002		0.002	0.000
	(0.012)		(0.015)	(0.016)
Share of landless households	1.252		1.240	1.350
	(0.817)		(0.876)	(0.916)
Quality of village road (1 = tarred)	-1.202***		-1.414***	-1.804***
	(0.419)		(0.519)	(0.612)
Meeting on needs basis (1 = yes)	2.071***		2.453***	2.697***
	(0.463)		(0.608)	(0.670)
Presence of self-help group (1 = yes)	1.446***		1.616***	1.845***
	(0.465)		(0.521)	(0.571)
Membership size		0.013	0.003	-0.008
		(0.019)	(0.024)	(0.025)
Square of membership size		0.000	0.000	0.001*
		(0.000)	(0.000)	(0.000)
Share of female members in committee		1.569	1.951	2.503
		(1.180)	(1.558)	(1.604)
Share contributed by villagers at inception		1.451**	-0.606	-0.567
		(0.684)	(1.172)	(1.239)
Ethnic diversity index		-0.661	0.166	4.839*
		(0.943)	(1.099)	(2.574)
Grain bank established 1993-94 (1 = yes)		-0.137	0.121	0.214
		(0.294)	(0.460)	(0.518)
Membership size*Ethnic diversity index				-0.122*
				(0.062)
Constant	-2.589***	-1.636*	-3.627**	-4.366***
	(0.822)	(0.860)	(1.495)	(1.635)
LR χ^2	38.82	8.72	42.92	47.71
p-value	0.000	0.190	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.079	0.387	0.430
Adjusted McFadden's Pseudo-R^2	0.224	-0.048	0.171	0.196

Notes: $N = 80$. Standard errors are reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

However, unlike the MLE probit model, in specification 4, the quadratic term for membership size, which was significant at the 10 percent level in the MLE probit model, loses statistical significance. In addition, the share of female members in the grain bank committee, which was not significant in the MLE probit model, gains significance at the 10 percent level in specification 3 and 4 in the WESMLE probit estimates. However, since this finding is not robust across specifications in the WESMLE probit model, we discount this finding. Finally, we also find that a higher share of landless households has a significant positive effect on grain bank survival.

In both the MLE and WESMLE probit estimates, we find a negative relationship between the index of ethnic diversity and likelihood of grain bank survival in specification 2. This finding is similar to other studies that have looked at ethnic diversity and the provision of public goods (e.g., Alesina et al. 1999, Alesina and La Ferrera 2000, Miguel and Gugerty 2004) or ethnic diversity and economic development (e.g., Easterly and Levine 1997). However, the relationship is not statistically significant. This finding is robust to alternate specifications where the ethnic diversity index is substituted by a dummy variable for uniethnic grain bank membership. This result is not surprising, given that there is a high degree of socioeconomic homogeneity in tribal Orissa. However, the result may not be generalizable to other regions where there is greater socioeconomic diversity.[42] The sign on the coefficient for ethnic diversity flips in specifications 3 and 4, and although it is statistically significant at the 10 percent level in the latter (only in the case of the MLE probit estimates), we discount this finding as the estimate is not robust across specifications.

[42]Note, however, that even in the presence of social heterogeneity, it is possible that the benefits from grain banks outweigh the costs associated with interacting with other ethnic groups. For example, Wade (1994) shows that socially diverse farming communities in India that face greater crop risks due to lack of irrigation are more likely to develop effective collective action mechanisms to deal with this problem than communities with similar levels of heterogeneity but less crop risk.

Table 2.15: Determinants of grain bank survival: Estimated coefficients
WESMLE binomial probit regression estimates ($\omega_1 = 0.31$)

Dependent variable: Surviving grain bank (1 = yes)

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.008			
	(0.005)			
Distance from block headquarters	0.001		0.002	0.000
	(0.010)		(0.013)	(0.014)
Share of landless households	1.325*		1.330*	1.370*
	(0.736)		(0.753)	(0.804)
Quality of village road (1 = tarred)	-1.204***		-1.437***	-1.764***
	(0.428)		(0.451)	(0.513)
Meeting on needs basis (1 = yes)	2.102***		2.417***	2.611***
	(0.494)		(0.502)	(0.534)
Presence of self-help group (1 = yes)	1.524***		1.660***	1.822***
	(0.514)		(0.491)	(0.516)
Membership size		0.013	0.006	-0.005
		(0.018)	(0.024)	(0.024)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		1.543	2.090*	2.484*
		(1.081)	(1.256)	(1.323)
Share contributed by villagers at inception		1.440**	-0.488	-0.464
		(0.657)	(1.090)	(1.214)
Ethnic diversity index		-0.683	-0.041	4.252
		(0.936)	(1.152)	(2.866)
Grain bank established 1993-94 (1 = yes)		-0.169	0.129	0.195
		(0.287)	(0.417)	(0.470)
Membership size*Ethnic diversity index				-0.111*
				(0.062)
Constant	-3.166***	-2.108***	-4.301***	-4.827***
	(0.865)	(0.778)	(1.463)	(1.416)
Wald χ^2	22.62	8.28	36.62	37.42
p-value	0.001	0.219	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.077	0.386	0.420
Adjusted McFadden's Pseudo-R^2	0.208	-0.065	0.142	0.156

Notes: $N = 80$. Standard errors are reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

We find that membership size (both the linear and quadratic terms) has a very small impact on grain bank survival, and it is not statistically significant, except in specification 4 in the MLE probit estimates, where we find that the quadratic term has a positive effect on the probability of grain bank survival, though the magnitude is small. This indicates that membership size is not an important determinant of grain bank survival. In addition, village size is also found to have a very small, positive but statistically insignificant effect on the probability of grain bank survival. This may be an indicator that village or grain bank membership size are not important determinants of grain bank survival in the data. This is not surprising given the small population size of settlements, economic and social homogeneity across households, and the high degree of observability of household economic and social behavior by others in the community in the sample villages.

Model diagnostics

Robustness to alternative population proportions of grain bank survival

To check if the results are sensitive to the use of alternative population proportions of surviving grain banks, we reestimate the WESMLE probit model with a population proportion lower than the reported population proportion of survival, namely 0.25, and a population proportion slightly higher than the reported population proportion of survival, namely 0.33. Tables 2.A5 and 2.A6 present coefficients estimated from maximizing a weighted log likelihood using the lower and higher population proportions of survival, respectively. We find that all the estimated coefficients have the same direction and comparable magnitudes as the estimated coefficients from the WESMLE probit model using the reported population proportion of survival of 0.31.

Goodness of fit

McFadden's likelihood ratio index (LRI) or pseudo-*R*-squared is provided at the bottom of Table 2.14 to compare the fit of the different specifications. We find that the LRI is higher for specification 1, 3 and 4 than 2, and highest for specification 4. The LRI increases as the fit of the model improves (Greene 2002). This implies that specification 4 (which we call the "full" model) provides the best fit. However, the LRI also increases as the number of explanatory variables increase. So, we also present the adjusted pseudo-*R*-squared measure that takes into account the number of explanatory variables. We find that the adjusted pseudo-*R*-squared is highest for specification 1, which also suggests that the explanatory power of the full model is driven by the village-level variables in the specification.

Multicollinearity

Multicollinearity can result in large estimated standard errors, thus impairing statistical inference. Variance inflation factors are a widely used measure of the degree of multicollinearity. It is essentially based on the *R*-squared value obtained by regressing the *i*th explanatory variable on the rest of the explanatory variables in the regression model.[43] We find that the variance inflation factors for the set of variables used in each of the three specifications is found to be around one. The rule of thumb commonly applied is that a value of 10 or higher is a sign of severe multicollinearity. Our findings indicate that multicollinearity is most likely not a problem in our model, even by conservative measures.

[43] The diagnostic information for multicollinearity is obtained by estimating specifications 1-4 via ordinary least squares even though the dependent variable is dichotomous. Menard (2002) argues that this is permissible as the relationship under scrutiny is that between the independent variables in the model and the functional form for the model of the dependent variable is not relevant.

Percentage of correct predictions

As an alternate measure of the predictive ability of the model, we look at the percentage of correct predictions made by specification 4. Tables 2.A7 and 2.A8 in the appendix provide a summary of the predictive ability of specification 4 fitted using the MLE and WESMLE probit models respectively. For the former model (which has a balanced sample of surviving and failed grain banks), we use the threshold value of 0.5, on the basis that we should predict a one if the model says a one is more likely than a zero.[44] We find that specification 3 predicts 65 out of 80, or about 81 percent of observations, correctly.

To examine the predictive ability of a probit model for an unbalanced sample, Heckman and Smith (1999) discuss the inappropriateness of the use of 0.5 as the cutoff, and recommend instead the use of the population probability as the threshold value. Therefore, in the case of the WESMLE probit model, we use the population probability (which in our case is 71/232, or approximately 0.31) as the threshold value. Thus, we predict an observation to be a one if it has $\hat{p} \geq 0.31$ and a zero otherwise. Using this criterion, we find that 63 out of 80, or about 79 percent of observations are predicted correctly.

Hypothesis testing using linear restrictions

We use the likelihood ratio (LR) test to see if all the point estimates in the probit model are zero. We place restrictions on the village level variables to test the hypothesis that the coefficients for all the village variables are simultaneously equal to zero. We do this by nesting specification 1 in specification 4 (which we call the full

[44] Use of 0.5 as the threshold may not be a good choice if the sample is unbalanced (Greene 2002). In such cases, the percentage of correct predictions may not be a good measure of the predictive ability of the model as the result may be driven by a disproportionate number of either zeros or ones being predicted correctly. However, the sample used in this study has an equal number of zeros and ones, and we also find that almost an equal number of zeros and ones are predicted correctly.

model). From the results of the chi-squared test, we reject the null hypothesis that the village level variables are simultaneously equal to zero.

Similarly, we perform a LR test of the hypothesis that all the grain bank variables are simultaneously equal to zero. However, we are unable to reject the null hypothesis. This indicates that the results are driven by the village-level variables. This is also corroborated by the fit measures provided at the bottom of Tables 2.14 and 2.15. From the *chi*-squared test, we find that the model in specification 2 (grain bank level explanatory variables only) is not statistically significant, but the models in specifications 1, 3 and 4 are.

Correction for potential heteroskedasticity

To see if the estimates are biased due to heteroskedasticity, we estimate the MLE probit model with heteroskedasticity-robust standard errors (using the Huber-White sandwich estimator). The estimates are presented in Table 2.A9. We find that the results are qualitatively and quantitatively similar to those presented in the MLE probit models in Table 2.14, which gives us some confidence that heteroskedasticity may not be a problem in our data. However, given that the robust estimator is an asymptotic correction and our analysis is based on a small sample, we recognize that diagnostic check does not rule out the presence of heteroskedasticity.

B. Cox proportional hazards estimates

To estimate the hazards model, we create a duration variable. This variable is uncensored for grain banks that have ceased to function and is right-censored for grain banks that were operational at the time of the survey, that is, for those that have not yet failed. For failed grain banks, the duration variable is the number of years of operation (survival) until failure. For surviving grain banks, the duration variable is

the number of years between inception and the year of the survey. The explanatory variables used in this model are the same as in the probit model. Further, as with the probit regression, four alternate model specifications are estimated, which we continue to refer to as specifications 1, 2, 3 and 4 respectively.

The coefficients estimated by the Cox proportional hazards model are presented in Table 2.16. For continuous variables, coefficients can be interpreted as follows: if the coefficient for distance from block headquarters is –0.015 (as in specification 4), then we say that a unit increase in the distance lowers the hazard of failure by roughly 1.5 percent, because *exp*(-0.015) = 0.985. For dummy variables, coefficients can be interpreted as follows: if the estimated coefficient for presence of a self-help group is –1.996, then we say that villages with a functional self-help group (i.e., presence of self-help group = 1) face a hazard roughly 86 percent lower than villages where the self-help group is not functional, because *exp*(-1.996) = 0.14. For ease of interpretation, we present hazard ratios associated with the estimated coefficients from the Cox model in Table 2.A10 in the appendix

As presented in Table 2.16, we find that village size and grain bank membership size do not have a statistically significant effect on the hazard of grain bank failure. Consistent with the probit estimates, the dummy variables for quality of village roads and social cohesion (indicated by the presence of self-help groups and meeting on as-needed basis) all have highly statistically, as well as economically, significant effects. Interpreting the estimates in specification 4, we find that villages with *pucca* roads face more than four times the hazard of failure faced by villages with *kachcha* roads. The hazard faced by villages with self-help group is about 86 percent lower, while the hazard faced by villages which hold meetings whenever need arises is 90 percent lower. These findings suggest the importance of social cohesion in extending time until grain bank failure.

The share of landless households does not appear to be an important determinant of time until failure. This finding is the same as the results from the MLE probit estimates. The share of female members in the grain bank committee has a large and highly statistically significant effect in specifications 3 and 4, but not in specification 3. As with the probit estimates, the share of the contribution made by villagers at the time of grain bank inception is found to be statistically significant in the second specification, but the significance drops away with the addition of village-level variables in the third specification. Thus, we claim that these variables are not as important in determining grain bank survival as the village-level variables discussed earlier.

Model diagnostics and robustness checks

We perform a range of diagnostics to test the model specification and the underlying assumptions. We also compare the results from the different survival models to test the robustness of the estimates to model specification.

Regression misspecification test

We use the link test to verify that the estimated coefficient on the squared linear predictor is insignificant. As discussed by Cleves et al. (2004), the basis for this test is to first estimate the $\boldsymbol{\beta}_{\mathbf{x}}$ vector from the model being fitted, and then estimate β_1 and β_2 from a second round model $\text{LRH} = \beta_1(\mathbf{x}\hat{\boldsymbol{\beta}}_{\mathbf{x}}) + \beta_2(\mathbf{x}\hat{\boldsymbol{\beta}}_{\mathbf{x}})^2$. Under the assumption that $\mathbf{x}\boldsymbol{\beta}_{\mathbf{x}}$ is correctly specified, we have that β_1=1 and $\beta_2 = 0$.[45] We find this to be the case in all four specifications.

[45] This result holds not only for survival models, but for all models.

Table 2.16: Determinants of time until grain bank failure: Estimated coefficients
Partial-likelihood Cox proportional hazards model estimates

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	-0.004			
	(0.005)			
Distance from block headquarters	-0.016		-0.016	-0.015
	(0.011)		(0.014)	(0.014)
Share of landless households	-0.465		-0.216	-0.268
	(0.808)		(0.854)	(0.845)
Quality of village road (1 = tarred)	1.154***		1.302***	1.479***
	(0.384)		(0.415)	(0.416)
Meeting on needs basis (1 = yes)	-2.135***		-2.436***	-2.560***
	(0.419)		(0.505)	(0.526)
Presence of self-help group (1 = yes)	-1.536***		-1.771***	-1.996***
	(0.473)		(0.544)	(0.570)
Membership size		-0.020	0.008	0.013
		(0.020)	(0.022)	(0.023)
Square of membership size		0.000	0.000	-0.000*
		(0.000)	(0.000)	(0.000)
Share of female members in committee		-1.935	-3.066**	-3.039**
		(1.219)	(1.393)	(1.373)
Share contributed by villagers at inception		-1.361*	0.430	0.612
		(0.694)	(1.122)	(1.128)
Ethnic diversity index		0.349	0.293	-2.704
		(1.007)	(0.992)	(2.064)
Grain bank established 1993-94 (1 = yes)		0.096	-0.057	0.032
		(0.325)	(0.409)	(0.407)
Membership size*Ethnic diversity index				0.084*
				(0.047)
LR χ^2	37.73	8.46	44.36	47.99
p-value	0.000	0.207	0.000	0.000
Log likelihood	-145.85	-160.49	-142.53	-140.72

Notes: $N = 80$. Standard errors are reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Comparing estimates across Cox and parametric model estimates

First, we compare the estimates from various parametric models to the Cox model estimates to see if they are robust across models. The estimates from the exponential and Weibull models are presented in the appendix in Tables 2.A11 and 2.A12, respectively. The latter are estimated in the proportional hazard metric. The advantage of using this metric is that the estimates are directly comparable to Cox estimates. We also present the estimates from the log-normal accelerated time metric in Table 2.A13. While the estimates from the log-normal model are not directly comparable, we can compare the significance level and direction of the estimates with the Cox estimates. If they prove not to be similar, then there is evidence of a misparametrized underlying baseline hazard.

We find that the estimates from all three parametric models are roughly similar to the Cox estimates. In particular, the coefficients on village road quality, the presence of functional self-help groups and the incidence of village meetings on a needs-basis are highly statistically significant across all specifications in the different survival models. The estimated coefficient on the variable denoting the share of females in the grain bank committee is statistically significant in specifications 3 and 4 (but not specification 2) across the Cox, Weibull and log-normal models, although it is not so in the exponential model.[46] The estimates from the log-normal model, while not directly comparable to the Cox estimates, show that almost all the same variables have a statistically significant effect on the hazard rate and are in the same direction as those from the Cox. For example, the incidence of meetings on a needs-basis and the

[46]This finding does not imply that the Cox estimates are questionable, as we find that the Weibull and log-normal models are a better fit for the data than the exponential model. For the Weibull estimates of specification 3 and 4, a Wald test for $H0 : \ln(p) = 0$ for which the test statistic is 5.20 and 5.52, indicates that we can reject the null hypothesis that the hazard is a constant. The likelihood ratio *chi*-square test at the bottom of tables A2.11-A2.13 also indicate that, of the parametric models, the Weibull and log-normal models may be a better fit for the data.

presence of a functional self-help group has the effect of slowing down time (or delaying grain bank failure), while the effect of a tarred village road is to accelerate time (or hasten failure). Thus, in general, we find that the results from the Cox model are similar to those from the parametric models.

Estimating the baseline cumulative hazard and hazard contributions

Although the Cox model does not produce a direct estimate of the relative hazard of grain bank failure, we can estimate the baseline cumulative hazard function from the vector of coefficients estimated by the model. On examining the cumulative hazard function, we find it to be increasing, initially at a slightly increasing rate and then at a decreasing rate (see Figure 2.A1 in the appendix). Since the baseline hazard is the derivative of the cumulative hazard, this may imply that the baseline hazard is increasing. However, when the discontinuities associated with the step functions in the cumulative hazard function are smoothed out in order to estimate the baseline hazard, we find that the baseline hazard is first increasing and then falling (see Figure 2.A2).[47] While the shape of the hazard function is important for deciding which of the parametric models is more appropriate for our data, the accuracy of the Cox estimates does not depend on any assumption on the shape of the underlying hazard. However, this exercise is informative in that it provides guidance on which of the parametric models is more appropriate as a test of robustness of the Cox model. We find that the exponential model, which assumes a constant hazard, is not a good fit for the data and therefore, in this study, it is more appropriate to examine if the estimates from the Cox model are similar to the estimates from the log-normal or Weibull models, which we find to be the case.

[47] To estimate the hazard, a standard kernel-smoothing methodology is used, as discussed in Cleves et al. (2004).

Tests of the proportional hazards assumption

In order to test the underlying assumption of proportional hazards for each covariate, which underlies the Cox model, we conduct two tests of specification 4.

Interacting analysis time with covariates

First, we interact analysis time with the covariates (individually) to examine if the effects of the covariates on the hazard change with time, because the proportional hazards assumption states that the effects of the covariates do not change with time except in ways which we have already parameterized (Cleves et al. 2004). The basis for this test is to first estimate the $\boldsymbol{\beta}_{\mathbf{x}}$ vector from the model being fitted, and then estimate second round models separately for each covariate as follows: $\text{LRH} = \mathbf{x}\boldsymbol{\beta}_{\mathbf{x}} + \beta_1(x_1 t)$ and test $\beta_1 = 0$, $\text{LRH} = \mathbf{x}\boldsymbol{\beta}_{\mathbf{x}} + \beta_2(x_2 t)$ and test $\beta_2 = 0$, and so on. We find this to be the case with the exception of the variable denoting distance to block headquarters. In other words, the effects of the other interacted variables are not different from zero.

Test of Schoenfeld residuals

Second, we conduct a test of nonzero slope in a generalized linear regression of the scaled Schoenfeld residuals on functions of time. As discussed by Cleves et al. (2004), the Schoenfeld residual for covariate $x_u, u = 1,...., p,$ and for observation *j* observed to fail is given by the residual

$$r_{uj} = x_{uj} - \frac{\sum_{i \in R_j} x_{ui} \exp(\mathbf{x_i}\hat{\boldsymbol{\beta}}_{\mathbf{x}})}{\sum_{i \in R_j} \exp(\mathbf{x_i}\hat{\boldsymbol{\beta}}_{\mathbf{x}})},$$

where $\hat{\beta}_x$ is the vector of estimated coefficients. In words, r_{uj} is the difference between the covariate value for the failed observation and the weighted average of the covariate values (weighted according to the relative hazard from a Cox model) over all the subjects at risk of failure when subject *j* failed. If we assume that the coefficient

on x_u does vary with time (that is, the proportional hazards assumption is violated), β_u can be written as

$$\beta_u(t) = \beta_u + q_j g(t),$$

where q_j is some coefficient and $g(t)$ is some function of time. If the assumption of proportional hazards is valid, $q_j = 0$. In addition, if we graph a scaled Schoenfeld residual against t_j and the curve has zero slope, we cannot reject the null hypothesis of proportional hazards.

Having estimated the Schoenfeld residuals after fitting the Cox model to the data, we find no evidence that our specification violates the proportional hazards assumption. As with the previous set of tests where we interacted time with each covariate, we find that only the variable denoting distance from the block headquarters appears to violate the proportional hazards assumption. This holds at the 10 percent significance level. We also graph the scaled residuals against time for each covariate, and find that the curves are roughly linear (see Figures 2.A3-2.A12 in the appendix). Thus, we cannot reject the assumption of proportional hazards.

Unobserved heterogeneity ("frailty")

As discussed in Kiefer (1988), heterogeneity arises when differences remain in the distribution after controlling the effect of observable variables. In the presence of unobserved heterogeneity (or "frailty"), the non-frailty model leads to downward biased estimates of the duration dependence in the true baseline hazard. Thus, heterogeneity due to omitted variables, either observable but which were not captured during the data collection, or unobservable, can give rise to biased coefficients. To address this problem, we can estimate a parametric survival model which also specifies a distribution for the unobserved effect. Although Heckman and Singer (1984) demonstrate that the heterogeneity term is sensitive to the distributional

assumption, Kiefer (1988) discusses that this finding may instead be due to incorrect specification of the survivor function. In our case, we estimate all three parametric models with a Gamma as well as inverse-Gaussian distribution for the heterogeneity term, to examine if our results are affected by unobserved heterogeneity. The results are qualitatively similar, and in the appendix we present the results for the former (see Tables 2.A14-2.A16). We find, with both the Gamma and inverse-Gaussian distributions for the heterogeneity term, that the estimated coefficients are very close to those from the model without frailty in specifications 1, 3 and 4. The frailty distribution variances are close to zero and the *p*-values for the likelihood ratio test (that frailty variance equals zero) are close to one, indicating that there is likely negligible unobserved heterogeneity. In the Weibull and lognormal models, in the case of specification 2 (i.e., grain bank variables only), we find evidence of unobserved heterogeneity, implying that there are omitted variables in this specification. This problem, however, is rectified when we add in village-level variables as in specifications 3 and 4. Thus, we can conclude that our estimates most likely do not suffer from a problem of unobserved heterogeneity.

2.6 Summary and conclusion

In this chapter, we present an overview of 80 villages surveyed as part of a study of community grain banks in Dasmantpur, Koraput. We also examine the institutional and environmental factors behind the survival of community grain banks introduced as part of a food security project in the 1990s. Such an analysis is important in light of the fact that the majority of grain banks established in this region have ceased to function. Yet, they continue to be seen as an integral part of the food security strategy of the government as well as rural development NGOs in tribal Orissa. In addition,

the government of India is expanding its grain bank scheme in tribal villages across the country to address the problem of food shortages in tribal areas.

In the first part of this chapter, using data from a recent survey of grain banks, we provide an overview of the socioeconomic features of the survey region. The region is characterized by a rural subsistence economy. Agriculture is the main source of employment and livelihood. The sample villages are characterized by poor infrastructure and transport and communication facilities. About 90 percent of the sample villages did not have access to the electric power grid. For over half the villages, the distance to the block headquarters, which was the seat of government agencies, medical facilities and major markets, was reported to be over 25 kilometers (or over 15 miles). But due to poor or non-existent transport facilities and poor road conditions, traversing this distance often took the better part of the day. In the majority of villages, walking was reported to be the main mode of transport to the block headquarters. More than two-thirds of the sample villager reported that their access and internal roads are wholly *kachcha* or untarred.

Using self-reported measures of household food sufficiency, all the villages were found to have food insecure households for at least one month of the year. Thirty-seven villages reported having households that faced food insufficiency for almost half the year, while the rest reported having households that faced food insufficiency for 3-4 months. The months most commonly reported as food shortage months were the late summer and monsoon months of *Landi*, *Aashad* and *Shravana* (mid-May to mid-August).

In the second part of the chapter, we present an overview of the sample surviving and failed grain banks. Due to the fact that grain bank membership is confined to members of the same village, grain banks cannot help member households cope with covariate risk. However, in villages where grain banks continue to operate,

loans tend to be state-contingent: participating households who defaulted on loans due to crop failures typically did not lose grain bank membership; instead, their loan repayment period was extended. To the extent that community grain banks provide consumption loans to cope with illness shocks or consumption shocks such as weddings, they provide member households with insurance against idiosyncratic risk. However, these types of loans are not the primary service of grain banks. Rather, successfully-functioning grain banks provide loans to member households, typically in the summer and monsoon season, to alleviate short-term, seasonal food shortages. These loans are provided at an interest rate usually between 25-50 percent, which is lower than the interest rates charged by local moneylenders In addition, according to anecdotal evidence collected from grain bank beneficiaries, these institutions have also helped to cultivate a savings habit that did not exist prior to the intervention. The role of NGO field staff was reported as critical in this endeavor. This is important to keep in mind for future grain bank schemes. If the intervention is purely institutional in nature and its success depends on inducing a behavioral change, then simply replicating grain banks by providing an external grant, without the impetus and inspiration needed to influence behavior, may not be fruitful.

One commonly-reported problem in grain bank operation was the lack of storage facilities, with only one of the 80 sample villages having a dedicated storehouse for the grains. This shortcoming, if addressed in future grain bank interventions, can reduce grain loss due to poor storage facilities. In addition, taking into account the type of grain that is most suitable for storage given the climatic conditions may also help to reduce losses due to storage.

A preliminary, unconditional examination of grain bank failure by year of establishment does not reveal any pattern in the chronology of failure. Contrary to expectation, earlier established grain banks did not collapse (which would have

indicated support for the "learning by doing" hypothesis). However, grain bank failure is higher in years of poor rainfall and crop failure, indicating that these institutions are not able to cope well with covariate shocks. Proximate reasons reported for grain bank failure also included mismanagement by the grain bank committee, indicating that there was a lack of managerial skills on the part of the grain bank committee. Thus, future grain bank schemes should include the necessary training in order to enable grain bank committee members to perform their responsibilities in a satisfactory manner.

Next, we estimate a binomial probit model of the determinants of grain bank survival. We find that grain banks in a village at a lower level of development (as indicated by the village road quality) have a higher predicted probability of survival. We interpret this to mean that the net benefit from grain bank survival is higher in less developed villages. We also find that the probability of grain bank survival is higher in the presence of a credit and borrowing group, known as self-help groups, and in villages where meetings are held on an as-needed basis. This may indicate that the presence of other community-based programs and meetings increases social cohesion and enables the survival of grain banks, or that factors that support the successful functioning of a self-help group also support grain bank survival. These findings are robust to specification.

We also estimate a Cox proportional hazards model to examine the impact of selected village and grain bank covariates on the hazard of grain bank failure. The results are qualitatively similar to the probit estimates – social cohesion (represented by the incidence of needs-based village meetings and presence of self-help groups) is found to decrease the hazard of failure, while a higher level of development (represented by better village road quality) is found to increase the hazard of grain bank failure. We also find that a higher share of women in the grain bank committee

reduces the hazard of grain bank failure. Given the strong assumption of proportionality underlying the Cox model, we also estimate parametric survival models based on the exponential, Weibull and log-normal distributions which do not make that assumption. We find that, in general, the results are robust across model specifications. We also implement other robustness checks, including tests of the proportionality assumption, which we find that we cannot reject.

Grain banks have some unique features, such as the provision of consumption credit, membership-based management and ownership and the ability to respond to food shortages in a timely manner, that make them a promising complement to existing food security programs. However, over two-thirds of grain banks established in the survey region and adjoining districts have ceased to function. This raises questions about the efficacy of the intervention, at least in its current form. Firstly, if the survival of grain banks depends crucially on village-level factors, such as the level of social cohesion, then implementing the same design in different environments does not bode well for enabling grain bank sustainability. Thus, careful geographic targeting may be critical.

Secondly, if community grain banks cannot cope with covariate shocks, then establishing these institutions in their current form in drought-prone areas is a recipe for failure. A possible innovation that can mitigate grain bank collapse during weather shocks is including the provision of index-based crop insurance, where indemnity payments can be indexed to area yields or weather.[48] Such insurance can provide households with consumption credit to cope with food shortages that occur due to recurrent covariate shocks without being vulnerable to moral hazard problems.

[48] See Mahul and Skees (2006) for a description of an index-based livestock insurance program in Mongolia.

An issue that we are not able to examine is the impact of the level of capitalization on grain bank sustainability. The data used in this study comes from grain banks from a small region subject to the same covariate shocks. Yet, not all grain banks are observed to fail. Therefore, it is possible that there exists a critical threshold level of capitalization above which grain banks can withstand a shock. However, in the absence of data on the level of grain bank stock prior to collapse, we are not able to examine this. Such data could provide critical information for future interventions. Other potential research projects for the future involve the analysis of grain bank data from a larger geographical area as well as the analysis of data from community grain banks having sufficient variation in the design features. Grain bank data from a larger geographical area can also provide an estimate of the role played by weather shocks on survival. In addition, data from grain banks having variation in design features can help to identify design parameters that determine grain bank sustainability and success.

APPENDIX

Table 2.A1: Determinants of grain bank survival: Marginal effects at the mean
MLE binomial probit regression estimates

Dependent variable: Surviving grain bank (1 = yes)				
	Marginal effects evaluated at sample means (MEM)			
Independent variables	Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.003			
	(0.002)			
Distance from block headquarters	0.001		0.001	0.000
	(0.005)		(0.006)	(0.006)
Share of landless households	0.500		0.495	0.538
	(0.326)		(0.350)	(0.365)
Quality of village road (1 = tarred)	-0.443***		-0.505***	-0.600***
	(0.131)		(0.148)	(0.140)
Meeting on needs basis (1 = yes)	0.699***		0.779***	0.820***
	(0.108)		(0.114)	(0.107)
Presence of self-help group (1 = yes)	0.503***		0.543***	0.587***
	(0.125)		(0.129)	(0.123)
Membership size		0.005	0.001	-0.003
		(0.007)	(0.010)	(0.010)
Square of membership size		0.000	0.000	0.000*
		(0.000)	(0.000)	(0.000)
Share of female members in committee		0.626	0.778	0.998
		(0.471)	(0.622)	(0.639)
Share contributed by villagers at inception		0.579**	-0.242	-0.226
		(0.273)	(0.468)	(0.494)
Ethnic diversity index		-0.264	0.066	1.929*
		(0.376)	(0.438)	(1.024)
Grain bank established 1993-94 (1 = yes)		-0.055	0.048	0.085
		(0.117)	(0.183)	(0.205)
Membership size x Ethnic diversity index				-0.049*
				(0.025)
LR χ^2, full model	38.82	8.72	42.92	47.71
p-value	0.000	0.190	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.079	0.387	0.430
Adjusted McFadden's Pseudo-R^2	0.224	-0.048	0.171	0.196

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A2: Determinants of grain bank survival: Average marginal effects
MLE binomial probit regression estimates

Dependent variable: Surviving grain bank (1 = yes)				
	Average marginal effects (AME)			
Independent variables	Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.002			
	(0.001)			
Distance from block headquarters	0.001		0.001	0.000
	(0.003)		(0.004)	(0.004)
Share of landless households	0.322		0.297	0.296
	(0.203)		(0.203)	(0.191)
Quality of village road (1 = tarred)	-0.307***		-0.326***	-0.368***
	(0.094)		(0.098)	(0.091)
Meeting on needs basis (1 = yes)	0.562***		0.593***	0.591***
	(0.081)		(0.080)	(0.077)
Presence of self-help group (1 = yes)	0.358***		0.372***	0.384***
	(0.094)		(0.093)	(0.085)
Membership size		0.005	0.001	-0.002
		(0.007)	(0.006)	(0.006)
Square of membership size		0.000	0.000	0.000*
		(0.000)	(0.000)	(0.000)
Share of female members in committee		0.573	0.468	0.548*
		(0.415)	(0.362)	(0.332)
Share contributed by villagers at inception		0.530	-0.145**	-0.124
		(0.229)	(0.280)	(0.271)
Ethnic diversity index		-0.242	0.040	1.060*
		(0.342)	(0.263)	(0.523)
Grain bank established 1993-94 (1 = yes)		-0.050	0.029	0.047
		(0.108)	(0.110)	(0.111)
Membership size x Ethnic diversity index				-0.027*
				(0.013)
LR χ^2, full model	38.82	8.72	42.92	47.71
p-value	0.000	0.190	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.079	0.387	0.430
Adjusted McFadden's Pseudo-R^2	0.224	-0.048	0.171	0.196

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A3: Determinants of grain bank survival: Marginal effects at the mean WESMLE binomial probit regression estimates ($\omega = 0.31$)

Dependent variable: Surviving grain bank (1 = yes)

	Marginal effects evaluated at sample means (MEM)			
Independent variables	Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.002			
	(0.001)			
Distance from block headquarters	0.000		0.001	0.000
	(0.003)		(0.004)	(0.004)
Share of landless households	0.402*		0.386*	0.369*
	(0.214)		(0.209)	(0.216)
Quality of village road (1 = tarred)	-0.310***		-0.341***	-0.371***
	(0.091)		(0.087)	(0.093)
Meeting on needs basis (1 = yes)	0.622***		0.681***	0.698***
	(0.111)		(0.110)	(0.118)
Presence of self-help group (1 = yes)	0.348***		0.351***	0.344***
	(0.083)		(0.085)	(0.090)
Membership size		0.005	0.002	-0.001
		(0.006)	(0.007)	(0.007)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		0.527**	0.606	0.669*
		(0.368)	(0.369)	(0.367)
Share contributed by villagers at inception		0.492	-0.141	-0.125
		(0.222)	(0.321)	(0.332)
Ethnic diversity index		-0.233	-0.012	1.144
		(0.319)	(0.334)	(0.784)
Grain bank established 1993-94 (1 = yes)		-0.057	0.038	0.053
		(0.096)	(0.121)	(0.128)
Membership size x Ethnic diversity index				-0.030*
				(0.017)
Wald χ^2	22.62	8.28	36.62	37.42
p-value	0.001	0.219	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.077	0.386	0.420
Adjusted McFadden's Pseudo-R^2	0.208	-0.065	0.142	0.156

Notes: $N = 80$. Standard errors in parentheses.

* Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A4: Determinants of grain bank survival: Average marginal effects
WESMLE binomial probit regression estimates ($\omega = 0.31$)

Dependent variable: Surviving grain bank (1 = yes)

	Average marginal effects (AME)			
Independent variables	Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.002			
	(0.001)			
Distance from block headquarters	0.000		0.000	0.000
	(0.002)		(0.003)	(0.003)
Share of landless households	0.325*		0.305*	0.293*
	(0.169)		(0.164)	(0.169)
Quality of village road (1 = tarred)	-0.281***		-0.309***	-0.341***
	(0.083)		(0.076)	(0.075)
Meeting on needs basis (1 = yes)	0.533***		0.549***	0.555***
	(0.078)		(0.075)	(0.076)
Presence of self-help group (1 = yes)	0.344***		0.354***	0.361***
	(0.081)		(0.077)	(0.075)
Membership size		0.004	0.001	-0.001
		(0.006)	(0.005)	(0.005)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		0.512	0.479*	0.531*
		(0.353)	(0.283)	(0.282)
Share contributed by villagers at inception		0.478**	-0.112	-0.099
		(0.208)	(0.250)	(0.258)
Ethnic diversity index		-0.227	-0.009	0.909
		(0.308)	(0.264)	(0.621)
Grain bank established 1993-94 (1 = yes)			0.030	0.041
			(0.095)	(0.099)
Membership size x Ethnic diversity index				-0.024*
				(0.013)
Wald χ^2	22.62	8.28	36.62	37.42
p-value	0.001	0.219	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.077	0.386	0.420
Adjusted McFadden's Pseudo-R^2	0.208	-0.065	0.142	0.156

Notes: $N = 80$. Standard errors in parentheses.
* Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A5: Determinants of grain bank survival: Estimated coefficients
WESMLE binomial probit regression estimates ($\omega = 0.25$)

Dependent variable: Surviving grain bank (1 = yes)

	Coefficients			
Independent variables	Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.008			
	(0.005)			
Distance from block headquarters	0.001		0.002	0.000
	(0.010)		(0.013)	(0.014)
Share of landless households	1.344*		1.366*	1.385*
	(0.724)		(0.747)	(0.798)
Quality of village road (1 = tarred)	-1.202***		-1.446***	-1.754***
	(0.436)		(0.460)	(0.519)
Meeting on needs basis (1 = yes)	2.105***		2.400***	2.582***
	(0.501)		(0.508)	(0.540)
Presence of self-help group (1 = yes)	1.540***		1.671***	1.814***
	(0.514)		(0.493)	(0.515)
Membership size		0.013	0.007	-0.004
		(0.018)	(0.023)	(0.024)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		1.526	2.147*	2.492*
		(1.117)	(1.266)	(1.326)
Share contributed by villagers at inception		1.423**	-0.435	-0.420
		(0.650)	(1.116)	(1.243)
Ethnic diversity index		-0.682	-0.099	4.052
		(0.906)	(1.133)	(2.859)
Grain bank established 1993-94 (1 = yes)		-0.177	0.134	0.192
		(0.283)	(0.423)	(0.472)
Membership size x Ethnic diversity index				-0.107*
				(0.062)
Constant	-3.350***	-2.253***	-4.538***	-4.998***
	(0.860)	(0.783)	(1.488)	(1.439)
Wald χ^2	21.33	8.10	35.28	36.51
p-value	0.002	0.231	0.000	0.000
McFadden's Pseudo-R^2	0.347	0.075	0.383	0.414

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A6: Determinants of grain bank survival: Estimated coefficients
WESMLE binomial probit regression estimates ($\omega = 0.33$)

Dependent variable: Surviving grain bank (1 = yes)

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.008			
	(0.005)			
Distance from block headquarters	0.002		0.002	0.000
	(0.010)		(0.013)	(0.014)
Share of landless households	1.317*		1.317*	1.366*
	(0.740)		(0.756)	(0.808)
Quality of village road (1 = tarred)	-1.204***		-1.433***	-1.769***
	(0.425)		(0.448)	(0.510)
Meeting on needs basis (1 = yes)	2.099***		2.423***	2.623***
	(0.492)		(0.500)	(0.533)
Presence of self-help group (1 = yes)	1.516***		1.655***	1.825***
	(0.514)		(0.491)	(0.516)
Membership size		0.014	0.005	-0.006
		(0.019)	(0.024)	(0.024)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		1.549	2.069*	2.483*
		(1.066)	(1.252)	(1.322)
Share contributed by villagers at inception		1.445**	-0.507	-0.480
		(0.659)	(1.080)	(1.203)
Ethnic diversity index		-0.682	-0.016	4.332
		(0.947)	(1.160)	(2.865)
Grain bank established 1993-94 (1 = yes)		-0.165	0.128	0.196
		(0.289)	(0.415)	(0.470)
Membership size x Ethnic diversity index				-0.112*
				(0.062)
Constant	-3.091***	-2.048***	-4.208***	-4.761***
	(0.867)	(0.775)	(1.452)	(1.407)
Wald χ^2	23.11	8.33	37.12	37.77
p-value	0.001	0.215	0.000	0.000
McFadden's Pseudo-R^2	0.351	0.078	0.387	0.422

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A7: Predictive ability of MLE probit model
Specification 4

	Predicted		
Actual	$\Phi\left(x\hat{\beta}\right) < 0.5$	$\Phi\left(x\hat{\beta}\right) \geq 0.5$	Total
$D = 0$	33	7	40
$D = 1$	8	32	40
Total	41	39	80

Sensitivity	Pr(+| *D*)	80.00%
Specificity	Pr(-|~*D*)	82.50%
Positive predictive value	Pr(*D*| +)	82.05%
Negative predictive value	Pr(~*D*| -)	80.49%
False + rate for true ~D	Pr(+|~*D*)	17.50%
False – rate for true D	Pr(-| *D*)	20.00%
False + rate for classified +	Pr(~*D*| +)	27.95%
False – rate for classified -	Pr(*D*| -)	19.51%
Correctly classified		81.25%

Table 2.A8: Predictive ability of WESMLE probit model $(\omega = 0.31)$
Specification 4

	Predicted		
Actual	$\Phi\left(x\hat{\beta}\right) < 0.31$	$\Phi\left(x\hat{\beta}\right) \geq 0.31$	Total
$D = 0$	31	9	40
$D = 1$	8	32	40
Total	39	41	80

Sensitivity	Pr(+| *D*)	80.00%
Specificity	Pr(-|~*D*)	77.50%
Positive predictive value	Pr(*D*| +)	78.05%
Negative predictive value	Pr(~*D*| -)	79.49%
False + rate for true ~D	Pr(+|~*D*)	22.50%
False - rate for true D	Pr(-| *D*)	20.00%
False + rate for classified +	Pr(~*D*| +)	21.95%
False - rate for classified -	Pr(*D*| -)	20.51%
Correctly classified		78.75%

Table 2.A9: Determinants of grain bank survival: Estimated coefficients corrected for potential heteroskedasticity
MLE binomial probit regression estimates

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.007			
	(0.005)			
Distance from block headquarters	0.002		0.002	0.000
	(0.010)		(0.012)	(0.014)
Share of landless households	1.252		1.240	1.350
	(0.770)		(0.777)	(0.841)
Quality of village road (1 = tarred)	-1.202***		-1.414***	-1.804***
	(0.407)		(0.429)	(0.499)
Meeting on needs basis (1 = yes)	2.071***		2.453***	2.697***
	(0.481)		(0.491)	(0.533)
Presence of self-help group (1 = yes)	1.446***		1.616***	1.845***
	(0.514)		(0.488)	(0.528)
Membership size		-0.137	0.121	0.214
		(0.294)	(0.404)	(0.470)
Square of membership size		0.013	0.003	-0.008
		(0.019)	(0.024)	(0.024)
Share of female members in committee		0.000	0.000	0.001*
		(0.000)	(0.000)	(0.000)
Share contributed by villagers at inception		1.569	1.951	2.503*
		(0.965)	(1.223)	(1.305)
Ethnic diversity index		1.451**	-0.606	-0.567
		(0.667)	(1.020)	(1.141)
Grain bank established 1993-94 (1 = yes)		-0.661	0.166	4.839*
		(1.006)	(1.202)	(2.803)
Membership size*Ethnic diversity index				-0.122**
				(0.061)
Constant	-2.589***	-1.636**	-3.627***	-4.366***
	(0.880)	(0.758)	(1.381)	(1.360)
Wald χ^2	25.52	8.54	39.59	39.35
p-value	0.000	0.201	0.000	0.000
McFadden's Pseudo-R^2	0.350	0.079	0.387	0.430

Notes: $N = 80$. Heteroskedasticity-robust standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A10: Determinants of time until grain bank failure: Hazard ratios
Partial-likelihood Cox proportional hazards model estimates

Dependent variable: Hazard of grain bank failure				
	Hazard ratios			
Independent variables	Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.996			
	(0.005)			
Distance from block headquarters	0.984		0.984	0.985
	(0.011)		(0.014)	(0.014)
Share of landless households	0.628		0.806	0.765
	(0.508)		(0.688)	(0.646)
Quality of village road (1 = tarred)	3.170***		3.677***	4.387***
	(1.217)		(1.525)	(1.826)
Meeting on needs basis (1 = yes)	0.118***		0.088***	0.077***
	(0.049)		(0.044)	(0.041)
Presence of self-help group (1 = yes)	0.215***		0.170***	0.136***
	(0.102)		(0.093)	(0.078)
Membership size		0.980	1.008	1.013
		(0.020)	(0.022)	(0.023)
Square of membership size		1.000	1.000	1.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		0.144	0.047	0.048
		(0.176)	(0.065)	(0.066)
Share contributed by villagers at inception		0.256*	1.537	1.845
		(0.178)	(1.724)	(2.081)
Ethnic diversity index		1.418	1.340	0.067*
		(1.428)	(1.329)	(0.138)
Grain bank established 1993-94 (1 = yes)		1.101	0.945	1.032
		(0.358)	(0.387)	(0.420)
Membership size x Ethnic diversity index				1.088*
				(0.051)
LR χ^2, full model	37.73	8.46	44.36	47.99
p-value	0.000	0.207	0.000	0.000
Log likelihood	-145.85	-160.49	-142.53	-140.72

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A11: Determinants of time until grain bank failure: Estimated coefficients MLE exponential survival model estimates

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	-0.004			
	(0.005)			
Distance from block headquarters	-0.011		-0.009	-0.007
	(0.011)		(0.014)	(0.013)
Share of landless households	-0.564		-0.348	-0.427
	(0.796)		(0.829)	(0.825)
Quality of village road (1 = tarred)	1.011***		1.068***	1.170***
	(0.372)		(0.399)	(0.396)
Meeting on needs basis (1 = yes)	-1.911***		-2.104***	-2.147***
	(0.385)		(0.458)	(0.467)
Presence of self-help group (1 = yes)	-1.370***		-1.513***	-1.638***
	(0.452)		(0.510)	(0.524)
Membership size		-0.018	0.008	0.012
		(0.020)	(0.022)	(0.022)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		-1.703	-1.989	-1.876
		(1.154)	(1.238)	(1.216)
Share contributed by villagers at inception		-1.286*	0.610	0.751
		(0.690)	(1.099)	(1.103)
Ethnic diversity index		0.423	0.257	-2.238
		(1.005)	(0.986)	(2.106)
Grain bank established 1993-94 (1 = yes)		0.003	-0.067	0.013
		(0.325)	(0.404)	(0.402)
Membership size x Ethnic diversity index				0.067
				(0.047)
Constant	-0.568	-0.943	0.050	0.079
	(0.657)	(0.773)	(1.032)	(0.998)
LR χ^2, full model	35.28	7.70	39.36	41.66
p-value	0.000	0.261	0.000	0.000
Log likelihood	-73.142	-86.929	-71.101	-69.951

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A12: Determinants of time until grain bank failure: Estimated coefficients MLE Weibull survival model estimates

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	-0.004			
	(0.005)			
Distance from block headquarters	-0.020*		-0.020	-0.019
	(0.012)		(0.015)	(0.014)
Share of landless households	-0.442		-0.159	-0.236
	(0.825)		(0.874)	(0.864)
Quality of village road (1 = tarred)	1.348***		1.547***	1.755***
	(0.390)		(0.422)	(0.426)
Meeting on needs basis (1 = yes)	-2.492***		-2.926***	-3.107***
	(0.435)		(0.532)	(0.561)
Presence of self-help group (1 = yes)	-1.788***		-2.089***	-2.386***
	(0.486)		(0.564)	(0.598)
Membership size		-0.022	0.012	0.019
		(0.020)	(0.023)	(0.023)
Square of membership size		0.000	0.000	-0.000**
		(0.000)	(0.000)	(0.000)
Share of female members in committee		-2.016*	-3.399**	-3.451**
		1.202)	(1.401)	(1.369)
Share contributed by villagers at inception		-1.438**	0.637	0.818
		(0.704)	(1.121)	(1.131)
Ethnic diversity index		0.419	0.117	-3.235
		(1.010)	(1.005)	(2.068)
Grain bank established 1993-94 (1 = yes)		-0.086	-0.220	-0.143
		(0.331)	(0.419)	(0.415)
Membership size x Ethnic diversity index				0.095**
				(0.047)
Constant	-1.690**	-1.529*	-0.523	-0.472
	(0.769)	(0.866)	(1.119)	(1.063)
LR χ^2, full model	47.25	9.17	55.48	60.18
p-value	0.000	0.164	0.000	0.000
Log likelihood	-65.069	-84.109	-60.953	-58.604

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A13: Determinants of time until grain bank failure: Estimated coefficients MLE log-normal survival model estimates

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.003			
	(0.003)			
Distance from block headquarters	0.010		0.010	0.009
	(0.007)		(0.007)	(0.007)
Share of landless households	0.443		0.375	0.376
	(0.462)		(0.449)	(0.447)
Quality of village road (1 = tarred)	-0.617***		-0.669***	-0.706***
	(0.234)		(0.239)	(0.242)
Meeting on needs basis (1 = yes)	1.341***		1.375***	1.389***
	(0.249)		(0.271)	(0.272)
Presence of self-help group (1 = yes)	0.960***		0.945***	0.962***
	(0.275)		(0.280)	(0.281)
Membership size		0.013	0.002	-0.001
		(0.015)	(0.012)	(0.013)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		1.429	1.399*	1.412*
		(0.921)	(0.769)	(0.765)
Share contributed by villagers at inception		0.973*	-0.282	-0.327
		(0.536)	(0.621)	(0.621)
Ethnic diversity index		-0.431	-0.321	0.814
		(0.745)	(0.589)	(1.291)
Grain bank established 1993-94 (1 = yes)		-0.189	0.047	0.023
		(0.240)	(0.232)	(0.234)
Membership size x Ethnic diversity index				-0.028
				(0.029)
Constant	0.536	0.919	-0.071	-0.064
	(0.419)	(0.604)	(0.645)	(0.641)
LR χ^2, full model	39.98	8.94	46.81	47.82
p-value	0.000	0.1768	0.000	0.000
Log likelihood	-65.554	-81.070	-62.136	-61.634

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

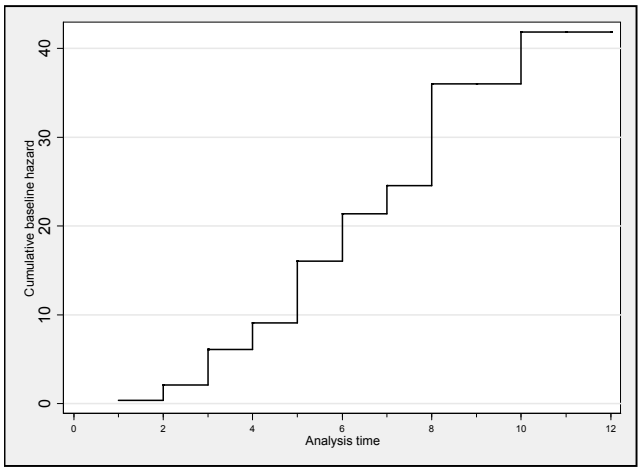


Figure 2.A1: Estimated baseline cumulative hazard

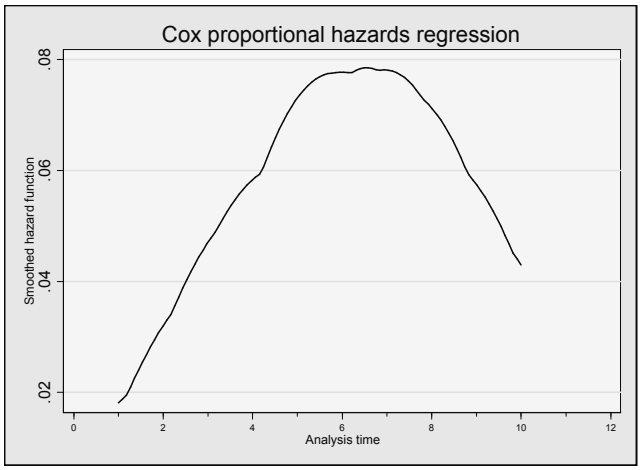


Figure 2.A2: Estimated hazard function

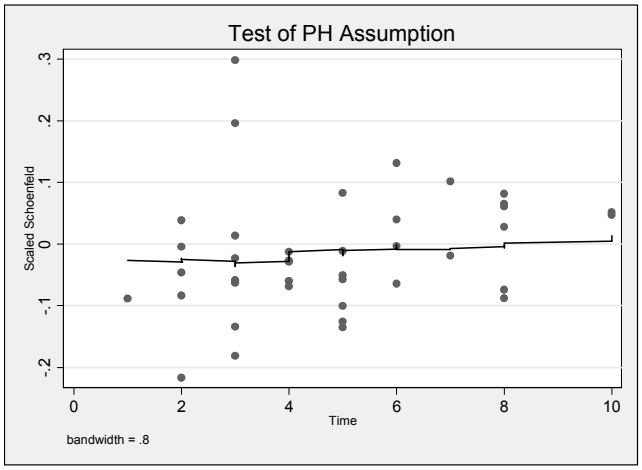


Figure 2.A3: Variable – Distance from block headquarters

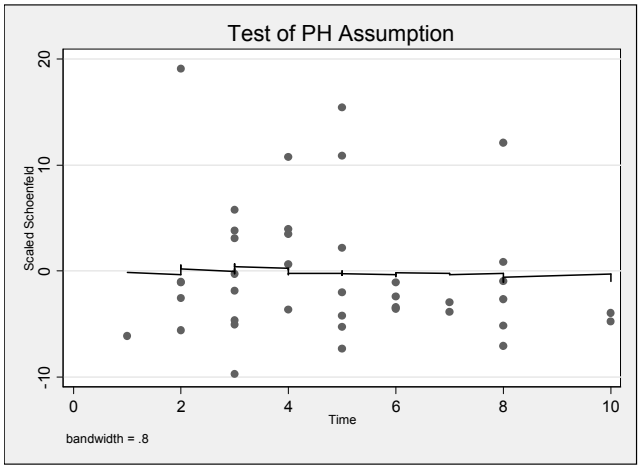


Figure 2.A4: Variable – Share of landless households

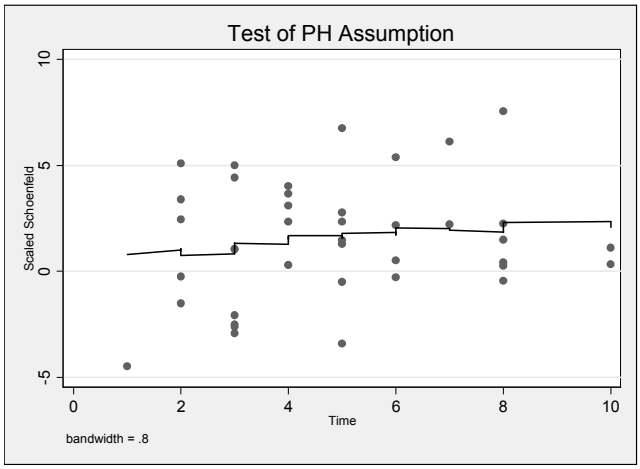


Figure 2.A5: Variable – Village road quality

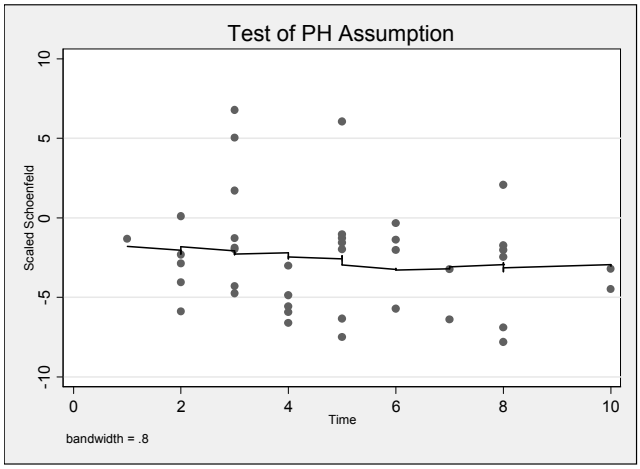


Figure 2.A6: Variable – Distance from block headquarters

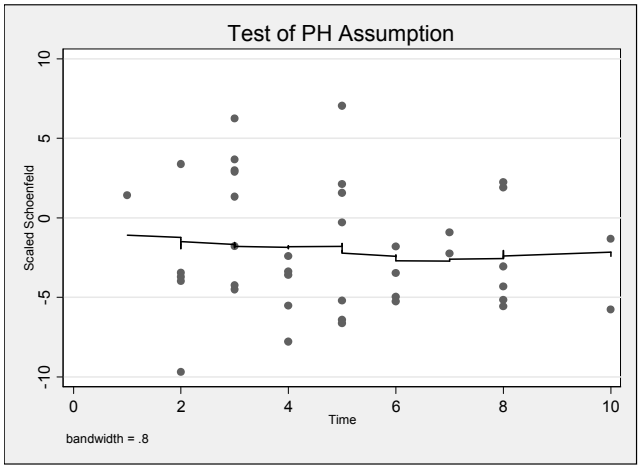


Figure 2.A7: Variable – Presence of SHG group

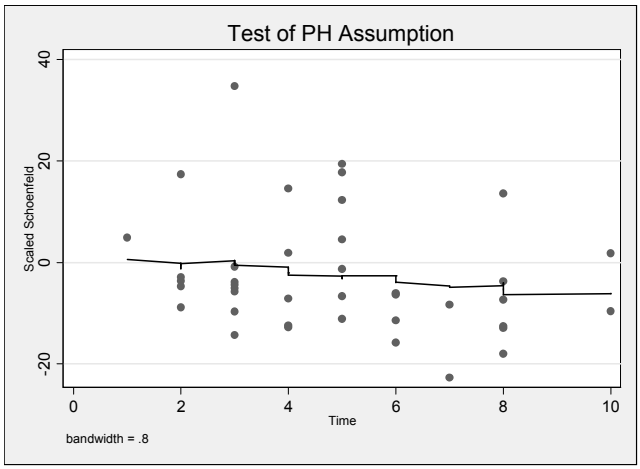


Figure 2.A8: Variable – Membership size

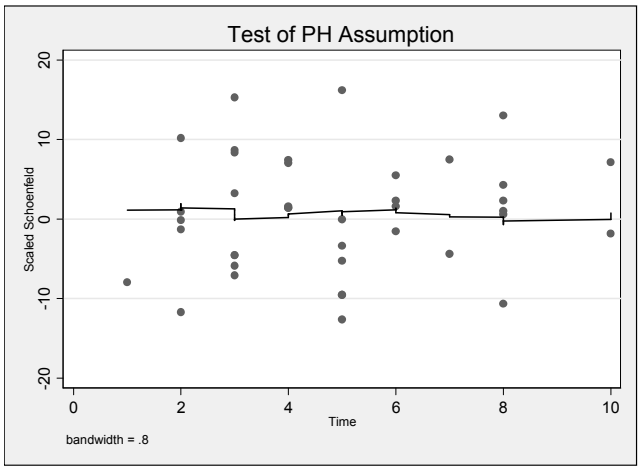


Figure 2.A9: Variable – Share contributed by villagers

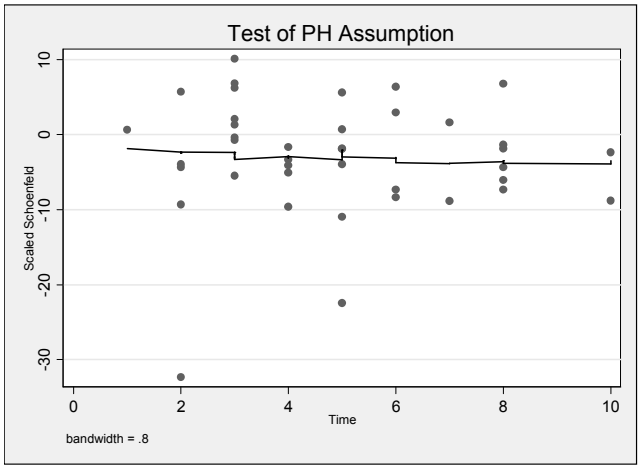


Figure 2.A10: Variable – Share of female committee members

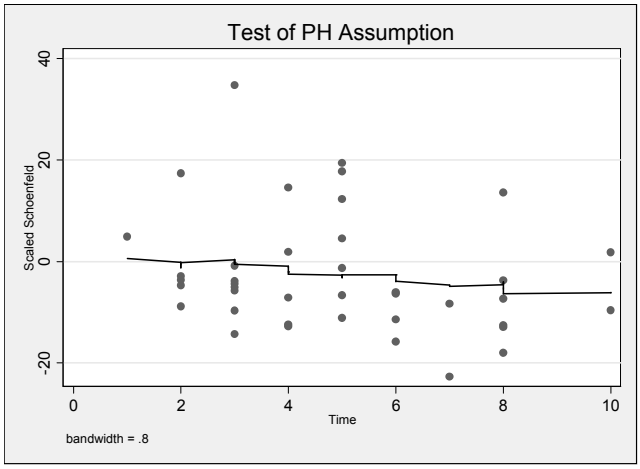


Figure 2.A11: Variable – Ethnic diversity index

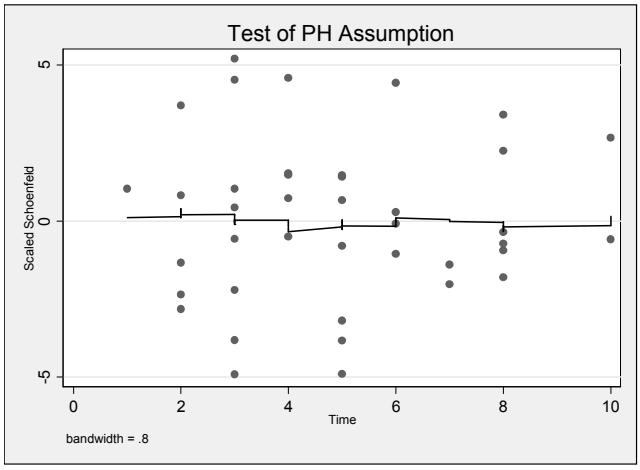


Figure 2.A12: Variable – Grain bank establishment year

Table 2.A14: Determinants of time until grain bank failure: Estimated coefficients MLE exponential survival model with Gamma frailty

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	-0.004			
	(0.005)			
Distance from block headquarters	-0.011		-0.009	-0.007
	(0.011)		(0.014)	(0.013)
Share of landless households	-0.564		-0.348	-0.428
	(0.796)		(0.829)	(0.825)
Quality of village road (1 = tarred)	1.011***		1.068***	1.170***
	(0.372)		(0.399)	(0.396)
Meeting on needs basis (1 = yes)	-1.911***		-2.104***	-2.148***
	(0.385)		(0.458)	(0.467)
Presence of self-help group (1 = yes)	-1.370***		-1.513***	-1.638***
	(0.452)		(0.510)	(0.524)
Membership size		-0.018	0.008	0.012
		(0.020)	(0.022)	(0.022)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		-1.703	-1.989	-1.876
		(1.154)	(1.238)	(1.216)
Share contributed by villagers at inception		-1.286*	0.610	0.751
		(0.690)	(1.099)	(1.103)
Ethnic diversity index		0.423	0.257	-2.238
		(1.005)	(0.986)	(2.106)
Grain bank established 1993-94 (1 = yes)		0.003	-0.067	0.013
		(0.325)	(0.404)	(0.402)
Membership size x Ethnic diversity index				0.067
				(0.047)
Constant	-0.568	-0.943	0.050	0.079
	(0.657)	(0.773)	(1.032)	(0.998)
LR χ^2, full model	35.28	7.70	39.36	41.66
p-value	0.000	0.261	0.000	0.000
Log likelihood	-73.142	-86.929	-71.101	-69.951
Frailty distribution variance θ	0.000	0.000	0.000	0.000
p-value (LR test of θ = 0)	1.000	1.000	1.000	1.000

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A15: Determinants of time until grain bank failure: Estimated coefficients MLE Weibull survival model with Gamma frailty

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	-0.007			
	(0.007)			
Distance from block headquarters	-0.019		-0.020	-0.019
	(0.013)		(0.015)	(0.014)
Share of landless households	-0.770		-0.159	-0.236
	(1.008)		(0.874)	(0.864)
Quality of village road (1 = tarred)	1.409***		1.547***	1.755***
	(0.463)		(0.422)	(0.426)
Meeting on needs basis (1 = yes)	-2.797***		-2.926***	-3.107***
	(0.639)		(0.532)	(0.561)
Presence of self-help group (1 = yes)	-2.066***		-2.090***	-2.386***
	(0.650)		(0.564)	(0.598)
Membership size		-0.031	0.012	0.019
		(0.055)	(0.023)	(0.023)
Square of membership size		0.000	0.000	-0.000**
		(0.000)	(0.000)	(0.000)
Share of female members in committee		-7.028	-3.399**	-3.451**
		(6.013)	(1.401)	(1.369)
Share contributed by villagers at inception		-3.859	0.637	0.818
		(3.197)	(1.121)	(1.131)
Ethnic diversity index		3.834	0.117	-3.236
		(3.071)	(1.005)	(2.068)
Grain bank established 1993-94 (1 = yes)		2.323*	-0.22	-0.143
		(1.351)	(0.419)	(0.415)
Membership size x Ethnic diversity index				0.095**
				(0.047)
Constant	-1.495*	-1.689	-0.523	-0.472
	(0.891)	(3.026)	(1.119)	(1.063)
LR χ^2, full model	37.37	10.90	45.01	49.71
p-value	0.000	0.092	0.000	0.000
Log likelihood	-64.773	-78.008	-60.953	-58.604
Frailty distribution variance θ	0.409	7.910	0.000	0.000
p-value (LR test of θ = 0)	0.221	0.00	1.000	1.000

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 2.A16: Determinants of time until grain bank failure: Estimated coefficients MLE log-normal survival model with Gamma frailty

Dependent variable: Hazard of grain bank failure

Independent variables	Coefficients Specification 1	Specification 2	Specification 3	Specification 4
Total households	0.003			
	(0.003)			
Distance from block headquarters	0.010		0.010	0.009
	(0.007)		(0.007)	(0.007)
Share of landless households	0.443		0.375	0.376
	(0.462)		(0.449)	(0.447)
Quality of village road (1 = tarred)	-0.617***		-0.669***	-0.706***
	(0.234)		(0.239)	(0.242)
Meeting on needs basis (1 = yes)	1.341***		1.375***	1.389***
	(0.249)		(0.271)	(0.272)
Presence of self-help group (1 = yes)	0.960***		0.945***	0.962***
	(0.275)		(0.280)	(0.281)
Membership size		0.006	0.002	-0.001
		(0.012)	(0.012)	(0.013)
Square of membership size		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Share of female members in committee		1.500	1.399*	1.412*
		(1.028)	(0.769)	(0.765)
Share contributed by villagers at inception		0.849	-0.282	-0.327
		(0.575)	(0.621)	(0.621)
Ethnic diversity index		-0.832	-0.321	0.814
		(0.635)	(0.589)	(1.291)
Grain bank established 1993-94 (1 = yes)		-0.589**	0.047	0.023
		(0.242)	(0.232)	(0.234)
Membership size x Ethnic diversity index				-0.028
				(0.029)
Constant	0.536	0.687	-0.071	-0.064
	(0.419)	(0.574)	(0.645)	(0.641)
LR χ^2, full model	35.76	11.36	42.59	43.60
p-value	0.000	0.078	0.000	0.000
Log likelihood	-65.554	-77.753	-62.136	-61.634
Frailty distribution variance θ	0.000	2.448	0.000	0.000
p-value (LR test of $\theta = 0$)	1.000	0.005	1.000	1.000

Notes: $N = 80$. Standard errors in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

REFERENCES

Alesina, Alberto, and Eliana La Ferrera. 2000. Participation in Heterogeneous Communities. *Quarterly Journal of Economics*. v115 (n3):857-904.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. Public Goods and Ethnic Divisions. *Quarterly Journal of Economics*. v114 (n4):1243-1284.

Anderson, Roy M., and Robert M. May. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.

Anderson, Siwan, and Jean-Marie Baland. 2002. The Economics of ROSCAS and Intra-household Resource Allocation. *Quarterly Journal of Economics*. v117 (n3):963-95.

Armendariz de Aghion, Beatriz, and Jonathan Morduch. 2004. Microfinance: Where do we Stand? in Charles Goodhart (Ed.) *Financial Development and Economic Growth: Explaining the Links.* Basingstoke, Hampshire: Palgrave Macmillan.

Ashraf, Nava, Nathalie Gons, Dean Karlan, and Wesley Yin. 2003. A Review of Commitment Savings Products in Developing Countries. *Asian Development Bank Economics and Research Department Working Paper* #45.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quarterly Journal of Economics*. v121 (n2):635-72.

Behrman, Jere R. 1996. The Impact of Health and Nutrition on Education. *World Bank Research Observer*. v11 (n1):23-37.

Besley, Timothy, Stephen Coate, and Glenn Loury. 1993. The Economics of Rotating Savings and Credit Associations. *American Economic Review*. v83 (n4):792-810.

Besley, Timothy, and Stephen Coate. 1995. Group Lending, Repayment Incentives and Social Collateral. *Journal of Development Economics*. v46 (n1):1-18.

Carter, Michael. 1997. Environment, Technology, and the Social Articulation of Risk in West African Agriculture. *Economic Development and Cultural Change*. v45 (n3):557-590.

Census of India. Various years. Government of India.

Chamberlin, John. 1974. Provision of Collective Goods as a Function of Group Size. *American Political Science Review*. v68 (n2):707-16.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*. v72 (n5):1409-43.

Cleves, Mario, William Gould, and Roberto G. Gutierrez. 2004. *An Introduction to Survival Analysis using Stata*. College Station, Texas: Stata Corporation.

Cox, David R. 1972. Regression Models And Life-Tables (with discussion). *Journal of the Royal Statistical Society,* Series B 34.

Dercon, Stefan, Joachim De Weerdt, Tessa Bold, and Alula Pankhurst. 2005. Membership based Indigenous Insurance Association in Ethiopia and Tanzania. Mimeo.

Duflo, Esther. 2003. Grandmothers and Granddaughters: Old-Age Pensions andIntrahousehold Allocation in South Africa. *World Bank Economic Review*. v17 (n1):1-25.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2005. Why are Farms Not Using Fertilizer: Fertilizer in Western Kenya, Preliminary Results from Field Experiments. MIT, Mimeo.

Easterly, William, and Ross Levine. 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics*. v112 (n4):1203-1250.

Ghatak, Maitreesh, and Timothy W. Guinnane. 1999. The Economics of Lending with Joint Liability: Theory and Practice. *Journal of Development Economics*. v60 (n1):195-228.

Greene, William. 2002. *Econometric Analysis*. New Jersey: Prentice Hall.

Gugerty, Mary K. 2003. You Can't Save Alone: Testing Theories of Rotating Savings and Credit Associations in Kenya. Mimeo.

Kiefer, Nicholas M. 1988. Economic Duration Data and Hazard Functions. *Journal of Economic Literature*. v26 (n2):646-79.

Heckman, James, and Burton Singer. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data. *Econometrica*. v52 (n2):271-320.

Heckman, James, and Jeffrey A. Smith. 1999. The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation. *Economic Journal*. v109 (n457):313-48.

Heltberg, Rasmus, and Carlo del Ninno. 2006. Household Risk Management and Social Safety Nets in Pakistan – or How we Engaged the Pakistanis on Social Protection. Human Development Week presentation, Novemeber 9. World Bank: Washington D.C.

Laibson, David. 1997. Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics*. v112 (n2):443-77.

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*. v114 (n498):750-77.

Mahul, Olivier, and Jerry Skees. 2006. Piloting Index-Based Livestock Insurance in Mongolia. *Financial Sector Vice-Presidency Newsletter No. 10* (March). World Bank: Washington D.C.

Menard, Scott W. 2002. *Applied Logistic Regression Analysis*. 2nd edition. London: Sage Publications Inc.

Messer, E. 1989. Seasonality In Food Systems: An Anthropological Perspective on Household Food Security, In David. E. Sahn (Ed.) *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Miguel, Edward, and Mary K. Gugerty. 2004. Ethnic Diversity, Social Sanctions and Public Goods in Kenya. Mimeo.

Morduch, Jonathan. 1999. Between the State and the Market: Can Informal Insurance Patch the Safety Net? *World Bank Research* Observer. v14 (n2):187-207.

Mude, Andrew G. 2006. Making Loans to Make Friends: Explaining the Dismal Financial Performance of Financial Service Associations. *Agricultural Finance Review*. v66 (n2):267-81.

New Concept Information Systems (NCIS). *Breaking the Shackles of Hunger: Process Documentation on the Orissa Household Food Security Project (1993-1998).*

O'Donoghue, Ted, and Matthew Rabin. 1999. Doing It Now or Doing It Later. *The American Economic Review*. v89 (n1):103-121.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge: Harvard University Press.

Pitt, Mark M., Shahidur Khandker, Omar Haider Chowdhury, and Daniel L. Millimet. 2001. Credit Programs for the Poor and the Health Status of Children in Rural Bangladesh. Brown University: Mimeo.

Pitt, Mark. M., and Shahidur R. Khandker. 1998. The Impact of Group-Based Credit on Poor Households in Bangladesh: Does the Gender of Participants Matter? *Journal of Political Economy*. v106 (n5):958-966.

Rawlings, Laura B., and Gloria M. Rubio. 2005. Evaluating the Impact of Conditional Cash Transfer Programs. *The World Bank Research Observer*. v20 (n1):29-55.

Rutherford, Stuart and others. 1999. *Savings and the Poor: The Methods, Use and Impact of Savings by The Poor of East Africa.* Report prepared for MicroSave-Africa, May 1999.

Sahn, David E. 1989 (ed.) *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Thomas, Duncan. 1990. Intra-household Resource Allocation: An Inferential Approach. *Journal of Human Resources*. v25 (n4):635-64.

Townsend, Robert M. 1994. Risk and Insurance in Village India. *Econometrica*. v62 (n3):539-591.

Udry, Christopher. 1990. Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy. *The World Bank Economic Review*. v4 (n3):251-269.

CHAPTER 3

GRAIN BANKS AND CHILDREN'S HEALTH

3.1 Introduction

In this chapter, we use non-experimental data collected through a small-scale household sample survey in Rayagada, Orissa to examine the impact of grain banks on young children's health. Grain banks are community-level institutions that provide households with in-kind consumption credit during the "hungry" season. In the absence of formal credit markets, grain banks provide an alternative source of consumption credit at interest rates lower than those charged by informal lenders such as local moneylenders, without the risk of loss of physical collateral or engaging in tied labor-credit contracts under exploitative terms.[49]

Grain banks have been enthusiastically implemented by non-governmental organizations (NGOs) promoting food security in this area. In recent years, grain banks have also been adopted by the Indian government's ministry of tribal welfare with the purpose of combating short-term, seasonal food shortages in villages where the majority of the population belongs to tribal communities (Ministry of Tribal Affairs 2002). In such times, grain banks are supposed to function as a safety net against deterioration in the nutritional status of particularly the most susceptible populations, namely children and lactating and pregnant women (ibid.). To date,

[49] While the average annual interest charged by grain banks is about 25 percent, local, private moneylenders charge at least double that rate. In addition, unlike borrowing from private moneylenders, borrowing from grain banks does not involve risk of loss of collateral. As a result, poor households who may otherwise be vulnerable to risk-rationing can become grain bank borrowers. In risk rationing, lenders shift so much of the contractual risk to borrowers that the latter voluntarily withdraw from the credit market even when they have the collateral wealth to participate in a credit contract. For empirical evidence on risk rationing, see, e.g., Boucher et al. (2005) and Boucher and Guirkinger (2005).

however, no rigorous evaluation of the impact of grain banks on health outcomes exists.

The main contribution of this chapter is to fill this gap in knowledge by measuring the impact of grain bank participation by households on children's health outcomes. We do this using propensity score matching estimators. Given that the tribal community forms one of the most socioeconomically disadvantaged groups in Indian society and has very poor nutritional status, evidence on the impact of grain banks on the health of this population would be particularly important.[50] If grain banks indeed improve the health outcomes of participants, the case for scaling up this initiative would be strengthened.[51] On the other hand, if grain banks are found not to have any quantifiable impact on health outcomes, then the case for scaling up would be significantly weakened and alternative interventions should be explored.

Reduced nutrient and food intake can retard skeletal growth and reduce the accumulation of muscle and fat tissues in children. Therefore, anthropometric measurements can be used to detect levels of malnutrition in a population.[52] We examine the following outcomes: height-for-age (or *haz* scores) and weight-for-height standardized *z*-scores (or *whz* scores) as well as the rate of growth in height between children from participant households and children from non-participant households.[53] We do not use weight-for-age standardized scores as they cannot distinguish between

[50] A brief overview of the the tribal community in India, as well as notes on the two main tribes in the survey region, are provided in a separate appendix.
[51] The strongest case would be made if the positive impacts that grain banks have are shown to be more cost-effective than the impacts from other interventions. Currently, we do not possess data on the cost of the grain bank intervention. However, given that grain banks are a community-based intervention, they are likely to have much lower storage, transportation and overhead costs relative to more centralized food security programs.
[52] See Morris (1999) for further discussion on how anthropometric data can detect malnutrition levels.
[53] The *z*-score is defined as the difference between the value for an individual and the median value of the reference population for the same age or height, divided by the standard deviation of the reference population (Cogill 2003).

chronic and acute malnutrition, and do not provide any information beyond what height-for-age and weight-for-height indicators provide (Alderman 2000).[54]

Height-for-age standardized *z*-scores are commonly used in detecting levels of long-term or chronic malnutrition in young children.[55] Low height-for-age relative to a child of the same sex and age in a reference population (or *haz* scores below -2) is referred to as "stunting". Since *haz* scores reflect the quality and not just the quantity of food intake and are useful in detecting long-term impacts, a preferred indicator of short-term changes in nutritional status that may occur due to seasonal fluctuations in food consumption is the weight-for-height standardized *z*-score (Morris 1999). Low weight-for-height relative to a child of the same sex and age in a reference population (or *whz* scores below -2) is referred to as "wasting". Growth in height is also a good indicator of underlying health status. Children experiencing slow growth in height are found to have poorer cognitive development, interact less frequently with their environment, have lower activity levels and acquire skills at a slower pace (Grantham-McGregor et al. 1999; Lasky et al. 1981).

Given that the majority of grain banks were established in the last decade, we focus on anthropometric outcomes of children below the age of 6 years or the children of the "grain bank era". We assume that the impact of grain banks on their health status is likely to be more pronounced than on older children or adults. We examine the impact of household participation in grain banks on children's health outcomes by comparing children of participating households in villages with grain banks to children

[54]These measures have been developed by the United States National Center for Health Statistics (NCHS) and are recommended for international use by the World Health Organization (WHO). The reference population chosen by the NCHS is a statistically valid random population of healthy infants and children. While questions regarding the validity of using a US-based reference standard for other ethnic populations are commonly raised, evidence suggests that until about 10 years of age, children from well-nourished and healthy households in both developing and industrialized countries grow at the same rate and have comparable height and weight, regardless of their ethnicity (Cogill 2003).
[55]See Alderman (2000) for a survey article on the use of anthropometric survey data.

in non-grain bank villages, within a matched sample. We also examine the impact of household participation in grain banks on children's health outcomes by comparing children in villages where grain banks have been in operation for at least 10 years to children in non-grain bank villages, again within a matched sample. The latter comparison allows us to test the hypothesis that grain banks which have been in operation for a longer duration may have a more pronounced health impact due to intergenerational transmission effects through the improved health of mothers.

Using local linear regression and kernel propensity score matching estimators, we find that household participation in grain banks does not have has a statistically significant effect on any of the children's health outcomes that we examine. We also find no statistically significant impact on these outcomes when we compare children in participant households in villages where grain banks have survived for a longer duration compared to children in non-grain bank villages. Our findings raise questions about the efficacy of the current generation of grain banks in improving the nutritional outcomes of young children.

The rest of this chapter is organized as follows. In Section 3.2, we provide a brief review of the relevant literature. In Section 3.3, we provide some background information on the survey location. In Section 3.4, we provide an overview of the socioeconomic characteristics of the survey sample villages. In Section 3.5, we describe the empirical methodology. In Section 3.6, we discuss the data and main findings. In Section 3.7, we provide some concluding remarks.

3.2 Literature review

A number of empirical studies have documented the adverse impact of households' inability to smooth consumption across the agricultural cycle on both short- and long-term health outcomes. For example, using ICRISAT data for southern rural India,

Behrman and Deolalikar (1989) find that reduced food intake during the lean season has a negative impact on future agricultural productivity. Thus, a household that experiences seasonal food insecurity in one year faces the prospect of lowered wages and farm profits in the next, which can lead to a cycle of even lower nutrient intake in the following year. Reardon and Matlon (1989) find evidence of reduced food intake during the cropping season in the Sudano-Sahel agroclimatic region in Burkina Faso, when demands for energy expenditure are the highest. Lawrence et al. (1989) find for the Gambia that seasonal food shortages lead to losses in female body weight as well as low birth weights. Given that lower birth weights are associated with lower height attainments in childhood and adulthood, and a consequent reduction in potential earnings, seasonal fluctuations in consumption not only have negative short-term effects but also negative long-term impacts (Martorell 1995, 1999; Glewwe and Jacoby 1995).

Health outcomes of young children are of particular interest not only because of concern over their immediate welfare, but also due to the fact that nutrition in early childhood plays a crucial role in future physical and mental development, thereby affecting health status as adults and future labor productivity.[56] A number of empirical studies, for example, show the relationship between early childhood nutrition and adult height and wages/productivity (see, e.g., Behrman and Deolalikar 1989 for rural India, Haddad and Bouis 1991 for rural Philippines, Thomas and Strauss 1997 for Brazil). Martorell (1999) and Maluccio et al. (2003) also provide general evidence on the impact of early childhood nutrition on adult nutritional status and cognitive achievement.

[56] See World Bank (2006) for a review of the empirical evidence on the impact of early childhood malnutrition on various development outcomes, including schooling, productivity and income poverty.

There is also a growing body of literature that documents the impact of adverse shocks on children's health outcomes, both concurrent and in the long-term. For example, Hoddinott and Kinsey (2001) use panel data from Zimbabwe to show that children aged 12-24 months, especially those from poorer households, experience a slowdown in growth due to a drought. They find that this cohort of children is not able to make up over time for this lost growth. Given that adult height is strongly correlated with height achieved by age 3 and is also correlated with labor earnings and productivity, the authors posit that the inability to smooth consumption increases the likelihood of chronic poverty. In addition, given the fact that taller women have, on average, healthier children, they posit that the impact of drought on female children can lead to the intergenerational persistence of poverty. Similarly, Dercon and Hoddinott (2005) find that recent droughts in Zimbabwe resulted in reduced growth for children of preschool age. They too find that this cohort is not able to catch up in terms of lost growth, as a result of which lifetime earnings are estimated to be roughly 7 percent lower. Alderman, Hoddinott and Kinsey (2006) and Hoddinott (2006) provide further evidence of both the immediate as well as persistent impact of adverse shocks on children's health outcomes. Using longitudinal data from Zimbabwe between 1983-84, 1987 and 2000, Alderman, Hoddinott and Kinsey (2006) show that improved nutritional status among children of pre-school age (as measured by height-for-age standardized *z* scores) is associated with increased height as a young adult, as well as better schooling attainment and a lower age at which the child started school. The authors use the 1982-83 and 1983-84 droughts as well as the civil war in the late 1970s to identify differences in pre-school nutritional status. Again using longitudinal data from Zimbabwe, Hoddinott (2006) finds that a severe drought in 1994-95 slowed down the rate of growth in height by 1.5-2.0 cm for children below 2 years of age, which is equivalent to a 15-20 percent loss in growth velocity. He shows that even

four years after the drought, these children remain shorter than children of the same ago who did not experience the drought between 12-24 months age. He finds that this result is particularly pronounced for children in less-wealthy households (as measured by livestock holdings).

Given the above evidence, we examine whether grain banks indeed improve children's health outcomes. By enabling households to smooth consumption over the agricultural cycle, they have the potential to not only improve children's health in the short term but to also mitigate the long-term adverse impacts of seasonal malnutrition.

3.3 Background

The data used in this study come from a village and household sample survey conducted in Kashipur block, Rayagada district in the tribal belt of southwest Orissa, one of the poorest regions in the state.[57] According to a BPL (Below Poverty Line) census conducted by the government of Orissa in 1997, 72 percent of households in Kashipur were classified as poor (Sethi 2003).[58] In terms of occupational distribution, 49.9 percent of rural workers are agricultural laborers, followed by 31.7 percent as self-employed cultivators (Census 2001). The overall literacy rate in rural areas is 29.9 percent. The female literacy rate is even lower, at 18.3 percent. Roughly 62 percent of the rural population of Kashipur is tribal, with the Kandha and the Soura constituting the two main tribes.

As described in Chapter 2 for the adjoining block of Dasmantapur, food consumption in this region is tied closely to the agricultural calendar and the composition, quality and quantity of food consumption changes with the seasons.

[57] See chapter 2 for a district map of Orissa depicting the tribal belt.
[58] A BPL household is defined as a family with an annual income below Rs. 11,000 which is roughly equivalent to a dollar a day (1992 rupees). The percentage of households classified as BPL in Koraput district by the same survey is 84 percent.

Between October-February, when produce is harvested, the tribal population consumes staples such as rice, millet and maize.[59] Between March-May, consumption depends on food stocks as well as purchases using earnings from daily wage labor (mainly in labor-intensive public works programs). Food shortages are experienced in the 'hungry' season (June-September), during which time food stocks tend to be low and opportunities for daily wage labor shrink due to the monsoon rains. During this period, meals are limited to gruel made from millet flour or flour from dried seeds (tamarind and mango), food consumption levels generally fall, and a decrease in the average body mass is observed. For example, the average weight of adult women in the survey sample declined from 42.3 kilograms in the winter to 40.9 kilograms towards the end of the monsoon season. This inter-seasonal difference in body weight is statistically significant at the 1 percent level.

Against this backdrop, grain banks were introduced in Kashipur in 1981 by Agragamee, an NGO, as a potential community-level solution to recurring seasonal food shortages. Initially, Agragamee established grain banks in villages in the vicinity of its headquarters in Kashipur town by providing a one-time grant (in the form of grains). In some cases, villagers also provided a matching contribution in order to increase initial grain bank holdings. Once established, these grain banks were managed internally by a grain bank committee comprised of individuals from grain bank member households. During the agricultural lean season, member households could borrow from the grain bank. These loans were to be returned with interest (also in the form of grain) after the following harvest season. Following the droughts of the early 1990s, Agragamee expanded the grain bank initiative to all villages in Kashipur

[59] Cereals form the most important part of tribal diets. Pulses (mainly *kandula*, which is grown locally) are also consumed. However, few fruits and vegetables are consumed, as these are grown mostly for marketing purposes. Meat is only consumed during festivals or when there are guests. Milk and eggs are generally not consumed.

block using funding provided by the state government as part of the Orissa Household Food Security Project (OHFSP).[60] In the next section, we present a brief overview of the sample survey villages.

3.4 Overview of sample villages

Socioeconomic features of sample villages

Table 3.1 presents selected socioeconomic outcomes across India, Orissa and the surveyed villages, in order to contextualize the development levels in the sample villages relative to all of Orissa and India. We see that households in sample villages are at extremely low levels of development, both in absolute as well as relative terms. For example, the share of girls aged 6-14 years attending school in the sample villages is a mere 32 percent, compared to over 70 percent in Orissa and India.

Table 3.1: Selected socioeconomic variables

Statistic	India [a]	Orissa	Sample villages
Percent of females age 6-14 attending school	73.7	75.1	31.9
Percent of households with electricity	60.1	33.8	0.9
Percent of births in medical institutions [b]	33.6	22.6	0.6
Percent of households with no toilet/latrine facility	64.0	86.5	99.6

Source: India and Orissa statistics are from the 1998-99 National Family Health Survey (NFHS-2). Statistics for sample villages are own calculations from household survey data collected in winter 2005.
Notes: [a] Excludes the state of Tripura. [b] India and Orissa figures for births in the three years preceding the NFHS-2.

The infrastructure in the sample villages is extremely poor or deficient. Only 1 of the 28 sample villages had an electricity connection. Village roads were unpaved

[60] For more details on the grain bank movement in Kashipur, see Chapter 2.

and the vast majority of homes have mud walls and floors.[61] Almost no households had toilet/latrine facilities.

Food security

In all sample villages, villagers replied in the affirmative to a question on whether any households residing in the village faced food insufficiency during at least 1 month of the year. The majority of villages included households that faced food insufficiency for at least 3 months out of the year, while some villages included households that faced food shortages for as much as 4-5 months. The months most commonly cited as food shortage months were the monsoon months of *Aashad*, *Shravana* and *Bhadrab* (mid-June to mid-September).

Health infrastructure

At the time of the survey, 11 sample villages reported not having either an *anganwadi* center or sub-center.[62] Of the remaining 17 villages which had an *anganwadi* center or sub-center, only 4 reported that the *anganwadi* worker (AWW) attended regularly – either daily or weekly. Eleven reported monthly visits, while the rest reported irregular attendance. AWW were reported to be from the tribal communities in only 7 sample villages, although the populations of these villages are predominately tribal. In more than half of the sampled villages, caste differences as well as the lack of

[61] As discussed in Catteneo et al. (2006), this is an important public health issue as mud flooring can increase the spread of infectious diseases and result in poor hygiene conditions, especially following heavy rains as is common during the monsoon months in this region.

[62] The *anganwadi* center forms the nodal point for the provision of services under the Integrated Child Development Services (ICDS) program. According to the design of the ICDS program, one *anganwadi* is established for every settlement of 1,000 people (roughly 200 families in rural areas) for the provision of nutrition, health and pre-school educational services (CIRCUS 2006). In practice, however, the coverage and functioning of the ICDS program is uneven across the nation. Villages with a majority SC/ST population are often found to lack well-functioning *anganwadi* centers (see, e.g., the box titled "India's forgotten forest children" describing the lack of *anganwadi* facilities in Dongriguda, an extremely impoverished tribal settlement in Nabarangpur district, Orissa (CIRCUS 2006, p. 50).

involvement of the AWW with the village population, resulted in tribal mothers and children reporting that they were not comfortable in dealing with the AWW. These findings mirror evidence from a recent survey of the Integrated Child Development Services (ICDS) program conducted in various Indian states (CIRCUS 2006). For example, in the state of Uttar Pradesh, almost half of the surveyed *anganwadis* were rated to be functioning poorly or very poorly. Caste differences were also found to result in discrimination in the states of Uttar Pradesh and Rajasthan (ibid).

Inhabitants of the majority of sample villages reported that they visit the government primary health care (PHC) center in the *gram panchayat* headquarters during medical emergencies. The average travel time was over 70 minutes by walking, which was the main mode of transport for the majority of villagers. In some sample villages, the time taken by inhabitants to reach the nearest PHC was between 3-6 hours, indicating the extreme level of difficulty in accessing medical facilities. Eight of the sample villages reported that the majority of villagers could not afford to pay for the medical facilities.

Economy

Agriculture was reported to be the main occupation in sample villages, followed by seasonal daily wage labor. The main crops cultivated in this area are cereals including paddy rice and different varieties of millet. Other crops include corn, oilseeds, pulses and vegetables (such as eggplant, potato, chillies and tomato).

Village and grain bank indicators, by grain bank status

An equal number of grain bank and non-grain bank villages were selected for the survey. Care was taken so that the set of grain bank and non-grain bank villages selected were, on average, similar across various dimensions.

Using data from the village survey conducted as part of the first wave of the household survey and comparing the means between the two sets of villages along various indicators, we indeed find no statistically significant differences up to the 10 percent level, with the exception of the variable denoting distance from the block headquarters.

Table 3.2: Means of continuous variables, by village grain bank status

Variable	GBV (1)	NGBV (2)	GBV-NGBV (1)-(2)
Village-related variables			
Distance from main road (km)	3.04	3.75	-0.71
Distance from block headquarters (km)	17.07	25.57	-8.50***
Time taken to travel to block headquarters by main mode of transport (minutes)	143.57	163.93	-20.36
Distance from closest Agragamee field office (km)	5.42	5.78	-0.36
Distance from the closest weekly market (km)	8.86	7.61	1.25
Distance from the closest post office (km)	3.14	3.64	-0.50
Total number of households	40.36	51.62	-11.26
Share of landless households	0.25	0.41	-0.16
Share of ST households	0.92	0.79	0.12
Share of households that have reported food inadequacy for at least 1 month in past year	0.55	0.74	-0.19
Grain bank-related variables (at inception)			
Grain bank household membership size	29.07	29.64	-0.57
Share of grains contributed by villages	0.23	0.14	0.09
Share of ST members in grain bank	0.86	0.83	0.03
Number of grain bank committee members	5.57	6.29	-0.72
Share of female grain bank committee members	0.47	0.43	0.04
N	14	14	--

Notes: Means for village-related variables for 14 villages with functional grain banks (GBVs) in (1) and 11 villages with failed grain banks and 3 where grain banks were never established (NGBVs) in (2). Means for grain bank-related variables for 14 villages with functional grain banks in (1) and for 11 villages where grain banks failed in (2). * Statistically significant at the 10 percent level; ** at the 5 percent level; and *** at the 1 percent level.

As shown in Table 3.2, we find no statistically significant differences between the means of grain bank villages (GBVs) and non-grain bank villages (NGBVs) in terms of the distance from the main road and number of households. This is expected as we stratified the sample across these two dimensions. In addition, there are no significant differences in terms of the time taken to travel to the seat of local

administrative offices, distance from the closest Agragamee field office, distance from the closest weekly market, share of landless households, share of tribal households and the share of households that have reported food inadequacy during the 12 months preceding the survey. Although the difference in the average distance to the block headquarters is statistically significant, the time taken to reach it is not (reflecting differences in topography and varying ease of access).

Table 3.3: Means of dichotomous variables, by village grain bank status

Variable	GBV (1)	NGBV (2)	GBV-NGBV (1)-(2)
Share of villages (in percent) having			
Functional primary school	42.86	71.43	-28.57
Midday meals at school	35.71	64.29	-28.58
Anganwadi center or sub-center	64.29	57.14	7.15
Regular (daily/weekly) visits by *Anganwadi* worker	21.43	7.14	14.29
Vaccination drive in last 5 years	42.86	21.43	21.43
Electrical connection	7.14	0.00	7.14
Tarred village road	14.29	7.14	7.15
New road in last 5 years	57.14	28.57	28.57
Walking as main mode of transport to block headquarters	71.43	64.29	7.14
Self-help groups	71.43	42.86	28.57
Village-level meetings on a need-basis	35.71	21.43	14.28
Ward member in last 5 years	64.29	64.29	0.00
Watershed management program in last 10 years	42.86	0.00	42.86***
Food-for-work program in last 10 years	71.43	71.43	0.00
Grain bank established prior to OHFSP [1]	71.43	54.55	16.88
N	14	14	

Notes: Shares reported for 14 villages with functional grain banks (GBVs) in (1) and 11 villages with failed grain banks and 3 where grain banks were never established (NGBVs) in (2). [1] Share in (2) does not reflect 3 villages where grain banks were never established. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

In Table 3.3, we find no statistically significant differences between the shares of grain bank and non-grain bank villages having a functional primary school, *anganwadi* center, tarred village roads, elected ward member in the 5 years preceding the survey, 'self-help' (credit) groups and almost all other important indicators of physical and social infrastructure.

Tables 3.4 and 3.5 present the means for the same characteristics, but for the set of 3 villages where grain banks were never established (NEGBVs) and the 11 villages where grain banks failed (FGBVs). In these two sets of villages too, we find almost no difference between the means. Thus, we find that the three sets of villages – where grain banks continue to be operational, where they failed and where they were never established – are alike, on average, across a range of important characteristics, including village size, distance from important geographical markers and physical and social infrastructure.

Table 3.4: Means of continuous variables for non-grain bank villages, by whether grain bank failed or was never established

Variable	FGBV (1)	NEGBV (2)	FGBV-NEGBV (1)-(2)
Village-related variables			
Distance from main road (km)	3.86	3.33	0.53
Distance from block headquarters (km)	24.64	29	-4.36
Time taken to travel to block headquarters by main mode of transport (minutes)	165	160	5
Distance from closest Agragamee field office (km)	6.59	2.83	3.76*
Distance from the closest weekly market (km)	7.09	9.50	-2.41
Distance from the closest post office (km)	4.18	1.67	2.51
Total number of households	53.90	44.00	9.90
Share of landless households	0.36	0.62	-0.26**
Share of ST households	0.77	0.89	-0.13
Share of households that have reported food inadequacy for at least 1 month in past year	0.74	0.73	0.01
N	11	3	

Notes: Means reported for 11 villages where grain banks failed (FGBVs) in (1) and 3 villages where grain banks were never established (NEGBVs) in (2). * Statistically significant at the 10 percent level; ** at the 5 percent level; and *** at the 1 percent level.

3.5 Empirical methodology

The evaluation problem

The central problem in evaluation arises because we cannot observe outcomes for the same observation in the counterfactual state (Heckman and Robb 1985). For example, in this study, we can observe children in households that are grain bank participants (the treatment group) or children in households that are grain bank non-participants

(the comparison group), but we cannot observe outcomes for the same child in both states.

Table 3.5: Means of dichotomous variables for non-grain bank villages, by whether grain bank failed or was never established

Variable	FGBV (1)	NEGBV (2)	FGBV-NEGBV (1)-(2)
Share of villages (in percent) having			
Functional primary school	72.73	66.67	6.06
Midday meals at school	63.64	66.67	-3.03
Anganwadi center or sub-center	54.55	66.67	-12.12
Regular (daily/weekly) visits by *Anganwadi* worker	9.09	0.00	9.09
Vaccination drive in last 5 years	18.18	33.33	-15.15
Tarred village road	9.09	0.00	9.09
New road in last 5 years	27.27	33.33	-6.06
Walking as main mode of transport to block headquarters	63.64	66.67	-3.03
Self-help groups	54.55	0.00	54.55*
Village-level meetings on a need-basis	18.18	33.33	-15.15
Ward member in last 5 years	63.64	66.67	-3.03
Food-for-work program in last 10 years	63.66	100.00	-36.34
N	11	3	

Notes: Shares reported for 11 villages where grain banks failed (FGBVs) in (1) and 3 villages where grain banks were never established (NEGBVs) in (2). * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

The cleanest solution to this missing data problem is a social experiment where the comparison group is constructed from a random subset of the eligible population. However, since we do not have experimental data but rather observational data, we have to rely on non-experimental methods to estimate the impact of grain bank participation on children's health outcomes.

Matching compared with other non-experimental estimators

In recent years, an estimator that has gained wide usage in evaluation using non-experimental data is matching.[63] Matching estimators are based on the premise that

[63] For an evaluation of matching estimators, see, among others, Dehejia and Wahba (1999, 2002), Heckman et al. (1997, 1998a), Heckman et al. (1998), and Smith and Todd (2001, 2005a). For an overview of different non-experimental estimators, see Blundell and Costa Dias (2002).

the most appropriate estimate of the counterfactual untreated outcome for a treated unit is the outcome of an untreated unit or units most similar to the treated unit from the identified comparison group, under the assumption that selection into program participation is based on observables. While this is a strong assumption, it is also made by alternate non-experimental estimators such as linear regression. Although instrumental variable (IV) estimation can address bias due to selection on unobservables, the validity of the estimates is sensitive to the choice of the IV (a variable which has to be correlated with the decision to participate, but uncorrelated with any unobserved factors that affect the outcome). However, the difficulties of finding a valid IV are well-known.

If we assume that selection is indeed based on observables (or that the variables observable to the researcher span almost all of those used by the agent in deciding whether or not to participate), matching estimators provide impact estimates that can approximate estimates provided by randomized experiments. To do this, matching estimators first generate a propensity score. The propensity score is simply a predicted probability, based on observed characteristics, that the individual (or household) will participate in the program. This is used to match treated units with untreated units that are similar in every (observable) respect except in their treatment status.[64] The difference in the outcome of interest between matched units can then be attributed to the program.

Matching has some important advantages over regression. The latter imposes a linear functional form to identify the counterfactual outcome.[65] However, unlike

[64] We discuss the matching algorithms used to match treated and untreated units in the latter half of this section.

[65] If we include a sufficient number of higher order terms, however, the linear model can approximate a given non-linear function of the set of conditioning variables X arbitrarily well. In practice, however, including a large number of higher order terms for each variable, results in a degrees of freedom problem.

regression, matching methods are either semi-parametric or non-parametric, depending on the particular method used (Black and Smith 2004). In matching, after units with comparable propensity scores are matched, the difference in the outcome between the units provides the impact estimate. This is particularly important if participants and non-participants differ substantially in terms of their observed characteristics, as regression assumes the same functional form for both groups.

In addition, matching only pairs untreated and treated units having similar propensity scores. For a given untreated unit, if no comparable unit (i.e., having a similar propensity score) is available in the treated group, then it is discarded from the analysis. This area of overlap between the propensity scores is known as the "common support" region (Smith and Todd 2005a). However, in regression, all treated and untreated units are compared, regardless of whether they are comparable. While matching does not solve the support problem, it highlights it in a way that regression does not (Black and Smith 2004). In other words, matching exposes the common support problem, i.e., whether or not comparable untreated units are available for each treated unit (Smith 2004). After estimating the propensity scores for the treated and untreated groups, it is possible to compare the densities in order to examine the extent of the common support problem.

In addition, matching weights observations differently than ordinary least squares (OLS) regression in calculating the expected counterfactual for each treated observation. In OLS regression, all untreated units play a role in determining the expected counterfactual for any given treated unit and receive equal weight. In matching, however, only untreated units similar to each treated unit have a positive weight in determining the expected counterfactual, and these weights can vary depending on the distance in propensity scores between the matched units.

Performance of matching compared to experimental estimates

A number of recent studies have compared the performance of matching by comparing experimental estimates with non-experimental estimates using matching, mostly using data from voluntary employment and job training programs in the United States. Comparing experimental estimates to matching estimates using non-experimental data from the Job Training Partnership Act (JPTA) program, Heckman et al. (1997, 1998a) and Heckman et al. (1998b) provide evidence that propensity score matching methods perform well relative to experimental estimators, provided the following conditions are met: (1) the presence of a rich set of conditioning variables; (2) use of the same survey instruments for participants and non-participants; and (3) participants and non-participants face the same economic conditions. Using data from mandatory welfare-to-work training programs, Michalopoulos et al. (2004) come to a similar conclusion: matching estimators do not provide reliable estimates if the data for the comparison group come from a different geographic and labor market. Using National Support Work (NSW) data, Smith and Todd (2005a) highlight the fact that the performance of matching estimators depends critically on the quality of the data – specifically, that the conditions established by Heckman et al. (1997, 1998a) and Heckman et al. (1998b) are met. A more recent paper by Diaz and Handa (2006) provides evidence on the performance of matching estimators using non-US data. Comparing experimental and non-experimental estimates of the impact of Progresa, a voluntary anti-poverty program in Mexico, they find that matching estimators using the latter perform well when the outcomes of interest are measured comparably across treatment and comparison groups and a rich set of covariates is available.

In our data, a number of conditioning variables that potentially identify program participation and outcomes are available; the same survey questionnaire is used in grain bank and non-grain bank villages as a result of which the outcomes of

interest are measured identically; and participants and non-participants are in the same geographical area, namely within the same administrative unit within Rayagada district in the state of Orissa, and face the same economic and ecological conditions. Therefore, we argue that matching estimates can provide reliable program impact estimates in our study. In the next section, we discuss the conditions that need to hold for us to obtain reliable matching estimates.

Matching methods

Following Heckman et al. (1997) and Smith and Todd (2001, 2005a), let Y_1 denote the outcome of interest for children in grain bank participant households (the treatment group), Y_0 the corresponding outcome for children in non-grain bank households (the comparison group), and D an indicator variable denoting grain bank participation. Let X denote the set of conditioning variables. The parameter of interest – the average impact of the treatment on the treated (ATT) – is given by

$$\begin{aligned} & E(Y_1 - Y_0 \mid D = 1, X) \\ & = E(Y_1 \mid D = 1, X) - E(Y_0 \mid D = 1, X) \\ & = E(Y_1 \mid D = 1, X) - E(Y_0 \mid D = 0, X) \end{aligned}$$

Under the assumption of conditional independence (CIA), i.e., treatment status is independent of the outcome conditional on a set of observed covariates, we can estimate the parameter of interest (Imbens 2004).[66] Thus, the first condition required for matching estimates to be valid is $(Y_0 \perp D) \mid X$ (Condition 1). In other words, selection into the program is based only on observables. Put differently, this assumes that the analyst can observe the complete set of variables used by the agent in making the decision to participate.

[66] Smith and Todd (2005a) show that only a weaker assumption, $E(Y_0 \mid X, D = 1) = E(Y_0 \mid X, D = 0)$ (i.e., the conditional mean assumption), is needed to estimate the parameter of interest.

The second condition required for matching to be valid is $0 < \Pr(D = 1 | X) < 1$ (Condition 2). In other words, for all X there is a positive probability of either participating ($D = 1$) or not participating ($D = 0$).

Condition 1 runs contrary to those invoked by many economic models of self-selection, and its validity depends crucially on how well the observed data capture all the variables that affect program participation and outcomes of interest. The validity of matching estimates thus depends crucially on a thorough understanding of the selection process and the availability of a rich set of conditioning variables that affect participation and outcomes.

However, as the number of conditioning variables increases, we have to contend with what is commonly referred to as matching's version of the "curse of dimensionality" (Heckman et al. 1997). In other words, as the number of conditioning variables increases, it is possible to have many cells without matches. In this context, Rosenbaum and Rubin (1983) show that if we can match on *X*, we can also match on $P(X) = \Pr(D = 1 | X)$, the conditional participation probability or the so-called propensity score.

If the propensity score is estimated non-parametrically, we again have to contend with the problem of dimensionality (Heckman et al. 1998a). If the propensity score is estimated using parametric (such as logit or probit regression) or semi-parametric methods, however, then the dimensionality of the matching problem is reduced. We can then match on the univariate propensity score. Given Conditions 1 and 2, we can then estimate the parameter of interest by

$$E(Y_1 | D = 1, P(X)) - E(Y_0 | D = 0, P(X)) .$$

In experimental data, Conditions 1 and 2 are satisfied by random assignment to treatment. For non-experimental data, there may or may not exist a set of variables X such that these conditions are satisfied. Thus, it is important to be aware that our

estimates depend crucially on our assumptions, in particular, that we have a set of variables X such that Conditions 1 and 2 are satisfied.

Smith and Todd (2005a) also draw attention to the important fact that matching estimates are only valid if the support of X overlaps for the $D = 1$ and $D = 0$ groups. In other words, the treatment effect should be redefined as the program impact on participants whose propensity scores lie within the common support region (Smith and Todd 2005a).

Practical issues in propensity score matching

Assuming that Conditions 1 and 2 hold, estimation via matching entails a number of practical steps. First, how to choose the propensity score model and which variables to use? Second, which matching estimator to use? Third, what criteria to use in implementing the common support restriction? Fourth, which balancing test to use for implementing the propensity score method? We examine these issues, as well as related ones (such as specifying the kernel and bandwidth for certain estimators) below.

Choosing the propensity score model

As discussed in Smith (1997), there is little guidance on the functional form to use in specifying the propensity score model. In principle, any discrete choice model can be used. In practice, for the binary treatment case, the binomial logit or probit models are used. These are usually preferred to the linear probability model, due to the latter's shortcomings, such as predictions which are outside the [0,1] range. However, given that the purpose of the propensity score model is classification into treatment and comparison groups rather than the estimation of structural coefficients, model choice is not a critical issue (Smith 1997). And since both the logit and probit models yield

similar results, both are commonly used in the empirical literature for predicting propensity scores.[67] In our study, we estimate the propensity score model using a binomial probit.

Choosing which covariates to include in the propensity score model

The CIA assumption depends critically on including a set of covariates which captures the variables involved in selection into program participation. Todd and Smith (2005) show that matching estimates are biased unless the set of covariates X that satisfies Conditions 1 and 2 is included in the propensity score model specification. Thus, in order to obtain reliable matching estimates, one needs to capture all observable variables that potentially affect program participation as well as the outcome of interest. Thus, the issue of choosing which covariates to include in the propensity score model is important.

While theory provides some guidance on which variables are likely to affect both participation and outcomes, there is little guidance on how to correctly specify X in practice. Heckman et al. (1997) and Heckman et al. (1998b) show that there are larger biases if a 'coarse' set of conditioning variables is used relative to a 'rich' one. Rubin and Thomas (1996, p.253) argue in favor of including more covariates in the propensity score model specification, unless there is consensus that a variable is "unrelated to the outcome or not a proper covariate". However, including more covariates in X can also be a problem, as it comes at the cost of reducing the region of common support (Smith and Todd 2005a). In fact, a model that predicts perfectly results in observations with propensity scores that do not have any common support.

[67] Model choice is more important in the case of multiple treatments, but this issue is not relevant for our study (see Lechner 2001 for examples of the propensity score model in the case of multiple treatments).

Bryson et al. (2002, as cited in Millimet and Tchernis 2006) also argue that the inclusion of "irrelevant" variables can increase the variance of the treatment effect indicator.[68] However, Millimet and Tchernis (2006) use Monte Carlo simulations to compare the performance of matching estimators when relevant higher order terms are excluded and irrelevant higher order terms are included in the specification of the propensity score model. They find that overspecification of the propensity score model does not result in less efficient estimates, using different propensity score matching estimators including kernel estimators. Underspecification, on the other hand, results in worse estimates. Their results are corroborated in an application testing the impact of the World Trade Organization (WTO) on the environment as well as that of currency unions on international trade.

In general, the choice of variables usually depends on economic theory and previous empirical findings. However, Heckman et al. (1997), Heckman et al. (1998b) and Black and Smith (2004) also suggest some formal tests for choosing *X*. Heckman et al. (1997) suggest choosing the set of variables that maximizes the within-sample prediction rates using the hit-or-miss method.[69] Heckman et al. (1998b) suggest a test of statistical significance. They start with a parsimonious specification, and then add new variables iteratively. If the new variable is statistically significant at conventional levels, then it is included in the final *X*; otherwise it is discarded. Black and Smith (2004) suggest the 'leave-one-out cross validation method' to specify the propensity score model. In this method, they begin the model selection procedure by starting with a minimally specified model, and then successively add blocks of variables,

[68] These may be higher order terms of variables that are relevant (but only at a lower order) or variables that are wholly irrelevant.

[69] In this method, an observation is classified as '1' if the estimated propensity score is greater than the sample proportion of persons who receive the treatment, and 0 otherwise. This maximizes the overall classification rate, assuming that the costs for misclassification are the same for both treatment and control groups (Heckman et al. 1997).

comparing the resulting mean squared errors. Essentially, this method involves examining the fit of the model. However, in the case that this method suggests leaving out covariates that theory and previous empirical evidence suggest should be included, they advise giving more consideration to the latter.

In our study, we include a number of child, mother, household and community-level variables that we believe affect both participation and outcomes. We use the 'leave-one-out cross-validation method'. First, we specify a model with household-level variables only. We then add blocks of mother, child and village variables in successive specifications and compare their fit.

Choosing the matching estimator

There exist a variety of matching estimators, which differ mainly in how they assign weights to each comparison group observation. Asymptotically, they produce similar results, as in a large sample, they only compare exact matches (Black and Smith 2004). However, in finite samples, the estimates produced by different matching methods differ because of differences in the weighting function they use, as also how they address the common support problem (ibid).

Following Smith and Todd (2005a), for simplicity of notation, we rewrite $P = \Pr\left(D = 1 \middle| X\right)$. A typical matching estimator, $\hat{\alpha}_{\mathrm{M}}$, takes the form

$$\hat{\alpha}_{\mathrm{M}} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} \left[Y_{1i} - \hat{\mathrm{E}}\left(Y_{0i} \middle| D_i = 1, P_i\right) \right],$$

where

$$\hat{\mathrm{E}}\left(Y_{0i} \middle| D_i = 1, P_i\right) = \sum_{j \in I_0} W\left(i, j\right) Y_{0j}$$

and I_1 denotes the set of program participants, I_0 the set of non-participants, S_P the region of common support, n_1 the number of persons who are in the set $I_1 \cap S_P$. Every participant $i \in I_1 \cap S_P$ is matched with a weighted average over the outcomes of non-

participants, where the weights $W(i, j)$ depend on the distance between P_i and P_j. For each $i \in I_1$, $C(P_i)$ is defined to be a neighborhood. Individuals matched to i are those people in set A_i such that $\{j \in I_0 \mid P_j \in C(P_i)\}$.

Matching estimators differ from one another mainly in how they construct the weights $W(i, j)$ and define the neighborhood $C(P_i)$. Broadly speaking, they can be classified into two main types – traditional pair or one-to-one matching estimators, and more recently developed non-parametric matching estimators which match most (or all) the control observations using a pre-defined weighting function.

In one-to-one (or one-to-many) matching, outcomes are compared to the most observably similar untreated unit. Untreated units that do not have sufficiently close propensity scores are discarded. This reduces bias (due to better matches) but increases variance (due to fewer matches) (Smith and Todd 2005a). A common example of traditional matching estimators include nearest neighbor matching. In this method, the untreated unit j is matched to that treated unit i such that distance between them is the smallest. The neighborhood is defined as

$$C(P_i) = \min_j \|P_i - P_j\|, \; j \in I_0 .$$

Nearest neighbor matching, which is traditionally performed without replacement, constructs $W(i, j) = 1$ such that all matched units receive equal weight. Nearest neighbor matching can also be performed with replacement, such that untreated units can be matched to more than one treated unit, with accompanying trade-offs between reduced bias and increased variance compared to matching without replacement.

However, nearest neighbor matching is clearly inefficient. More efficient alternatives to nearest neighbor matching include a family of non-parametric, kernel-weighted matching estimators such as kernel matching and local linear matching

(Heckman et al. 1997, 1998a).[70] Relative to pair-wise matching, these estimators reduce the asymptotic mean squared error (Smith and Todd 2005a). In a study of finite sample properties of various matching estimators using Monte Carlo simulations, Fröhlich (2004) finds that the nearest neighbor estimator performs poorly relative to non-parametric kernel-weighted matching estimators. The latter match a treated unit with the weighted average score of *all* untreated units within a certain distance, referred to as the bandwidth. The weight given to the untreated unit is inversely proportionate to the distance between i and j and depends on the weighting function that is used. Relative to pair-wise matching, the use of more untreated units reduces the variance of the matching estimates; however, it increases the bias.

The most commonly used kernel-weighted estimators include the kernel estimator and the local linear estimator. Heckman et al. (1997) find that the local linear matching estimator has a slight advantage over the kernel matching estimator because of some desirable statistical properties; namely, it converges at a faster rate at boundary points and adapts better to different data densities. Fan (1992) also shows that local linear matching is better able to adapt to survey design. Therefore, one of the estimators used in this study is based on local linear matching. In local linear matching, the weighting function is given by

$$W(i,j) = \frac{G_{ij}\sum_{k\in I_0} G_{ik}\left(P_k - P_i\right)^2 - \left[G_{ij}((P_j - P_i)\right]\left[\sum_{k\in I_0} G_{ik}(P_k - P_i)\right]}{\sum_{j\in I_0} G_{ij}\sum_{k\in I_0} G_{ij}\left(\left(P_k - P_i\right)^2 - \sum_{k\in I_0} G_{ik}\left(P_k - P_i\right)\right)^2},$$

where $G_{ij} = G\left(\left(P_j - P_i\right)/a_n\right)$, and G denotes the kernel function and a_n the parameter determining the kernel bandwidth.

However, Heckman et al. (1997, 1998a) find that the local linear matching estimator does not perform well in small samples when there are regions of sparse data

[70] We do not discuss the difference-in-differences matching estimator proposed by Heckman et al. (1997) and Heckman et al. (1998b) since it requires before-after data which is not available to us. This estimator is preferred since it can control for time-invariant unobservables.

density. A common solution is to implement a trimming procedure in regions where the density of the propensity in the comparison population is small. However, Fröhlich (2004) points out that there is little practical guidance on the optimal level of trimming. From the distribution of propensity scores in our estimations, we find that while there is a large region of overlap in the propensity scores of treated and control observations (i.e., the region of common support), the distribution of the propensity scores for the two groups in our data is quite different, and so there are regions of sparse density. In addition, our sample size is small. For this reason, we also estimate the kernel matching model to examine if the estimates from the local linear matching analysis are comparable.

The weighting function for the kernel estimator is given by

$$G((P_j - P_i)/a_n) / \sum_{k \in I_0} G((P_k - P_i)/a_n).$$

Following Heckman et al. (1997) and Smith and Todd (2005a), the kernel-weighted matching estimate of program impact takes the form

$$ATT = \frac{1}{n_1} \sum_{i \in I_0} \left\{ Y_{1i} - \frac{\sum_{j \in I_0} Y_{0j} G\left(\frac{P_j - P_i}{a_n}\right)}{\sum_{k \in I_0} G\left(\frac{P_k - P_j}{a_n}\right)} \right\}.$$

The main difference between local linear matching and kernel matching is that the former includes a linear term in P_i in addition to the intercept. According to Smith and Todd (2005a, p. 317), this is helpful when "comparison group observations are distributed asymmetrically around the participant observations, as would be the case at a boundary point of *P* or at any point where there are gaps in the distribution of *P*". Thus, the local linear regression estimator is a more generalized version of the kernel estimator.

For both the local linear and kernel matching estimators, the neighborhood $C(P_i)$ depends on the choice of the kernel function. Two commonly used kernels in

the empirical matching literature include the Epanechnikov and Gaussian kernel. Using data from the National Longitudinal Survey of Youth (NLSY), in a study of the effects of college quality, Black and Smith (2004) find that the Epanechnikov kernel estimator performs slightly better than the Gaussian kernel estimator, independent of the size of the bandwidth. They find that the former converges faster than the Gaussian kernel and implicitly imposes the support condition through the choice of the bandwidth. Given this, we use the Epanechnikov kernel here with variable bandwidth. The latter, relative to a fixed bandwidth, has the advantage of varying the bandwidth depending on the data density at that point, i.e., it uses a small bandwidth in regions where the probability mass is dense and a large bandwidth when the probability mass is sparse (Ham et al. 2005).

A related choice that has important implications for the tradeoff between variance and bias, especially for matching methods based on kernel regression, is the size of the bandwidth or smoothing parameter. While a smaller bandwidth results in smaller bias but larger variance, a larger bandwidth results in smaller variance but larger bias (Galdo, Smith, and Black 2006). Specifically, in the case of matching, a smaller bandwidth leads to the use of few untreated units for each treated unit while a larger bandwidth leads to the use of untreated units that may be rather different from each treated unit. The standard approach in the matching literature to guide bandwidth choice is minimizing some quadratic loss function, such as the mean squared error (MSE) or integrated mean squared error (IMSE), as a measure of fit. However, as discussed by Galdo et al.(2006), the bandwidth that minimizes the MSE for the regression function of the untreated outcome is not that which minimizes the MSE for the object of interest, which in our case is the average treatment effect. In addition, if the distribution of the conditioning variables is not balanced for the treated and comparison observations, then the optimal bandwidth in the region of low propensity

scores (where most comparison units lie) will be different from the optimal bandwidth in the region of high propensity scores (where most treated units lie). This results in biased estimates.

In our study, we examine the MSE for varying bandwidths between 0.01 and 0.1 (in 0.01 increments). We present estimates for bandwidth size 0.06, which, while not having the minimum MSE in the range examined, enables us to include a more inclusive X vector of balanced covariates. We also present estimates for different bandwidth sizes (specifically, 0.05 and 0.07) to examine the sensitivity of our impact estimates to the choice of bandwidth.

Choosing the criteria to restrict matching to the common support region

In order to obtain credible matching estimates, only those comparison and treatment observations whose propensity scores fall within the region of common support should be included. This is particularly important for kernel-weighted matching since it uses all comparison observations to estimate the counterfactual outcome, unlike nearest neighbor matching. Implementing the common support restriction can improve the quality of the matches used to estimate the *ATT*, although it comes at the cost of reduced sample size as observations at the boundaries of the common support are excluded.

In practice, one of two methods is commonly implemented for ensuring common support. The first involves the comparison of the minimum and maximum propensity scores of treatment and comparison observations. All observations in the treatment (or comparison) group whose propensity score is smaller than the minimum and larger than the maximum of the propensity scores in the opposite group are discarded.

The second method for implementing the common support is a trimming procedure, as suggested by Smith and Todd (2005a). They determine the common support region by

$$\hat{S}_P = \{P : \hat{f}(P \mid D=1) > 0 \text{ and } \hat{f}(P \mid D=0) > 0\},$$

where $\hat{f}(P \mid D=d)$, $d \in \{0,1\}$ are non-parametric density estimators given by

$$\hat{f}(P \mid D=d) = \sum_{k \in I_d} G((P_k - P)/a_n).$$

They then define a trimming level q and require that densities are strictly positive and exceed zero by q. They exclude observations with P for which the density is zero, and then an additional q percent of remaining P points for which the estimated density is low (although positive). Thus, the set of eligible matches is given by

$$\hat{S}_q = \{P \in I_1 \cap \hat{S}_P : \hat{f}(P \mid D=1) > c_q \text{ and } \hat{f}(P \mid D=0) > c_q\}.$$

In our study, we use the min-max method to implement the common support region as there are no guidelines on how to arrive at the optimal trimming level given the data (Fröhlich 2004).

Choosing a balancing test for the set of covariates included in the propensity score model

If the CIA assumption is valid, after we condition on P, additional conditioning on any of the X's should not provide any new information about the treatment decision. This implies that all the Xs should be "balanced" across the treatment and matched comparison groups. Thus, in order to assess the quality of the matching estimates, we need to conduct a 'balancing test' of the characteristics of the matched samples. There are a few formal tests that are commonly implemented in the literature, though there is no consensus on which one is definitive.

Proposed by Rosenbaum and Rubin (1985), one balancing test is the examination of standardized differences (see, e.g., Sianesi 2004). In words, the

standardized difference for a variable X_k is the difference in means between the treatment and matched comparison group samples, as a percentage of the square root of the average of the sample variance in both groups. Intuitively, this provides the size of the difference in means of a conditioning variable between the treatment and matched comparison group, scaled down by the average of the variances.[71] In this method, for each of the conditioning variables, the standardized bias before matching is

$$SB_{before} = 100 \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{\frac{\text{var}_1(X) + \text{var}_0(X)}{2}}},$$

where $\overline{X}_0$, $\text{var}_0(X)$, $\overline{X}_1$, and $\text{var}_1(X)$ denote the means and variances in the treatment and comparison groups *before* matching, respectively. The standardized bias after matching is

$$SB_{after} = 100 \frac{\overline{X}_{1M} - \overline{X}_{0M}}{\sqrt{\frac{\text{var}_{1M}(X) + \text{var}_{0M}(X)}{2}}},$$

where $\overline{X}_{0M}$, $\text{var}_{0M}(X)$, $\overline{X}_{1M}$, and $\text{var}_{1M}(X)$ denote the means and variances in the treatment and comparison groups *after* matching, respectively. If the covariates are balanced, we expect a reduction in the standardized bias (although there is no formal criteria for how much reduction should occur).

A second balancing test proposed by Rosenbaum and Rubin (1985) is to examine if there are significant differences in covariate means between the treatment and matched comparison groups, using two-sample *t*-tests (see, e.g., Ham et al. 2005). While we expect differences to exist before matching, there should be no significant differences after matching as the covariates should be balanced in both groups.

[71] The shortcoming of this method is that there are no formal criteria for which the standardized bias is too large to pass the balancing test (Smith and Todd 2005b). Smith and Todd (2005b) also point out that the standardized bias can be manipulated by the researcher by adding additional observations to the comparison group if these additional observations increase the second variance term in the denominator.

A third test is a Hotelling *T*-squared test of the joint null of equal means of all the *X*'s between the treatment and the matched or reweighted comparison group (see, e.g., Smith and Todd 2005b).

A fourth test is comparing the pseudo-*R*-squared of the original model used to create propensity scores with a re-estimated model using only the matched observations (Sianesi 2004). Since the pseudo-*R*-squared indicates how well the *X*'s predict the probability of participation, the second pseudo-*R*-squared value should be low since we do not expect systematic differences in the distribution of covariates between the treatment and matched comparison groups.

A fifth test proposed by Dehejia and Wahba's (1999, 2002) is balancing the *X*'s within certain strata. They first divide the treatment and comparison observations into an arbitrary number of strata based on their propensity scores, such that within each stratum, there is no statistically difference in the mean propensity scores between the groups. Then, within each stratum, for each covariate, they use a series of *t*-tests to examine if there is any significant difference in the means between the two groups. If they find any significant difference, they add higher order and interaction terms in the specification of the propensity score until no differences appear.

A sixth test is based on a regression framework. For example, Smith and Todd (2005b) first estimate the following regression:

$$\begin{aligned} X_k &= \beta_0 + \beta_1 \hat{P}(X) + \beta_2 \hat{P}(X)^2 + \beta_3 \hat{P}(X)^3 + \beta_4 \hat{P}(X)^4 \\ &+ \beta_5 D + \beta_6 D\hat{P}(X) + \beta_7 D\hat{P}(X)^2 + \beta_8 D\hat{P}(X)^3 + \beta_9 D\hat{P}(X)^4 + \eta \end{aligned}$$

The joint null is that that the coefficient on all the terms with the treatment dummy *D* equals zero. In other words, after conditioning on the *X*'s, *D* should provide no information on X_k .

In our study, we examine two-sample *t*-tests for differences between means of the treatment and matched comparison observations. We also compare the

standardized bias before and after matching, and examine the pseudo-*R*-squared of the propensity score model re-estimated after matching using reweighted observations on the common support only.

Sensitivity analysis: Unobserved heterogeneity

Given that the validity of the matching estimates hinges on whether the CIA assumption holds, an analysis of the sensitivity of our results to departures from this identifying assumption is crucial. In this context, a bounding approach suggested by Rosenbaum (2002) is increasingly used in matching applications. While the bounding approach cannot test the CIA assumption per se, it examines the extent to which the statistical significance the results hinges on this untestable assumption. Suppose the probability of participation P_i is given by

$$\mathrm{P}(D_i = 1 \mid X_i) = F(\beta X_i + \gamma U_i),$$

where X_i are the observed characteristics for individual i, U_i are unobserved characteristics, and β and γ are the impacts of X_i and U_i on the participation decision. If there are no unobservable characteristics that affect participation, i.e., $\gamma = 0$, then two individuals with the same set of observable characteristics X have the same probability of participation. However, if $\gamma \neq 0$, i.e., there are unobservable characteristics that affect participation, then two individuals with the same X have differing probabilities of participation. Assuming, for simplicity, that F is the logistic distribution, the odds that two individuals i and j participate are given by $\mathrm{P}_i / (1-\mathrm{P}_i)$ and $\mathrm{P}_j / (1-\mathrm{P}_j)$ respectively. Then, the odds ratio can be written as

$$\frac{\mathrm{P}_i / (1-\mathrm{P}_i)}{\mathrm{P}_j / (1-\mathrm{P}_j)} = \mathrm{P}_i(1-\mathrm{P}_j) / \mathrm{P}_j(1-\mathrm{P}_i) = \frac{\exp(\beta X_i + \gamma U_i)}{\exp(\beta X_j + \gamma U_j)}.$$

If i and j form a matched pair, then the X vector cancels out and the odds ratio can be simply written as $\exp[\gamma(U_i - U_j)]$. If there are no differences in unobservables, i.e.

$U_i = U_j$, or the unobservable factors do not affect the probability of participating, i.e. $\gamma = 0$, the odds ratio equals 1, implying that the matching estimates do not suffer from unobserved selection bias. However, if this is not the case, then the matching estimates are said to suffer from a "hidden bias". In this context, Rosenbaum (2002) shows that the following bounds can be placed on the odds ratio:

$$\frac{1}{e^{\gamma}} \leq \frac{\mathrm{P}_i / (1-\mathrm{P}_i)}{\mathrm{P}_j / (1-\mathrm{P}_j)} \leq e^{\gamma}.$$

As e^{γ} increases, the bounds move apart reflecting uncertainty due to the presence of unobserved selection bias. Thus, e^{γ} is a measure of the extent to which the analysis suffers from this bias.

In our study, we follow Rosenbaum's approach to test the sensitivity of the significance of our results, if any, to violations of the CIA assumption.

3.6 Data and results

In this section, we first discuss the data. We then present mean *haz*- and *whz*-scores and proportions of stunting and wasting outcomes in the agricultural lean season for children in participant households in all grain bank villages (GBVs), children in participant households in long-lived grain bank villages (long-lived GBVs), and children in non-grain bank villages (NGBVs). We then implement a matching-based analysis of the impact on children's *haz*-scores and *whz*-scores in the agricultural lean season, as well as the change in height between the post-harvest and lean seasons, using local linear regression and kernel estimators.

3.6.1 Analysis sample

Analysis of levels of haz-scores and whz-scores

The data for the analysis of the levels of *haz*-scores and *whz*-scores come from the second wave of a small-scale household and grain bank survey implemented in Kashipur block in rural Rayagada, Orissa, corresponding with the agricultural lean season when food shortages are at their peak.[72] Of the 499 households sampled in the second wave of data collection, 375 had 1 or more children below 6 years of age. They include 263 children in 167 households in participant households in GBVs and 308 children in 208 non-participating households. Of the latter, 23 children are from 15 households in GBVs who were not grain bank members. More than half of these children were in households that reported being excluded as they were either not creditworthy or because of caste-related discrimination, while the rest of the children were in households that reported not joining grain banks as they did not have food shortages.[73] We drop these observations, restricting our analysis only to children in participating households in GBVs and children in households in NGBVs.[74] Thus, the comparison group is confined to the 285 children in households in NGBVs only, bringing the total sample under consideration to 548 children.

Of these, height data is missing in 35 cases, and weight (or both height and weight) in 42 cases. The missing anthropometric data are split fairly evenly between the comparison and treatment groups, with height data missing for 16 children in

[72] Details of the survey and sampling methodology are provided in a separate appendix.

[73] The first set of households (numbering 13) was distributed across 6 villages, with the majority from Dhobasil. The second set of households was distributed fairly evenly across 4 villages.

[74] The share of households that do not participate in villages with operational grain banks is small (19 out of 250 households, or 7.6 percent), and comprises two main categories: the majority that state that they did not or could not participate because they were not creditworthy or were objects of caste discrimination, and close to a third that reported not having food shortages. These households represent the two tails of the distribution of households in terms of food-security – while the former is probably more food-insecure than the households that do participate, the latter are potentially better-off than those that selected into the program. Therefore, we exclude the small number of observations of non-participants in grain bank villages. We argue that our estimates are valid for the vast majority of households that do participate and are representative of almost the entire distribution excluding the tails.

participant households and 19 children in NGBVs and weight data missing for 21 children in participating households and 21 children in NGBVs. Thus, the usable sample for the calculation of the standardized scores is 513 and 506 for *haz*-scores and *whz*-scores respectively. Of these, 3 observations, whose absolute value of the *haz*-scores was above 6, are discarded, bringing the effective sample for the analysis of *haz*-scores to 510.[75]

Birthdates from *anganwadi* records and immunization cards are available in only about 30 percent of cases. However, even when available, the reliability of this data is questionable. In all other cases, it is based on the report by the mother of the child.[76] For children below 2 years, recumbent length instead of height was measured.

For the calculation of *whz*-scores, 16 observations, whose height fell below 77 cm, the permissible floor for the calculation of these scores, are discarded. [77] While *wlz*-scores could be calculated for these observations, their height, not recumbent length, had been measured, as their age was reported to be above 24 months. This reduces the sample size for the calculation of *whz*-scores (and *wlz*-scores) to 490 observations. Two observations whose recumbent length was measured as slightly below 45 cm are also dropped, as is 1 observation whose measured height was above the permissible ceiling. Finally, 2 observations for whom the absolute value of the standardized score was above 6 are dropped, bringing the effective sample of children with *whz*-scores to 485.

---

[75] The Center for Disease Control (CDC)'s 2000 reference growth charts were used for deriving height-for-age *z*-scores. For children 2 years or less, length-for-age *z*-scores were estimated to take into account the fact that recumbent length, not height, was measured.
[76] A discussion of the quality of the age data is included in the appendix on survey notes.
[77] The CDC's 2000 reference growth charts were used for deriving weight-for-height *z*-scores. For children above years, weight-for-height *z*-scores were estimated, while for children 2 years or less, weight-for-length *z*-scores were estimated to take into account the fact that recumbent length, not height, was measured. The permissible range for height was 77 to 121.5 cm. The permissible range for length was 45 to 103.5 cm.

Analysis of change in height

For the analysis of the change in height, data for children in households that were sampled in both waves of the household survey are used. The total number of such children is 405. Of these, 19 observations corresponding to children in non-participant households in GBVs are dropped. In addition, height data is missing in 53 observations in at least one of the waves (36 observations are missing from the first wave and 21 from the second wave, including 4 from both). One additional observation is dropped since the change is height is found to be negative (-0.01 cm). As a result, the usable sample for the analysis of change in height is 332 children. Note that the effective sample sizes for the empirical estimation are smaller than the usable sample sizes discussed above, as the matching estimation is restricted to observations whose propensity score belongs to the intersection of the supports of the propensity scores of treatment and comparison units. We discuss this feature, and other details of the empirical framework, in the next section. The effective sample sizes are included at the bottom of the tables which present the average treatment effect on the treated as estimated by matching.

3.6.2 Naïve results

In this subsection, we present our estimates of the impact of grain bank participation on the short- and long-term health outcomes of children without the use of matching. The findings are presented largely to serve as a comparison to our main matching-based estimates.

Table 3.6 presents the mean *haz*-score and proportion of stunting. The majority of children in participant households as well as households in NGBVs are found to be at least two standard deviations below the healthy reference for *haz*-scores, indicating the high incidence of stunting in the population. The mean *haz*-

score and share of stunted children is found to be very similar across children in participant households in all GBVs (including those in long-lived GBVs only) and children in NGBVs. Further, as indicated by the 95 percent confidence intervals, the point estimates are not statistically significantly different from each other. These results in themselves are not too surprising as grain banks are not designed to have a long-term impact on health and nutrition outcomes.

Table 3.6: Mean *haz*-scores and share of stunted children in agricultural lean season

Children in	Mean	Share
participant households in all GBVs	-2.020 [-2.217, -1.823]	0.586 [0.521, 0.651]
participant households in long-lived GBVs	-2.004 [-2.219, -1.788]	0.587 [0.514, 0.632]
households in NGBVs	-2.067 [-2.258, -1.875]	0.564 [0.496, 0.632]

Notes: Sampling weights used to adjust means and proportions. 95 percent confidence intervals in brackets.

Table 3.7 presents the mean *whz*-score and share of wasting of children across participating households and households in NGBVs.

Table 3.7: Mean *whz*-scores and share of wasted children in agricultural lean season

Children in	Mean	Share
participant households in all GBVs	-1.716 [-1.896, -1.536]	0.370 [0.305, 0.436]
participant households in long-lived GBVs	-1.739 [-1.937, -1.542]	0.366 [0.293, 0.440]
households in NGBVs	-1.543 [-1.702, -1.384]	0.347 [0.282, 0.413]

Notes: Means and proportions are corrected for sampling weights. 95 percent confidence intervals in brackets.

Contrary to expectation, we find that the mean *whz*-score is lower and the share of wasted children higher in participant households in GBVs than in households in NGBVs during the agricultural lean season, though this difference is not statistically significant. We obtain the same result when we compare children in participant households in long-lived GBVs with children in NGBVs. Thus, even when we examine an indicator of short-term health, we find no difference between children in participant and non-participant households.

Table 3.8 presents the average change in height of young children below 6 years. While children in participant households have a slightly higher rate of growth than children in NGBVs, we find no statistically discernible difference between them. We now turn to a matching-based analysis to see if our results change when we match on various individual, household and community level characteristics.

Table 3.8: Average change in height between post-harvest and lean seasons

Children in	Mean
participant households in all GBVs	0.052 [0.041, 0.062]
participant households in long-lived GBVs	0.050 [0.037, 0.062]
households in NGBVs	0.044 [0.036, 0.052]

Notes: Means and proportions are corrected for sampling weights. 95 percent confidence intervals in brackets.

3.6.3. Matching-based results

In this section, we examine the impact of grain bank participation on children's health outcomes using propensity score matching methods. Below, we discuss the steps involved.

Calculating propensity scores

Using a binomial probit regression model, we first estimate propensity scores to match children in participant households in grain bank villages to children in households in non-grain bank villages. We estimate the model for two different samples: one where the treatment group comprises of children in participant households in grain bank villages (Sample 1) and the other where the treatment group comprises of children in participant households in long-lived grain bank villages (Sample 2). The comparison group in both samples is the same for both samples, namely children in households in non-grain bank villages. We estimate the models separately for each outcome of interest as Sample 1 and 2 are slightly different in the case of *whz*-scores, *haz*-scores and change in height (depending on the availability of weight and height data).

In specifying the propensity score model, we include a number of covariates which predict both the decision to participate as well as the outcome. In choosing the set of covariates, we have to balance the benefit of improving the predictive ability of the model with the cost of reducing the region of common support. In specification 1, we include a number of variables that capture different aspects of household wealth which we consider to be important determinants of the outcome variable. In specification 2, we also include variables that capture relevant characteristics of the child's mother. In specification 3, we also include variables capturing relevant child characteristics. In specification 4, we also include village-level variables which we consider to have an important effect on grain bank survival (thereby determining participation).

Discussion of control variables

The control variables chosen to estimate the propensity scores include various community, household and individual-level characteristics. We do not include

village-level dummies in the estimation as the treatment and comparison observations belong to distinct villages. Including village dummies would result in no matches being made, as no control-treatment matches are possible within the same village. The control variables used in the different specifications can be roughly grouped as follows:

1. household-level characteristics (e.g., whether or not the household is tribal, household size, number of children, number of adult females; size of landholdings by quality of the land, dwelling floor quality, number of agricultural tools and livestock owned, amount of gold owned;
2. mother characteristics (e.g., height, age, years of education);
3. child characteristics (e.g., age, sex); and
4. village-level characteristics (e.g., distance from the market, distance from the closest Agragamee field office, an indicator for the frequency of village-level meetings).

Quadratic terms of some of the continuous variables above are also included to improve the fit of the model.

Tables 3.9-3.10, 3.11-3.12, and 3.13-3.14 present the model of participation used to create propensity scores for the matching algorithm for Sample 1 and Sample 2 respectively, for each of the outcomes, namely, *haz*-score, *whz*-score and change in height. In each table, column (1) presents the results for the estimation for specification 1, i.e., household variables only. Column (2) presents the results for the estimation for specification 2, i.e., household and mother variables. Column (3) presents the results for the estimation for specification 3, i.e., household mother, and child variables. Column (4) presents the results for the estimation for specification 4, i.e., household, mother, child and community variables. Across the different samples and outcomes under consideration, we find that specification 4 has the highest

likelihood ratio index (referred to as McFadden's pseudo-*R*-squared in our tables). Therefore, we estimate the remaining results using specification 4. Selected statistics of fit are presented at the bottom of the tables.

Distribution of predicted propensity scores

Figures 3.1-3.3 depict the distribution of the predicted propensity scores for children from participant households as well as children from households in non-grain bank villages for Sample 1 for each of the outcomes, namely, *haz*-score, *whz*-score and growth, respectively. Figures 3.4-3.6 depict the distribution of the predicted propensity scores for children from participant households as well as children from households in non-grain bank villages for Sample 2 for each of the outcomes mentioned above. We see that children in non-grain bank villages have a higher probability mass at lower levels of the propensity score compared to children in participant households for both Samples 1 and 2. This indicates that based on the set of observable characteristics used to create the propensity scores, children in the treatment group differ from children in the comparison group. Thus, there is a potential gain from using matching estimators compared to regression.

Common support constraint and balancing tests

After the propensity scores are generated, the common support restriction is implemented, so that the test of the balancing property is performed only on observations whose propensity score belongs to the intersection of the supports of the propensity score of treatment and comparison units. To do this, treatment observations whose propensity score is higher that the maximum or less than the minimum propensity score of comparison observations are dropped.

Table 3.9: Determinants of grain bank participation estimated for creating propensity scores to examine impact on haz scores (Sample 1)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Estimated coefficients			
	(1) Specification A	(2) Specification B	(3) Specification C	(4) Specification D
Household characteristics				
Social group (1=tribal)	0.317	0.416	0.395	0.0692
	(0.27)	(0.31)	(0.31)	(0.27)
Household size	0.622**	0.576	0.557	0.353
	(0.32)	(0.36)	(0.36)	(0.34)
Square of household size	-0.0718***	-0.0719***	-0.0709***	-0.0515*
	(0.024)	(0.027)	(0.027)	(0.027)
Number of children (14 years or less)	0.128	0.203	0.209	0.183
	(0.14)	(0.15)	(0.15)	(0.17)
Number of adult females (15 years or more)	0.0294	0.145	0.160	0.295
	(0.19)	(0.21)	(0.21)	(0.25)
Low-lying fertile (bila) land owned (acres)	0.309**	0.438***	0.399**	0.648***
	(0.14)	(0.16)	(0.16)	(0.18)
Low-lying fertile land, squared	-0.0776***	-0.0964***	-0.0880***	-0.132***
	(0.026)	(0.031)	(0.032)	(0.043)
Semi-fertile (goda) land owned (acres) – cubic term	-0.00255***	-0.00235**	-0.00220**	-0.00283***
	(0.00093)	(0.00093)	(0.00091)	(0.00089)
Infertile (dongar) upland owned (acres)	-0.0167	-0.0409	-0.0300	0.134
	(0.13)	(0.13)	(0.13)	(0.15)
Infertile (dongar) upland, squared	-0.0314	-0.0243	-0.0255	-0.0557*
	(0.025)	(0.027)	(0.026)	(0.029)
Gold holdings (gm)	0.0328	0.0410*	0.0410*	0.0602***
	(0.020)	(0.023)	(0.023)	(0.023)
Floor quality (1 = pucca)	0.171	0.0890	0.151	-0.175
	(0.26)	(0.31)	(0.31)	(0.39)
Number of ploughs owned	0.0536	0.116	0.0894	0.0215
	(0.19)	(0.21)	(0.21)	(0.23)
Number of crowbars owned	-0.0836	-0.0280	-0.0510	0.0682
	(0.13)	(0.15)	(0.15)	(0.15)
Number of spades owned	0.0861	0.0788	0.0892	0.121
	(0.093)	(0.11)	(0.11)	(0.12)
Number of sickles owned	0.258***	0.258***	0.258***	0.157**
	(0.062)	(0.066)	(0.066)	(0.072)
Number of cows owned	-0.148***	-0.180***	-0.176***	-0.264***
	(0.045)	(0.058)	(0.059)	(0.061)
Number of bullocks owned	0.195**	0.128	0.137	0.200
	(0.099)	(0.12)	(0.12)	(0.12)

Table 3.9 (Continued)

Number of goats owned	-0.0343	-0.0396	-0.0427	-0.0370
	(0.025)	(0.028)	(0.028)	(0.028)
Number of buffaloes owned	-0.366***	-0.517***	-0.506***	-0.484***
	(0.13)	(0.18)	(0.18)	(0.18)
Mother's characteristics				
Mother's education (years)		0.00374	0.00599	0.0868
		(0.080)	(0.079)	(0.073)
Mother's height (cm)		0.00410	0.00562	0.00478
		(0.013)	(0.013)	(0.014)
Mother's age (years)		-0.273***	-0.266***	-0.167*
		(0.10)	(0.10)	(0.10)
Square of mother's age		0.00429***	0.00421***	0.00273*
		(0.0016)	(0.0016)	(0.0016)
Child characteristics				
Child's age (months)			0.00137	0.0104
			(0.013)	(0.014)
Child's age, squared			-0.0000198	-0.000145
			(0.00018)	(0.00020)
Child's sex (1 = male)			0.281**	0.326**
			(0.14)	(0.15)
Community characteristics				
Distance from market (km)				0.0501***
				(0.014)
Distance from closest Agragamee field office (km)				0.00624
				(0.048)
Distance from closest Agragamee field office (km), squared				-0.00133
				(0.0024)
Frequency of village meetings (1 = frequent/ as needed)				1.794***
				(0.20)
Constant	-2.866***	0.318	-0.145	-2.799
	(0.84)	(2.29)	(2.31)	(2.49)
N	510	430	430	430
Wald χ^2	86.06	76.82	81.17	194.97
p-value	0.0000	0.0000	0.0000	0.0000
McFadden's Pseudo-R^2	0.1441	0.1731	0.1803	0.3809

Notes: Estimates are corrected for sampling weights. Standard errors reported in parentheses.
*Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.10: Determinants of grain bank participation estimated for creating propensity scores to examine impact on *haz* scores (Sample 2)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Estimated coefficients			
	(1) Specification A	(2) Specification B	(3) Specification C	(4) Specification D
Household characteristics				
Social group (1=tribal)	0.125	0.176	0.156	-0.230
	(0.28)	(0.33)	(0.33)	(0.29)
Household size	0.762**	0.836**	0.808*	0.498
	(0.36)	(0.42)	(0.43)	(0.40)
Square of household size	-0.0794***	-0.0920***	-0.0898***	-0.0584*
	(0.027)	(0.033)	(0.033)	(0.032)
Number of children (14 years or less)	0.0363	0.121	0.128	0.0581
	(0.15)	(0.17)	(0.17)	(0.18)
Number of adult females (15 years or more)	-0.0750	0.0331	0.0467	0.0761
	(0.21)	(0.23)	(0.23)	(0.26)
Low-lying fertile (bila) land owned (acres)	0.402***	0.567***	0.522***	0.789***
	(0.15)	(0.18)	(0.18)	(0.20)
Low-lying fertile land, squared	-0.0866***	-0.112***	-0.103***	-0.162***
	(0.028)	(0.036)	(0.038)	(0.048)
Semi-fertile (goda) land owned (acres) – cubic term	-0.00373**	-0.00319***	-0.00295***	-0.00400***
	(0.0015)	(0.0012)	(0.0011)	(0.00099)
Infertile (dongar) upland owned (acres)	-0.0409	-0.115	-0.114	0.149
	(0.14)	(0.15)	(0.15)	(0.19)
Infertile (dongar) upland, squared	-0.0417	-0.0305	-0.0293	-0.0748*
	(0.029)	(0.031)	(0.030)	(0.039)
Gold holdings (gm)	0.0421**	0.0479**	0.0486**	0.0670***
	(0.021)	(0.024)	(0.024)	(0.025)
Floor quality (1 = pucca)	0.161	0.0945	0.154	-0.0791
	(0.27)	(0.33)	(0.33)	(0.43)
Number of ploughs owned	-0.0744	0.0382	-0.00223	0.329**
	(0.22)	(0.26)	(0.26)	(0.16)
Number of crowbars owned	-0.316**	-0.304*	-0.330*	-0.110
	(0.16)	(0.17)	(0.17)	(0.17)
Number of spades owned	0.154	0.212*	0.223*	0.253*
	(0.10)	(0.12)	(0.12)	(0.13)
Number of sickles owned	0.302***	0.307***	0.306***	0.178**
	(0.070)	(0.074)	(0.074)	(0.080)
Number of cows owned	-0.186***	-0.236***	-0.238***	-0.287***
	(0.050)	(0.067)	(0.069)	(0.075)
Number of bullocks owned	0.325***	0.253*	0.271*	
	(0.12)	(0.14)	(0.14)	

Table 3.10 (Continued)

Number of goats owned	-0.0342	-0.0500*	-0.0526*	-0.0424
	(0.025)	(0.027)	(0.027)	(0.027)
Number of buffaloes owned	-0.415***	-0.569***	-0.555**	-0.640***
	(0.16)	(0.22)	(0.22)	(0.22)
Mother's characteristics				
Mother's education (years)		-0.0119	-0.00931	0.100
		(0.084)	(0.083)	(0.078)
Mother's height (cm)		-0.00132	0.000470	0.00628
		(0.014)	(0.014)	(0.016)
Mother's age (years)		-0.342***	-0.343***	-0.212*
		(0.11)	(0.11)	(0.11)
Square of mother's age		0.00536***	0.00541***	0.00343*
		(0.0018)	(0.0018)	(0.0018)
Child characteristics				
Child's age (months)			0.00371	0.0165
			(0.014)	(0.015)
Child's age, squared			-0.0000274	-0.000222
			(0.00019)	(0.00021)
Child's sex (1 = male)			0.301*	0.342**
			(0.15)	(0.16)
Community characteristics				
Distance from market (km)				0.0272*
				(0.015)
Distance from closest Agragamee field office (km)				0.494**
				(0.20)
Distance from closest Agragamee field office (km), squared				-0.0543***
				(0.020)
Frequency of village meetings (1 = frequent/ as needed)				1.453***
				(0.21)
Constant	-2.979***	1.756	1.330	-2.700
	(0.93)	(2.55)	(2.58)	(2.90)
N	454	382	382	382
Wald χ^2	94.67	99.3	103.1	214.27
p-value	0.0000	0.0000	0.0000	0.0000
McFadden's Pseudo-R^2	0.1894	0.2424	0.2504	0.4131

Notes: Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.11: Determinants of grain bank participation estimated for creating propensity scores to examine impact on *whz* scores (Sample 1)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Estimated coefficients			
	(1) Specification A	(2) Specification B	(3) Specification C	(4) Specification D
Household characteristics				
Social group (1=tribal)	0.312	0.429	0.408	0.0957
	(0.28)	(0.31)	(0.31)	(0.27)
Household size	0.472	0.496	0.474	0.321
	(0.33)	(0.37)	(0.37)	(0.35)
Square of household size	-0.0656***	-0.0709**	-0.0699**	-0.0531*
	(0.025)	(0.028)	(0.028)	(0.028)
Number of children (14 years or less)	0.218	0.285*	0.298*	0.260
	(0.14)	(0.16)	(0.16)	(0.18)
Number of adult females (15 years or more)	0.115	0.232	0.255	0.401
	(0.20)	(0.21)	(0.22)	(0.25)
Low-lying fertile (bila) land owned (acres)	0.298**	0.410***	0.370**	0.611***
	(0.14)	(0.16)	(0.16)	(0.18)
Low-lying fertile land, squared	-0.0687***	-0.0859***	-0.0770**	-0.118***
	(0.026)	(0.030)	(0.031)	(0.041)
Semi-fertile (goda) land owned (acres) – cubic term	-0.00338**	-0.00300**	-0.00288**	-0.00317***
	(0.0016)	(0.0013)	(0.0013)	(0.0011)
Infertile (dongar) upland owned (acres)	-0.0710	-0.105	-0.0987	0.0774
	(0.13)	(0.13)	(0.13)	(0.15)
Infertile (dongar) upland, squared	-0.0188	-0.0104	-0.0103	-0.0457*
	(0.025)	(0.026)	(0.026)	(0.027)
Gold holdings (gm)	0.0103	0.0133	0.0115	0.0373
	(0.021)	(0.023)	(0.023)	(0.024)
Floor quality (1 = pucca)	0.234	0.157	0.233	-0.129
	(0.26)	(0.31)	(0.31)	(0.39)
Number of ploughs owned	0.0404	0.0911	0.0736	-0.0360
	(0.20)	(0.22)	(0.22)	(0.24)
Number of crowbars owned	-0.0719	-0.0348	-0.0532	
	(0.13)	(0.15)	(0.15)	
Number of spades owned	0.0770	0.0634	0.0718	0.114
	(0.093)	(0.11)	(0.10)	(0.12)
Number of sickles owned	0.260***	0.252***	0.248***	0.151**
	(0.065)	(0.068)	(0.069)	(0.075)
Number of cows owned	-0.134***	-0.165***	-0.166***	-0.250***
	(0.046)	(0.060)	(0.062)	(0.063)
Number of bullocks owned	0.238**	0.171	0.176	0.248*
	(0.10)	(0.12)	(0.12)	(0.13)

Table 3.11 (Continued)

Number of goats owned	-0.0349	-0.0397	-0.0416	-0.0382
	(0.026)	(0.029)	(0.029)	(0.029)
Number of buffaloes owned	-0.372***	-0.449***	-0.438***	-0.410**
	(0.13)	(0.16)	(0.17)	(0.17)
Mother's characteristics				
Mother's education (years)		-0.00908	-0.00603	0.0683
		(0.079)	(0.079)	(0.073)
Mother's height (cm)		0.00526	0.00683	0.00646
		(0.013)	(0.013)	(0.014)
Mother's age (years)		-0.289***	-0.294***	-0.201*
		(0.10)	(0.10)	(0.10)
Square of mother's age		0.00454***	0.00463***	0.00323**
		(0.0016)	(0.0016)	(0.0016)
Child characteristics				
Child's age (months)			0.00327	0.00910
			(0.014)	(0.014)
Child's age, squared			-0.0000242	-0.000105
			(0.00019)	(0.00020)
Child's sex (1 = male)			0.318**	0.348**
			(0.14)	(0.15)
Community characteristics				
Distance from market (km)				0.0474***
				(0.014)
Distance from closest Agragamee field office (km)				-0.00791
				(0.048)
Distance from closest Agragamee field office (km), squared				-0.000620
				(0.0024)
Frequency of village meetings (1 = frequent/ as needed)				1.741***
				(0.20)
Constant	-2.592***	0.570	0.205	-2.409
	(0.86)	(2.40)	(2.40)	(2.58)
N	485	409	409	409
Wald χ^2	77.43	71.29	75.2	182.44
p-value	0.0000	0.0000	0.0000	0.0000
McFadden's Pseudo-R^2	0.1414	0.1637	0.1733	0.3656

Notes: Dependant variable equals 1 if child is from a participant household in grain bank village and 0 if child is from a household in a non-grain bank village. Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.12: Determinants of grain bank participation estimated for creating propensity scores to examine impact on *whz* scores (Sample 2)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Estimated coefficients			
	(1) Specification A	(2) Specification B	(3) Specification C	(4) Specification D
Household characteristics				
Social group (1=tribal)	0.134	0.207	0.187	-0.0459
	(0.28)	(0.33)	(0.33)	(0.27)
Household size	0.613*	0.767*	0.728*	0.509
	(0.37)	(0.43)	(0.44)	(0.42)
Square of household size	-0.0761***	-0.0954***	-0.0930***	-0.0605*
	(0.028)	(0.034)	(0.034)	(0.034)
Number of children (14 years or less)	0.142	0.237	0.260	0.0952
	(0.16)	(0.18)	(0.18)	(0.18)
Number of adult females (15 years or more)	0.0233	0.153	0.182	0.142
	(0.22)	(0.24)	(0.24)	(0.26)
Low-lying fertile (bila) land owned (acres)	0.376**	0.519***	0.472***	0.764***
	(0.15)	(0.17)	(0.18)	(0.20)
Low-lying fertile land, squared	-0.0679**	-0.0882***	-0.0785**	-0.132***
	(0.027)	(0.032)	(0.033)	(0.049)
Semi-fertile (goda) land owned (acres) – cubic term	-0.00860***	-0.00740***	-0.00695***	-0.00900**
	(0.0026)	(0.0022)	(0.0021)	(0.0038)
Infertile (dongar) upland owned (acres)	-0.109	-0.207	-0.218	0.0588
	(0.14)	(0.15)	(0.15)	(0.19)
Infertile (dongar) upland, squared	-0.0241	-0.00900	-0.00509	-0.0568
	(0.028)	(0.029)	(0.027)	(0.038)
Gold holdings (gm)	0.0183	0.0170	0.0145	0.0467
	(0.022)	(0.025)	(0.025)	(0.028)
Floor quality (1 = pucca)	0.293	0.240	0.314	-0.00876
	(0.28)	(0.34)	(0.33)	(0.43)
Number of ploughs owned	-0.0748	-0.00124	-0.0316	
	(0.23)	(0.27)	(0.27)	
Number of crowbars owned	-0.303*	-0.325*	-0.351*	-0.0958
	(0.16)	(0.18)	(0.18)	(0.18)
Number of spades owned	0.125	0.167	0.178	0.253*
	(0.11)	(0.13)	(0.13)	(0.13)
Number of sickles owned	0.322***	0.319***	0.313***	0.163**
	(0.081)	(0.082)	(0.082)	(0.082)
Number of cows owned	-0.170***	-0.217***	-0.224***	-0.222***
	(0.051)	(0.072)	(0.075)	(0.074)
Number of bullocks owned	0.380***	0.325**	0.343**	
	(0.12)	(0.14)	(0.15)	

Table 3.12 (Continued)

Number of goats owned	-0.0387	-0.0529*	-0.0543*	-0.0375
	(0.027)	(0.028)	(0.028)	(0.027)
Number of buffaloes owned	-0.380***	-0.481**	-0.463**	-0.559***
	(0.14)	(0.20)	(0.20)	(0.21)
Mother's characteristics				
Mother's education (years)		-0.0281	-0.0251	0.0943
		(0.084)	(0.082)	(0.077)
Mother's height (cm)		-0.000255	0.00162	0.0120
		(0.015)	(0.015)	(0.016)
Mother's age (years)		-0.378***	-0.399***	-0.254**
		(0.12)	(0.12)	(0.12)
Square of mother's age		0.00589***	0.00623***	0.00394**
		(0.0019)	(0.0019)	(0.0019)
Child characteristics				
Child's age (months)			0.00759	0.0190
			(0.015)	(0.015)
Child's age, squared			-0.0000479	-0.000232
			(0.00021)	(0.00022)
Child's sex (1 = male)			0.345**	0.350**
			(0.16)	(0.16)
Community characteristics				
Distance from market (km)				0.0152
				(0.015)
Distance from closest Agragamee field office (km)				0.533***
				(0.19)
Distance from closest Agragamee field office (km), squared				-0.0612***
				(0.020)
Frequency of village meetings (1 = frequent/ as needed)				1.376***
				(0.21)
Constant	-2.699***	2.314	2.065	-2.722
	(0.97)	(2.74)	(2.77)	(3.04)
N	430	363	363	363
Wald χ^2	85.95	91.95	94.01	190.73
p-value	0.0000	0.0000	0.0000	0.0000
McFadden's Pseudo-R^2	0.1916	0.2412	0.2533	0.3926

Notes: Dependant variable equals 1 if child is from a participant household in grain bank village and 0 if child is from a household in a non-grain bank village. Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.13: Determinants of grain bank participation estimated for creating propensity scores to examine impact on growth (Sample 1)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Estimated coefficients			
	(1) Specification A	(2) Specification B	(3) Specification C	(4) Specification D
Household characteristics				
Social group (1=tribal)	0.438	0.529	0.524	0.0284
	(0.36)	(0.50)	(0.48)	(0.47)
Household size	0.258	0.461	0.330	0.295
	(0.40)	(0.47)	(0.47)	(0.48)
Square of household size	-0.0548*	-0.0845**	-0.0743**	-0.0824**
	(0.030)	(0.037)	(0.037)	(0.041)
Number of children (14 years or less)	0.322*	0.529**	0.550**	0.628**
	(0.19)	(0.22)	(0.22)	(0.25)
Number of adult females (15 years or more)	0.0521	0.423	0.437	0.812**
	(0.26)	(0.30)	(0.30)	(0.33)
Low-lying fertile (bila) land owned (acres)	0.173	0.485**	0.401*	0.905***
	(0.17)	(0.21)	(0.21)	(0.24)
Low-lying fertile land, squared	-0.0478	-0.0892**	-0.0705*	-0.138**
	(0.031)	(0.036)	(0.037)	(0.054)
Semi-fertile (goda) land owned (acres)	0.208*	0.168	0.189	0.253*
	(0.12)	(0.13)	(0.13)	(0.15)
Semi-fertile (goda) land owned, squared	-0.0448**	-0.0449**	-0.0444**	-0.0561***
	(0.018)	(0.018)	(0.018)	(0.020)
Infertile (dongar) upland owned (acres)	-0.0649	0.0137	0.0394	0.416
	(0.17)	(0.22)	(0.20)	(0.28)
Infertile (dongar) upland, squared	-0.0467	-0.0606	-0.0564	-0.125*
	(0.041)	(0.050)	(0.043)	(0.064)
Gold holdings (gm)	0.0165	0.0295	0.0239	0.0269
	(0.026)	(0.028)	(0.028)	(0.032)
Floor quality (1 = pucca)	0.968***	1.449***	1.580***	1.546**
	(0.32)	(0.46)	(0.48)	(0.64)
Number of ploughs owned	0.486***	0.374**	0.368**	
	(0.16)	(0.18)	(0.18)	
Number of crowbars owned	0.0759	0.220	0.148	0.294
	(0.17)	(0.21)	(0.21)	(0.19)
Number of spades owned	-0.0289	-0.207	-0.179	-0.203
	(0.12)	(0.13)	(0.13)	(0.15)
Number of sickles owned	0.370***	0.378***	0.398***	0.350***
	(0.083)	(0.086)	(0.087)	(0.098)
Number of cows owned	-0.0927	-0.162*	-0.170**	-0.228**
	(0.064)	(0.083)	(0.085)	(0.093)

Table 3.13 (Continued)

Number of goats owned	-0.134**	-0.172**	-0.188**	-0.172*
	(0.064)	(0.085)	(0.086)	(0.096)
Number of buffaloes owned	-0.566**	-1.096***	-1.130***	-1.088***
	(0.23)	(0.27)	(0.25)	(0.26)
Mother's characteristics				
Mother's education (years)		-0.0857	-0.0399	0.00254
		(0.097)	(0.096)	(0.085)
Mother's height (cm)		0.0287*	0.0305*	0.0454***
		(0.017)	(0.016)	(0.017)
Mother's age (years)		-0.431***	-0.373***	-0.389***
		(0.14)	(0.14)	(0.15)
Square of mother's age		0.00670***	0.00584***	0.00602***
		(0.0021)	(0.0022)	(0.0023)
Child characteristics				
Child's age (months)			0.0195	0.0163
			(0.022)	(0.024)
Child's age, squared			-0.000292	-0.000261
			(0.00027)	(0.00030)
Child's sex (1 = male)			0.483**	0.408*
			(0.19)	(0.22)
Community characteristics				
Distance from market (km)				0.0715***
				(0.020)
Distance from closest Agragamee field office (km)				0.0177
				(0.070)
Distance from closest Agragamee field office (km), squared				-0.00213
				(0.0037)
Frequency of village meetings (1 = frequent/ as needed)				1.882***
				(0.30)
Constant	-2.537**	-1.416	-2.834	-6.619**
	(1.05)	(2.98)	(2.97)	(3.27)
N	332	279	279	279
Wald χ^2	78.71	74.78	86.9	120.28
p-value	0.0000	0.0000	0.0000	0.0000
McFadden's Pseudo-R^2	0.2018	0.266	0.2891	0.4532

Notes: Dependant variable equals 1 if child is from a participant household in grain bank village and 0 if child is from a household in a non-grain bank village. Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.14: Determinants of grain bank participation estimated for creating propensity scores to examine impact on growth (Sample 2)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Estimated coefficients			
	(1) Specification A	(2) Specification B	(3) Specification C	(4) Specification D
Household characteristics				
Social group (1=tribal)	0.198	0.224	0.216	-0.0247
	(0.37)	(0.52)	(0.50)	(0.53)
Household size	0.378	0.802	0.709	0.453
	(0.43)	(0.51)	(0.52)	(0.54)
Square of household size	-0.0693**	-0.122***	-0.113***	-0.0858*
	(0.034)	(0.042)	(0.042)	(0.049)
Number of children (14 years or less)	0.352*	0.618**	0.621**	0.501*
	(0.21)	(0.25)	(0.25)	(0.26)
Number of adult females (15 years or more)	0.0662	0.440	0.444	0.577
	(0.29)	(0.34)	(0.33)	(0.35)
Low-lying fertile (bila) land owned (acres)	0.268	0.648***	0.577**	1.012***
	(0.18)	(0.24)	(0.25)	(0.29)
Low-lying fertile land, squared	-0.0535	-0.108**	-0.0920**	-0.163**
	(0.034)	(0.043)	(0.044)	(0.064)
Semi-fertile (goda) land owned (acres)	0.261	0.276	0.319	0.345
	(0.20)	(0.23)	(0.25)	(0.23)
Semi-fertile (goda) land owned, squared	-0.0698	-0.0812*	-0.0868	-0.0932*
	(0.044)	(0.049)	(0.056)	(0.049)
Infertile (dongar) upland owned (acres)	-0.0257	-0.0000605	0.00864	0.355
	(0.21)	(0.25)	(0.23)	(0.29)
Infertile (dongar) upland, squared	-0.0715	-0.0711	-0.0579	-0.102
	(0.056)	(0.059)	(0.053)	(0.064)
Gold holdings (gm)	0.0131	0.0303	0.0254	0.0418
	(0.028)	(0.030)	(0.030)	(0.036)
Floor quality (1 = pucca)	0.978***	1.479***	1.584***	
	(0.33)	(0.44)	(0.46)	
Number of ploughs owned	0.617***	0.554***	0.543**	
	(0.19)	(0.21)	(0.21)	
Number of crowbars owned	-0.0993	0.100	0.0609	0.185
	(0.20)	(0.22)	(0.23)	(0.22)
Number of spades owned	0.0655	-0.0829	-0.0704	0.109
	(0.13)	(0.15)	(0.14)	(0.16)
Number of sickles owned	0.415***	0.445***	0.468***	0.309***
	(0.094)	(0.093)	(0.095)	(0.11)
Number of cows owned	-0.128*	-0.243***	-0.249***	-0.274***
	(0.070)	(0.092)	(0.093)	(0.10)

Table 3.14 (Continued)

Number of goats owned	-0.160**	-0.232**	-0.264***	-0.249**
	(0.071)	(0.10)	(0.099)	(0.10)
Number of buffaloes owned	-0.547**	-1.117***	-1.153***	-1.116***
	(0.24)	(0.26)	(0.25)	(0.27)
Mother's characteristics				
Mother's education (years)		-0.0763	-0.0258	0.177*
		(0.10)	(0.10)	(0.100)
Mother's height (cm)		0.0255	0.0286	0.0539***
		(0.019)	(0.019)	(0.019)
Mother's age (years)		-0.496***	-0.441***	-0.267
		(0.16)	(0.16)	(0.17)
Square of mother's age		0.00776***	0.00699***	0.00402
		(0.0025)	(0.0025)	(0.0027)
Child characteristics				
Child's age (months)			0.0188	0.0250
			(0.025)	(0.027)
Child's age, squared			-0.000297	-0.000397
			(0.00031)	(0.00033)
Child's sex (1 = male)			0.490**	0.441*
			(0.22)	(0.23)
Community characteristics				
Distance from market (km)				0.0573***
				(0.018)
Distance from closest Agragamee field office (km)				0.397*
				(0.22)
Distance from closest Agragamee field office (km), squared				-0.0491**
				(0.023)
Frequency of village meetings (1 = frequent/ as needed)				1.652***
				(0.31)
Constant	-2.952***	-1.263	-2.922	-10.48***
	(1.14)	(3.37)	(3.44)	(3.73)
N	297	251	251	251
Wald χ^2	71.75	79.31	87.85	131.45
p-value	0.0000	0.0000	0.0000	0.0000
McFadden's Pseudo-R^2	0.2357	0.3271	0.3499	0.4692

Notes: Dependant variable equals 1 if child is from a participant household in grain bank village and 0 if child is from a household in a non-grain bank village. Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

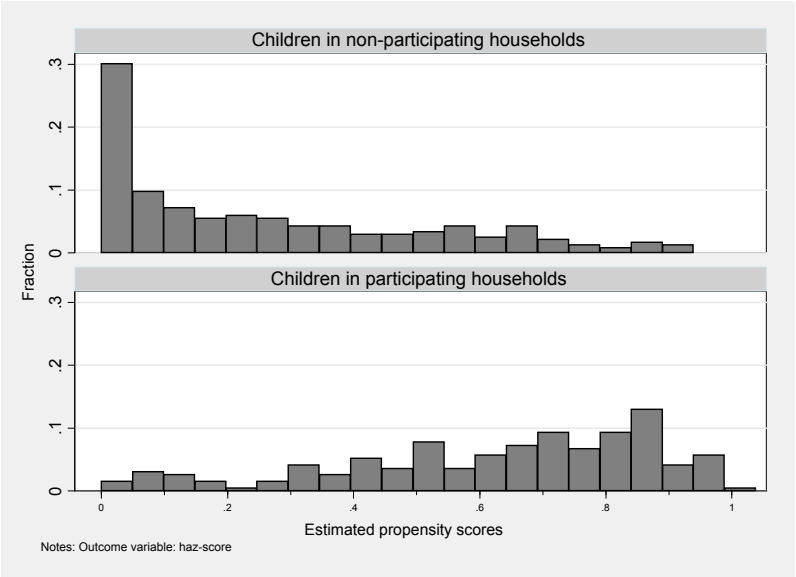


Figure 3.1: Distribution of predicted propensity scores (Sample 1)

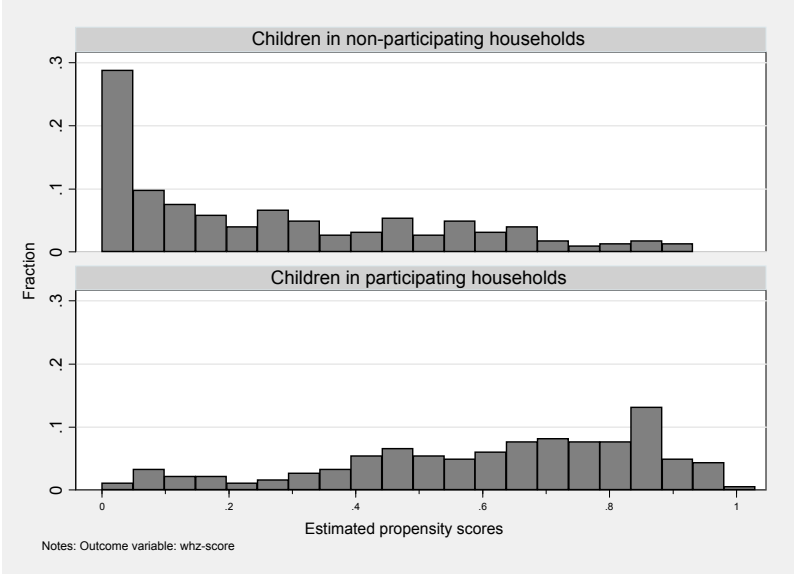


Figure 3.2: Distribution of predicted propensity scores (Sample 1)

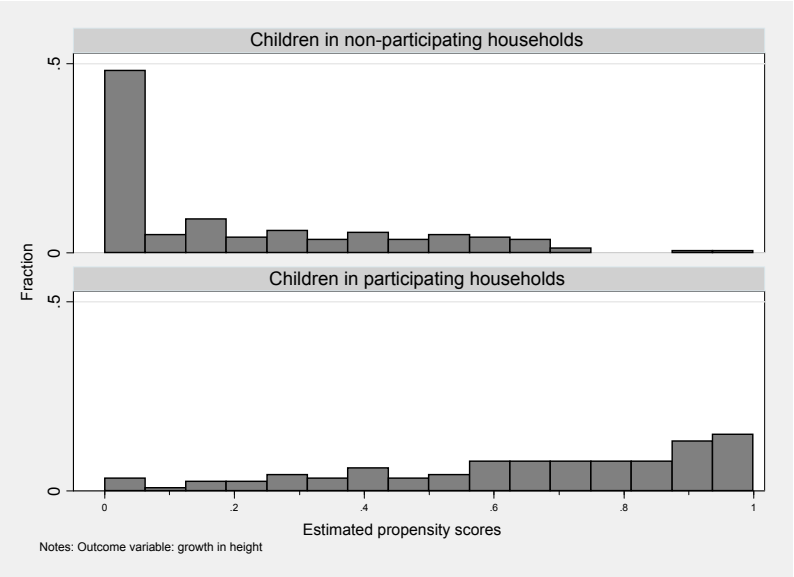


Figure 3.3: Distribution of predicted propensity scores (Sample 1)

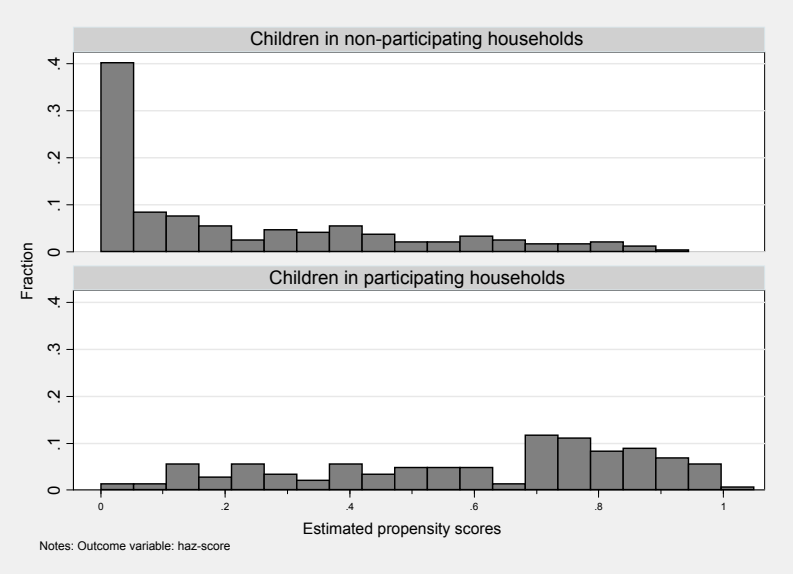


Figure 3.4: Distribution of predicted propensity scores (Sample 2)

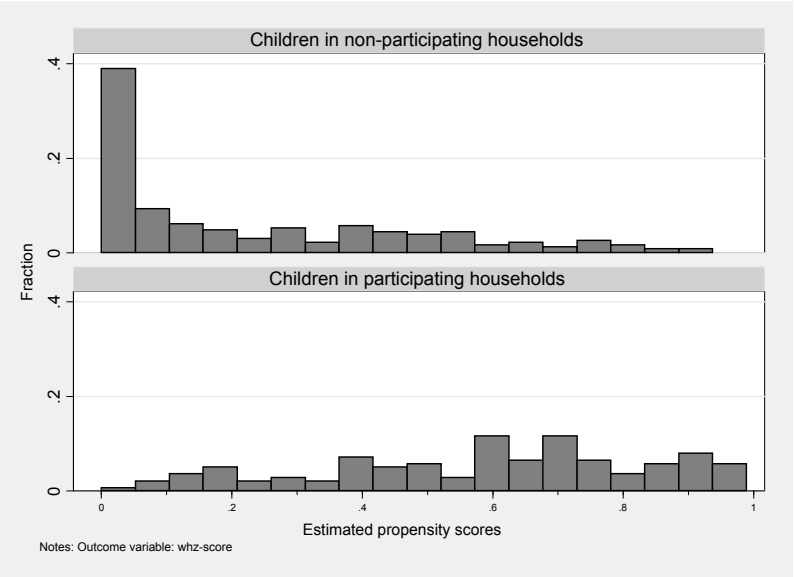


Figure 3.5: Distribution of predicted propensity scores (Sample 2)

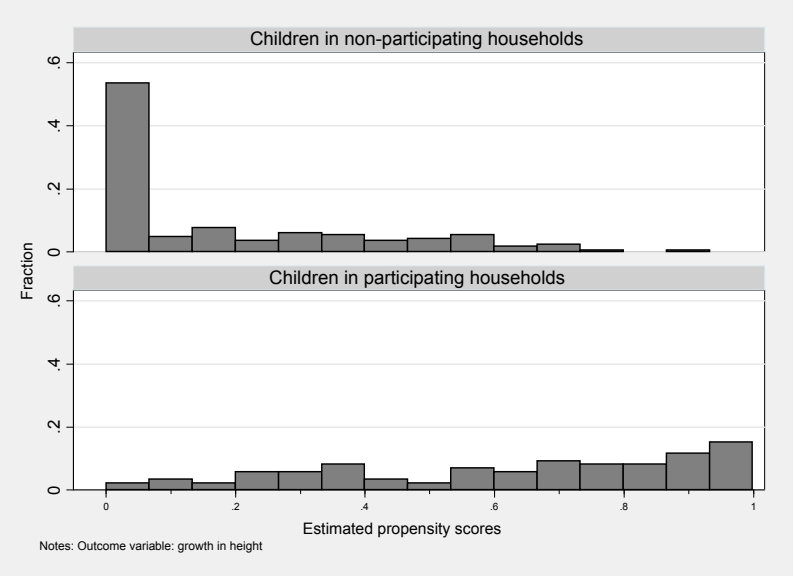


Figure 3.6: Distribution of predicted propensity scores (Sample 2)

The balancing property of the different specifications is examined using three difference tests: (1) *t*-tests for difference in covariate means between the matched treatment and comparison samples; (2) standardized bias before and after matching; and (3) pseudo-*R*-squared of the propensity score model *after* matching, using observations in the common support region only as well as weights generated from the matching algorithm.

Results of balanced covariates using *t*-tests are presented in Tables 3.A1-3.A12 in the appendix. Examining the *t*-test results, we only include those variables in the final specification that have no statistically significant difference in means.

The results using measures of pseudo *R*-squared and standardized bias are presented in Tables 3.A13-3.A15 in the appendix. Examining the pseudo *R*-squared by re-estimating the propensity score model after matching, we find that in all cases, the pseudo *R*-squared generated is much lower than the pseudo *R*-squared generated prior to matching. In addition, the joint significance of the covariates in the model is always rejected. Prior to matching, the joint significance of the covariates in the model was never rejected. Finally, examining the median standardized bias before and after matching, we find that that it is always lower after matching, and never above a value of 8, which is an acceptable value (Smith and Todd 2005b).

Average impact of grain bank participation

Table 3.15 presents local linear regression matching estimates of the average impact of participation in grain banks on individual children's *haz*-scores. Column (1) presents the matching results for Sample 1, where the treatment sample is composed of children from participant households from all grain bank villages. Column (2) presents the matching results for Sample 2, where the treatment sample is confined to children from participant households from long-lived grain bank villages. In both

cases, we find that, on average, children in participant households have a slightly higher *haz*-score. However, this difference is not statistically significant.

Table 3.15: Average treatment on the treated (ATT): Impact of grain bank participation on children's *haz*-scores
Local linear regression matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	-1.963		-1.930	
Average outcome, non-participants	-2.050		-1.998	
Difference in average outcomes (ATT)	0.087		0.0968	
	(0.228)		(0.232)	
	Off support	On support	Off support	On support
No. of treated units	16	176	15	130
No. of comparison units	0	236	0	236

Notes: Standard errors are provided in parentheses. Estimates generated using bandwidth size=0.06 for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.16 presents local linear regression matching estimates of the average impact of participation in grain banks on individual children's *whz*-scores. These estimates reflect the pattern observed from the naïve estimates of wasting in the agricultural lean season. Estimates for both Sample 1 and Sample 2 indicate that children in participant households have a slightly lower *whz*-score, although this result is also not statistically significant. The inconsistency in the signs of the ATT estimates for *haz*-scores and *whz*-scores is puzzling, however.

Finally, in Table 3.17, examining the impact on the change in height between the post-harvest and agricultural lean season, we see that, on average, children in participant households have a slightly faster rate of growth than children in the matched comparison sample. However, the estimates for neither Sample 1 not Sample 2 are statistically significant.

Table 3.16: Average treatment on the treated (ATT): Impact of grain bank participation on children's *whz*-scores
Local linear regression matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	-1.712		-1.710	
Average outcome, non-participants	-1.577		-1.612	
Difference in average outcomes (ATT)	-0.135		-0.099	
	(0.189)		(0.188)	
	Off support	On support	Off support	On support
No. of treated units	11	172	14	123
No. of comparison units	0	226	0	226

Notes: Standard errors are provided in parentheses. Estimates generated using bandwidth size=0.06 for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.17: Average treatment on the treated (ATT): Impact of grain bank participation on growth
Local linear regression matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	0.052		0.053	
Average outcome, non-participants	0.045		0.043	
Difference in average outcomes (ATT)	0.007		0.010	
	(0.009)		(0.010)	
	Off support	On support	Off support	On support
No. of treated units	39	74	33	52
No. of comparison units	0	166	0	166

Notes: Standard errors are provided in parentheses. Estimates generated using bandwidth size=0.06 for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Robustness to alternative specifications

Estimates of the average impact of participation in grain banks on children's *haz*-score, *whz*-score and growth using kernel matching estimation are presented in Tables 3.18-3.20 respectively. We find that these results are similar to those using local linear matching, indicating that the estimates are robust across the two specifications.

Table 3.18: Average treatment on the treated (ATT): Impact of grain bank participation on children's *haz*-scores
Kernel matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	-1.963		-1.930	
Average outcome, non-participants	-2.049		-2.034	
Difference in average outcomes (ATT)	0.086		0.104	
	(0.220)		(0.226)	
	Off support	On support	Off support	On support
No. of treated units	16	176	15	130
No. of comparison units	0	236	0	236

Notes: Standard errors are provided in parentheses. Estimates generated using bandwidth size=0.06 for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.19: Average treatment on the treated (ATT): Impact of grain bank participation on children's *whz*-scores
Kernel matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	-1.712		-1.710	
Average outcome, non-participants	-1.582		-1.606	
Difference in average outcomes (ATT)	-0.130		-0.104	
	(0.188)		(0.185)	
	Off support	On support	Off support	On support
No. of treated units	11	172	11	172
No. of comparison units	0	226	0	226

Notes: Standard errors are provided in parentheses. Estimates generated using bandwidth size=0.06 for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Sensitivity to choice of bandwidth

The results in Tables 3.15-3.20 were estimated using a bandwidth size of 0.06. We re-estimate the local linear regression and kernel matching models with varying bandwidth size (0.05 and 0.07). The results are presented in Tables 3.A16-3.A18 in the appendix. We find that there is some fluctuation in the magnitude of the estimates.

However, the results continue to be statistically insignificant, regardless of the bandwidth size.

Table 3.20: Average treatment on the treated (ATT): Impact of grain bank participation on growth
Kernel matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	0.051		0.049	
Average outcome, non-participants	0.040		0.040	
Difference in average outcomes (ATT)	0.011		0.008	
	(0.012)		(0.011)	
	Off support	On support	Off support	On support
No. of treated units	30	83	23	62
No. of comparison units	0	166	0	166

Notes: Standard errors are provided in parentheses. Estimates generated using bandwidth size=0.06 for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

3.7 Conclusion

In the past two decades, grain banks have been enthusiastically adopted by tribal development NGOs promoting food security in rural Orissa. However, to date, no quantitative analysis of the impact of grain banks on food security outcomes exist. In this chapter, we attempt to fill this gap in knowledge by measuring the impact on individual children's short- and long-term health outcomes, using local linear regression and kernel propensity score matching. Specifically, we examine three outcomes, namely, children's *haz*-scores and *whz*-scores in the agricultural lean season as well as the change in height between the post-harvest and lean seasons.

Our matching estimates indicate that while the average *haz*-score is slightly higher for children in participant households, the estimate is statistically insignificant. We obtain a similar result when we use matching methods to examine children's growth in height. In addition, examining the average *whz*-scores for children in

participant households and children in non-grain bank villages, we find that the matching-based estimate is slightly lower for the former. While the inconsistency in the signs of the ATT estimates for *haz*-scores and *whz*-scores is puzzling, neither set of estimates is statistically significant. We also do not obtain any statistically significant results when we compare children in participant households in villages where grain banks have survived for a longer duration compared to children in non-grain bank villages.

Given that a large number of our observations have missing height or weight information and that the age data are of somewhat poor quality, we are hesitant about making any firm conclusions based on our analysis. On the one hand, height or weight data do not appear to be missing systematically across treatment and comparison samples and the age data may be no poorer than similar data from comparable settings. In this case, our estimates may well be valid and reflect the fact that grain banks do not, indeed, have a quantifiable impact on children's health outcomes. However, it is also quite possible that the average treatment effect on the treated is heterogeneous across the treated units. Thus, our results, which are based on the entire distribution of households, may simply be disguising the fact that the average treatment effect varies over the distribution. Unfortunately, we are not able to estimate the treatment effect at different points or sections of the distribution due to sample size limitations. Alternately, parents may be acting as a buffer against volatility in consumption of their children in non-grain bank villages, thereby masking any effect that grain banks may have in reducing volatility in consumption and health outcomes of children.

Given that the Indian government proposes to expand the grain bank initiative across tribal villages in different Indian states, we suggest three projects for future research which we hope can provide more conclusive information. First,

implementing a baseline survey before establishing grain banks in target villages will permit difference-in-difference propensity score matching estimation of program impact. This methodology is superior to cross-sectional analysis as it can help to correct for selection bias based on time-invariant unobservable factors. In addition, data on a sufficiently large number of observations will permit estimation of the average treatment effect for different points or sections of the distribution. Second, survey data collection from target villages as part of a phased implementation of grain banks can act as a source of experimental data which will provide the most reliable estimates of program impact. Third, data which enables the researcher to compare the cost-effectiveness and impact of grain banks compared to other food security interventions can provide critical information to a policymaker on how best to allocate scarce resources.

APPENDIX

Table 3.A1: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on local linear regression propensity score matching

Outcome variable: Impact on children's *haz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.97	0.99	-0.02	-1.2
Household size	5.50	5.48	0.02	0.11
Square of household size	32.22	32.35	-0.13	-0.07
Number of children (14 years or less)	2.94	2.89	0.05	0.41
Number of adult females (15 years or more)	1.34	1.33	0.01	0.1
Low-lying fertile (bila) land owned (acres)	0.62	0.65	-0.03	-0.31
Low-lying fertile land, squared	1.25	1.08	0.17	0.53
Semi-fertile (goda) land owned (acres) – cubic term	14.16	21.86	-7.70	-0.9
Infertile (dongar) upland owned (acres)	1.45	1.52	-0.07	-0.56
Infertile (dongar) upland, squared	3.38	3.76	-0.38	-0.67
Gold holdings (gm)	2.75	2.90	-0.15	-0.37
Floor quality (1 = pucca)	0.08	0.10	-0.02	-0.52
Number of ploughs owned	0.82	0.80	0.01	0.26
Number of crowbars owned	1.18	1.10	0.08	1.52
Number of spades owned	2.15	2.11	0.04	0.53
Number of sickles owned	2.87	2.92	-0.05	-0.39
Number of cows owned	0.79	0.63	0.16	1.39
Number of bullocks owned	1.45	1.53	-0.07	-0.61
Number of goats owned	0.84	0.90	-0.06	-0.23
Number of buffaloes owned	0.07	0.06	0.02	0.48
Mother's education (years)	0.13	0.12	0.01	0.19
Mother's height (cm)	148.71	148.03	0.68	1.11
Mother's age (years)	28.36	27.49	0.88	1.39
Square of mother's age	842.48	786.96	55.52	1.45
Child's age (months)	35.84	34.13	1.70	0.74
Child's age, squared	1722.90	1641.50	81.40	0.5
Child's sex (1 = male)	0.56	0.55	0.00	0.08
Distance from market (km)	8.66	8.67	-0.01	-0.01
Distance from closest Agragamee field office (km)	4.78	4.46	0.32	1.14
Distance from closest Agragamee field office (km), squared	31.17	25.52	5.65	1.43
Frequency of village meetings (1 = frequent/ as needed)	0.89	0.87	0.02	0.71

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A2: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on kernel propensity score matching

Outcome variable: Impact on children's *haz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.97	0.99	-0.02	-1.19
Household size	5.50	5.51	-0.01	-0.09
Square of household size	32.22	32.69	-0.48	-0.26
Number of children (14 years or less)	2.94	2.90	0.04	0.29
Number of adult females (15 years or more)	1.34	1.35	-0.01	-0.14
Low-lying fertile (bila) land owned (acres)	0.62	0.63	-0.02	-0.18
Low-lying fertile land, squared	1.25	1.06	0.19	0.61
Semi-fertile (goda) land owned (acres) – cubic term	14.16	21.54	-7.38	-0.86
Infertile (dongar) upland owned (acres)	1.45	1.47	-0.02	-0.18
Infertile (dongar) upland, squared	3.38	3.54	-0.16	-0.29
Gold holdings (gm)	2.75	2.86	-0.11	-0.27
Floor quality (1 = pucca)	0.08	0.11	-0.03	-0.84
Number of ploughs owned	0.82	0.80	0.02	0.40
Number of crowbars owned	1.18	1.11	0.07	1.33
Number of spades owned	2.15	2.11	0.04	0.54
Number of sickles owned	2.87	2.92	-0.05	-0.39
Number of cows owned	0.79	0.64	0.15	1.30
Number of bullocks owned	1.45	1.50	-0.05	-0.39
Number of goats owned	0.84	0.89	-0.05	-0.18
Number of buffaloes owned	0.07	0.06	0.02	0.45
Mother's education (years)	0.13	0.12	0.01	0.14
Mother's height (cm)	148.71	148.08	0.63	1.03
Mother's age (years)	28.36	27.53	0.83	1.31
Square of mother's age	842.48	790.75	51.73	1.34
Child's age (months)	35.84	34.23	1.61	0.70
Child's age, squared	1722.90	1657.90	65.00	0.40
Child's sex (1 = male)	0.56	0.53	0.03	0.55
Distance from market (km)	8.66	9.00	-0.34	-0.44
Distance from closest Agragamee field office (km)	4.78	4.43	0.36	1.26
Distance from closest Agragamee field office (km), squared	31.17	25.38	5.79	1.42
Frequency of village meetings (1 = frequent/ as needed)	0.89	0.86	0.03	0.82

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A3: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on local linear regression propensity score matching

Outcome variable: Impact on children's *haz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.97	0.99	-0.02	-1.2
Household size	5.50	5.48	0.02	0.11
Square of household size	32.22	32.35	-0.13	-0.07
Number of children (14 years or less)	2.94	2.89	0.05	0.41
Number of adult females (15 years or more)	1.34	1.33	0.01	0.1
Low-lying fertile (bila) land owned (acres)	0.62	0.65	-0.03	-0.31
Low-lying fertile land, squared	1.25	1.08	0.17	0.53
Semi-fertile (goda) land owned (acres) – cubic term	14.16	21.86	-7.70	-0.9
Infertile (dongar) upland owned (acres)	1.45	1.52	-0.07	-0.56
Infertile (dongar) upland, squared	3.38	3.76	-0.38	-0.67
Gold holdings (gm)	2.75	2.90	-0.15	-0.37
Floor quality (1 = pucca)	0.08	0.10	-0.02	-0.52
Number of ploughs owned	0.82	0.80	0.01	0.26
Number of crowbars owned	1.18	1.10	0.08	1.52
Number of spades owned	2.15	2.11	0.04	0.53
Number of sickles owned	2.87	2.92	-0.05	-0.39
Number of cows owned	0.79	0.63	0.16	1.39
Number of bullocks owned	1.45	1.53	-0.07	-0.61
Number of goats owned	0.84	0.90	-0.06	-0.23
Number of buffaloes owned	0.07	0.06	0.02	0.48
Mother's education (years)	0.13	0.12	0.01	0.19
Mother's height (cm)	148.71	148.03	0.68	1.11
Mother's age (years)	28.36	27.49	0.88	1.39
Square of mother's age	842.48	786.96	55.52	1.45
Child's age (months)	35.84	34.13	1.70	0.74
Child's age, squared	1722.90	1641.50	81.40	0.5
Child's sex (1 = male)	0.56	0.55	0.00	0.08
Distance from market (km)	8.66	8.67	-0.01	-0.01
Distance from closest Agragamee field office (km)	4.78	4.46	0.32	1.14
Distance from closest Agragamee field office (km), squared	31.17	25.52	5.65	1.43
Frequency of village meetings (1 = frequent/ as needed)	0.89	0.87	0.02	0.71

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A4: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on kernel propensity score matching

Outcome variable: Impact on children's *haz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.95	0.98	-0.02	-0.93
Household size	5.44	5.31	0.13	0.71
Square of household size	31.56	30.39	1.17	0.57
Number of children (14 years or less)	2.82	2.72	0.10	0.69
Number of adult females (15 years or more)	1.36	1.33	0.03	0.46
Low-lying fertile (bila) land owned (acres)	0.70	0.69	0.01	0.07
Low-lying fertile land, squared	1.53	1.26	0.28	0.62
Semi-fertile (goda) land owned (acres) – cubic term	12.39	15.92	-3.53	-0.50
Infertile (dongar) upland owned (acres)	1.33	1.37	-0.04	-0.29
Infertile (dongar) upland, squared	2.83	3.11	-0.28	-0.54
Gold holdings (gm)	2.93	2.96	-0.03	-0.06
Floor quality (1 = pucca)	0.09	0.14	-0.05	-1.15
Number of ploughs owned	0.89	0.75	0.14	1.94
Number of crowbars owned	1.12	1.12	0.01	0.13
Number of spades owned	2.20	2.13	0.07	0.76
Number of sickles owned	2.96	2.95	0.01	0.09
Number of cows owned	0.79	0.62	0.17	1.21
Number of goats owned	0.89	0.65	0.24	0.72
Number of buffaloes owned	0.06	0.07	-0.01	-0.19
Mother's education (years)	0.15	0.11	0.04	0.41
Mother's height (cm)	148.28	148.31	-0.03	-0.05
Mother's age (years)	27.63	26.99	0.64	0.84
Square of mother's age	802.11	763.97	38.14	0.84
Child's age (months)	36.66	35.54	1.12	0.41
Child's age, squared	1793.50	1750.30	43.20	0.22
Child's sex (1 = male)	0.58	0.54	0.03	0.56
Distance from market (km)	8.65	8.96	-0.31	-0.34
Distance from closest Agragamee field office (km)	4.48	4.22	0.26	1.01
Distance from closest Agragamee field office (km), squared	24.42	21.88	2.54	0.99
Frequency of village meetings (1 = frequent/ as needed)	0.85	0.83	0.02	0.42

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A5: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on local linear regression propensity score matching

Outcome variable: Impact on children's *whz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.97	0.98	-0.02	-1.14
Household size	5.48	5.47	0.01	0.06
Square of household size	32.03	32.16	-0.13	-0.07
Number of children (14 years or less)	2.94	2.89	0.05	0.39
Number of adult females (15 years or more)	1.33	1.34	-0.01	-0.09
Low-lying fertile (bila) land owned (acres)	0.63	0.63	-0.01	-0.09
Low-lying fertile land, squared	1.27	1.09	0.18	0.55
Semi-fertile (goda) land owned (acres) – cubic term	12.12	19.91	-7.79	-0.99
Infertile (dongar) upland owned (acres)	1.42	1.47	-0.05	-0.39
Infertile (dongar) upland, squared	3.32	3.56	-0.24	-0.43
Gold holdings (gm)	2.57	2.38	0.19	0.55
Floor quality (1 = pucca)	0.08	0.11	-0.03	-0.9
Number of ploughs owned	0.81	0.80	0.01	0.18
Number of spades owned	2.15	2.11	0.04	0.49
Number of sickles owned	2.88	2.91	-0.04	-0.28
Number of cows owned	0.78	0.64	0.14	1.16
Number of bullocks owned	1.45	1.47	-0.01	-0.11
Number of goats owned	0.80	0.93	-0.14	-0.51
Number of buffaloes owned	0.09	0.07	0.01	0.29
Mother's education (years)	0.16	0.15	0.01	0.11
Mother's height (cm)	148.81	148.18	0.63	1.04
Mother's age (years)	28.40	27.46	0.95	1.48
Square of mother's age	845.53	784.70	60.83	1.57
Child's age (months)	36.36	33.50	2.86	1.22
Child's age, squared	1765.20	1619.30	145.90	0.89
Child's sex (1 = male)	0.57	0.53	0.04	0.66
Distance from market (km)	8.54	9.41	-0.87	-1.1
Distance from closest Agragamee field office (km)	4.74	4.32	0.42	1.47
Distance from closest Agragamee field office (km), squared	30.59	24.37	6.22	1.53
Frequency of village meetings (1 = frequent/ as needed)	0.89	0.88	0.01	0.35

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A6: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on kernel propensity score matching

Outcome variable: Impact on children's *whz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.97	0.98	-0.02	-1.15
Household size	5.48	5.53	-0.05	-0.32
Square of household size	32.03	32.90	-0.87	-0.47
Number of children (14 years or less)	2.94	2.92	0.02	0.14
Number of adult females (15 years or more)	1.33	1.35	-0.02	-0.28
Low-lying fertile (bila) land owned (acres)	0.63	0.65	-0.02	-0.22
Low-lying fertile land, squared	1.27	1.10	0.17	0.53
Semi-fertile (goda) land owned (acres) – cubic term	12.12	19.50	-7.38	-1.02
Infertile (dongar) upland owned (acres)	1.42	1.47	-0.05	-0.41
Infertile (dongar) upland, squared	3.32	3.53	-0.21	-0.37
Gold holdings (gm)	2.57	2.43	0.14	0.40
Floor quality (1 = pucca)	0.08	0.12	-0.03	-1.05
Number of ploughs owned	0.81	0.79	0.02	0.33
Number of crowbars owned	1.18	1.11	0.07	1.39
Number of spades owned	2.15	2.11	0.04	0.55
Number of sickles owned	2.88	2.93	-0.05	-0.36
Number of cows owned	0.78	0.64	0.14	1.16
Number of bullocks owned	1.45	1.48	-0.03	-0.22
Number of goats owned	0.80	0.92	-0.12	-0.45
Number of buffaloes owned	0.09	0.07	0.02	0.38
Mother's education (years)	0.16	0.16	0.00	0.05
Mother's height (cm)	148.81	148.15	0.66	1.09
Mother's age (years)	28.40	27.60	0.81	1.24
Square of mother's age	845.53	794.23	51.30	1.31
Child's age (months)	36.36	33.77	2.59	1.11
Child's age, squared	1765.20	1632.60	132.60	0.81
Child's sex (1 = male)	0.57	0.53	0.04	0.83
Distance from market (km)	8.54	9.14	-0.60	-0.76
Distance from closest Agragamee field office (km)	4.74	4.34	0.40	1.41
Distance from closest Agragamee field office (km), squared	30.59	24.73	5.86	1.41
Frequency of village meetings (1 = frequent/ as needed)	0.89	0.88	0.01	0.35

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A7: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on local linear regression propensity score matching

Outcome variable: Impact on children's *whz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.94	0.97	-0.03	-1.22
Household size	5.39	5.42	-0.03	-0.16
Square of household size	31.07	31.43	-0.37	-0.18
Number of children (14 years or less)	2.80	2.85	-0.05	-0.32
Number of adult females (15 years or more)	1.33	1.31	0.01	0.19
Low-lying fertile (bila) land owned (acres)	0.66	0.60	0.05	0.45
Low-lying fertile land, squared	1.36	1.11	0.24	0.58
Semi-fertile (goda) land owned (acres) – cubic term	9.63	10.38	-0.75	-0.24
Infertile (dongar) upland owned (acres)	1.31	1.38	-0.06	-0.45
Infertile (dongar) upland, squared	2.83	3.09	-0.26	-0.5
Gold holdings (gm)	2.67	2.58	0.09	0.18
Floor quality (1 = pucca)	0.09	0.14	-0.05	-1.14
Number of spades owned	1.12	1.08	0.04	0.64
Number of sickles owned	2.15	2.15	0.01	0.07
Number of cows owned	2.90	2.88	0.02	0.15
Number of bullocks owned	0.72	0.60	0.12	0.9
Number of goats owned	0.86	0.72	0.14	0.41
Number of buffaloes owned	0.07	0.08	-0.02	-0.35
Mother's education (years)	0.14	0.15	-0.01	-0.12
Mother's height (cm)	148.42	148.33	0.09	0.14
Mother's age (years)	27.72	27.74	-0.01	-0.02
Square of mother's age	808.15	809.25	-1.10	-0.02
Child's age (months)	37.41	34.64	2.77	1
Child's age, squared	1847.80	1687.60	160.20	0.82
Child's sex (1 = male)	0.58	0.56	0.02	0.33
Distance from market (km)	8.83	8.89	-0.06	-0.06
Distance from closest Agragamee field office (km)	4.40	4.24	0.16	0.63
Distance from closest Agragamee field office (km), squared	23.89	21.68	2.21	0.94
Frequency of village meetings (1 = frequent/ as needed)	0.85	0.83	0.01	0.27

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A8: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on kernel propensity score matching

Outcome variable: Impact on children's *whz*-score

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.94	0.97	-0.03	-1.16
Household size	5.39	5.41	-0.02	-0.09
Square of household size	31.07	31.27	-0.21	-0.1
Number of children (14 years or less)	2.80	2.85	-0.04	-0.28
Number of adult females (15 years or more)	1.33	1.31	0.02	0.22
Low-lying fertile (bila) land owned (acres)	0.66	0.59	0.06	0.54
Low-lying fertile land, squared	1.36	1.07	0.29	0.7
Semi-fertile (goda) land owned (acres) – cubic term	9.63	10.74	-1.11	-0.27
Infertile (dongar) upland owned (acres)	1.31	1.39	-0.08	-0.58
Infertile (dongar) upland, squared	2.83	3.13	-0.30	-0.57
Gold holdings (gm)	2.67	2.68	-0.01	-0.02
Floor quality (1 = pucca)	0.09	0.12	-0.03	-0.78
Number of crowbars owned	1.12	1.08	0.04	0.67
Number of spades owned	2.15	2.14	0.01	0.12
Number of sickles owned	2.90	2.85	0.05	0.31
Number of cows owned	0.72	0.60	0.12	0.88
Number of goats owned	0.86	0.72	0.14	0.4
Number of buffaloes owned	0.07	0.08	-0.01	-0.29
Mother's education (years)	0.14	0.13	0.00	0.04
Mother's height (cm)	148.42	148.34	0.08	0.13
Mother's age (years)	27.72	27.72	0.00	0
Square of mother's age	808.15	808.56	-0.41	-0.01
Child's age (months)	37.41	35.13	2.28	0.83
Child's age, squared	1847.80	1713.00	134.80	0.69
Child's sex (1 = male)	0.58	0.56	0.02	0.29
Distance from market (km)	8.83	8.71	0.12	0.12
Distance from closest Agragamee field office (km)	4.40	4.25	0.15	0.56
Distance from closest Agragamee field office (km), squared	23.89	21.94	1.96	0.8
Frequency of village meetings (1 = frequent/ as needed)	0.85	0.83	0.02	0.4

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A9: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on local linear regression propensity score matching

Outcome variable: Impact on child growth

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.99	0.98	0.00	0.09
Household size	5.73	5.70	0.03	0.1
Square of household size	34.97	34.42	0.55	0.18
Number of children (14 years or less)	3.07	3.08	-0.01	-0.07
Number of adult females (15 years or more)	1.42	1.38	0.04	0.35
Low-lying fertile (bila) land owned (acres)	0.58	0.58	0.00	0.01
Low-lying fertile land, squared	1.07	1.19	-0.12	-0.22
Semi-fertile (goda) land owned (acres)	1.45	1.54	-0.09	-0.4
Semi-fertile (goda) upland owned, squared)	3.53	4.62	-1.09	-0.69
Infertile (dongar) upland owned (acres)	1.47	1.55	-0.07	-0.4
Infertile (dongar) upland, squared	3.23	3.36	-0.13	-0.18
Gold holdings (gm)	2.64	3.04	-0.39	-0.58
Floor quality (1 = pucca)	0.03	0.02	0.00	0.12
Number of crowbars owned	1.12	1.08	0.04	0.54
Number of spades owned	2.04	2.08	-0.04	-0.36
Number of sickles owned	2.86	2.82	0.04	0.21
Number of cows owned	0.80	0.86	-0.06	-0.29
Number of goats owned	0.74	0.78	-0.04	-0.16
Number of buffaloes owned	0.03	0.06	-0.04	-0.69
Mother's education (years)	0.00	0.03	-0.03	-0.53
Mother's height (cm)	148.89	148.64	0.25	0.22
Mother's age (years)	29.89	29.82	0.07	0.06
Square of mother's age	932.35	929.28	3.07	0.04
Child's age (months)	39.85	41.27	-1.42	-0.44
Child's age, squared	1906.40	2050.80	-144.40	-0.57
Child's sex (1 = male)	0.54	0.53	0.01	0.12
Distance from market (km)	8.74	7.89	0.86	0.78
Distance from closest Agragamee field office (km)	5.00	4.69	0.31	0.64
Distance from closest Agragamee field office (km), squared	33.57	27.98	5.59	0.81
Frequency of village meetings (1 = frequent/ as needed)	0.84	0.75	0.09	1.22

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A10: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on kernel propensity score matching

Outcome variable: Impact on child growth	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.99	0.99	0.00	-0.1
Household size	5.69	5.77	-0.08	-0.4
Square of household size	34.22	35.04	-0.83	-0.33
Number of children (14 years or less)	3.07	3.26	-0.18	-1.0
Number of adult females (15 years or more)	1.39	1.27	0.12	1.5
Low-lying fertile (bila) land owned (acres)	0.49	0.62	-0.12	-1.02
Low-lying fertile land, squared	0.88	0.92	-0.05	-0.12
Semi-fertile (goda) land owned (acres)	1.52	1.38	0.14	0.66
Semi-fertile (goda) upland owned, squared)	3.92	3.76	0.16	0.11
Infertile (dongar) upland owned (acres)	1.46	1.50	-0.04	-0.26
Infertile (dongar) upland, squared	3.17	3.18	0.00	-0.01
Gold holdings (gm)	2.58	2.12	0.46	0.87
Floor quality (1 = pucca)	0.07	0.05	0.02	0.46
Number of ploughs owned	0.80	0.80	0.00	-0.01
Number of spades owned	2.13	2.23	-0.09	-0.87
Number of sickles owned	2.80	3.06	-0.26	-1.51
Number of cows owned	0.75	0.60	0.15	0.87
Number of goats owned	0.61	0.61	0.00	0
Number of buffaloes owned	0.02	0.04	-0.02	-0.38
Mother's education (years)	0.14	0.02	0.13	1.16
Mother's height (cm)	149.23	149.01	0.22	0.23
Mother's age (years)	29.63	28.63	0.99	1.06
Square of mother's age	919.22	850.36	68.86	1.18
Child's age (months)	40.87	44.60	-3.73	-1.32
Child's age, squared	1989.30	2317.90	-328.60	-1.44
Child's sex (1 = male)	0.57	0.48	0.08	1.05
Distance from market (km)	8.39	8.24	0.14	0.15
Distance from closest Agragamee field office (km)	4.83	5.31	-0.48	-1.15
Distance from closest Agragamee field office (km), squared	31.53	33.97	-2.44	-0.4
Frequency of village meetings (1 = frequent/ as needed)	0.87	0.84	0.03	0.51

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A11: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on local linear regression propensity score matching

Outcome variable: Impact on child growth

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.98	0.99	-0.01	-0.34
Household size	5.58	5.71	-0.13	-0.45
Square of household size	33.31	34.60	-1.29	-0.38
Number of children (14 years or less)	2.87	3.03	-0.17	-0.67
Number of adult females (15 years or more)	1.42	1.38	0.04	0.37
Low-lying fertile (bila) land owned (acres)	0.54	0.49	0.05	0.3
Low-lying fertile land, squared	0.97	0.95	0.02	0.03
Semi-fertile (goda) land owned (acres)	1.44	1.44	0.00	0.02
Semi-fertile (goda) land owned, squared	3.51	3.70	-0.19	-0.18
Infertile (dongar) upland owned (acres)	1.43	1.40	0.04	0.17
Infertile (dongar) upland, squared	3.01	3.01	0.00	0
Gold holdings (gm)	2.79	3.94	-1.15	-1.23
Number of crowbars owned	1.08	1.07	0.01	0.1
Number of spades owned	2.12	2.10	0.01	0.09
Number of sickles owned	2.83	2.77	0.06	0.24
Number of cows owned	0.65	0.79	-0.13	-0.65
Number of goats owned	0.62	0.72	-0.10	-0.45
Number of buffaloes owned	0.04	0.06	-0.03	-0.41
Mother's education (years)	0.10	0.04	0.05	0.54
Mother's height (cm)	148.63	147.77	0.86	0.71
Mother's age (years)	28.27	28.64	-0.37	-0.29
Square of mother's age	833.31	865.12	-31.81	-0.42
Child's age (months)	39.19	38.33	0.86	0.24
Child's age, squared	1859.40	1803.40	56.00	0.2
Child's sex (1 = male)	0.52	0.53	-0.01	-0.08
Distance from market (km)	7.79	9.52	-1.73	-1.24
Distance from closest Agragamee field office (km)	4.42	4.54	-0.11	-0.26
Distance from closest Agragamee field office (km), squared	25.08	24.42	0.66	0.16
Frequency of village meetings (1 = frequent/ as needed)	0.79	0.73	0.06	0.69

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A12: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on kernel propensity score matching

Outcome variable: Impact on child growth

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Social group (1=tribal)	0.98	0.99	0.00	-0.22
Household size	5.61	5.66	-0.05	-0.21
Square of household size	33.42	33.70	-0.28	-0.1
Number of children (14 years or less)	2.94	3.16	-0.22	-1.01
Number of adult females (15 years or more)	1.40	1.28	0.12	1.14
Low-lying fertile (bila) land owned (acres)	0.68	0.66	0.02	0.11
Low-lying fertile land, squared	1.43	1.14	0.29	0.51
Semi-fertile (goda) land owned (acres)	1.48	1.42	0.05	0.25
Semi-fertile (goda) land owned, squared	3.50	3.48	0.01	0.01
Infertile (dongar) upland owned (acres)	1.48	1.56	-0.09	-0.47
Infertile (dongar) upland, squared	3.06	3.63	-0.57	-0.77
Gold holdings (gm)	3.02	3.13	-0.11	-0.14
Floor quality (1 = pucca)	0.11	0.10	0.02	0.3
Number of crowbars owned	1.06	1.03	0.04	0.55
Number of spades owned	2.26	2.23	0.02	0.17
Number of sickles owned	2.84	2.88	-0.04	-0.17
Number of cows owned	0.90	0.62	0.28	1.28
Number of goats owned	0.55	0.59	-0.04	-0.21
Number of buffaloes owned	0.03	0.05	-0.01	-0.26
Mother's education (years)	0.08	0.02	0.06	0.77
Mother's height (cm)	148.50	149.55	-1.05	-1.01
Mother's age (years)	28.24	28.00	0.25	0.23
Square of mother's age	835.47	815.98	19.49	0.3
Child's age (months)	39.69	43.25	-3.56	-1.08
Child's age, squared	1889.10	2218.00	-328.90	-1.24
Child's sex (1 = male)	0.56	0.55	0.01	0.12
Distance from market (km)	8.56	8.76	-0.20	-0.17
Distance from closest Agragamee field office (km)	4.45	4.70	-0.25	-0.6
Distance from closest Agragamee field office (km), squared	24.77	27.15	-2.37	-0.53
Frequency of village meetings (1 = frequent/ as needed)	0.82	0.74	0.09	1.14

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 3.A13: Balancing tests using measure of pseudo-*R*-squared and standardized bias

Outcome variable: Impact on children's *haz*-score					
		Pseudo-R^2 (after matching)	Wald test *p*-value	Standardized bias Before matching	After matching
Local linear regression matching					
	Sample (1)	0.05	0.788	13.13	4.80
	Sample (2)	0.052	0.948	16.95	5.21
Kernel matching					
	Sample (1)	0.047	0.853	13.13	4.25
	Sample (2)	0.043	0.988	16.95	5.31

Notes: Tests conducted on observations in matched sample having common support.

Table 3.A14: Balancing tests using measure of pseudo-*R*-squared and standardized bias

Outcome variable: Impact on children's *whz*-score					
		Pseudo-R^2 (after matching)	Wald test *p*-value	Standardized bias Before matching	After matching
Local linear regression matching					
	Sample (1)	0.045	0.877	12.59	4.18
	Sample (2)	0.031	0.999	18.95	3.49
Kernel matching					
	Sample (1)	0.053	0.765	12.47	4.38
	Sample (2)	0.025	1.000	18.95	3.62

Notes: Tests conducted on observations in matched sample having common support.

Table 3.A15: Balancing tests using measure of pseudo-*R*-squared and standardized bias

Outcome variable: Impact on growth					
		Pseudo-R^2 (after matching)	Wald test *p*-value	Standardized bias Before matching	After matching
Local linear regression matching					
	Sample (1)	0.073	0.995	21.00	4.50
	Sample (2)	0.064	1.000	23.73	4.50
Kernel matching					
	Sample (1)	0.079	0.954	21.00	7.13
	Sample (2)	0.090	0.987	26.28	4.65

Notes: Tests conducted on observations in matched sample having common support.

Table 3.A16: Average treatment on the treated (ATT): Impact of grain bank participation on children's *haz*-scores
Sensitivity to choice of bandwidth

	(1) Sample 1	(2) Sample 2
Local linear regression matching estimates		
Bandwidth = 0.05	0.094	0.085
	(0.234)	(0.233)
Bandwidth = 0.07	0.086	0.056
	(0.224)	(0.231)
Kernel matching estimates		
Bandwidth = 0.05	0.084	0.079
	(0.222)	(0.214)
Bandwidth = 0.07	0.090	0.081
	(0.218)	(0.211)

Notes: Standard errors in parentheses. Estimates for matched sample having common support. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level

Table 3.A17: Average treatment on the treated (ATT): Impact of grain bank participation on children's *whz*-scores
Sensitivity to choice of bandwidth

	(1) Sample 1	(2) Sample 2
Local linear regression matching estimates		
Bandwidth = 0.05	-0.175	-0.113
	(0.203)	(0.193)
Bandwidth = 0.07	-0.116	-0.093
	(0.189)	(0.187)
Kernel matching estimates		
Bandwidth = 0.05	-0.137	-0.104
	(0.188)	(0.187)
Bandwidth = 0.07	-0.131	-0.102
	(0.184)	(0.184)

Notes: Standard errors in parentheses. Estimates for matched sample having common support. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level

Table 3.A18: Average treatment on the treated (ATT): Impact of grain bank participation on growth
Sensitivity to choice of bandwidth

	(1) Sample 1	(2) Sample 2
Local linear regression matching estimates		
Bandwidth = 0.05	0.009	0.000
	(0.010)	(0.009)
Bandwidth = 0.07	0.007	0.011
	(0.009)	(0.010)
Kernel matching estimates		
Bandwidth = 0.05	0.009	0.008
	(0.012)	(0.011)
Bandwidth = 0.07	0.011	0.009
	(0.012)	(0.011)

Notes: Standard errors in parentheses. Estimates for matched sample having common support. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level

REFERENCES

Alderman, Harold. 2000. Anthropometry. In Paul W. Glewwe and Margaret Grosh, ed., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Development Study*. Washington, D.C.: World Bank.

Alderman, Harold, John Hoddinott, and Bill H. Kinsey. 2006. Long Term Consequences of Early Childhood Malnutrition. *Oxford Economic Papers*. v58 (n3):450-74.

Behrman, Jere, and John Hoddintt. 2000. An Evaluation of the Impact of PROGRESA on Pre-School Child Height. International Food Policy Research Institute. Mimeo.

Behrman, Jere R., and Anil B. Deolalikar. 1989. Agricultural Wages in India: The Role of Health, Nutrition and Seasonality. In David E. Sahn, ed., *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Black, Dan A., and Jeffrey A. Smith. 2004. How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching. *Journal of Econometrics*. v121 (n1-2):99-124.

Blundell, Richard, and Monica Costa Dias. 2002. Alternative Approaches to Evaluation in Empirical Economics. *Cemmap Working Paper* 10. University College London: Institute for Fiscal Studies.

Boucher, Stephen R., and Catherine Guirkinger. 2005. Credit Constraints and Productivity in Peruvian Agriculture. Department of Agricultural Economics, University of California-Davis.

Boucher, Stephen R., Michael R. Carter, and Catherine Guirkinger. 2005. Risk Rationing and Wealth Effects in Credit Markets. Agricultural and Resource Economics Working paper number 05-010, University of California-Davis.

Catteneo, Matias D., Sebastian Galiani, Paul J. Gertler, Sebastian Martinez, and Rocio Titinuik. 2006. Housing, health and happiness. Mimeo.

Census of India. 2001. New Delhi: Government of India.

Citizens' Initiative for the Rights of Children Under Six (CIRCUS). 2006. *Focus on Children Under Six*. Abridged Report, December. New Delhi: Secretariat of the Right to Food Campaign.

Cogill, Bruce. 2003. *Anthropometric Indicators Measurement Guide*. Food and Nutrition Technical Assistance Project. Washington, D.C.: Academy for Educational Development.

Coleman, Brett E. 2002. Microfinance in North-East Thailand: Who Benefits and How Much? ERD Working Paper 9, Asian Development Bank.

Dehejia, Rajeev H., and Sadek Wahba. 1999. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*. v94 (n448):1053-62.

Dehejia, Rajeev H., and Sadek Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics*. v84 (n1):151-61.

Dercon, Stefan, and John Hoddinott. 2005. Health, Shocks and Poverty Persistence. In Stefan Dercon, ed., *Insurance against Poverty*. Oxford: Oxford University Press.

Diaz, Juan Jose, and Sudhanshu Handa. 2006. An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA Program. *Journal of Human Resources*. v41 (n2):319-45.

Fan, Jianqing. 1992. Design-Adaptive Nonparametric Regression. *Journal of the American Statistical Association*. v87 (n420):998-1004.

Fröhlich, Markus. 2004. Finite Sample Properties of Propensity-score Matching and Weighting Estimators. *Review of Economics and Statistics.* v86 (n1):77-90.

Galdo, Jose, Jeffrey A. Smith, and Dan A. Black. 2006. Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data. Mimeo.

Gilligan, Daniel O., and John Hoddinott. 2007. Is There Persistence in the Impact of Emergency Food Aid? Evidence on Consumption, Food Security, and Assets in Rural Ethiopia. *American Journal of Agricultural Economics*. v89 (n2):225-42.

Glewwe, Paul, and Hanan Jacoby. 1995. An Economic Analysis of Delayed Primary School Enrollment and Childhood Malnutrition in a Low Income Country. *Review of Economics and Statistics*. v77 (n1):156-69.

Grantham-McGregor, Sally M., Lia Fernald, and Kavita Sethuraman. 1999. Effects of Health and Nutrition on Cognitive and Behavioral Development in Children in the First Three Years of Life: Part 1: Low Birthweight, Breastfeeding and Protein-Energy Malnutrition. *Food and Nutrition Bulletin.* v20:53-75

Haddad, Lawrence J., and Howarth E. Bouis. 1991. The Impact of Nutritional Status on Agricultural Productivity: Wage Evidence from the Philippines. *Oxford Bulletin of Economics and Statistics*. v53 (n1):45-68.

Ham, John C., Xianghong Li, and Patricia B. Reagan. 2005. *Propensity score matching, a distance-based measure of migration, and the wage growth of young men*. Federal Reserve Bank of New York, Staff Reports: 212.

Heckman, James, and Richard Robb. 1985. Alternative Methods for Evaluating the Impact of Interventions. In James Heckman and Burton Singer (ed.) *Longitudinal Analysis of Labor Market Data*. Cambridge University.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies*. v64 (n4): 605-654.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1998a. Matching as an Econometric Evaluation Estimator. *Review of Economic Studies.* v65 (n223):261-294.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998b. *Characterizing Selection Bias Using Experimental Data*. National Bureau of Economic Research, Inc, NBER Working Papers: 6699.

Hoddinott, John. 2006. Shocks and their Consequences Across and Within Households in Rural Zimbabwe. *Journal of Development Studies*. v42 (n2):301-21.

Hoddinott, John, and Bill Kinsey. 2001. Child Growth in the Time of Drought. *Oxford Bulletin of Economics and Statistics.* v63 (n4):409:436.

Imbens, Guido W. 2004. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *Review of Economics and Statistics.* v86 (n1):4–29.

Lasky, Robert E., Robert E. Klein, Charles Yarbrough, Patricia L. Engle, Aaron Lechtig, and Reynaldo Martorell. 1981. The Relationship between Physical Growth and Infant Behavioral Development in Rural Guatemala. *Child Development*. v52 (n1):219-226.

Lawrence, Mark, F. Lawrence, T. J. Cole, W. A. Coward, J. Singh and R. G. Whitehead. 1989. Seasonal Pattern of Activity and its Nutritional Consequence in Gambia. In David E. Sahn, ed., *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Lechner, Michael. 2001. Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption. In Michael Lechner and Friedhelm Pfeiffer (eds.) *Econometric Evaluation of Labour Market Policies.* Physica, Heidelberg and New York.

Leuven, Edwin, and Barbara Sianesi. 2003. Psmatch2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing. http://ideas.repec.org/c/boc/bocode/s432001.html., version 3.0.0.

Maluccio, John, John Hoddinott, Jere Behrman, Reynaldo Martorell, Agnes Quisumbing, and Aryeh Stein. 2003. *The Impact of Nutrition during Early Childhood on Education among Guatemalan Adults*. Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, PIER Working Paper Archive.

Martorell, Reynaldo. 1995. Results and Implications of the INCAP Follow-up Study. *Journal of Nutrition*125 (Suppl): 1127S - 1138S.

Martorell, Reynaldo. 1999. The Nature of Child Malnutrition and its Long-Term Implications. *Food and Nutrition Bulletin* 20: 288-292.

Michalopoulos, Charles, Howard Bloom, and Carolyn Hill. 2004. Can Propensity Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?" *Review of Economics & Statistics*. v86 (n1):156-179.

Millimet, Daniel, and Rusty Tchernis. 2006. *On the Specification of Propensity Scores: with an Application to the WTO-Environment Debate*. Center for Applied Economics and Policy Research, Economics Department, Indiana University Bloomington, Caepr Working Papers: 2006-013.

Ministry of Tribal Affairs. 2002. Grain banks. Ministry of Tribal Affairs, Government of India website, 2002, http://tribal.nic.in/grain bank1.html. Accessed 5th April 2004

Morris, Saul S. 1999. *Measuring Nutritional Dimensions of Household Food Security*. IFPRI Technical Guide 5. Washington, D.C.: IFPRI.

Pitt, Mark. M., and Shahidur R. Khandker. 1998. The Impact of Group-Based Credit on Poor Households in Bangladesh: Does the Gender of Participants Matter? *Journal of Political Economy*. v106 (n5):958-966.

Reardon, Thomas, and Peter Matlon. 1989. Food Insecurity in Burkina Faso. In David E. Sahn, ed., *Seasonal Variability in Third World Agriculture*. Baltimore: Johns Hopkins University Press.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer.

Rosenbaum, Paul R., and Donald Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. v70(n1): 41-55.

Rubin, Donald B. and Neal Thomas. 1996. Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics*. v52 (n1): 249-264.

Sianesi, Barbara. 2004. An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *Review of Economics and Statistics*. v86 (n1):133-55.

Sethi, Bishnupada. *Vikash Sathi in Rayagada district*, June 2003. <http://ruralinformatics.nic.in/files/1_0_0_129.doc> (5 April 2005), Insights into Rural Informatics.

Smith, Herbert. 1997. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. *Sociological Methodology*. v27 (n1):325-353.

Smith, Jeffrey A. 2004. Evaluation Local Economic Development Policies: Theory and Practice. Mimeo.

Smith, Jeffrey A., and Petra E. Todd. 2001. Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods. *American Economic* Review. v91 (n2):112-18.

Smith, Jeffrey A., and Petra E. Todd. 2005a. Does Matching Overcome LaLonde's Critique of Non-experimental Estimators? *Journal of Econometrics*. v125 (n1-2):305-53.

Smith, Jeffrey, and Petra Todd. 2005b. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? Rejoinder. *Journal of Econometrics*. v125 (n1-2):365-75.

Thomas, Duncan, and John Strauss. 1997. Health and Wages: Evidence on Men and Women in Urban Brazil. *Journal of Econometrics*. v77 (n1):159-85.

World Bank. 2006. *Repositioning Nutrition as Central to Development: A Strategy for Large-Scale Action.* Washington DC: The World Bank.

CHAPTER 4

DO GRAIN BANKS DISPLACE LOCAL MONEYLENDERS?

4.1 Introduction

In this chapter, we examine the impact of household participation in grain banks on the incidence of borrowing from informal, private moneylenders using propensity score matching. We also examine how this impact varies by the lifespan of the grain bank, specifically by examining the impact in long-lived grain bank villages. Grain banks are a community level institution which provides grain loans to participant households to meet their food consumption needs typically during the agricultural lean season. The data we use are non-experimental and come from a village and household sample survey conducted in the lean season in rural Rayagada, Orissa, India. This is a small region where grain banks were established en masse over the 1990s by Agragamee, one of the more prominent tribal development NGOs in Orissa.

The study makes two main contributions. First, grain banks have been increasingly promoted by rural development NGOs as well as the Indian government – via its national and state tribal development programs – based almost entirely on anecdotal evidence on their effectiveness in enhancing household food security and reducing dependence on local moneylenders. These moneylenders are often viewed as offering credit under exploitative terms and conditions. To date, however, the impact of grain bank participation on either mentioned outcome has not been carefully quantified. While we examined the impact of grain bank participation on young children's health and nutrition in Chapter 3, in this chapter, we hope to provide more reliable estimates of the impact of grain bank participation on household borrowing from moneylenders. Given the current and envisaged increase in the extent of grain

bank development activity in India, this evidence should be of interest to government as well as NGOs.

Second, and more broadly, this study contributes to the economics literature on the impact of competition in rural financial markets.[78] While some recent empirical studies have examined the impact of competition between microfinance providers in rural areas, few have examined the impact of competition on existing credit providers such as local, informal moneylenders.[79] The displacement of households' reliance on moneylenders by grain banks – the latter providing loans at more favorable terms – can potentially be welfare-improving for participant households. The welfare effects for the wider population in the areas affected by increased credit market competition are however unclear.

The remaining sections of the paper are organized as follows. In the next section, Section 4.2, we review the relevant theoretical and empirical literature on the features of rural financial markets and the effects of (increased) competition in financial markets. We then provide some background information on the credit market in rural Rayagada and on the design and operation of the grain banks introduced by Agragamee. In Section 4.3, we present our empirical methodology. We first discuss the potential advantages of matching compared to other non-experimental estimators, as well as the existing evidence on the performance of matching compared to experimental estimators. We then discuss the key assumptions underlying matching, followed by a description of different matching estimators commonly used in the economics literature. In this section, we also justify the use of matching given our data, explain our selection of matching estimators from the available set of

[78] Given evidence from other contexts that competition among lenders alters the characteristic of rural financial markets, we would have also liked to examine how grain banks have affected local moneylender operations. Due to data limitations, however, we are unable to examine this rigorously.
[79] A notable exception is Kaboski and Townsend (2005), who examine this issue in rural Thailand. This study is discussed in the background and literature review section (Section 4.2).

options, and discuss our implementation steps and sensitivity analysis methodology. In Section 4.4, we present our findings on the impact of grain bank participation on borrowing from local moneylenders. In presenting our impact results, we first present naïve impact estimates where participant households are compared to households in non-grain bank villages without the use of matching. We then present our matching-based results, where we use local linear and kernel matching. We end the section by presenting our sensitivity analysis findings. Section 4.5 summarizes our main results and concludes.

4.2 Background and literature review

Features of rural financial markets

In a financially isolated rural community as in rural Rayagada, credit markets are subject to some of the same problems that plague rural credit markets in other developing countries.[80] An influential body of theoretical literature illustrates the potential causes behind these deficiencies in financial markets. Seminal studies by Akerlof (1970) and Stiglitz (1974) have shown how informational asymmetries between borrowers and lenders can lead to thin or missing markets. In another seminal contribution, Stiglitz and Weiss (1981) show how informational asymmetries can result in a backward-bending supply curve and credit rationing in equilibrium, whereby even potential borrowers who are willing to pay a high price cannot obtain credit as it is not profitable for financial institutions to set market-clearing interest rates. In addition, due to the absence of well-defined property rights and well-functioning judicial courts, mechanisms for contract enforcement are costly or non-

[80] For examples of recent theoretical and empirical research on rural financial markets in developing countries, see Besley (1994), Stiglitz (1994), Morduch (1995), Meyer and Nagarajan (2000), Banerjee (2003) and relevant book chapters in Basu (1997) and Ray (1998). For a recent survey of rural financial markets in developing countries, see Conning and Udry (2005).

existent. In the absence of credible sanctions, problems of limited commitment can arise causing individuals to renege on contracts (Ligon et al. 2002). In a theoretical model of agrarian credit markets in developing economies, Carter (1988) shows how adverse incentives and selection problems, in addition to 'statistical' discrimination by lenders against small farms (whereby lenders use farm size as an indicator of farm "quality") can lead to credit rationing of small farmers by profit-maximizing, competitive lenders.

A few empirical studies also provide evidence on the extent of asymmetric information and enforcement problems and/or how they affect financial market characteristics. For example, using data from a randomized credit market intervention in South Africa, Karlan and Zinman (2007) find that between 7-16 percent of loan defaults can be attributed to moral hazard or adverse selection. In another example, using data from a fishing village in southern India, Gine and Klonner (2005) find that informational asymmetries regarding borrower type result in binding credit constraints, as a result of which individuals below a certain wealth threshold cannot adopt a profitable but costly technological innovation. Using data from rural Pakistan, Aleem (1990) finds that asymmetric information and enforcement problems contribute to the high costs of operation by informal market moneylenders, which contribute, in turn, to the high interest rates charged by them. He also finds that informational problems lead to product differentiation, resulting in widely ranging interest rates.

Using household survey data on credit applications, another set of empirical studies provide evidence on the presence of credit rationing. For example, Zeller (1994) finds using data from rural Madagascar that about one in four individuals face credit (quantity) constraints, whereby they either did not apply for formal credit because they felt that they had no chance of receiving it or their credit applications were subject to rationing or complete rejection. Diagne (1999) and Zeller et al. (1998)

arrive at similar findings using data from Malawi. Barham et al. (1996), Boucher and Guirkinger (2005), and Boucher et al. (2005) find using data from Guatemala, Peru, Honduras, and Nicaragua that informational asymmetries (together with the riskiness of agricultural production) leads to credit risk rationing (especially of less wealthy farmers), in addition to conventional price and quantity rationing. In risk rationing, lenders shift so much of the contractual risk to borrowers that the latter voluntarily withdraw from the credit market even when they have the collateral wealth to participate in a credit contract.

Impact of competition on rural financial markets

To date, there are few empirical studies of how group-based lending institutions affect traditional credit sources, such as moneylenders. One exception is Kaboski and Townsend (2005), who use data from Thailand to study the impact of four different microfinance institutions (MFIs) – namely, production credit groups, women's groups, buffalo banks and rice banks (which are identical to grain banks) – on reliance on informal moneylenders, among other welfare outcomes. They use the presence of various kinds of MFI within a village as an identifying instrument, arguing that these institutions are promoted by different agencies and ministries and have large variations in lending, borrowing and membership policies. They find that the presence of any MFI reduces the probability of becoming a customer of a moneylender by 8 percentage points. When they examine the differential impact of each of the 4 MFIs on moneylender reliance, however, they find that their results are statistically significant only for women's groups.

A few recent theoretical studies look at the broader issue of how certain types of unregulated competition between potential financial providers can, contrary to initial expectations, reduce borrower welfare. While competition can induce

innovation, it can also alter market structures such that there is a loss of scale economies among incumbent suppliers. Given information asymmetries, this can translate into a shrinking market, higher prices and lower welfare for all borrowers, especially poorer borrowers. For example, Hoff and Stiglitz (1997) predict how the introduction of subsidized credit into a monopolistically-competitive market with high enforcement costs can lead to the exit of incumbent lenders and higher prices for borrowers. McIntosh and Wydick (2005) show how competition among microfinance lenders reduces cross-subsidization among clients of each institution and increase asymmetric information problems, resulting in less favorable contracts for all borrowers.

In addition, there are a few empirical studies which examine the impact of competition on lending relationships. For example, Petersen and Rajan (1995) show how creditors are less likely to provide loans to credit-constrained firms in more competitive markets. Using data from small-business lending in the United States, they find that loan sizes are larger in areas where there is less competition as financial institutions are able to enter into long-term lending relationships with firms when there are fewer competitors who take away their more successful clients. Navajas et al. (2003) use data from Bolivia to test their theoretical predictions that in the presence of information asymmetries, unregulated competition among MFIs can lead to adverse effects on borrowers. They find that as a result of competition, poorer borrowers are worst affected as they cannot avail of credit as creditors wish to serve only the most profitable clients. Using data from the Foundation for International Community Assistance (FINCA), the largest incumbent MFI in Uganda, McIntosh et al. (2005) show how the entry of competing lenders induces a decline in the repayment rate and the savings deposited with the incumbent. They find that increased competition does not lead to an increase in the dropout rate or client enrollment rate, but suggest that

this might indicate that the market is not fully saturated and the more adverse impacts of competition predicted by theory are yet to be observed. Kranton and Swamy (1999) show how a positive institutional change affected the structure of credit markets, which, in turn, had an *adverse* impact on borrower welfare. Using data from agricultural credit markets in the Bombay Deccan in colonial India, they find that the introduction of civil courts increased competition among lenders, thereby reducing their incentives to subsidize farmers' investments in bad periods and making them more vulnerable during crises. These studies indicate the importance of examining the impact of introducing competition into rural credit markets.

Features of credit markets in Rayagada

While our household survey data do not permit the examination of the causes behind thin or missing financial markets in rural Rayagada, we find that a majority of households are unable to smooth consumption in the face of shocks, suggesting potential credit rationing and/or missing credit and insurance markets. Table 4.1 provides the share of households that use different risk-coping strategies in order to manage up to the 3 recent-most shocks faced in the 5 years preceding the survey. 301 out of 499 households surveyed (or 60 percent) reported facing one or more shocks. 57 percent of these households reported that they reduced consumption to cope with the shock. The second-most important coping strategy was reported to be borrowing from local, private informal moneylenders (23 percent), followed by transfers from friends and family (10 percent), and sale of productive assets (7 percent) .

Table 4.1: Coping strategies used to manage shocks experienced by households

Coping strategies	Share of households (in percent)
Reduce consumption	56.9
Borrow from local moneylender	22.6
Transfers from friends and family	10.0
Sale of assets (agricultural tools, livestock, land, other)	7.0
Other (increase labor supply, transfers from govt. and NGO, etc.)	3.5

Notes: Shares reflect reports of coping strategies ever used for up to the 3 most recent shocks, conditional on experiencing a shock in the 5 years preceding the survey. Out of the 301 households who reported facing shocks, 41 percent report using more than 1 risk-coping mechanism while 7 percent report using 3 risk-coping mechanisms. Estimates are corrected for sampling weights.

Who are the moneylenders in this region and what is their modus operandi? [81]

Moneylenders in this region are typically local shopkeepers or traders. They provide loans in the form of cash as well as in-kind (mostly in the form of grains). Due to the nature of their relationships with borrowers, which normally involve repeated interactions over long periods of time, contract enforcement is reported not to be a problem. Informal interviews with villagers in Kashipur revealed that, after harvest time, moneylenders often collect loans in the form of cash crops by visiting the homes of their borrowers. In fact, the vast majority of loans were reportedly returned in-kind, regardless of whether the loans were originally made in the form or cash or in-kind. In the event that a borrowing household is unable to return a loan, moneylenders employ them as unpaid agricultural labor (such as for grazing animals) or domestic labor. This indicates the presence of tied labor-credit contracts in the area, often with onerous labor requirements. Such tied labor-credit contracts are commonly observed across rural economies in the developing world (Hoff et al. 1993).

[81] This brief description is based on group interviews conducted in a number of villages in Kashipur, separate from the grain bank village and household surveys. The villages, which include both grain bank and non-grain bank villages, are Paraja Sila, Telengiri, Similiguda, Ranjuguda, Patamanda and Bajansil.

What are the loan sizes and interest rates offered by the moneylenders?

There appears to be a considerable variation in both the loan amounts offered and the interest rates charged by moneylenders. The loan amounts ranged between 50 rupees (slightly over US$1, at the current exchange rate) and 7,000 rupees (about US$175). The average loan size was 1,112 rupees (about US$29), with a standard deviation of 1,185 rupees (about US$30). The mean annual interest rate charged by moneylenders on outstanding cash loans (which formed 98 percent of outstanding loans) was reported to be close to 50 percent, with a standard deviation of 30 percent and a range of 5 percent to 250 percent.[82] Due to the high level of occupational homogeneity in our data – virtually all household heads are employed in small-scale farming – we are unable to examine if there is credit market fragmentation by occupation (i.e., different interest rates for different occupations), as, for example, Meyer et al. (1997) do for rural informal credit markets in the Philippines.

What are the main reasons that households borrow from moneylenders?

Table 4.2 shows that the vast majority of loans from the moneylender were consumption loans. Seventy-two percent of households cited that the reason for borrowing was to meet food and household expenses. Another 12 percent of households cited expenses related to medical treatment or wedding expenses as the reason for borrowing. Clearly, borrowing for consumption purposes, especially to meet food expenses, is by far the most common reason for borrowing from moneylenders.

[82] Given the flexibility and unwritten nature of these credit contracts, the interest rate figures reported above are unlikely to accurately take into account the value of in-kind payments (including unpaid household or farm labor in lieu of credit payments, pointing to interlinked labor-credit contracts), and are likely to be biased downwards.

Table 4.2: Reason for loan from moneylender

Reason	Share of loans (percent)
Food/ household expenses	72.2
Agricultural expenses	9.0
Home repairs	6.8
Medical treatment	6.0
Other (Trade, Wedding expenses, etc.)	6.0
Total	100.0

Notes: Shares are conditional on reporting an outstanding loan from the moneylender at the time of the survey.

Community grain banks

Given this context, grain banks are perceived as a welcome alternative source of credit for consumption smoothing purposes by households in the study area. Agragamee first introduced grain banks in Kashipur in 1981 and later expanded the initiative to all villages in Kashipur during the 1990s as part of the Orissa Household Food Security Project (OHFSP).[83]

What are grain banks? Grain banks provide loans in the form of grains, typically during the lean season when food shortages are at a peak, to be returned with interest (also in the form of grains), typically during the harvest season. They are community-based institutions: they are village-level (or hamlet-level, in the case that there is more than one hamlet in a village). They are member-owned and member-managed: after an initial external grant, grain bank operations are managed by an elected committee, comprising of grain bank members themselves, and operational decisions (such as setting interest rates) are usually made in open meetings comprising all grain bank members.

Though they are not group-based, grain banks employ some of the same mechanisms used by group-based credit institutions to mitigate informational and enforcement problems. Given that tribal, agrarian societies are bound by close ties of

[83] For more details on the grain bank initiative in Kashipur, see Chapter 2.

clan and kinship, individuals typically possess a rich set of information regarding fellow members. Thus, peer monitoring can be an effective and low-cost instrument for attenuating informational asymmetries and enabling contract enforcement. This has been shown both theoretically and empirically in the case of German credit cooperatives by Banerjee et al. (1994).

Community-based institutions such as grain banks can also take advantage of the role of the community in enforcing non-opportunistic behavior through the threat of social sanctions. Besley et al. (1993) and Besley and Coate (1995) show that the threat of social sanctions can help maintain high repayment rates and also help overcome free rider problems in activities with a public good character, such as peer monitoring and auditing. This has been shown empirically by Miguel and Gugerty (2004) in the provision of public goods in Kenya.[84]

Even in the absence of group-based lending, dynamic incentives have been shown to reduce default risk. For example, Morduch (1999) and Alexander (2006) show that when a borrower has continual credit needs, access to future loans can provide a strong reason to avoid default on a current loan. This situation is also true of grain bank borrowers: food shortages occur year after year, and the threat of losing access to future loans may be sufficient to prevent opportunistic default. Moreover, since members of a small, traditional rural community typically interact with the same individuals on a repeated basis over long periods of time, informal contracts can become self-enforcing as the short-term benefits from reneging are much smaller than the long-term costs (Posner 1980, Coate and Ravallion 1993). As a result, even in the

[84] Mude (2006) however shows how the informational advantages provided by close kinship ties in small communities are diminished if incentives for patronage and favor-peddling are present, providing a cautionary note on the limitations of peer monitoring in small, traditional communities. In addition, the personal nature of community-based arrangements (together with the absence of checks and balances in isolated rural communities) can also make them more vulnerable to capture by local elites or non-traditional leaders, as members of the community may be afraid of retribution or unwilling to act as whistle-blowers (see, e.g., Mansuri and Rao 2004 or Conning and Kevane 2002).

absence of formal legal courts, informal contracts such as in grain banks can be self-enforcing and address informational and enforcement problems (Platteau 2000).

The recent establishment of grain banks in rural Rayagada has introduced an element of competition into the credit market where none existed earlier. However, given that grain banks provide a very limited loan variety (consumption credit, in the form of grains), it is unclear how their presence affect private, informal moneylenders, the most common source of credit in these areas. In addition, since not all households become members of grain banks, it would be important to examine how grain bank establishment affects the ability of, not only members, but also non-members to obtain credit from moneylenders.

While grain banks have been designed to address food security concerns, they are likely to have an impact on the structure of the credit market. Clearly, grain banks offer a more attractive alternative by setting lower interest rates (which, at 20-25 percent, are lower than standard moneylender interest rates). Anecdotal evidence from grain bank members indicates that grain banks have indeed reduced dependence on moneylenders, the incumbent lender in these markets. Yet, it is unclear if this impact is restricted to grain bank members or also extends to non-members. Economic theory would predict that a price-discriminating moneylender would charge higher interest rates to grain bank non-members in order to compensate for the loss in client base. However, informal interviews indicate that there have been no changes in interest rates charged to non-grain bank members in either grain bank villages or non-grain bank villages. While our data do not permit a rigorous analysis of the impact of grain banks on interest rates charged by moneylenders, we present naïve estimates in our empirical results section (Section 4.4).

4.3 Empirical methodology[85]

The evaluation problem

The central problem in evaluation arises because we cannot observe outcomes for the same observation in the counterfactual state (Heckman and Robb 1985). For example, in this study, we can observe households that are grain bank participants (the treatment group) or households that are grain bank non-participants (the comparison group), but we cannot observe outcomes for the same household in both states. The cleanest solution to this missing data problem is a social experiment where the comparison group is constructed from a random subset of the eligible population. However, since we do not have experimental data but rather observational data, we have to rely on non-experimental methods to estimate the impact of grain bank participation on borrowing from local moneylenders.

Matching compared with other non-experimental estimators

In recent years, an estimator that has gained wide usage in evaluation using non-experimental data is matching.[86] Matching estimators are based on the premise that the most appropriate estimate of the counterfactual untreated outcome for a treated unit is the outcome of an untreated unit or units most similar to the treated unit from the identified comparison group, under the assumption that selection into program participation is based on observables. While this is a strong assumption, it is also made by alternate non-experimental estimators such as linear regression. Although instrumental variable (IV) estimation can address bias due to selection on unobservables, the validity of the estimates is sensitive to the choice of the IV (a

[85] This section largely borrows from the corresponding section in Chapter 3.

[86] For an evaluation of matching estimators, see, among others, Dehejia and Wahba (1999, 2002), Heckman et al. (1997, 1998a), Heckman et al. (1998), and Smith and Todd (2001, 2005a). For an overview of different non-experimental estimators, see Blundell and Costa Dias (2002).

variable which has to be correlated with the decision to participate, but uncorrelated with any unobserved factors that affect the outcome). However, the difficulties of finding a valid IV are well-known.

If we assume that selection is indeed based on observables (or that the variables observable to the researcher span almost all of those used by the agent in deciding whether or not to participate), matching estimators provide impact estimates that can approximate estimates provided by randomized experiments. To do this, matching estimators first generate a propensity score. The propensity score is simply a predicted probability, based on observed characteristics, that the individual (or household) will participate in the program. This is used to match treated units with untreated units that are similar in every (observable) respect except in their treatment status.[87] The difference in the outcome of interest between matched units can then be attributed to the program.

Matching has some important advantages over regression. The latter imposes a linear functional form to identify the counterfactual outcome.[88] However, unlike regression, matching methods are either semi-parametric or non-parametric, depending on the particular method used (Black and Smith 2004). In matching, after units with comparable propensity scores are matched, the difference in the outcome between the units provides the impact estimate. This is particularly important if participants and non-participants differ substantially in terms of their observed characteristics, as regression assumes the same functional form for both groups.

[87] We discuss the matching algorithms used to match treated and untreated units in the latter half of this section.

[88] If we include a sufficient number of higher order terms, however, the linear model can approximate a given non-linear function of the set of conditioning variables X arbitrarily well. In practice, however, including a large number of higher order terms for each variable, results in a degrees of freedom problem.

In addition, matching only pairs control and treatment units having similar propensity scores. For a given control unit, if no comparable unit (i.e., having a similar propensity score) is available in the treatment group, then it is discarded from the analysis. This area of overlap between the propensity scores is known as the "common support" region (Smith and Todd 2005a). However, in regression, all treatment and control units are compared, regardless of whether they are comparable. While matching does not solve the support problem, it highlights it in a way that regression does not (Black and Smith 2004). In other words, matching exposes the common support problem, i.e., whether or not comparable untreated units are available for each treated unit (Smith 2004). After estimating the propensity scores for the treatment and comparison groups, it is possible to compare the densities in order to examine the extent of the common support problem.

In addition, matching weights observations differently than ordinary least squares (OLS) regression in calculating the expected counterfactual for each treated observation. In OLS regression, all untreated units play a role in determining the expected counterfactual for any given treated unit and receive equal weight. In matching, however, only untreated units similar to each treated unit have a positive weight in determining the expected counterfactual, and these weights can vary depending on the distance in propensity scores between the matched units.

Performance of matching compared to experimental estimates

A number of recent studies have compared the performance of matching by comparing experimental estimates with non-experimental estimates using matching, mostly using data from voluntary employment and job training programs in the United States. Comparing experimental estimates to matching estimates using non-experimental data from the Job Training Partnership Act (JPTA) program, Heckman et al. (1997, 1998a)

and Heckman et al. (1998b) provide evidence that propensity score matching methods perform well relative to experimental estimators, provided the following set of conditions are met: (1) the presence of a rich set of conditioning variables; (2) use of the same survey instruments for participants and non-participants; and (3) participants and non-participants face the same economic conditions. Using data from mandatory welfare-to-work training programs, Michalopoulos et al. (2004) come to a similar conclusion: matching estimators do not provide reliable estimates if the data for the comparison group come from a different geographic and labor market. Using National Support Work (NSW) data, Smith and Todd (2005a) highlight the fact that the performance of matching estimators depends critically on the quality of the data – specifically, that the conditions established by Heckman et al. (1997, 1998a), and Heckman et al. (1998b), are met. A more recent paper by Diaz and Handa (2006) provides evidence on the performance of matching estimators using non-US data. Comparing experimental and non-experimental estimates of the impact of Progresa, a voluntary anti-poverty program in Mexico, they find that matching estimators using the latter perform well when the outcomes of interest are measured comparably across treated and untreated groups and a rich set of covariates is available.

In our data, a number of conditioning variables that potentially identify program participation and outcomes are available; the same survey questionnaire is used in grain bank and non-grain bank villages as a result of which the outcomes of interest are measured identically; and participants and non-participants are in the same geographical area, namely within the same administrative unit within Rayagada district in the state of Orissa, and face the same economic and ecological conditions. Therefore, we argue that matching estimates can provide reliable program impact estimates in our study. In the next section, we discuss the conditions that need to hold to obtain valid matching estimates.

Matching methods

Following Heckman et al. (1997) and Smith and Todd (2001, 2005a), let Y_1 denote the outcome of interest for grain bank participant households (the treatment group), Y_0 the corresponding outcome for non-grain bank households (the comparison group), and $D \in \{0,1\}$ an indicator variable denoting grain bank participation. Let X denote the set of conditioning variables. The parameter of interest – the average impact of the treatment on the treated (ATT) – is given by

$$\begin{aligned} & E(Y_1 - Y_0 \mid D = 1, X) \\ & = E(Y_1 \mid D = 1, X) - E(Y_0 \mid D = 1, X) \\ & = E(Y_1 \mid D = 1, X) - E(Y_0 \mid D = 0, X) \end{aligned}$$

Under the assumption of conditional independence (CIA), i.e., treatment status is independent of the outcome conditional on a set of observed covariates, we can estimate the parameter of interest (Imbens 2004).[89] Thus, the first condition required for matching estimates to be valid is $(Y_0 \perp D) \mid X$ (Condition 1). In other words, selection into the program is based only on observables. Put differently, this assumes that the analyst can observe the complete set of variables used by the agent in making the decision to participate.

The second condition required for matching to be valid is $0 < \Pr(D = 1 \mid X) < 1$ (Condition 2). In other words, for all X there is a positive probability of either participating ($D = 1$) or not participating ($D = 0$).

Condition 1 runs contrary to those invoked by many economic models of self-selection, and its validity depends crucially on how well the observed data capture all the variables that affect program participation and outcomes of interest. The validity of matching estimates thus depends crucially on a thorough understanding of the

[89] Smith and Todd (2005a) show that only a weaker assumption, $E(Y_0 \mid X, D = 1) = E(Y_0 \mid X, D = 0)$ (i.e., the conditional mean assumption), is needed to estimate the parameter of interest.

selection process and the availability of a rich set of conditioning variables that affect participation and outcomes.

However, as the number of conditioning variables increases, we have to contend with what is commonly referred to as matching's version of the "curse of dimensionality" (Heckman et al. 1997). In other words, as the number of conditioning variables increases, it is possible to have many cells without matches. In this context, Rosenbaum and Rubin (1983) show that if we can match on *X*, we can also match on $P(X) = \Pr(D = 1 | X)$, the conditional participation probability or the so-called propensity score.

If the propensity score is estimated non-parametrically, we again have to contend with the problem of dimensionality (Heckman et al. 1998a). If the propensity score is estimated using parametric (such as logit or probit regression) or semi-parametric methods, however, then the dimensionality of the matching problem is reduced. We can then match on the univariate propensity score. Given Conditions 1 and 2, we can then estimate the parameter of interest by

$$E(Y_1 | D = 1, P(X)) - E(Y_0 | D = 0, P(X)) .$$

In experimental data, Conditions 1 and 2 are satisfied by random assignment to treatment. For non-experimental data, there may or may not exist a set of variables X such that these conditions are satisfied. Thus, it is important to be aware that our estimates depend crucially on our assumptions, in particular, that we have a set of variables X such that Conditions 1 and 2 are satisfied.

Smith and Todd (2005a) also draw attention to the important fact that matching estimates are only valid if the support of X overlaps for the $D = 1$ and $D = 0$ groups. In other words, the treatment effect should be redefined as the program impact on participants whose propensity scores lie within the common support region (Smith and Todd 2005a).

Practical issues in propensity score matching

Assuming that Conditions 1 and 2 hold, there are a number of practical steps involved in estimation using matching methods. First, how to choose the propensity score model and which variables to use? Second, which matching estimator to use? Third, what criteria to use in implementing the common support restriction? Fourth, which balancing test to use for implementing the propensity score method? We examine these issues, as well as related ones (such as specifying the kernel and bandwidth for certain estimators) below.

Choosing the propensity score model

As discussed in Smith (1997), there is little guidance on the functional form to use in specifying the propensity score model. In principle, any discrete choice model can be used. In practice, for the binary treatment case, the binomial logit or probit models are used. These are usually preferred to the linear probability model, due to the latter's shortcomings, such as predictions which are outside the [0,1] range. However, given that the purpose of the propensity score model is classification into treatment and comparison groups rather than the estimation of structural coefficients, model choice is not a critical issue (Smith 1997). And since both the logit and probit models yield similar results, both are commonly used in the empirical literature for predicting propensity scores.[90] In our study, we estimate the propensity score model using a binomial probit.

[90] Model choice is more important in the case of multiple treatments, but this issue is not relevant for our study (see Lechner 2001 for examples of the propensity score model in the case of multiple treatments).

Choosing which covariates to include in the propensity score model

The CIA assumption depends critically on including a set of covariates which captures the variables involved in selection into program participation. Todd and Smith (2005) show that matching estimates are biased unless the set of covariates X that satisfies Conditions 1 and 2 is included in the propensity score model specification. Thus, in order to obtain reliable matching estimates, one needs to capture all observable variables that potentially affect program participation as well as the outcome of interest. Thus, the issue of choosing which covariates to include in the propensity score model is important.

While theory provides some guidance on which variables are likely to affect both participation and outcomes, there is little guidance on how to correctly specify X in practice. Heckman et al. (1997) and Heckman et al. (1998b) show that there are larger biases if a 'coarse' set of conditioning variables is used relative to a 'rich' one. Rubin and Thomas (1996, p.253) argue in favor of including more covariates in the propensity score model specification, unless there is consensus that a variable is "unrelated to the outcome or not a proper covariate." However, including more covariates in X can also be a problem, as it comes at the cost of reducing the region of common support (Smith and Todd 2005a). In fact, a model that predicts perfectly results in observations with propensity scores that do not have any common support.

Bryson et al. (2002, as cited in Millimet and Tchernis 2006) also argue that the inclusion of "irrelevant" variables can increase the variance of the treatment effect indicator.[91] However, Millimet and Tchernis (2006) use Monte Carlo simulations to compare the performance of matching estimators when relevant higher order terms are excluded and irrelevant higher order terms are included in the specification of the

[91] These may be higher order terms of variables that are relevant (but only at a lower order) or variables that are wholly irrelevant.

propensity score model. They find that overspecification of the propensity score model does not result in less efficient estimates, using different propensity score matching estimators including kernel estimators. Underspecification, on the other hand, results in worse estimates. Their results are corroborated in an application testing the impact of the World Trade Organization (WTO) on the environment as well as that of currency unions on international trade.

In general, the choice of variables usually depends on economic theory and previous empirical findings. However, Heckman et al. (1997), Heckman et al. (1998b) and Black and Smith (2004) also suggest some formal tests for choosing *X*. Heckman et al. (1997) suggest choosing the set of variables that maximizes the within-sample prediction rates using the hit-or-miss method.[92] Heckman et al. (1998b) suggest a test of statistical significance. They start with a parsimonious specification, and then add new variables iteratively. If the new variable is statistically significant at conventional levels, then it is included in the final *X*; otherwise it is discarded. Black and Smith (2004) suggest the 'leave-one-out cross validation method' to specify the propensity score model. In this method, they begin the model selection procedure by starting with a minimally specified model, and then successively add blocks of variables, comparing the resulting mean squared errors. Essentially, this method involves examining the fit of the model. However, in the case that this method suggests leaving out covariates that theory and previous empirical evidence suggest should be included, they advise giving more consideration to the latter.

In our study, we include a number of household and community-level variables that we believe affect both participation and outcomes. We use the 'leave-one-out

[92] In this method, an observation is classified as '1' if the estimated propensity score is greater than the sample proportion of persons who receive the treatment, and 0 otherwise. This maximizes the overall classification rate, assuming that the costs for misclassification are the same for both treatment and control groups (Heckman et al. 1997).

cross-validation method'. First, we specify a model with household-level variables only. We then add village-level variables and compare the fit of the models.

Choosing the matching estimator

There exist a variety of matching estimators, which differ mainly in how they assign weights to each comparison group observation. Asymptotically, they produce similar results, as in a large sample, they only compare exact matches (Black and Smith 2004). However, in finite samples, the estimates produced by different matching methods differ because of differences in the weighting function they use, as also how they address the common support problem (ibid).

Following Smith and Todd (2005a), for simplicity of notation, we rewrite $P = \Pr\left(D = 1 \middle| X\right)$. A typical matching estimator, $\hat{\alpha}_{\mathrm{M}}$, takes the form

$$\hat{\alpha}_{\mathrm{M}} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} \left[Y_{1i} - \hat{\mathrm{E}}\left(Y_{0i} \middle| D_i = 1, P_i\right) \right],$$

where

$$\hat{\mathrm{E}}\left(Y_{0i} \middle| D_i = 1, P_i\right) = \sum_{j \in I_0} W\left(i, j\right) Y_{0j}$$

and I_1 denotes the set of program participants, I_0 the set of non-participants, S_P the region of common support, n_1 the number of persons who are in the set $I_1 \cap S_P$. Every participant $i \in I_1 \cap S_P$ is matched with a weighted average over the outcomes of non-participants, where the weights $W(i, j)$ depend on the distance between P_i and P_j. For each $i \in I_1$, $C(P_i)$ is defined to be a neighborhood. Individuals matched to i are those people in set A_i such that $\left\{ j \in I_0 \mid P_j \in C(P_i) \right\}$.

Matching estimators differ from one another mainly in how they construct the weights $W\left(i, j\right)$ and define the neighborhood $C\left(P_i\right)$. Broadly speaking, they can be classified into two main types – traditional pair or one-to-one matching estimators, and

more recently developed non-parametric matching estimators which match most (or all) the control observations using a pre-defined weighting function.

In one-to-one (or one-to-many) matching, outcomes are compared to the most observably similar untreated unit. Untreated units that do not have sufficiently close propensity scores are discarded. This reduces bias (due to better matches) but increases variance (due to fewer matches) (Smith and Todd 2005a). A common example of traditional matching estimators include nearest neighbor matching. In this method, the untreated unit j is matched to that treated unit i such that distance between them is the smallest. The neighborhood is defined as

$$C\left(P_i\right) = \min_j \| P_i - P_j \|, \; j \in I_0 .$$

Nearest neighbor matching, which is traditionally performed without replacement, constructs $W(i, j) = 1$ such that all matched units receive equal weight. Nearest neighbor matching can also be performed with replacement, such that untreated units can be matched to more than one treated unit, with accompanying trade-offs between reduced bias and increased variance compared to matching without replacement.

However, nearest neighbor matching is clearly inefficient. More efficient alternatives to nearest neighbor matching include a family of non-parametric, kernel-weighted matching estimators such as kernel matching and local linear matching (Heckman et al. 1997, 1998a).[93] Relative to pair-wise matching, these estimators reduce the asymptotic mean squared error (Smith and Todd 2005a). In a study of finite sample properties of various matching estimators using Monte Carlo simulations, Fröhlich (2004) finds that the nearest neighbor estimator performs poorly relative to non-parametric kernel-weighted matching estimators. The latter match a treated unit with the weighted average score of *all* untreated units within a certain

[93] We do not discuss the difference-in-differences matching estimator proposed by Heckman et al. (1997) and Heckman et al. (1998b) since it requires before-after data which is not available to us. This estimator is preferred since it can control for time-invariant unobservables.

distance, referred to as the bandwidth. The weight given to the untreated unit is inversely proportionate to the distance between i and j and depends on the weighting function that is used. Relative to pair-wise matching, the use of more untreated units reduces the variance of the matching estimates; however, it increases the bias.

The most commonly used kernel-weighted estimators include the kernel estimator and the local linear estimator. Heckman et al. (1997) find that the local linear matching estimator has a slight advantage over the kernel matching estimator because of some desirable statistical properties; namely, it converges at a faster rate at boundary points and adapts better to different data densities. Fan (1992) also shows that local linear matching is better able to adapt to survey design. Therefore, one of the estimators used in this study is based on local linear matching. In local linear matching, the weighting function is given by

$$W(i,j)=\frac{G_{ij}\sum_{k\in I_0}G_{ik}\left(P_k-P_i\right)^2-\left[G_{ij}((P_j-P_i)\right]\left[\sum_{k\in I_0}G_{ik}(P_k-P_i)\right]}{\sum_{j\in I_0}G_{ij}\sum_{k\in I_0}G_{ij}\left(\left(P_k-P_i\right)^2-\sum_{k\in I_0}G_{ik}\left(P_k-P_i\right)\right)^2},$$

where $G_{ij}=G\left(\left(P_j-P_i\right)/a_n\right)$, and G denotes the kernel function and a_n the parameter determining the kernel bandwidth.

However, Heckman et al. (1997, 1998a) find that the local linear matching estimator does not perform well in small samples when there are regions of sparse data density. A common solution is to implement a trimming procedure in regions where the density of the propensity in the untreated population is small. However, Fröhlich (2004) points out that there is little practical guidance on the optimal level of trimming. From the distribution of propensity scores in our estimations, we find that while there is a large region of overlap in the propensity scores of treated and untreated observations (i.e., the region of common support), the distribution of the propensity scores for the two groups in our data is quite different, and so there are regions of sparse density. In addition, our sample size is small. For this reason, we

also estimate the kernel matching model to examine if the estimates from the local linear matching analysis are comparable.

The weighting function for the kernel estimator is given by

$$G((P_j - P_i)/a_n)/\sum_{k \in I_0} G((P_k - P_i)/a_n).$$

Following Heckman et al. (1997) and Smith and Todd (2005a), the kernel-weighted matching estimate of program impact takes the form

$$ATT = \frac{1}{n_1}\sum_{i \in I_0}\left\{ Y_{1i} - \frac{\sum_{j \in I_0} Y_{0j} G\left(\frac{P_j - P_i}{a_n}\right)}{\sum_{k \in I_0} G\left(\frac{P_k - P_j}{a_n}\right)} \right\}.$$

The main difference between local linear matching and kernel matching is that the former includes a linear term in P_i in addition to the intercept. According to Smith and Todd (2005a, p. 317), this is helpful when "comparison group observations are distributed asymmetrically around the participant observations, as would be the case at a boundary point of *P* or at any point where there are gaps in the distribution of *P*. Thus, the local linear regression estimator is a more generalized version of the kernel estimator.

For both the local linear and kernel matching estimators, the neighborhood $C(P_i)$ depends on the choice of the kernel function. Two commonly used kernels in the empirical matching literature include the Epanechnikov and Gaussian kernel. Using data from the National Longitudinal Survey of Youth (NLSY), in a study of the effects of college quality, Black and Smith (2004) find that the Epanechnikov kernel estimator performs slightly better than the Gaussian kernel estimator, independent of the size of the bandwidth. They find that the former converges faster than the Gaussian kernel and implicitly imposes the support condition through the choice of the bandwidth. Given this, we use the Epanechnikov kernel here with variable bandwidth.

The latter, relative to a fixed bandwidth, has the advantage of varying the bandwidth depending on the data density at that point, i.e., it uses a small bandwidth in regions where the probability mass is dense and a large bandwidth when the probability mass is sparse (Ham et al. 2005).

A related choice that has important implications for the tradeoff between variance and bias, especially for matching methods based on kernel regression, is the size of the bandwidth or smoothing parameter. While a smaller bandwidth results in smaller bias but larger variance, a larger bandwidth results in smaller variance but larger bias (Galdo, Smith, and Black 2006). Specifically, in the case of matching, a smaller bandwidth leads to the use of few untreated units for each treated unit while a larger bandwidth leads to the use of untreated units that may be rather different from each treated unit. The standard approach in the matching literature to guide bandwidth choice is minimization of some quadratic loss functions, such as the mean squared error (MSE) or integrated mean squared error (IMSE), as a measure of fit. However, as discussed by Galdo, Smith and Black (2006), the bandwidth that minimizes the MSE for the regression function of the untreated outcome is not that which minimizes the MSE for the object of interest, which in our case is the ATT. In addition, if the distribution of the conditioning variables is not balanced for the treated and untreated observations, then the optimal bandwidth in the region of low propensity scores (where most untreated units lie) will be different from the optimal bandwidth in the region of high propensity scores (where most treated units lie). This results in biased estimates.

In our study, we examine the MSE for varying bandwidths between 0.01 and 0.1 (in 0.01 increments). We present estimates for bandwidth size 0.06, which, while not having the minimum MSE in the range examined, enables us to include a more inclusive X vector of balanced covariates. We also present estimates for different

bandwidth sizes (specifically, 0.05 and 0.07) to examine the sensitivity of our impact estimates to the choice of bandwidth.

Choosing the criteria to restrict matching to the common support region

In order to obtain credible matching estimates, only those comparison and treatment observations whose propensity scores fall within the region of common support should be included. This is particularly important for kernel-weighted matching since it uses all comparison observations to estimate the counterfactual outcome, unlike nearest neighbor matching. Implementing the common support restriction can improve the quality of the matches used to estimate the *ATT*, although it comes at the cost of reduced sample size as observations at the boundaries of the common support are excluded.

In practice, one of two methods is commonly implemented for ensuring common support. The first involves the comparison of the minimum and maximum propensity scores of treatment and comparison observations. All observations in the treatment (or comparison) group whose propensity score is smaller than the minimum and larger than the maximum of the propensity scores in the opposite group are discarded.

The second method for implementing the common support is a trimming procedure, as suggested by Smith and Todd (2005a). They determine the common support region by

$$\hat{S}_P = \{P : \hat{f}(P \mid D=1) > 0 \text{ and } \hat{f}(P \mid D=0) > 0\},$$

where $\hat{f}(P \mid D=d)$, $d \in \{0,1\}$ are non-parametric density estimators given by

$$\hat{f}(P \mid D=d) = \sum_{k \in I_d} G((P_k - P)/a_n).$$

They then define a trimming level q and require that densities are strictly positive and exceed zero by q. They exclude observations with P for which the density is zero,

and then an additional q percent of remaining P points for which the estimated density is low (although positive). Thus, the set of eligible matches is given by

$$\hat{S}_q = \{P \in I_1 \cap \hat{S}_P : \hat{f}(P \mid D=1) > c_q \text{ and } \hat{f}(P \mid D=0) > c_q\}.$$

In our study, we use the min-max method to implement the common support region as there are no guidelines on how to arrive at the optimal trimming level given the data (Fröhlich 2004).

Choosing a balancing test for the set of covariates included in the propensity score model

If the CIA assumption is valid, after we condition on P, additional conditioning on any of the X's should not provide any new information about the treatment decision. This implies that all the Xs should be "balanced" across the treated and matched untreated groups. Thus, in order to assess the quality of the matching estimates, we need to conduct a 'balancing test' of the characteristics of the matched samples. There are a few formal tests that are commonly implemented in the literature, though there is no consensus on which one is definitive.

Proposed by Rosenbaum and Rubin (1985), one balancing test is the examination of standardized differences (see, e.g., Sianesi 2004). In words, the standardized difference for a variable X_k is the difference in means between the treated and matched untreated group samples, as a percentage of the square root of the average of the sample variance in both groups. Intuitively, this provides the size of the difference in means of a conditioning variable between the treated and matched untreated group, scaled down by the average of the variances.[94] In this method, for each of the conditioning variables, the standardized bias before matching is

[94] The shortcoming of this method is that there are no formal criteria for which the standardized bias is too large to pass the balancing test (Smith and Todd 2005b). Smith and Todd (2005b) also point out that the standardized bias can be manipulated by the researcher by adding additional observations to the

$$SB_{before} = 100 \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{\frac{\text{var}_1(X) + \text{var}_0(X)}{2}}},$$

where $\overline{X}_0$, $\text{var}_0(X)$, $\overline{X}_1$, and $\text{var}_1(X)$ denote the means and variances in the treated and untreated groups *before* matching, respectively. The standardized bias after matching is

$$SB_{after} = 100 \frac{\overline{X}_{1M} - \overline{X}_{0M}}{\sqrt{\frac{\text{var}_{1M}(X) + \text{var}_{0M}(X)}{2}}},$$

where $\overline{X}_{0M}$, $\text{var}_{0M}(X)$, $\overline{X}_{1M}$, and $\text{var}_{1M}(X)$ denote the means and variances in the treated and untreated groups *after* matching, respectively. If the covariates are balanced, we expect a reduction in the standardized bias (although there is no formal criteria for how much reduction should occur).

A second balancing test proposed by Rosenbaum and Rubin (1985) is to examine if there are significant differences in covariate means between the treated and matched untreated groups, using two-sample *t*-tests (see, e.g., Ham et al. 2005). While we expect differences to exist before matching, there should be no significant differences after matching as the covariates should be balanced in both groups.

A third test is a Hotelling *T*-squared test of the joint null of equal means of all the *X*'s between the treated and the matched or reweighted untreated group (see, e.g., Smith and Todd 2005b).

A fourth test is comparing the pseudo-*R*-squared of the original model used to create propensity scores with a re-estimated model using only the matched observations (Sianesi 2004). Since the pseudo-*R*-squared indicates how well the *X*'s predict the probability of participation, the second pseudo-*R*-squared value should be

---

comparison group if these additional observations increase the second variance term in the denominator.

low since we do not expect systematic differences in the distribution of covariates between the treated and matched untreated groups.

A fifth test proposed by Dehejia and Wahba's (1999, 2002) is balancing the *X*'s within certain strata. They first divide the treated and untreated observations into an arbitrary number of strata based on their propensity scores, such that within each stratum, there is no statistically difference in the mean propensity scores between the groups. Then, within each stratum, for each covariate, they use a series of *t*-tests to examine if there is any significant difference in the means between the two groups. If they find any significant difference, they add higher order and interaction terms in the specification of the propensity score until no differences appear.

A sixth test is based on a regression framework. For example, Smith and Todd (2005b) first estimate the following regression:

$$X_k = \beta_0 + \beta_1 \hat{P}(X) + \beta_2 \hat{P}(X)^2 + \beta_3 \hat{P}(X)^3 + \beta_4 \hat{P}(X)^4$$
$$+\beta_5 D + \beta_6 D\hat{P}(X) + \beta_7 D\hat{P}(X)^2 + \beta_8 D\hat{P}(X)^3 + \beta_9 D\hat{P}(X)^4 + \eta$$

The joint null is that that the coefficient on all the terms with the treatment dummy *D* equals zero. In other words, after conditioning on the *X*'s, *D* should provide no information on X_k .

In our study, we examine two-sample *t*-tests for differences between means of the treated and matched untreated observations. We also compare the standardized bias before and after matching, and examine the pseudo-*R*-squared of the propensity score model re-estimated after matching, using reweighted observations on the common support only.

Sensitivity analysis: Unobserved heterogeneity

Given that the validity of the matching estimates hinges on whether the CIA assumption holds, an analysis of the sensitivity of our results to departures from this

identifying assumption is crucial. In this context, a bounding approach suggested by Rosenbaum (2002) is increasingly used in matching applications. While the bounding approach cannot test the CIA assumption per se, it examines the extent to which the statistical significance the results hinges on this untestable assumption. Suppose the probability of participation P_i is given by

$$\mathrm{P}(D_i = 1 \mid X_i) = F(\beta X_i + \gamma U_i),$$

where X_i are the observed characteristics for individual i, U_i are unobserved characteristics, and β and γ are the impacts of X_i and U_i on the participation decision. If there are no unobservable characteristics that affect participation, i.e., $\gamma = 0$, then two individuals with the same set of observable characteristics X have the same probability of participation. However, if $\gamma \neq 0$, i.e., there are unobservable characteristics that affect participation, then two individuals with the same X have differing probabilities of participation. Assuming, for simplicity, that F is the logistic distribution, the odds that two individuals i and j participate are given by $\mathrm{P}_i / (1-\mathrm{P}_i)$ and $\mathrm{P}_j / (1-\mathrm{P}_j)$ respectively. Then, the odds ratio can be written as

$$\frac{\mathrm{P}_i / (1-\mathrm{P}_i)}{\mathrm{P}_j / (1-\mathrm{P}_j)} = \mathrm{P}_i(1-\mathrm{P}_j) / \mathrm{P}_j(1-\mathrm{P}_i) = \frac{\exp(\beta X_i + \gamma U_i)}{\exp(\beta X_j + \gamma U_j)}.$$

If i and j form a matched pair, then the X vector cancels out and the odds ratio can be simply

written as $\exp[\gamma(U_i - U_j)]$. If there are no differences in unobservables, i.e. $U_i = U_j$, or the unobservable factors do not affect the probability of participating, i.e. $\gamma = 0$, the odds ratio equals 1, implying that the matching estimates do not suffer from unobserved selection bias. However, if this is not the case, then the matching estimates are said to suffer from a "hidden bias". In this context, Rosenbaum (2002) shows that the following bounds can be placed on the odds ratio:

$$\frac{1}{e^{\gamma}} \leq \frac{\mathrm{P}_i / (1-\mathrm{P}_i)}{\mathrm{P}_j / (1-\mathrm{P}_j)} \leq e^{\gamma}.$$

As e^{γ} increases, the bounds move apart reflecting uncertainty due to the presence of unobserved selection bias. Thus, e^{γ} is a measure of the extent to which the analysis suffers from this bias.

In our study, we follow Rosenbaum's approach to test the sensitivity of the significance of our results, if any, to violations of the CIA assumption. For binary outcomes as in our study, Aakvik (2001) suggests using the Mantel and Haneszel (MH) test statistic. The treatment effect on outcome Y is said to be significant if it crosses some test statistics $t(D,Y)$, where D is a dummy variable denoting program participation. Let n_1 and n_0 be the numbers of treated and non-treated units, where $n = n_1 + n_0$. Let y_1 and y_0 be the numbers of treated and untreated units where the binary outcome variable takes value 1, and where $y = y_1 + y_0$. The test statistic Q_{MH}, which asymptotically follows the normal distribution, is given by

$$Q_{MH} = \frac{|y_1 - E(y_1)| - 0.5}{\sqrt{Var(y_1)}} = \frac{\left|y_1 - \frac{n_1 y}{n}\right| - 0.5}{\sqrt{\frac{n_1 n_0 y(n-y)}{n^2(n-1)}}}.$$

Rosenbaum (2002) shows that Q_{MH} can be bounded by two known distributions. If $e^{\gamma} = 1$, the bounds are the same as the base scenario (i.e. no bias due to unobservables). However, as e^{γ} increases, the bounds move apart. Let $Q_{MH}{}^{+}$ be the test statistic in the case that we have overestimated the treatment effect and $Q_{MH}{}^{-}$ if we have underestimated the treatment effect. The two bounds are then given by:

$$Q_{MH}{}^{+} = \frac{\left|y_1 - \tilde{E}^{+}\right| - 0.5}{\sqrt{Var\left(\tilde{E}^{+}\right)}}$$

and

$$Q_{MH}{}^{-} = \frac{\left|y_1 - \tilde{E}^{-}\right| - 0.5}{\sqrt{Var\left(\tilde{E}^{-}\right)}},$$

where $\tilde{E}$ and $Var\left(\tilde{E}\right)$ are approximations of the expectation and variance of the number of treated units where the outcome variable takes value 1 in the population for given values of γ.

4.4 Data, sample, and findings

4.4.1 Data and sample

The data come from the second wave of a small-scale household and grain bank survey implemented in Kashipur block in rural Rayagada, Orissa.[95] The second wave was conducted towards the end of the "hungry" season, when the need for consumption credit is likely to be at its peak. 499 households across a total of 26 villages were sampled during this wave. The villages included 13 villages with operational grain banks at the time of the survey (referred to as grain bank villages) and 13 where grain banks were non-operational at the time of the survey (referred to as non-grain bank villages).

In grain bank villages, grain banks have been continuous operation for an average duration of 16.6 years; with a range of 8 to 23 years. Among these villages, 10 had grain banks that had been in continuous operation for over 10 years, which we refer to as "long-lived" grain bank villages. The non-grain bank villages include 10 villages where grain banks had failed at least 5 years prior to the survey, and 3 villages where grain banks were never established. The average operational duration of the failed grain banks was 4.5 years, with a maximum duration of 8 years. Table 4.3 provides descriptive statistics for our grain bank sample.

Of the total sample of 499 households in our survey, 250 households were in grain bank villages and 249 households in non-grain bank villages. Of the 250

[95] Details of the survey and sampling methodology are provided in a separate appendix.

households in grain bank villages, 194 households were in long-lived grain bank villages.

Table 4.3: Duration of grain bank survival (in years)

Grain bank status	*N*	Mean	S.D.	Min.	Max.
Operational grain banks	13	16.6	4.7	8	23
Long-lived operational grain banks	10	18.8	2.9	13	23
Non-operational grain banks	10	4.5	2.4	0	8

Of the total sample of households, 224 households reported having at least one currently outstanding loan, with 9 of these households having two loans, bringing the total number of outstanding loans to 233. Table 4.4 presents the distribution of loans by source. Fifty-seven percent of outstanding loans were provided by private moneylenders, making them the most commonly used source of credit. This finding applies equally for both grain bank and non-grain bank villages – no statistically significant difference is observed in the distribution of credit by source between the two types of villages.

Table 4.4: Source of loan by grain bank status

Source	All loans	Loans in GBVs	Loans in NGBVs
	Share of loans (in percent)		
Local moneylender	57.1	53.0	60.2
Government scheme	40.3	43.0	38.3
Other (including self-help group, friend/family, etc.)	2.6	4.0	1.5
Total	100.0	100.0	100.0

Notes: Statistics are conditional on households having one or more outstanding loans at the time of the survey. GBV and NGBV denote grain bank and non-grain bank village, respectively.

4.4.2 Naïve results

In this subsection, we present our estimates of the impact of grain bank participation on the incidence of borrowing from moneylenders without the use of matching. The

findings are presented largely to serve as a comparison to our main matching-based estimates.

We implement two related unconditional analyses. First, we compare the incidence of borrowing from moneylenders between participant households in *all* grain bank villages with households in non-grain bank villages. Second, we compare the incidence of borrowing between participant households in long-lived grain bank villages to households in non-grain bank villages. Our hypothesis is that the estimate of the difference in the incidence from the second analysis should be quantitatively larger, since we expect a more pronounced impact on borrowing from moneylenders in villages where grain banks have been in continuous operation for a longer duration.

Table 4.5 presents the incidence of borrowing from moneylenders by grain bank availability and participation status for all villages. Twenty-one percent of households in grain bank villages report having an outstanding loan from the local moneylender at the time of the survey, whereas 22 percent of households in non-grain bank villages do the same. We find that the two shares do not appear to be statistically different from each other.

This result is virtually unchanged when we eliminate households in grain bank villages that are non-participants (19 observations). We do so because we believe that the moneylender can potentially price discriminate between households within the same village (and indeed, as presented in Section 4.2, moneylenders are observed to charge a wide range of interest rates to households within the same region).

Therefore, although we do not have the data to examine our claim rigorously, we believe that the impact of grain bank establishment on non-participants may be different than for participants within grain bank villages. We find that roughly 16 percent of non-participant households have outstanding moneylender loans. This

share does not appear to be statistically different from the share of participant households with moneylender loans.

Table 4.5: Incidence of borrowing by households from local moneylenders, by grain bank availability and participation status

Village type	Share of households (percent)
Households in grain bank villages (1)	20.97
Participant households in grain bank villages (2)	21.29
Households in non-grain bank villages (3)	22.18
Participant households in long-lived grain bank villages (4)	17.70
H_0: Share (1) – Share (3) = 0. H_A: Share (1) – Share (3) < 0. *t*-statistic = 0.57; *p*-value = 0.2956.	
H_0: Share (2) – Share (3) = 0. H_A: Share (2) – Share (3) < 0. *t*-statistic = 0.38; *p*-value = 0.3516.	
H_0: Share (4) – Share (3) = 0. H_A: Share (4) – Share (3) < 0. *t*-statistic = 1.95; *p*-value = 0.0259.	

Notes: Estimates are corrected for sampling weights. Two households in non-grain banks villages reported having 2 outstanding loans from moneylenders, at the time of survey.

Table 4.5 also presents the incidence of borrowing from moneylenders separately for participant households in grain bank villages and households in non-grain bank villages. We find that 21 percent of participant households have outstanding moneylender loans. Once again, we find that the share does not appear to be statistically different from the corresponding share for households in non-grain bank villages.

Finally, Table 4.5 also presents the incidence of borrowing from moneylenders for households in non-grain bank villages. Of the 194 households in long-lived grain bank villages, 18 are grain bank non-participants. Thus, the sample size of participant households is 176. We find that 18 percent of participant households in long-lived grain bank villages have outstanding moneylender loans. As expected, this share is smaller than the corresponding share for participant households from all grain banks villages. Furthermore, we find that the difference in the shares between participant

households in long-lived grain bank villages and households in non-grain bank villages is statistically significant at the 5 percent level.

The question that follows is how grain banks affect the terms of moneylender loan contracts offered to borrower households. Due to the small sample size, we are unable to implement a rigorous analysis of the terms of the contracts offered by moneylenders. However, in Tables 4.6 and 4.7, we present selected summary statistics on loan amounts and interest rates on moneylender loans. We present these statistics for three groups of borrower households: participant households, non-participant households in grain bank villages and households in non-grain bank villages.

In Table 4.6, we present descriptive statistics for the amounts borrowed from moneylenders. While we do not find a significant difference between the average loan size from participant and non-participant households in grain bank villages, we find that the average loan size for participant households is statistically significantly lower than that for households in non-grain bank villages at the 1 percent level, using a one-sided *t*-test of difference in means.

Table 4.6: Loan amounts from moneylenders, by grain bank availability and participation status

Participation status	*N*	Mean	S.D.	Min.	Max.
Participant households in grain bank villages (1)	53	763.87	637.94	100.00	3000.00
Non-participant households in grain bank villages (2)	3	878.92	294.92	400.00	1000.00
Households in non-grain bank villages (3)	76	1329.74	1323.86	50.00	7000.00

H_0: Mean (1) – Mean (2) = 0. H_A: Mean (1) – Mean (2) < 0.
t-statistic = 0.33; *p*-value = 0.3697.

H_0: Mean (1) – Mean (3) = 0. H_A: Mean (1) – Mean (3) < 0.
t-statistic = 3.92; *p*-value = 0.0001.

Notes: Statistics are conditional on having an outstanding loan at the time of survey. Estimates are corrected using sampling weights. Loan amounts are in Indian rupees. Loans were provided in cases, except in 2 cases in non-grain bank villages where loans were provided in the form of millet grain.

As a point of comparison, we calculate the cash value of current grain stocks in grain banks. The prices used for this exercise are taken from the post-harvest season. Dividing the total cash value by the number of grain bank member households, we find it to be about 330 Indian rupees per household. This can be interpreted as the base amount by which grain bank loans can displace moneylender loans, as the displacement amount is potentially larger as the price of grains is higher in the lean season.

In Table 4.7, we present descriptive statistics for the annual interest rate charged by moneylenders. We find that the mean interest rates charged by moneylenders to grain bank participants and non-participants (whether in grain bank villages or in non-grain bank villages) are not statistically different. However, our sample sizes are too small, especially for non-participants in grain bank villages, to draw any firm conclusions regarding the impact of grain bank presence and participation on the terms of contracts offered by moneylenders.

Table 4.7: Interest rates charged on moneylender loans, by grain bank availability and participation status

Participation status	*N*	Mean	S.D.	Min.	Max.
Participant households in grain bank villages (1)	45	50.51	29.02	10.00	150.00
Non-participant households in grain bank villages (2)	3	51.26	3.07	50.00	56.25
Households in non-grain bank villages (3)	62	46.45	27.68	5.00	250.00

H_0: Mean (1) – Mean (2) = 0. H_A: Mean (1) – Mean (2) < 0.
t-statistic = 0.09; *p*-value = 0.4629.

H_0: Mean (1) – Mean (3) = 0. H_A: Mean (1) – Mean (3) < 0.
t-statistic = -0.90; *p*-value = 0.8154.

Notes: Statistics are conditional on having an outstanding loan at the time of survey. Estimates are corrected using sampling weights. Annual interest rates calculated using data on loan amounts, interest payments and loan duration, as reported by borrower households, conditional on data availability on interest payments. Such data were missing for 8 loans borrowed by participant households and 14 loans borrowed by households in non-grain bank villages.

Again, as a point of comparison, using data from member households, we find that the interest paid on consumption loans taken from grain banks during the lean season is about half the interest rate that they would pay to moneylenders for the same loans.

To summarize, our naïve estimates indicate that grain bank participation reduces dependence on borrowing from moneylenders. The effect is however only statistically significant for participant households from long-lived grain bank villages. We find that the average loan amount is significantly smaller for participant households compared to households in non-grain bank villages. However, we do not find any statistical difference when we compare the average interest rates for participant and non-participant households.

4.4.3 Matching-based results

In this section, we examine the impact of grain bank participation on the incidence of borrowing from moneylenders using propensity score matching methods. Below, we discuss the steps involved.

Calculating propensity scores

Using a binomial probit regression model, we first estimate propensity scores to match participant households in grain bank villages to households in non-grain bank villages. We estimate the model for two different samples: one where the treatment group comprises of participant households in grain bank villages (Sample 1) and the other where the treatment group comprises of participant households in long-lived grain bank villages (Sample 2). The comparison group is the same for both samples, namely households in non-grain bank villages.

In specifying the propensity score model, we include a number of covariates which predict both the decision to participate as well as the outcome. In choosing the

set of covariates, we have to balance the benefit of improving the predictive ability of the model with the cost of reducing the region of common support. In specification 1, we include a number of variables that capture different aspects of household wealth which we consider to be important determinants of the outcome variable, i.e., borrowing from the moneylender.

Household asset variables have the potential to predict grain bank participation in grain bank villages, as nonparticipants in grain bank villages report that their main reasons for not participating are either that they did not face food shortages, or that they were not creditworthy. However, since we eliminate non-participant households in grain bank villages from our estimation, household assets cannot explain participation for the remaining households. Village-level and grain bank-level variables which explain grain bank survival, are more important for predicting participation. However, three of the sample non-grain bank villages never had grain banks and therefore all households from these villages have missing information for grain bank-level variables. In addition, data on other grain bank features (e.g., share of women in the grain bank management committee at inception) are missing for 7 out of the 11 sample non-grain bank villages. Therefore, including grain bank variables would reduce the sample of control observations considerably and so we choose not to include them. In addition, from Chapter 2, we find that grain bank level variables, for the most part, do not have a significant impact on the likelihood of grain bank survival. We do, however, include village-level variables that may be important predictors of grain bank participation in a separate specification. We refer to this as specification 2. Below, we discuss the covariates that are included in the estimation of propensity scores.

Discussion of control variables

The control variables chosen to estimate the propensity scores include a number of variables indicating different aspects of household wealth, including human capital and physical assets, as well as village-level variables. Since our interest here is not statistical inference but rather the predictive accuracy of the probit model, we are not concerned about the potential presence of multicollinearity from say including different measures of household wealth in additive fashion in the same specification. The control variables included can be grouped as follows:

1. Human capital assets, such as household size, number of adult males (aged 15 years or more), number of adult females (age 15 years or more), the age of the head of the household, the highest level of formal education of any member within the household (in years), whether the household belongs to a Scheduled Tribe or not; and

2. Physical assets, such as the household's holdings of agricultural land of different qualities (infertile *dongar* land, moderately fertile *goda* land and fertile *bila* land), share of the agricultural land that is irrigated, amount of gold holdings, the number of rooms in the household's dwelling, flooring quality (whether *pucca* (cemented) or not), number of ploughs owned, number of spades owned, number of axes owned, number of cows owned, number of goats owned, number of bullocks owned.

3. Village-level variables, such as the level of isolation of the village, as measured by the time taken to travel to the seat of local government (i.e., the block headquarters); distance of the village from the closest Agragamee field office; whether any member of the village had been elected as a political

representative in the *gram panchayat* system (namely, as a Ward member) in the 5 years preceding the survey; and whether village-level meetings are held on a frequent/as needed basis or not.[96]

Since this is an agricultural society, some of the most important assets are landholdings and agricultural tools. The latter are simple hand tools. For this reason, we also include the number of adults as an important human capital asset. We include the number of adult males and females separately as they perform different agricultural activities. Two other important indicators of wealth which we include are livestock holdings as well as gold holdings (typically, in the form of jewelry). We also include a variable for flooring quality, which is a binary indicator variable for whether the floor was cemented (*pucca*) or made of mud/earth.

Based on findings from Chapter 2 that an increase in the share of women on the grain bank committee increases the probability of grain bank survival, we also hypothesize that the number of adult females in the household may impact the participation decision. We also hypothesize that the different village-level variables affect grain bank survival (which determines the participation decision in our case).

Tables 4.8 and 4.9 present the model of participation used to create propensity scores for the matching algorithm for samples 1 and 2 respectively. In each table, column (1) presents the results for the estimation for specification 1, i.e., household asset variables only. Column (2) presents the results for the estimation for specification 2, i.e., household asset and village-level variables. From the fit statistics presented at the bottom of the tables, we see that in both cases, the likelihood-ratio

[96] We do not include village-level variables in the final specification which do not satisfy the property of balanced means for the matched control and treatment samples. These include variables such as the total number of households, share of households that do not own any deeded land, distance from the closest weekly market, and the quality of the village road.

index (or McFadden's pseudo-R^2) is higher for specification 2. This implies that the model is a better fit when village-level variables are included. Therefore, we estimate the remaining results using specification 2.

Distribution of predicted propensity scores

Figures 4.1 and 4.2 depict the distribution of the propensity scores for participant households as well as households in non-grain bank villages for Samples 1 and 2, respectively. We see that participant households have a higher probability mass at higher levels of the propensity score, while households in non-grain bank villages have a higher probability mass at lower levels of the propensity score. This indicates that based on the set of observable characteristics used to create the propensity scores, households in the treatment group differ from households in the comparison group. Thus, there is a potential gain from using matching estimators compared to ordinary least squares.

Common support constraint and balancing tests

After the propensity scores are generated, the common support restriction is implemented, so that the test of the balancing property is performed only on observations whose propensity score belongs to the intersection of the supports of the propensity score of treatment and comparison units. To do this, treatment observations whose propensity score is higher that the maximum or less than the minimum propensity score of comparison observations are dropped.

The balancing property of the different specifications is examined using three difference tests: (1) *t*-tests for difference in covariate means between the matched treatment and comparison samples; (2) standardized bias before and after matching; and (3) pseudo-*R*-squared of the propensity score model *after* matching, using.

Table 4.8: Determinants of grain bank participation estimated for propensity score matching (Sample 1)
Pseudo-MLE probit regression estimates

Dependent variable**:** Grain bank participant (1 = yes)

	Coefficients	
	Specification 1	Specification 2
Household human capital assets		
Household size	0.417**	0.369**
	(0.20)	(0.19)
Square of household size	-0.0395**	-0.0366**
	(0.018)	(0.017)
Number of adult females (15 years or older)	-0.151	-0.0440
	(0.12)	(0.11)
Number of adult males (15 years or older)	-0.0947	-0.125
	(0.12)	(0.11)
Age of head of household	-0.0406	0.000417
	(0.044)	(0.0081)
Square of age of head of household	0.000538	
	(0.00057)	
Highest level of education of any household member (years)	0.0966	0.0235
	(0.062)	(0.023)
Square of highest level of education of any household member (years)	-0.00584	
	(0.0067)	
Social group (1 = tribal)	0.588**	0.325
	(0.30)	(0.26)
Household physical assets		
Amount of fertile *(bila)* agricultural land (acres)	0.341**	0.0803
	(0.15)	(0.083)
Square of amount of *bila* agricultural land owned	-0.0861***	
	(0.030)	
Amount of moderately fertile (*goda*) agricultural land (acres)	-0.115	-0.116***
	(0.080)	(0.045)
Square of amount of *goda* agricultural land owned	0.00203	
	(0.010)	
Amount of infertile (*dongar*) agricultural land (acres)	-0.113	-0.111*
	(0.13)	(0.058)
Square of amount of *dongar* agricultural land owned	-0.00226	
	(0.030)	
Amount of gold holdings (gm)	0.0144	0.0359*
	(0.020)	(0.019)
Number of rooms in dwelling	-0.0655	-0.149*
	(0.090)	(0.090)

Table 4.8 (Continued)

Flooring quality (1 = *pucca*)	0.373	0.328
	(0.25)	(0.25)
Number of ploughs owned	0.0747	-0.0389
	(0.15)	(0.16)
Number of spades owned	0.151*	0.150*
	(0.084)	(0.085)
Number of axes owned	0.0344	0.0433
	(0.075)	(0.079)
Number of cows owned	-0.144***	-0.158***
	(0.044)	(0.046)
Number of goats owned	-0.000595	-0.00129
	(0.026)	(0.024)
Number of bullocks owned	0.221***	0.225***
	(0.081)	(0.085)
Village-level variables		
Time taken to travel to seat of local government (minutes)		-0.00365***
		(0.00075)
Distance to closest Agragamee field office (*km*)		-0.0221
		(0.017)
Village member elected as local political representative (1 = yes)		0.144
		(0.15)
Frequency of village meetings (1 = frequent/as needed)		-0.593***
		(0.12)
Constant	-1.142	0.335
	(0.96)	(0.69)
Wald χ^2	56.75	122.54
p-value	0.0002	0.0000
McFadden's Pseudo-R^2	0.0976	0.1983
N	425	

Notes: Outcome variable is an indicator of borrowing from the private local moneylender. Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 4.9: Determinants of grain bank participation estimated for propensity score matching (Sample 2)
Pseudo-MLE probit regression estimates

Dependent variable: Grain bank participant (1 = yes)

	Coefficients	
	Specification 1	Specification 2
Household human capital assets		
Household size	0.512**	0.418*
	(0.23)	(0.24)
Square of household size	-0.0482**	-0.0422*
	(0.020)	(0.022)
Number of adult females (15 years or older)	-0.153	0.0423
	(0.12)	(0.13)
Number of adult males (15 years or older)	-0.0917	-0.0908
	(0.13)	(0.14)
Age of head of household	-0.0508	-0.00213
	(0.048)	(0.0093)
Square of age of head of household	0.000657	
	(0.00061)	
Highest level of education of any household member (years)	0.0959	0.0448*
	(0.067)	(0.026)
Square of highest level of education of any household member (years)	-0.00425	
	(0.0072)	
Social group (1 = tribal)	0.534*	-0.211
	(0.30)	(0.32)
Household physical assets		
Amount of fertile *(bila)* agricultural land (acres)	0.341**	0.194**
	(0.16)	(0.092)
Square of amount of *bila* agricultural land owned	-0.0783**	
	(0.032)	
Amount of moderately fertile (*goda*) agricultural land (acres)	-0.230***	-0.236***
	(0.081)	(0.061)
Square of amount of *goda* agricultural land owned	0.00934	
	(0.0089)	
Amount of infertile (*dongar*) agricultural land (acres)	-0.174	-0.171**
	(0.14)	(0.082)
Square of amount of *dongar* agricultural land owned	0.00283	
	(0.035)	
Amount of gold holdings (gm)	0.0148	0.0459**
	(0.022)	(0.022)
Number of rooms in dwelling	0.0589	-0.0503
	(0.098)	(0.10)

Table 4.9 (Continued)

Flooring quality (1 = *pucca*)	0.281	0.0823
	(0.26)	(0.27)
Number of ploughs owned	0.0796	-0.143
	(0.17)	(0.19)
Number of spades owned	0.157*	0.197*
	(0.090)	(0.10)
Number of axes owned	0.0115	-0.0656
	(0.083)	(0.093)
Number of cows owned	-0.157***	-0.202***
	(0.049)	(0.061)
Number of goats owned	0.00379	0.0196
	(0.025)	(0.024)
Number of bullocks owned	0.246***	0.312***
	(0.090)	(0.10)
Village-level variables		
Time taken to travel to seat of local government (minutes)		-0.0116***
		(0.0015)
Distance to closest Agragamee field office (*km*)		-0.107***
		(0.030)
Village member elected as local political representative (1 = yes)		0.0886
		(0.16)
Frequency of village meetings (1 = frequent/as needed)		-0.563***
		(0.13)
Constant	-1.386	1.714**
	(1.02)	(0.84)
Wald χ^2	58.88	162.88
p-value	0.0001	0.0000
McFadden's Pseudo-R^2	0.1199	0.3523
N	480	

Notes: Outcome variable is an indicator of borrowing from the private local moneylender. Estimates are corrected for sampling weights. Standard errors reported in parentheses. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

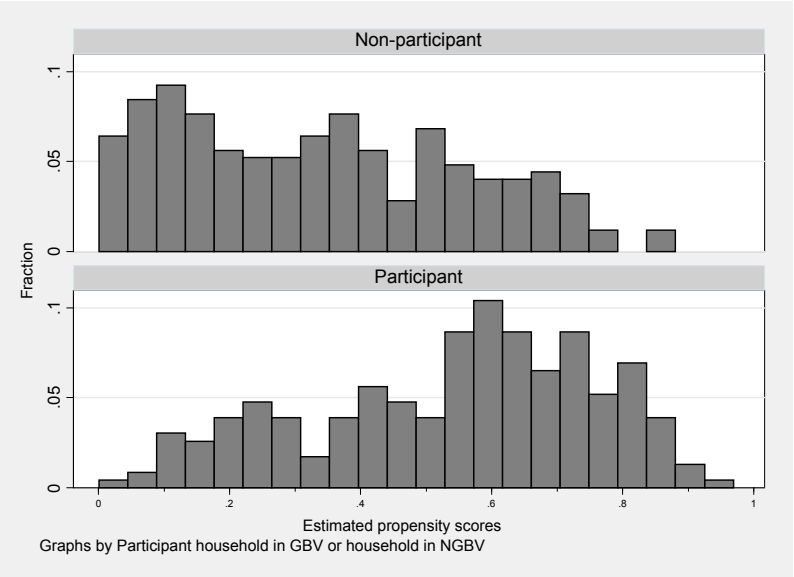


Figure 4.1: Distribution of predicted propensity scores (Sample 1)

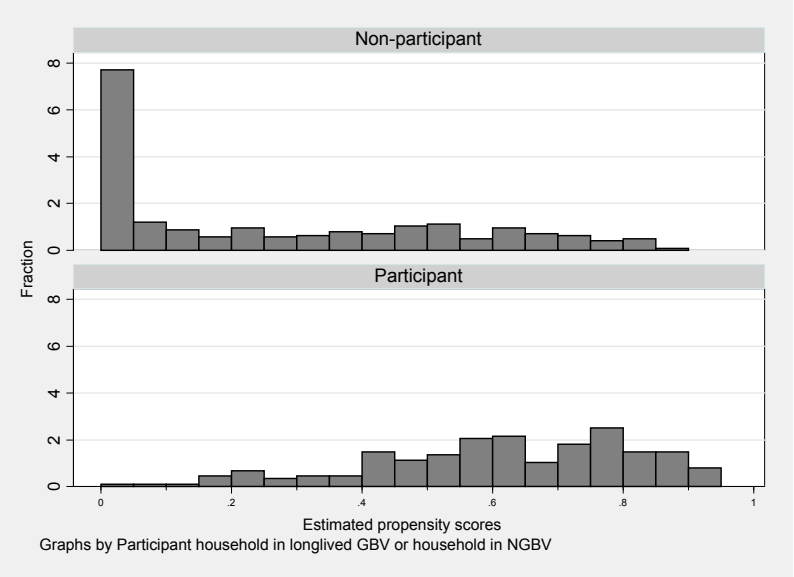


Figure 4.2: Distribution of predicted propensity scores (Sample 2)

observations in the common support region only as well as weights generated from the matching algorithm

Results of balanced covariates using *t*-tests are presented in Tables 4.A1-4.A4. Examining the *t*-test results, we only include those variables in the final specification that have no statistically significant difference in means.

The results using measures of pseudo *R*-squared and standardized bias are presented in Table 4.A5. Examining the pseudo *R*-squared by re-estimating the propensity score model after matching, we find that in all cases, the pseudo *R*-squared generated is much lower than the pseudo *R*-squared generated prior to matching. In addition, the joint significance of the covariates in the model is always rejected. Prior to matching, the joint significance of the covariates in the model was never rejected. Finally, examining the median standardized bias before and after matching, we find that that it is always lower after matching, and never above a value of 8, which is an acceptable value (Smith and Todd 2005b).

Average impact of grain bank participation

Table 4.10 presents local linear regression matching estimates of the average impact of participation in grain banks on the incidence of borrowing from moneylenders. Column (1) presents the matching results for Sample 1, where the treatment sample is composed of participant households from all grain bank villages. We find that, on average, the incidence of borrowing from moneylenders among participant households is 10 percentage points lower than among households in non-grain bank villages, with this difference statistically significant at the 10 percent level. Translating this absolute difference into relative terms, the incidence of borrowing from moneylenders is about 30 percent lower for participant households.

Table 4.10: Average treatment on the treated (*ATT*): Impact of grain bank participation on the incidence of borrowing from moneylenders
Local linear regression matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	0.238		0.198	
Average outcome, non-participants	0.338		0.388	
Difference in average outcomes (*ATT*)	-0.100*		-0.190***	
	(0.055)		(0.060)	
	Off support	On support	Off support	On support
No. of treated units	8	223	9	167
No. of comparison units	0	249	0	249

Notes: Standard errors are provided in parentheses. Estimates are for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Comparing the estimates in Column (1) with the difference in the mean incidences of borrowing that are presented in Table 4.5, we find that the estimate for the incidence of borrowing by non-participant households in much larger after matching. We find that very few observations are off the common support, indicating that it is not the support restriction that results in the larger estimate. This implies that the ATT estimate generated by the matching algorithm is driven by the reweighting process involved in kernel-weighted matching. Thus, a simple comparison of means between participant and non-participant households would not have revealed the program impact revealed when similar households are compared (or rather, households are weighted in proportion to how closely they "match" one another).

Column (2) presents the matching results for Sample 2, where the treatment sample is confined to participant households from long-lived grain bank villages. We find that, on average, the incidence of borrowing from moneylenders is 19 percentage points lower, and that this effect is highly statistically significant. In relative terms, the incidence of borrowing from moneylenders is about 49 percent lower for participant households in long-lived grain bank villages. This finding confirms our hypothesis that the negative effect on the incidence of borrowing from moneylenders

should be quantitatively larger when we examine participants in villages where grain banks have been continuously operational for a longer duration.

Again, we find that the program impact estimated in Column (2) is much larger than the difference in the mean incidences of borrowing that are presented in Table 5. We find, once again, that very few observations are off the common support, which implies that the large program impact that is estimated is driven by the reweighting of households during the matching process.

Robustness to alternative specifications

Estimates of the average impact of participation in grain banks on the incidence of borrowing from moneylenders using kernel matching estimation are presented in Table 4.11. We find that these results are similar to those in Table 4.10 using local linear matching.

Table 4.11: Average treatment on the treated (*ATT*): Impact of grain bank participation on the incidence of borrowing from moneylenders
Kernel matching estimates using propensity scores

	(1) Sample 1		(2) Sample 2	
Average outcome, participants	0.238		0.198	
Average outcome, non-participants	0.344		0.382	
Difference in average outcomes (ATT)	-0.106**		-0.184***	
	(0.050)		(0.058)	
	Off support	On support	Off support	On support
No. of treated units	8	223	9	167
No. of comparison units	0	249	0	249

Notes: Standard errors are provided in parentheses. Estimates are for matched sample having common support only. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Sensitivity to choice of bandwidth

The results in Tables 4.10 and 4.11 were estimated using a bandwidth size of 0.06. We re-estimate the local linear regression and kernel matching models with varying

bandwidth size of the kernel (0.05 and 0.07). The results are presented in Table 4.A6. We find that neither the magnitude nor the statistical significance of the results is sensitive to the choice of the bandwidth size in the range examined.

Sensitivity to unobserved heterogeneity

Table 4.A7 presents results of a sensitivity analysis of the matching estimates to unobserved heterogeneity using Rosenbaum bounds. We examine the cases for $1 \leq \Gamma \leq 5$ (in increments of 0.5), where $\Gamma > 1$ is a measure of the extent of unobserved selection bias. The Q_{MH}^{+} statistic adjusts the MH statistic downward for positive (unobserved) selection (i.e., villages where grain banks are likely to survive are also those where households are less likely to rely on borrowing from moneylenders, regardless of grain bank presence – therefore leading to an overestimation of the treatment effect). The Q_{MH}^{-} statistic adjusts the MH statistic upward for negative (unobserved) selection (i.e., villages where grain banks are likely to survive are also those where households are more likely to rely on borrowing from moneylenders, regardless of grain bank presence – therefore leading to an underestimation of the treatment effect). At $\Gamma = 1$ (i.e., no hidden bias), we find that our estimate is statistically significant (for both Samples 1 and 2). Since what is of concern to us is whether we have (incorrectly) overestimated the treatment effect, we examine the Q_{MH}^{+} statistic as Γ increases. We find that it continues to be significant, leading us to conclude that our treatment effect estimates are not sensitive to unobserved heterogeneity across the range of Γ values examined. However, given the asymptotic properties of the test and the small sample size on which our analysis is based, we recognize that it does not rule out the presence of unobserved heterogeneity.

4.5 Conclusion

In the past two decades, grain banks have been adopted by rural and tribal development NGOs in Orissa, and more recently, by the Indian government's Ministry of Tribal Welfare, with the stated objective of combating short-term, seasonal food shortages. However, widespread anecdotal evidence regarding the benefits of grain banks in displacing private, informal moneylenders suggests that they may also have an additional, if unintended, positive effect. To date, there has been no quantitative evaluation of the displacement effects of grain banks on alternative sources of credit. In this chapter, we attempt to fill this gap in knowledge by measuring the average treatment effect of grain banks on the incidence of borrowing from moneylenders, traditionally the main source of credit, using propensity score matching estimators.

Our matching-based results indicate that grain banks indeed have large displacement effects on private, informal moneylenders. Local linear matching estimates indicate that on average the incidence of borrowing from moneylenders among participants in grain bank villages relative to our matched sample of households in non-grain bank villages was 10 percentage points (30 percent) lower. Further, we find a larger impact among participant households in long-lived grain bank villages (those in which grain bank were in operation for 10 years or more): specifically, local linear matching estimates indicate that, on average, the incidence of borrowing is 19 percentage points (49 percent) lower. These estimates are much larger than the mean differences in the incidence of borrowing between participant households and households in non-grain bank villages. Thus, the reweighting process by which similar households are matched is revealing the impact of grain banks in a way that a simple difference in means does not. Our matching-based results are robust to the use of an alternative matching estimator (kernel matching), the choice of bandwidth size, and the potential presence of unobserved heterogeneity that affects

both program participation and the outcome of interest. Interpreted in the context of the difference in the average borrowing levels from moneylenders between participant households and households in non-grain bank villages, which we find to be statistically significant, we conclude that grain banks do indeed displace borrowing from moneylenders.

Given the importance of short-term consumption loans and the evidence that grain banks offer these loans at lower interest rates, grain banks can indeed provide an attractive alternative to private, informal moneylenders. However, important caveats remain. First, grain banks are designed in order to provide consumption credit in the form of grains. As a result, they cannot act as a broad-based credit alternative which would resolve a potential problem of thin or missing credit markets in the region. Second, the data in this study do not permit the analysis of how grain banks affect the terms of contracts offered by moneylenders to non-participants residing in either grain bank villages or non-grain bank villages. Given the theoretical and empirical evidence from previous studies on the differential welfare impacts on different groups following the introduction of competition into previously monopolistic credit markets, future studies of how grain banks affect access to credit for both participants as well as non-participants can provide relevant policy information. Third, our data enable us to examine the difference in borrowing incidence between participants and non-participants residing in neighboring villages. As a result, a likely hazard affecting our estimates is contamination in the control group due to potential spillover effects. Our data do not enable us to draw conclusions about the extent of the problem. To enable a cleaner analysis of impact on moneylender reliance, there is need for data from non-grain bank villages having similar characteristics to grain bank villages but geographically isolated from them, so as to reduce the concern of spillover effects.

More broadly, there is a pressing need for data examining the structure of rural credit markets in tribal Orissa, which would permit analysis of the causes behind missing and thin financial markets in rural Rayagada (and tribal Orissa in general), as policy prescriptions hinge crucially on understanding why these deficiencies occur. If they are a function of low levels of development, as theorized by a large literature on credit and growth, financial innovation and intermediaries may emerge spontaneously, as needed.[97] If, however, as theorized by the literature on poverty traps, non-convexities in the technologies associated with institutional innovation exist, such that the market is too small or there is insufficient local capital, then there exists a continuing need for effective outside intervention in credit markets.[98] In the latter case, the grain bank intervention may indeed be a valuable policy intervention.

[97] See, e.g., Acemoglu and Zilibotti (1997).
[98] See Azariadis and Stachurs ki (2004) for a recent review of the literature on poverty traps.

APPENDIX

Table 4.A1: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on local linear regression propensity score matching

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Household size	5.09	5.09	0.00	-0.03
Square of household size	28.36	28.62	-0.26	-0.16
Number of adult females (15 years or older)	1.42	1.41	0.01	0.17
Number of adult males (15 years or older)	1.36	1.37	-0.01	-0.13
Age of head of household	35.57	35.83	-0.26	-0.32
Highest level of education of any household member (years)	2.81	2.86	-0.04	-0.15
Social group (1 = tribal)	0.97	0.98	-0.01	-0.71
Amount of fertile *(bila)* agricultural land (acres)	0.53	0.58	-0.05	-0.47
Amount of moderately fertile (*goda*) agricultural land (acres)	1.52	1.76	-0.25	-1.59
Amount of infertile (*dongar*) agricultural land (acres)	1.35	1.50	-0.14	-1.33
Amount of gold holdings (gm)	2.85	3.22	-0.37	-0.96
Number of rooms in dwelling	2.44	2.56	-0.12	-1.44
Flooring quality (1 = *pucca*)	0.12	0.10	0.02	0.72
Number of ploughs owned	0.89	0.87	0.02	0.31
Number of spades owned	2.30	2.31	-0.02	-0.20
Number of axes owned	2.09	2.11	-0.03	-0.30
Number of cows owned	0.81	0.84	-0.03	-0.18
Number of goats owned	0.96	1.02	-0.05	-0.24
Number of bullocks owned	1.53	1.55	-0.02	-0.17
Time taken to travel to seat of local government (minutes)	126.19	116.10	10.09	1.45
Distance to closest Agragamee field office (*km*)	4.41	4.80	-0.39	-1.16
Village member elected as local political representative (1 = yes)	0.65	0.72	-0.07	-1.56
Frequency of village meetings (1 = frequent/ as needed)	1.54	1.52	0.02	0.43

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 4.A2: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on local linear regression propensity score matching

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Household size	5.10	4.91	0.19	1.09
Square of household size	28.41	26.67	1.74	0.96
Number of adult females (15 years or older)	1.43	1.37	0.06	0.87
Number of adult males (15 years or older)	1.37	1.36	0.01	0.1
Age of head of household	35.49	35.37	0.12	0.13
Highest level of education of any household member (years)	3.12	2.70	0.41	1.22
Social group (1 = tribal)	0.96	0.97	-0.01	-0.66
Amount of fertile *(bila)* agricultural land (acres)	0.52	0.58	-0.06	-0.59
Amount of moderately fertile (*goda*) agricultural land (acres)	1.34	1.56	-0.21	-1.21
Amount of infertile (*dongar*) agricultural land (acres)	1.25	1.38	-0.13	-1.01
Amount of gold holdings (gm)	2.74	2.49	0.25	0.59
Number of rooms in dwelling	2.56	2.58	-0.02	-0.23
Flooring quality (1 = *pucca*)	0.14	0.08	0.06	1.66
Number of ploughs owned	0.89	0.92	-0.04	-0.55
Number of s pades owned	2.29	2.25	0.04	0.37
Number of axes owned	2.06	2.07	-0.01	-0.13
Number of cows owned	0.77	0.92	-0.15	-0.86
Number of goats owned	0.98	1.19	-0.21	-0.72
Number of bullocks owned	1.52	1.51	0.00	0.01
Time taken to travel to seat of local government (minutes)	93.23	86.57	6.66	1.55
Distance to closest Agragamee field office (*km*)	3.99	3.72	0.27	1.21
Village member elected as local political representative (1 = yes)	0.63	0.68	-0.05	-1.01
Frequency of village meetings (1 = frequent/ as needed)	1.48	1.42	0.06	1

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 4.A3: Balancing *t*-tests for propensity score model covariates, Sample (1) Matched treatment and comparison samples, based on kernel propensity score matching

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Household size	5.09	5.08	0.01	0.04
Square of household size	28.36	28.49	-0.13	-0.08
Number of adult females (15 years or older)	1.42	1.40	0.02	0.4
Number of adult males (15 years or older)	1.36	1.38	-0.03	-0.39
Age of head of household	35.57	35.61	-0.04	-0.05
Highest level of education of any household member (years)	2.81	2.68	0.13	0.44
Social group (1 = tribal)	0.97	0.98	-0.01	-0.64
Amount of fertile *(bila)* agricultural land (acres)	0.53	0.61	-0.07	-0.76
Amount of moderately fertile (*goda*) agricultural land (acres)	1.52	1.74	-0.23	-1.43
Amount of infertile (*dongar*) agricultural land (acres)	1.35	1.46	-0.10	-0.93
Amount of gold holdings (gm)	2.85	2.91	-0.07	-0.18
Number of rooms in dwelling	2.44	2.54	-0.10	-1.23
Flooring quality (1 = *pucca*)	0.12	0.11	0.01	0.21
Number of ploughs owned	0.89	0.88	0.01	0.18
Number of spades owned	2.30	2.29	0.00	0.05
Number of axes owned	2.09	2.15	-0.06	-0.63
Number of cows owned	0.81	0.84	-0.03	-0.21
Number of goats owned	0.96	1.04	-0.08	-0.34
Number of bullocks owned	1.53	1.53	0.00	-0.01
Time taken to travel to seat of local government (minutes)	126.19	115.73	10.46	1.51
Distance to closest Agragamee field office (*km*)	4.41	4.75	-0.33	-1
Village member elected as local political representative (1 = yes)	0.65	0.71	-0.06	-1.31
Frequency of village meetings (1 = frequent/ as needed)	1.54	1.53	0.02	0.34

Notes. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 4.A4: Balancing *t*-tests for propensity score model covariates, Sample (2) Matched treatment and comparison samples, based on kernel propensity score matching

	Means (Treated)	Means (Control)	Difference in means	*t*-statistic
Household size	5.10	4.95	0.15	0.87
Square of household size	28.41	27.00	1.41	0.78
Number of adult females (15 years or older)	1.43	1.37	0.06	0.82
Number of adult males (15 years or older)	1.37	1.36	0.01	0.1
Age of head of household	35.49	35.09	0.41	0.43
Highest level of education of any household member (years)	3.12	2.70	0.42	1.25
Social group (1 = tribal)	0.96	0.97	-0.02	-0.79
Amount of fertile *(bila)* agricultural land (acres)	0.52	0.60	-0.08	-0.72
Amount of moderately fertile (*goda*) agricultural land (acres)	1.34	1.56	-0.22	-1.24
Amount of infertile (*dongar*) agricultural land (acres)	1.25	1.38	-0.13	-0.99
Amount of gold holdings (gm)	2.74	2.50	0.24	0.56
Number of rooms in dwelling	2.56	2.61	-0.05	-0.57
Flooring quality (1 = *pucca*)	0.14	0.09	0.05	1.33
Number of ploughs owned	0.89	0.92	-0.03	-0.5
Number of spades owned	2.29	2.28	0.01	0.1
Number of axes owned	2.06	2.11	-0.05	-0.43
Number of cows owned	0.77	0.92	-0.15	-0.86
Number of goats owned	0.98	1.22	-0.24	-0.81
Number of bullocks owned	1.52	1.50	0.02	0.15
Time taken to travel to seat of local government (minutes)	93.23	86.62	6.62	1.52
Distance to closest Agragamee field office (*km*)	3.99	3.73	0.26	1.17
Village member elected as local political representative (1 = yes)	0.63	0.67	-0.04	-0.85
Frequency of village meetings (1 = frequent/ as needed)	1.48	1.44	0.04	0.66

Notes: * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 4.A5: Balancing tests using measure of pseudo *R*-squared and standardized bias

		Pseudo-R^2 (after matching)	Wald test *p*-value	Standardized bias Before matching	After matching
Local linear regression matching					
	Sample (1)	0.025	0.870	11.97	3.14
	Sample (2)	0.047	0.538	13.19	7.72
Kernel matching					
	Sample (1)	0.022	0.938	11.97	3.49
	Sample (2)	0.042	0.675	13.19	7.48

Notes: Tests only on observations in matched sample having common support.

Table 4.A6: Average treatment on the treated (*ATT*): Impact of grain bank participation on the incidence of borrowing
Sensitivity to choice of bandwidth

	(1) Sample (1)	(2) Sample (2)
Local linear regression matching estimates		
Bandwidth = 0.05	-0.096*	-0.180***
	(0.052)	(0.059)
Bandwidth = 0.07	-0.100*	-0.192***
	(0.051)	(0.059)
Kernel matching estimates		
Bandwidth = 0.05	-0.111**	-0.187***
	(0.051)	(0.058)
Bandwidth = 0.07	-0.103**	-0.180***
	(0.050)	(0.057)

Notes. Standard errors in parentheses. Estimates for matched sample having common support. * Statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level.

Table 4.A7: Sensitivity of treatment effect estimates to unobserved heterogeneity

	Sample (1)				Sample (2)			
$\Gamma = e^{\gamma}$	Q_{MH}^{+}	Q_{MH}^{-}	p_{MH}^{+}	p_{MH}^{-}	Q_{MH}^{+}	Q_{MH}^{-}	p_{MH}^{+}	p_{MH}^{-}
1.0	1.445	1.445	0.074	0.074	2.246	2.246	0.012	0.012
1.5	3.412	0.284	<0.001	0.388	4.008	0.530	<0.001	0.298
2.0	4.839	1.661	<0.001	0.048	5.296	0.439	<0.001	0.331
2.5	5.974	2.739	<0.001	0.003	6.324	1.375	<0.001	0.085
3.0	6.925	3.631	<0.001	<0.001	7.188	2.147	<0.001	0.016
3.5	7.749	4.398	<0.001	<0.001	7.937	2.808	<0.001	0.002
4.0	8.481	5.074	<0.001	<0.001	8.602	3.390	<0.001	<0.001
4.5	9.141	5.680	<0.001	<0.001	9.201	3.912	<0.001	<0.001
5.0	9.745	6.232	<0.001	<0.001	9.748	4.386	<0.001	<0.001

Notes: Γ : odds of differential assignment due to unobserved factors.

Q_{MH}^{+} : Mantel-Haenszel statistic (assumption: overestimation of treatment effect).

Q_{MH}^{-} : Mantel-Haenszel statistic (assumption: underestimation of treatment effect).

p_{MH}^{+} : significance level (assumption: overestimation of treatment effect).

p_{MH}^{-} : significance level (assumption: underestimation of treatment effect).

REFERENCES

Aakvik, Arild. 2001. Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics*. v63 (n1):115-43.

Acemoglu, Daron and Fabrizio Zilibotti. 1997. Was Prometheus Unbound by Chance? Risk, Diversification, and Growth. *Journal of Political Economy*. v105 (n4):709-751.

Akerlof, George A. 1970. The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*. v84 (n3): 488-500.

Aleem, Irfan. 1990. Imperfect Information, Screening, and the Costs of Informal Lending: A Study of a Rural Credit Market in Pakistan. *World Bank Economic Review*. v4 (n3):329-349.

Alexander, Gwendolyn. 2006. Here Today, Gone Tomorrow: Can Dynamic Incentives Make Microfinance More Flexible? *Journal of Development Economics*. v80 (n1):84-105.

Azariadis, Costas, and John Stachurski. 2004. Poverty traps. In Philippe Aghion and Steven Durlauf (eds.) *Handbook of Economic Growth* Volume 1. Amsterdam: North-Holland.

Banerjee, Abhijit V. 2003. Contracting Constraints, Credit Markets, and Economic Development. In Mathias Dewatripont, Lars Peter Hansen and Stephen Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications*. Eighth World Congress, Volume III, Cambridge, U.K.: Cambridge University Press.

Banerjee, Abhijit, Timothy Besley, and Timothy Guinnane. 1994. Thy Neighbor's Keeper: The Design of Credit Cooperatives with Theory and a Test. *Quarterly Journal of Economics*. v109 (n2):491-515.

Barham, Bradford L, Stephen Boucher, and Michael R. Carter. 1996. Credit Constraints, Credit Unions, and Small-Scale Producers in Guatemala. *World Development*. v24 (n5):793-806.

Basu, Kaushik. 1997. *Analytic Development Economics: The Less Developed Economy Revisited*. Cambridge, MA: MIT Press.

Besley, Timothy, and Stephen Coate. 1995. Group Lending, Repayment Incentives and Social Collateral. *Journal of Development Economics*. v46 (n1):1-18.

Besley, Timothy, Stephen Coate, and Glenn Loury. 1993. The Economics of Rotating Savings and Credit Associations. *American Economic Review*. v83 (n4):792-810.

Besley, Timothy. 1994. Savings, Credit and Insurance. In Jere Behrman and T.N. Srinivasan (eds.), *Handbook of Development Economics*. v3A: 2123-2207.

Black, Dan A., and Jeffrey A. Smith. 2004. How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching. *Journal of Econometrics*. v121 (n1-2):99-124.

Blundell, Richard, and Monica Costa Dias. 2002. Alternative Approaches to Evaluation in Empirical Economics. *Cemmap Working Paper* 10. University College London: Institute for Fiscal Studies.

Boucher, Stephen R., and Catherine Guirkinger. 2005. Credit Constraints and Productivity in Peruvian Agriculture. Department of Agricultural Economics, University of California-Davis.

Boucher, Stephen R., Michael R. Carter, and Catherine Guirkinger. 2005. Risk Rationing and Wealth Effects in Credit Markets. Agricultural and Resource Economics Working paper number 05-010, University of California-Davis.

Carter, Michael R. 1988. Equilibrium Credit Rationing of Small Farmer Agriculture. *Journal of Development Economics*. v28 (n1):83-103.

Coate, Stephen, and Martin Ravallion. 1993. Reciprocity without Commitment: Characterization and Performance of Informal Insurance Arrangements. *Journal of Development Economics*. v40 (n1):1-24.

Conning, Jonathan, and Michael Kevane. 2002. Community-based Targeting Mechanisms for Social Safety Nets: A Critical Review. *World Development*. v30 (n3):375-394.

Conning, Jonathan H., and Udry, Christopher R. 2005. Rural Financial Markets in Developing Countries. Yale University Economic Growth Center Discussion Paper No. 914.

Dehejia, Rajeev H., and Sadek Wahba. 1999. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*. v94 (n448):1053-62.

Dehejia, Rajeev H., and Sadek Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics*. v84 (n1):151-61.

Diagne, Aliou. 1999. Determinants of Household Access to and Participation in Formal and Informal Credit Markets in Malawi. FCND Discussion Paper No. 67. Washington DC: International Food Policy Research Institute.

Fan, Jianqing. 1992. Design-Adaptive Nonparametric Regression. *Journal of the American Statistical Association*. v87 (n420):998-1004.

Fröhlich, Markus. 2004. Finite Sample Properties of Propensity-score Matching and Weighting Estimators. *Review of Economics and Statistics.* v86 (n1):77-90.

Galdo, Jose, Jeffrey Smith, and Dan Black. 2006. Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data. Mimeo.

Gine, Xavier, and Stefan Klonner. 2005. Learning about Oneself: Technology Financing in a Tamil Fishing Village. Mimeo.

Ham, John C., Xianghong Li, and Patricia B. Reagan. 2005. *Propensity score matching, a distance-based measure of migration, and the wage growth of young men*. Federal Reserve Bank of New York, Staff Reports: 212.

Heckman, James, and Richard Robb. 1985. Alternative Methods for Evaluating the Impact of Interventions. In James Heckman and Burton Singer (ed.) *Longitudinal Analysis of Labor Market Data*. Cambridge University.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies*. v64 (n4): 605-654.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1998a. Matching as an Econometric Evaluation Estimator. *Review of Economic Studies*. v65 (n2):261-294.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998b. *Characterizing Selection Bias Using Experimental Data*. National Bureau of Economic Research, Inc, NBER Working Papers: 6699.

Hoff, Karla, and Joseph E. Stiglitz. 1997. Moneylenders and Bankers: Price-Increasing Subsidies in a Monopolistically Competitive Market. *Journal of Development Economics*. v52 (n2): 429-462.

Hoff, Karla, Avishay Braverman, and Joseph Stiglitz, eds. 1993. *The Economics of Rural Organization: Theory, practice, and policy*. Oxford; New York; Toronto and Melbourne: Oxford University Press for the World Bank.

Kaboski, Joseph P., and Robert M. Townsend. 2005. Policies and Impact: An Analysis of Village-Level Microfinance Institutions. *Journal of the European Economic Association*. v3 (n1):1-50.

Karlan, Dean, and Jonathan Zinman. 2007. Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment. Mimeo.

Kranton, Rachel E., and Anand V. Swamy. 1990. The Hazards of Piecemeal Reform: British Civil Courts and the Credit Market in Colonial India. *Journal of Development Economics*. v58 (n1):1-24.

Lechner, Michael. 2001. Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption. In Michael Lechner and Friedhelm Pfeiffer (eds.) *Econometric Evaluation of Labour Market Policies.* Physica, Heidelberg and New York.

Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall. 2002. Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. *Review of Economic Studies*. v69 (n1):209-244.

Mansuri, Ghazala, and Vijayendra Rao. 2004. Community-Based and -Driven Development: A Critical Review. *World Bank Research Observer*. v19 (n1):1-39.

McIntosh, Craig, Alain de Janvry, and Elisabeth Sadoulet. 2005. How Rising Competition among Microfinance Institutions Affects Incumbent Lenders. *Economic Journal*. v115 (n506):987-1004.

McIntosh, Craig, and Bruce Wydick. 2005. Competition and Microfinance. *Journal of Development Economics*. v78 (n2):271-298.

Meyer, Richard L., and Geetha Nagarajan. 2000. *Rural financial markets in Asia : policies, paradigms, and performance, [A study of rural Asia ; v. 3]*. Oxford; New York: Oxford University Press.

Meyer, Richard L., Geetha Nagarajan, and Leroy J. Hushak. 1997. Segmentation in the Philippine Informal Credit Markets: A Multinomial Logit Analysis. In Roger Rose, Carolyn Tanner and Margot Bellamy (eds.), *Issues in agricultural competitiveness: Markets and policies*. International Association of Agricultural Economists Occasional Paper no. 7. Aldershot, U.K, and Brookfield, Vt.: Ashgate.

Miguel, Edward, and Mary K. Gugerty. 2004. Ethnic Diversity, Social Sanctions and Public Goods in Kenya. Mimeo.

Millimet, Daniel, and Rusty Tchernis. 2006. *On the Specification of Propensity Scores: with an Application to the WTO-Environment Debate*. Center for Applied Economics and Policy Research, Economics Department, Indiana University Bloomington, Caepr Working Papers: 2006-013.

Morduch, Jonathan. 1995. Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives*. v9 (n3):103-114.

Morduch, Jonathan. 1999. Between the State and the Market: Can Informal Insurance Patch the Safety Net? *World Bank Research Observer*. v14 (n2):187-207.

Mude, Andrew G. 2006. Making Loans to Make Friends: Explaining the Dismal Financial Performance of Financial Service Associations. *Agricultural Finance Review*. v66 (n2):267-81.

Navajas, Sergio, Jonathan Conning, and Claudio Gonzalez-Vega. 2003. Lending Technologies, Competition and Consolidation in the Market for Microfinance in Bolivia. *Journal of International Development*. v15 (n6):747-770.

Petersen, Mitchell A. and Raghuram G. Rajan. 1995. The Effect of Credit Market Competition on Lending Relationships. *Quarterly Journal of Economics*. v110 (n2):407-443.

Platteau, Jean-Philippe. 2000. *Institutions, Social Norms, and Economic Development*. Reading: Harwood Academic.

Posner, Richard A. 1980. A Theory of Primitive Society. *Journal of Law and Economics*. v23 (n1):1-53.

Ray, Debraj. 1998. *Development Economics*. Princeton, N.J.: Princeton University Press.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. v70 (n1):41-55.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer.

Rubin, Donald B. and Neal Thomas. 1996. Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics*. v52 (n1): 249-264.

Sianesi, Barbara. 2004. An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *Review of Economics and Statistics*. v86 (n1):133-55.

Smith, Herbert. 1997. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. *Sociological Methodology*. v27 (n1):325-353.

Smith, Jeffrey A. 2004. Evaluation Local Economic Development Policies: Theory and Practice. Mimeo.

Smith, Jeffrey A., and Petra E. Todd. 2001. Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods. *American Economic* Review. v91 (n2):112-18.

Smith, Jeffrey A., and Petra E. Todd. 2005a. Does Matching Overcome LaLonde's Critique of Non-experimental Estimators? *Journal of Econometrics*. v125 (n1-2):305-53.

Smith, Jeffrey, and Petra Todd. 2005b. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? Rejoinder. *Journal of Econometrics*. v125 (n1-2):365-75.

Stiglitz, Joseph E. 1974. Incentives and Risk Sharing in Sharecropping. *Review of Economic Studies*. v41 (n2):219-55.

Stiglitz, Joseph E. 1990. Peer Monitoring and Credit Markets. *World Bank Economic Review*. v4 (n3):351-366.

Stiglitz, Joseph E. 1994. The Role of the State in Financial Markets. In Proceedings of the World Bank Annual Conference on Development Economics, 1993: Supplement to The World Bank Economic Review and The World Bank Research Observer. Washington, D.C.: World Bank, 19-52.

Stiglitz, Joseph E., and Andrew Weiss. 1981. Credit Rationing in Markets with Imperfect Information. *American Economic Review*. v71 (n3):393-410.

Zeller, Manfred, Aliou Diagne, and Charles Mataya. 1998. Market Access by Smallholder Farmers in Malawi: Implications for Technology Adoption, Agricultural Productivity and Crop Income. *Agricultural Economics*. v19 (n1-2):219-229.

Zeller, Manfred. 1994. Determinants of Credit Rationing: A Study of Informal Lenders and Formal Credit Groups in Madagascar. *World Development*. v22 (n12):1895-1907.

APPENDIX 1

NOTES ON TRIBES OF KASHIPUR, ORISSA

In this appendix, we first provide a brief overview of the tribal population in India. We then present an overview of the social and economic practices and characteristics of the main tribes in the household survey site, Kashipur, using field observations as well as secondary sources.

1 Overview of tribal population in India[99]

Possibly barring Africa, India has the largest concentration of tribal people in the world (Ministry of Tribal Affairs 2001). According to the 2001 Census, about 84.3 million Indians, or 8.2 per cent of its total population, are classified as tribal. The tribes of India are often referred to as aboriginal, autochthonous or *adivasi* ("first settlers"). However, there is considerable racial diversity among tribal groups. In addition, there is no distinct separation in racial or physical traits between tribal and non-tribal populations, probably due to the coexistence and interpenetration between the two groups over a period of centuries (Beteille 1998). Nevertheless, to a large extent, tribes can be distinguished from non-tribal populations based on their habitat (mostly forest and hill areas and peripheries of the subcontinent, including the islands), language and level of social, economic and political marginalization (ibid.).[100]

Since the independence of India, the government's tribal development policy has followed Prime Minister Jawaharlal Nehru's *panchsheel* (five principles) of tribal development. Nehru called for integrating tribes with mainstream India and bringing

[99] The overview of the tribal population in India is taken from Bhattamishra (2007).
[100] For more details on the socioeconomic characteristics of the tribal population in India, see Bhattamishra (2007).

them to an equal footing economically while, at the same time, maintaining their cultural distinctness. Distinct policies for tribal development have earmarked resources from state and central budgets specifically for tribal areas to reduce poverty and improve physical and social infrastructure in these areas. However, tribal communities continue to suffer from very low levels of socioeconomic development.

Of the 533 communities currently recognized by the Indian government as Scheduled Tribes (ST) (Ministry of Tribal Affairs 2000-01), the largest number, 62 tribes, reside in the state of Orissa. With over 7.6 million tribal people in rural areas, Orissa also has one of the highest concentrations of tribal populations among Indian states.[101] While the ST population comprises about 8 percent of the national population, the tribal population comprises about a quarter of the total rural population of Orissa (Census 2001). In the next section, we provide brief details on two of the largest tribes of Orissa, the Kandha and the Paraja Jhodia.

2 Notes on tribes of Kashipur, Rayagada

The main tribes in Kashipur block are the Kandha and Paraja. The Kandha are the largest tribe of Orissa in terms of population (about 1 million according to the 1981 Census). The Kandha are descendants of Proto-Australoid and Mongoloid races. The Paraja Jhodia is another large tribe of Orissa (about 270,000 people by the 1981 Census) concentrated in undivided Koraput and Kalahandi. The following section documents the habitat, agricultural cycle and customs of the Kandha and Paraja tribes in Kashipur, Rayagada.

[101]The tribal population in India is classified under the Scheduled Tribe (ST) social group as defined by the Census of India. These numbers therefore reflect the population classified as ST.

Habitat

Both the Kandha and Paraja tribes in Rayagada are settled in areas near forests or at the foot of hills near a perennial water source (such as a stream or spring). In many villages in Rayagada, these two tribes live together (sometimes with other tribes and castes such as Loharas, Goudas, Paikas, Naiks) while in others, they have uniethnic settlements. In most cases, huts are constructed in two rows on either side of a street running through the length of the village. Typically, each hut has a veranda in front and in the back, and one room (or two rooms, both in the same line). Roofs are low and can be thatched or tiled, and there is not much light inside the huts. In the clearing that runs through the middle of the village, there is an altar to the Earth Goddess. There is also an elevated mud platform for meetings held by village elders. In some cases, there are youth dormitories for boys (*Dhangda basaghar*) and girls (*Dhangdi basaghar*). By day, these dormitories can be used to hold meetings of village elders. By night, young girls and boys go the dormitories to interact with each other through song and dance, though this tradition is disappearing in villages that are in contact with the non-tribal population.

Economic practices

Both the Kandha and the Paraja combine settled subsistence farming (mainly on unproductive hilly land) with forest collection, hunting and fishing. Land is classified into *Beda* (wet cultivation of crops such as paddy, on low-lying land), *Goda* (non-irrigated middle land for growing millets and paddy) and *Dongar* (non-irrigated hill slopes for growing millets, pulses and oilseeds). The concept of private land ownership does not exist. New land for farming is obtained by clearing forests, which are viewed as common property. The distribution of land is clan-based. Within a clan (which may comprise families across 8-10 villages), all families farm equal areas of

land. The unit of production is the family. Patrilineal inheritance is practiced and land is divided equally among all sons after a father's death. Simple implements, such as ax, hoe, spade and sickle are used in farming. During the summer months, these tribes also seek wage work (*e.g.*, in public works programs, mining, road construction). Cases of migration during the summer months in search of wage work are also found.

Both the Kandha and Paraja tribes own cattle, goats and chicken mainly for their meat as well as buffalo for offering in sacrifice. Bullocks are used for ploughing on low-lying plain agricultural land, for those that own such land. These tribes practice barter economy, where they exchange grains, livestock, liquor, brooms and crafts (such as wood carving in the case of the Paraja tribe) for clothes, salt, jewelry, and other items as well as grains and liquor.

Indebtedness is a common feature of many tribal households, as many are caught in annual debt cycles whereby they borrow grains or money to tide over the lean season that they are unable to return until the following harvest.

Over the past decade, in the face of opposition from the tribal population, the Utkal Alumina consortium of firms has been in the process of establishing the first large-scale industry in Kashipur. The Baphilimali, Kadingamali and Sasubohumali hills of Kashipur are rich in bauxite deposits, making the area lucrative for mining. However, the tribal population and their supporters fear that the mining enterprise will destroy their environment, as these hills are also sources of numerous perennial streams and forest products which are integral to the tribal way of living. In addition, the prevalent sentiment among the tribal population is that the plant will result in few or no job opportunities for them, as most jobs will require educational qualifications that they lack. Given the state of Orissa's poor rehabilitation record, it is likely that the tribal population in the vicinity of the aluminum plant will be faced with displacement, deforestation and adverse changes in their way of life, at least in the

short run. Whether the plant will offer new economic opportunities to the tribal population in the long run remains to be seen.

Agricultural calendar

The agricultural cycle begins in the *Chaitra* season (March-April) with the clearing of land by burning forest cover on a chosen patch. The land is prepared for sowing in the monsoon (May-June) through hoeing. Sowing is done in the rainy season (June-July), followed by weeding (August-September). These months are the most difficult in terms of food availability, as the tribes depend wholly on collecting foods from the forest at this time. The harvest season begins in the autumn, with short-duration crops (including paddy, millets) harvested in October-November and long-duration crops harvested in December-January. After this, threshing and preparing the crops for storage takes place. January-March is the time for weddings and merry-making as food after a good harvest is plentiful. In some villages, March-April also marks the time for collection of *mahua* flowers in order to make wine.

Social organization and practices

Among the Kandha and Paraja tribes, there is very little specialization in social roles, except for the *jani* or *muduli* (village headman), *pujari* or *disari* (priest) and *bejuni* or *gurumain* (shamans) (Government of Orissa 1990). Although there is role differentiation by gender, whereby men are responsible for heavier agricultural work and women are responsible for housework, women also engage in various agricultural activities alongside their men, such as sowing, weeding and threshing. Order is maintained by lineage heads and village heads and transgressions of social norms are handled according to the severity of the mistake (through physical punishments, social ostracization and excommunication from the tribe). The social and political

organization is coincident. Village heads are obtained by descent. The priest conducts rituals for communal festivals as well as weddings, births and death ceremonies. *Bejunis* are believed to be endowed with special powers enabling them to rid people of sicknesses.

Marriage rules prescribe marriage within the tribe (tribe endogamous) but outside one's own clan (clan exogamous). Marriages can occur by capture or consent. Brideprice is prevalent, and after marriage, the bride moves to the groom's house (patrilocal). Households typically comprise of the husband, wife and children, *i.e.*, nuclear households are more common than extended families found in other parts of Indian society.

Unlike the all-India sex ratio, the sex ratio in these tribes is skewed towards women, with about 1030 women for every 1000 men. The literacy rate among the Kandha and Paraja Jhodia, particularly among women, is very low (about 12 percent in 1981). This is partly due to the lack of properly functioning government primary schools in these areas, as well as instruction in Oriya instead of the native languages of these tribes.

The language of the Kandha is Kuie and that of the Paraja Jhodia is Paraja. As is common across the tribes of Orissa, many Paraja and some Kandha also speak Oriya when they interact with the non-tribal population.

The major festivals include *Chaitra Parba*, celebrated in the month of *Chaitra* (mid-March to mid-April), *Pous Parba*, celebrated in the month of *Pous* (mid-December to mid-January) and *Dhana Nuakhai*, celebrated in the month of *Kartik* (mid-October to mid-November). As in other agricultural societies, these festivals are closely tied with the agricultural calendar. *Chaitra Parba* marks the onset of a new agricultural cycle, while *Pous Parba* marks festivities following the harvest.

Religion

The Kandha and Paraja practice a mixture nature worship, ancestor worship and animism. Their rituals are closely related to agriculture and are seasonal in nature, coinciding with times of sowing, planting and harvesting. They believe that divine spirits can be benevolent or malevolent, and practice animal sacrifice rituals to appease the latter. The Kandha used to practice human sacrifice in a ritual called *Tokimara*, in which a pre-puberty girl would be sacrificed to appease the Earth Goddess in order to ensure a good harvest. In many of the villages, Hindu deities such as Siva are also worshipped, reflecting the extent of external influences.

Dietary practices

The diet of these tribes comprise mainly of grains such as millet and to a lesser extent, rice. Food habits change between the post-harvest and lean season due to the fluctuations in food availability. When food is plentiful, it is consumed three times a day, although cooking typically takes place once a day, in the morning. Common items in the diets of the Kandha and Paraja Jhodia tribes include millet (*mandia anda, mandia peja*), rice, tamarind gravy, and turmeric and chillies as condiments. Meat is consumed only at festivals or special occasions, such as visit by guests. Liquor (such as *salap* or sago palm wine, *todi* or date palm wine and *mahuli* or wine from *mahua* flowers) is an important part of the tribal diet, and is consumed by men and women alike. During the monsoon months, when food is scarce, forest foods are complemented by tamarind seeds and mango kernels which are stored from the summer months.

APPENDIX 2

SURVEY NOTES AND METHODOLOGY

1 Introduction

In this appendix, we describe the data collected in Orissa, India in 2005 for a project to evaluate community grain banks.[102] The project, which was directed by Ruchira Bhattamishra as part of her dissertation research, was undertaken in three phases: first, a household survey was implemented in January-March 2005 in 28 villages in Kashipur, Rayagada district in southwest Orissa (hereafter referred to as the post-harvest household survey). This household survey was accompanied by a village and grain bank survey. Second, a village and grain bank survey was implemented between April-May 2005 in 80 villages in Dasmantapur block, Koraput. Koraput and Rayagada districts are adjoining. Third, a second wave of the household survey was implemented in 26 of the 28 initially sampled villages in Kashipur, between August-early October 2005 (hereafter referred to as the lean season household survey).

Advice on undertaking a survey in this region as well as helpful logistical support were provided by the management and staff of Agragamee, a non-governmental organization (NGO) widely considered to be a pioneer of the grain bank movement in tribal Orissa.[103] Useful insights on the Government of India's Grain Bank program in Orissa were provided by Shibanarayan Mishra, Integrated Tribal Development Agency (ITDA) officer. Data on children's immunization and birthdate records were also provided by various *anganwadi* workers of surveyed villages in

[102] The survey instruments and data are available from the author upon request.

[103] Special thanks are due to Agragamee staff members Jitendra Mohanty, Nakula Bisoi, Ashok, Agastya and Atul for their support. Thanks are also due to Abhiram Jhodia, Adu Naik, Ganesh Nayak, Jaga Majhi, Kambhu Majhi, Phulsingh Majhi, Raghu Naik, Ramdhar Jhodia, Ramesh Chandra Majhi, Sashi Majhi, Shiba Pradhan, Surath Gopal, Tumbeshwar Majhi and Yudhisthir Jhodia, who served as guides during the survey.

Kashipur, under the guidance of the Integrated Child Development Scheme (ICDS) officer, Jhunu Kumari Patra.

Funding for a pre-survey visit was provided by the Mario Einaudi Center for International Studies, Cornell University as well as the Department of Applied Economics and Management, Cornell University. Funding for the first wave of the household survey as well as the grain bank and village survey in Dasmantapur was provided by the Graduate School, Cornell University and the Department of Applied Economics and Management, Cornell University. Funding for the second wave of the household survey was provided by the National Science Foundation SES-0518424. All the above sources of funding are gratefully acknowledged.

2 Background

Over the past two decades, several NGOs in Orissa, a state in southeast India, have established grain banks in order to directly confront cyclical episodes of hunger and food insecurity. A number of grain banks have been established in the "tribal belt" in southwest Orissa by Agragamee, one of the more prominent tribal and rural development NGOs in Orissa. Given that Agragamee was a key player in the grain bank movement in Orissa and had one of the longest associations with the establishment of grain banks, the researcher contacted Agragamee's management in order to secure their cooperation in implementing a household and grain bank and village survey in Kashipur and Dasmantapur. The main objective of this project was to collect grain bank- and village-level data to implement an institutional analysis; in particular, an analysis of the determinants of grain bank survival and duration, given that a large number of grain banks collapsed after establishment. A related objective was to examine what impact grain banks have on household food security and

consumption smoothing outcomes, in light of the fact that grain banks were established to confront food insecurity, especially of a seasonal nature.

The post-harvest and lean season household surveys in Kashipur serve as the main source of data for the impact evaluation. The village and grain bank survey in Dasmantapur serve as the main source of data for the institutional analysis. The following sections describe the survey methods and data of the household and village surveys.

3 POST-HARVEST HOUSEHOLD SURVEY

3.1 Survey methods

3.1.1 Preparations[104]

In order to ensure accuracy of the anthropometric measurements, standard equipment such as electronic mother-child weighing scales were obtained from SECA, Germany. In addition, height measurement boards were manufactured for accurately measuring the recumbent height of children below 2 years of age. This was made possible by Dr. Dilip Mahalanobis, Society for Applied Studies, Kolkata, whose help for constructing measurement boards of an excellent quality is gratefully acknowledged.

The survey data were collected on questionnaires written in Oriya, the language spoken by the majority of the population in Orissa, including most members of the Kandha and Paraja Jhodia tribes. The questionnaires were originally written in English by the researcher, with editing assistance from Christopher B. Barrett, International Professor, Department of Applied Economics and Management, Cornell University. The questionnaires were then translated into Oriya by the researcher in

[104] The principle guide consulted on conducting field research was Barrett and Cason (1997). For an overview on the Kandha and Paraja Jhodia tribal population of Orissa, the Tribal and Harijan Research-Cum-Training Institute's book 'Tribals of Orissa' (1990) was consulted.

collaboration with Manasi Satpathy, M. Phil., Anthropology, Utkal University, Bhubaneswar before leaving for the field.

The questionnaires were pre-tested in Kashipur and Dasmantapur outside of the survey sample, resulting in considerable revisions to improve clarity, to refine codes, to reduce the duration of the interviews and to ensure the usage of local terminology and local dialect translations.[105]

The enumerators were trained as a group on the Oriya questionnaires. Extensive instructions on how to implement the anthropometric module was provided to the entire survey team following Cogill (2003), although, finally, only two of the enumerators were entrusted with taking anthropometric measurements in all sample villages.

The household questionnaire included modules on household demographics, occupation, morbidity, mortality, food security, anthropometry, credit, grain bank membership, assets and agricultural output. The village and grain bank questionnaire included questions on relevant village and grain bank characteristics, such as village and grain bank membership size, ethnic composition, transport and communication, social and physical infrastructure. In villages where grain banks had failed, only retrospective information on grain bank design and functioning was gathered, whereas in villages with surviving grain banks, both contemporaneous and retrospective information on grain bank design parameters and functioning was gathered.

The following sections describe the sampling methods, survey management, survey sites and data for the household survey.

[105] The vast majority of respondents spoke and understood a local dialect of Oriya. In very few cases, the respondents spoke only Kuie or Paraja, the languages of the Kandha and Paraja Jhodia tribes respectively. In such cases, the local enumerators conducted the interview and translated responses.

3.1.2 Sampling

The sample frame of villages was drawn in consultation with Agragamee staff and management. Administrative divisions (locally called *gram panchayats*) undergoing unrest related to the establishment of the Utkal Alumina factory (such as Maikancha, Tikiri, Kucheipadar, Gorakhpur, Podapadi and Sanhkarada) were eliminated from this frame.[106] In addition, villages that were practically inaccessible due to the rocky terrain or lack of communication facilities or were home to isolated primitive tribal groups (such as the *Dongaria Kandhas*) were not included in the sampling frame.

From the remaining 11 *gram panchayats*, two separate lists of villages were constructed after consultation with Agragamee field officials and office records. These included 14 villages with functioning grain banks (hereafter referred to as grain bank villages, or GBVs) and 14 villages with failed grain banks or villages where grain banks were never set up (hereafter referred to as non-grain bank villages, or NBGVs). The villages were spread across 9 *gram panchayats*, namely, Mandibisi, Godibali and Siripai in the east, Dongasil and Kodipari in the southwest, Manusgaon and Chandragiri in the west and Kashipur and Renga in the central part of Kashipur block.

The villages were chosen so as to fulfill the following criteria: (1). grain banks in GBVs were continuously operational since inception for at least 5 years, and grain banks in NGBVs stopped functioning at least 5 years prior to the survey. (2). the two sets of villages shared as many characteristics as possible, such as socio-economic level of development, village size, distance from important geographical markers (such as seat of local government, NGO offices, local markets, etc), and (3). there was

[106]This was done mainly due to safety considerations for the survey team. In addition, inhabitants of these areas experience issues that are distinct from adjoining areas where the grain bank survey was implemented, including conflict and violence, loss of agricultural land, environmental degradation, as well as the potential for an increase in employment and earnings.

variation in village size (small, medium large), distance from the main road (close, distant) and distance from Kashipur town (close, distant), which housed the administrative offices for Kashipur block and was also the site of Agragamee's main office. The first criterion was adopted in order to make possible a comparison of the impact of grain banks on health outcomes of children below the age of 5 years. While grain banks can be expected to have some impact on the health outcomes of adults, it is unclear how to attribute this impact to grain banks vis-à-vis other factors prior to the establishment of grain banks. The second criterion was adopted to obtain a set of treatment and control villages that resembled each other closely. The third criterion was adopted in order to obtain variation in the characteristics of the sample villages.

A list of all households (with the name of the household head) in these villages was drawn up. Within each of the selected villages, 20 households (or less, if the total number of households in a village was less than 20) were randomly selected. Sampling was thus not proportionate to village size. The response rate was high, at 97.6 per cent, and the few non-responses were due to the unavailability of sampled households at the time of the survey.[107] A total of 557 households were sampled in 28 villages. Of these, 13 households could not be contacted due to non-availability at the time of the survey, bringing the total usable sample to 544 households. Thus, the response rate was over 97 per cent. Table A.1 shows the sample size and decay.

[107] Villagers were generally willing to share time as well as the detailed socio-economic data that was asked as part of the survey since the majority of the information was non-private in nature (this is better understood in context, as given the physical and social proximity of households, members of the same village tended to have close to full information on one another)

Table A.1: Sample size and decay (Post-harvest household survey)

Village name (Hamlet, if applicable)	Number of households sampled	Number of households that responded	Number of households that did not respond
Baliguda	20	20	
Chirikul	20	20	
Dhangdisil (Adivasi Sahi)	20	20	
Dhobasil	20	19	1
Gaimundtunda	20	20	
Godibali (Tala Sahi)	20	20	
Gulmijholla (Harijan Sahi)	20	20	
Huder	19	19	
Jhodipadar (Adivasi Sahi)	20	20	
Kasnadora	20	20	
Keshkeri	20	20	
Kodikitunda	20	20	
Kukuragud (Bhatipas)	20	20	
Mahulkuna (Ranjuguda)	20	20	
Mandibisi (Haliasahi)	18	18	
Manusgaon	20	20	
Paraja Sila (Puruna Sahi)	20	20	
Patesh	20	18	2
Patiasil	20	20	
Pipalpadar	20	20	
Potamund (Bajansil)	20	16	4
Renga (Upar Sahi)	20	20	
Runjimaska	20	14	6
Sanmatru	20	20	
Sargiguda	20	20	
Siriguda	20	20	
Sirlijodi	20	20	
Tharly	20	20	
TOTAL	**557**	**544**	**13**

Notes. The total number of households in Haliasahi and Huder were 18 and 19 respectively.

3.1.3 Survey sites

Fourteen villages where grain banks continue to be functional and 14 additional villages where grain banks collapsed at least five years prior to the survey date (or were never established) were surveyed.[108] In addition to the household survey, a

[108] Grain banks were not established in Godibali (Tala Sahi), Sanmatru and Baliguda. Conversations with Agragamee staff and villagers revealed that the reason in the case of the former was that they did not feel the need of a grain bank. In the case of the other two villages, they were afraid of getting indebted to Agragamee by accepting the initial grant.

village and grain bank survey was conducted in each of these 28 villages in order to gather village and grain bank information.

The following table provides the distribution of the usable sample of 544 households across grain bank presence. Out of these households, 269 lived in villages where grain banks were active, and 275 in villages where grain banks were absent.

Table A2: Distribution of usable sample by grain bank presence (Post-harvest household survey)

Grain bank present		Grain bank absent	
Village name (Hamlet, if applicable)	No. of households	Village name (Hamlet, if applicable)	No. of households
Dhobasil	19	Baliguda	20
Gaimundtunda	20	Chirikul	20
Keshkeri	20	Dhangdisil (Adivasi Sahi)	20
Kodikitunda	20	Godibali (Tala Sahi)	20
Kukuragud (Bhatipas)	20	Gulmijholla (Harijan Sahi)	20
Mahulkuna (Ranjuguda)	20	Huder	19
Mandibisi (Haliasahi)	18	Jhodipadar (Adivasi Sahi)	20
Paraja Sila (Puruna Sahi)	20	Kasnadora	20
Patesh	18	Manusgaon	20
Pipalpadar	20	Patiasil	20
Renga (Upar Sahi)	20	Potamund (Bajansil)	16
Runjimaska	14	Sanmatru	20
Siriguda	20	Sargiguda	20
Sirlijodi	20	Tharly	20
TOTAL	**269**	**TOTAL**	**275**

3.1.4 Survey management

The enumerators were divided into teams of two each. The survey team, which also included the researcher, was introduced to the respondents by local Agragamee staff. Before the survey, villagers were informed about the purpose of the survey and the time it would require to complete it. They were then requested to provide a time when it was convenient for them to remain in the village. No survey interviews were conducted on the day of the weekly market. All of these contributed to the low sample decay ratio.

The survey team, accompanied by the researcher or a supervisor, took care to arrive at the appointed time, usually early in the morning, before the men and women left for agricultural or household activities. Before any interviews were conducted, the researcher or a supervisor explained the purpose of the research, the randomness of selection into the survey and the lack of affiliation between the research team and Agragamee or the government. In addition, a prepared introduction to the survey was read to all respondents in order to explain the purpose of the research and assure them of the anonymity of the responses. An English translation is provided in Table A3 below.

Table A3: Oral consent script

Namaskar. My name is _____ and my colleague's name is _____. We are doing a study on grain banks and their impact on food security in Orissa. The purpose of this study is research. We hope that this study will help to improve the food security condition of your community. We would like to ask some questions regarding grain bank participation, household food consumption and other issues. We would also like to take height and weight measurements of you and your children as part of the study. This will help to analyse the nutritional status of members of your household. I can assure you that all information that you submit during the interview will be treated confidentially, and your identity will not be revealed to anyone outside this research team. Since we want to learn about food consumption and children's health, we would like to talk especially to the mother of the children. The survey will take about half an hour and we will be very grateful if you could talk to us now. Thank you for your cooperation. May we proceed with the interview?"

Respondents to the household questionnaire included all adults in the household. It was not possible to conduct the interviews privately due to the nature of the homes in the survey villages – there is insufficient light inside the homes to conduct the interview, and the verandahs of the homes are not private.

The village and grain bank questionnaire was administered in Oriya simultaneously to a group of inhabitants of the village where household surveys were also being administered. These individuals were identified by the villagers themselves

as being knowledgable about grain bank operations. They included women, who typically comprised half of grain bank management committees, as well as traditional village elders and younger individuals who had been elected as political leaders under the contemporary *panchayat* system, where relevant.

The flow of questionnaires was controlled through an identification number unique to each questionnaire. In order to ensure good data quality, completed questionnaires were passed through a series of quality control checks. First, when a survey team completed an interview, the researcher or a supervisor proofed the questionnaire before leaving the village. In addition, questionnaires were reviewed for feasibility of responses as well as missing responses.. When possible, corrections were made in consultation with the survey respondent before leaving the survey site.

3.2 Data

3.2.1 Data entry and cleaning

After the survey interviews were completed, all the data were entered into Excel files by data entry technicians in Bhubaneswar. Individual-level data in the different questionnaire modules are linked by unique village, household and personal identifiers. Household-level data are linked by unique village and household identifiers. The files were converted into Stata data format by the researcher. Extensive data cleaning was then implemented, including further checks for data inconsistencies and missing values. No records were discarded completely, but implausible responses were deleted. Missing values were not imputed. All data have been stripped of names to protect the privacy of respondents per the requirements of the Cornell University Committee on Human Subjects.

3.2.2 Data quality assessment

Some parts of the survey appear to have yielded quite useful and reliable data, such as the sections on demography and literacy, occupation, assets, dietary diversity and food security perceptions. On the other hand, the anthropometry module has a number of missing records. Of the total sample of 544 households, 400 had 1 or more children below 5 years of age, bringing the total sample of children to 599. Of these, height data were missing in 51 cases (or about 9 percent) and weight data were missing in 40 cases (about 7 percent). However, diagnostics by treatment status reveal that the latter has no impact on the probability of an observation missing anthropometric data. In addition, the missing observations are split fairly evenly between the treated and untreated groups.[109]

Data on age and date of birth collected as part of the anthropometry module are not complete. In a little over half the survey villages, birthdates were available from *anganwadi* records.[110] However, these were incomplete and suffer from inaccuracies.[111] Less than 11 percent of the observations for whom data were

---

[109] Height data are missing for 30 children in participating households and 21 children in non-participating households. Weight data are missing for 21 children in participating households and 19 children in non-participating households. Diagnostics of missing observations by village reveal that are distributed across the sample (e.g., the missing height observations are distributed across 20 of the 28 sample villages, with 1 observation missing in each of 6 villages, 2 missing in each of 5 villages, 3 missing in each of 3 villages, 4 missing in each of 5 villages and 6 missing in 1 village. Similarly, the missing weight observations are distributed across 19 villages, with 1 observation missing in each of 6 villages, 2 missing in each of 7 villages, 3 missing in each of 4 villages, and 4 missing in each of 2 villages).

[110] Records were complete (or near complete) in only 7 out of 28 survey villages, namely, Haliasahi, Paraja Sila, Bajansil, Pipalpadar, Renga, Sanmatru and Siriguda. Records were incomplete in the case of 7 other villages for which *anganwadi* records were available, namely, Keshkeri, Ranjumaska, Kodikitunda, Gaimundtunda, Dhobasil, Manusgaon and Ranjuguda.

[111] For example, in the villages of Bhatipas and Sirlijodi, although *anganwadi* records were available, there was a lack of agreement in names, sex and age of children from the household survey records and the former. Thus, from the point of view of collecting date of birth information for the anthropometry module, these records had to be discarded. In another example, in one sample village, *anganwadi* records indicated that all children born after January 2000 had exactly the same birth weight. Although the ICDS officer for Kashipur block was alerted to these issues, no further steps were taken by her. Unfortunately, poor quality and falsified records in the *anganwadi* system are not limited to these villages, as a survey of the ICDS program in six Indian states has revealed similar problems (CIRCUS 2006).

collected in the anthropometry module possessed birthdates information from immunization cards which were produced by the children's parents. Therefore, birthdates could be recorded in only few cases: of the 599 children who were reported 5 years of less, date of birth information was not available for more than 400. In order to increase the accuracy in estimating the age of children below 5 years of age, the enumerators were trained to probe the season, as well as month of the Oriya calendar when the child was born, followed by a question on how many times that season had recurred since birth, in order to estimate the years completed by the child.

Data on the terms of contract for outstanding loans are also not reliable, as respondents could not provide data on loan duration and rates of interest. This also reflects the nature of the informal credit markets that prevail in the survey region, which are characterized by unwritten and loosely-defined credit contracts.

Data collected with the objective of capturing the nature of preferences (whether exponential or time-inconsistent) are also not reliable, as in most cases respondents seemed unable to comprehend the relevant questions.

4 VILLAGE AND GRAIN BANK SURVEY

Since there is a large amount of overlap in the survey methods used for the surveys in Kashipur and Dasmantapur, we do not repeat them here. Only differences are noted.

4.1 Survey methods

4.1.1 Preparations

Preparations were implemented concurrently with the preparations for the household survey in Kashipur. The same set of enumerators used in the household survey was used as they were already familiar with the purpose of the research project. These enumerators were then trained extensively on the village and grain bank questionnaire.

4.1.2 Sampling

The survey was fielded in Dasmantapur in 40 villages where grain banks were operational (in other words, surviving grain bank villages or SGBVs) and 40 villages where grain banks were once present but were no longer operational at the time of survey (in other words, failed grain bank villages or FGBVs). First, a list of villages and their hamlets was obtained from Agragamee staff who have been active in the area for over a decade. From this list, two lists were made in consultation with Agragamee field staff – villages with operational and those with non-operational grain banks. From the list of villages having currently functional grain banks instituted by Agragamee, 40 villages were randomly chosen. Similarly, from the list of villages where grain banks had been instituted but collapsed, 40 more villages were randomly chosen. In 6 of the 40 FGVBs originally sampled, surveys could not be conducted.[112] In order to maintain the desired sample size, they were replaced by 6 villages randomly chosen from the list of FGBVs that had not already been included in the sample. These 6 replacement villages were located across 3 different *gram panchayats*, including Chikamba, Mujanga and Podaguda.

The sample numbers of surviving and failed grain bank villages were balanced and not proportionate to the population numbers of surviving and failed grain bank villages. According to the most recent records available from Agragamee, out of the 232 grain banks initially established, only 71 were still operational. Thus, the population proportion of operational grain banks was 30.6 percent. Proportional sampling would have resulted in a lower number of observations of surviving grain

[112]This was due to opposition by an NGO that was active in Chanabad and had adversarial relations with Agragamee.

banks, and given the objectives of the study to study the factors that influence grain bank sustainability, operational grain bank villages were intentionally oversampled.

4.1.3 Survey sites

Table A4 presents the survey sites for the village and grain bank survey.

4.1.4 Survey management

The enumerators were divided into teams of two each. Each team included at least one enumerator who was not from Dasmantapur and unrelated to likely survey respondents, in order to reduce respondent bias. Villagers were notified in advance regarding the purpose of the survey team's visit. Their availability for the survey was requested, and surveys were undertaken on days that were deemed convenient for most respondents. The survey team was also trained to keep detailed notes on field observations and information obtained through informal conversations during their visit (e.g., problems associated with grain bank, primary school and *anganwadi* operations; important crops; main source of livelihood) in order to complement the information obtained through the survey questionnaire.

The questionnaire was administered in Oriya simultaneously to a group of members of the sampled village who possessed knowledge on the resources of the village and more specific information on grain bank operations. The flow of questionnaires was controlled through an identification number unique to each questionnaire. After completion of interviews by the survey team, the questionnaires were reviewed for feasibility of responses as well as missing responses by the supervisor. Corrections were made in consultation with the survey team with the help of their notes.

.

Table A4: Distribution of sample villages (Village and grain bank survey)

Surviving grain bank village		***Failed grain bank village***	
Gram Panchayat	Village (hamlet)	*Gram Panchayat*	Village (hamlet)
Chanabad	Kitesh (Salapguda)	*Bejapadar*	Chandankhuti (Nuaguda)
	Kadamjhola (Kadamjhola)		Mangaraguda
	Kadamjhola (Tandiputa)	*Chanabad*	Kitesh (Lataput)
Chikamba	Dhalagadla		Chanabad
	Pirimachi		Kitesh (Depoguda)
	Phatkijam (Adivasi Sahi)	*Chikamba*	Dakamara (Barijholla)
Dasmantapur	Goudabarikanta (Goudabarikanta)		Upar Gadala
	Baghchema (Baghchema)		Dhunakhala
	Mundar (Mundar)	*Dasmantapur*	Chaulakanti (Chaulakanti)
	Goudabarikanta (Dudijhola)		Chaulakanti (Ladibeda)
	Paraja Barikanta (Paraja Barikanta)		Dandabada
	Mandiaguda (Mandiaguda)		Chaulakanti (Janiguda)
	Goudakanti		Chaulakanti (Khajuriguda)
Gadiaguda	Bilanosila (Lamatapur Bariguda)		Durkaguda
	Balighat (Balighat)		Pedisil (Pedisil)
	Padeiput (Khajuriput)	*Dumbaguda*	Bhandisil
	Bilanosila (Bramanasuku)	*Gadiaguda*	Paikapuki
	Gadiaguda		Parajapuki (Garudamunda)
	Malimunda (Malimunda)	*Girliguma*	Rohiamba (Bijimara)
Girliguma	Dakribeda (Harijan Sahi)		Chadri (Alachi)
	Dhalaghata (Bada Majhi Sahi)	*Kucheipadar*	Gambhariguda
	Chhotamba (Bada Adivasi Sahi)	*Malkangiri*	Malkangiri
	Girli (Sundhigirli)	*Mujanga*	Kankaraput (Upar sahi)
	Dakri (Fundaguda)		Runchaguda (Durakaguda)
	Rohiamba (Majhi Sahi)		Kankaraput (Dongaladhaput)
	Ratabandha (Banapadar)		Batisil
	Girliguma (Parajagirli)		Kilaro (Dumbaguda)
	Bhagalmati (Tikirapada)		Mujanga (Dhaiguda)
	Gadri (Gadri)	*Podaguda*	Naranga
	Bhagalmati		Rautaputa (Dudulaguda)

Table A4 (Continued)

Mujanga	Mujanga (Tentuliguda)	*Podaguda*	Mundar
	Kankaraput (Baliguda)		Mundar (Sundhiputguda)
	Kilaro (Daluguda)		Anchalguda
	Malingajodi (Badaguda)		Kantabeti
	Batisil (Keshabguda)		Patamaliguda (Bhanjapadar)
	Mujanga (Thelaguda)		Upar Naranga (Tala Naranga)
	Mujanga (Bhejapadar)		Majhiguda
	Tentuliguda (Punjisila)		Rautaputa (Chamakaliguda)
Podaguda	Rautaputa (Rautaputa)		Patamaliguda (Lukumari)
	Chakarjholla		Chotaguda (Majhiguda)

4.2 Data

4.2.1. Data entry and cleaning

After the survey interviews were completed, all the data were entered into Excel files by the researcher and then converted to Stata data format. Checks for data inconsistencies and missing values were implemented. No records were discarded completely, but implausible responses were deleted. Missing values were not imputed.

4.2.2 Data quality assessment

Most of the village-level data are fairly reliable. Distance data (e.g. distance of the village from the seat of local headquarters, from the main road, etc.) are based on reports by the team of survey respondents and reflect the distances actually perused by the majority of villagers (whose main mode of transport is walking and who typically use short-cuts on unpaved roads whenever possible).

For SGBVs, data on grain bank stocks were obtained in local units from one or more grain bank committee members, which were later converted to metric units for comparability. These are rough estimates, and it was not possible to crosscheck the actual weight of the stocks using standard weighing scales. The measurement units that were used locally were *puti, mana* and *ada*. According to the inhabitants of the region, 1 *puti* equaled 20 *mana* and 1 *mana* equaled 4 *ada*. Depending on the crop, 1 *mana* was converted into metric units, per the advice of grain bank committee members, as follows:

Crop name	Metric equivalent
Local millets (mandia, kosala)	4.00 kg
Paddy rice	4.00 kg
Local lentils (biri, kandula)	4.00 kg
Local oilseeds (alasi, mustand)	3.25 kg

For FGBVs, survey enumerators were instructed to obtain data from account books. However, these were typically not available and data were obtained using corroborating accounts from former committee members and former grain bank members.

5 LEAN SEASON HOUSEHOLD SURVEY

The second wave of the household sample survey was conducted in Kashipur towards the end of the monsoon season (August-early October of the same year). The timing of the survey corresponds with the agricultural lean season when food shortages are at their most critical. The same villages were surveyed, excluding Bhatipas and Gulmijholla. They could not be contacted due to flooding of the access bridge to the former and the outbreak of a cholera epidemic in the latter.

5.1 Survey methods

5.1.1 Preparations

The survey methods are similar to those of the post-harvest household survey. The same group of enumerators and supervisors were used in this wave to maintain comparability. An additional module related to shocks and safety nets was included to the original questionnaire. A pared down version of the village and grain bank survey

was administered in all the sample villages in order to complement the data already collected as part of the post-harvest survey.

5.1.2 Sampling

A rotating panel sample design was adopted in order to balance the objectives of obtaining a large number of panel observations without reducing the sample size. This was done keeping in mind that the survey was being conducted at a very critical time in the agricultural calendar. Since this was the sowing season, adult members of the household were likely to leave their village during the day to perform agricultural tasks in their fields.

Within each sample village, attempts were made to contact 15 out of the 20 sample households selected in the first wave and to contact 5 new households based on the original household listings. This was possible in all but 3 sample villages, where more new households had to be sampled due to the absence of more than 5 households from the previous wave. Households that could not be contacted had indeed left for agricultural tasks in spite of prior notification by the survey team. A total of 516 households were sampled. Contact was reestablished with 400 households from the first wave, and 99 new households were added. This brought the total usable sample to 499 households; 250 households were in grain bank villages, and 249 were in non-grain bank villages. The response rate was close to 97 per cent. Table A5 shows the sample size and decay.

5.2 Data

5.2.1. Data quality assessment

In most cases, the survey appears to have yielded quite useful and reliable data, such as information on household demographics, assets, and food security perceptions. However, there are a number of missing records in the anthropometry module.

Table A5: Sample size and decay (Lean season household survey)

Village name (Hamlet, if applicable)	Number of households censused	Total number of households responding	Number of households not responding
Mandibisi (Haliasahi)	17	17	
Patesh	20	18	2
Pipalpadar	20	20	
Dhobasil	20	19	1
Renga (Upar sahi)	20	20	
Sirlijodi	20	19	1
Keshkeri	20	20	
Kodikitunda	20	20	
Gaimundtunda	20	20	
Paraja Sila (Puruna sahi)	20	20	
Mahulakana (Ranjuguda)	20	20	
Siriguda	20	20	
Runjimaska	20	17	3
Tharly	20	20	
Potamunda (Bajansil)	20	15	5
Jhodipadar (Adivasi sahi)	20	20	
Dhangdisil (Adivasi sahi)	20	20	
Sanmatru	20	16	4
Chirikuli	20	20	
Baliguda	20	20	
Huder	19	18	1
Godibali (Tala sahi)	20	20	
Manusgaon	20	20	
Kasnadora	20	20	
Patiasil	20	20	
Sargiguda	20	20	
TOTAL	516	499	17

Notes. The total number of households in Haliasahi and Huder were 17 and 19 respectively. Since the time of the post-harvest survey, two households in Haliasahi had merged to form one household.

Of the total sample of 499 households, 375 had 1 or more children below 6 years of age, bringing the total sample of children to 571. Of these, height data were missing in 37 cases (about 7 percent) and weight data were missing in 44 cases (about 8 percent). However, diagnostics of missing observations by treatment status reveal that, as in the post-harvest survey, the latter has no impact on the probability of an observation missing anthropometric data. In addition, the missing observations are split fairly evenly between the treated and untreated groups. [113]

Unlike the earlier survey when anthropometric measurements were taken of children up to 5 years of age, measurements were taken for children up to 6 years of age. This was done so as not to lose observations of children who were under 5 in the previous survey but had crossed 5 years by the time of the monsoon survey.

As in the post-harvest household survey, data on age and date of birth are not complete. Birthdates from *anganwadi* records and immunization cards were available in only 30 percent of cases.[114] When available, the records appear to suffer from deficiencies.[115]

---

[113] Height data are missing for 16 children in participating households and 21 children in non-participating households. Weight data are missing for 23 children in participating households and 21 children in non-participating households. Examining the distribution of missing observations by village, we find that the missing height observations are distributed across 22 of the 26 sample villages, with 1 observation missing in each of 14 villages, 2 missing in each of 3 villages, 3 missing in each of 3 villages, and 4 missing in each of 2 villages. Similarly, the missing weight observations are distributed across 22 villages, with 1 observation missing in each of 12 villages, 2 missing in each of 3 villages, 3 missing in each of 3 villages, 4 missing in each of 3 villages and 5 missing in 1 village

[114] *Anganwadi* records were complete for the most part in 9 out of the 26 survey villages, namely, Haliasahi, Ranjuguda, Jhodipadar, Dhangdisil, Ranjumaska, Sargiguda, Siriguda, Patesh and Pipalpadar. Records were incomplete in 9 other villages for which *anganwadi* records were available, namely, Baliguda, Godibali (Tala Sahi), Gaimundtunda, Keshkeri, Renga, Huder, Sanmatru, Chirikul, Paraja Sila. Records were unavailable for the remaining villages.

[115] For example, in many of the villages in Godibali *panchayat,* namely Chirkul, Sanmatru, Huder, Godibali (Tala Sahi), the names of the children and parents in the *anganwadi* records did not match those collected from the household survey. In Jhodipadar, the birthdates of children within a household were interchanged for 2 households. In Dhangdisil, there were inconsistencies in the year of birth as entered in *anganwadi* records compared to the reported age as well as the weight of the child. In Ranjumaska, there were multiple entries for the date of birth for a child having the same name and the same parents.

REFERENCES

Barrett, Christopher B., and Jeffrey W. Cason. 1997. *Overseas Research: A Practical Guide*. Baltimore and London: Johns Hopkins University Press.

Beteille, Andre. 1998. The Idea of Indigenous People. *Current Anthropology*. v39 (n2):187-91.

Bhattamishra, Ruchira. 2005. Field notes, Kashipur. February-March 2005.

Bhattamishra, Ruchira. 2007. Tribal Development. In Kaushik Basu, (ed). *The Oxford Companion to Economics in India*. New Delhi: Oxford University Press.

Census of India. 2001. New Delhi: Government of India.

Cogill, Bruce. 2003. *Anthropometric Indicators Measurement Guide*. Food and Nutrition Technical Assistance Project. Washington D.C: Academy for Educational Development.

Government of Orissa. 1990. *Tribals of Orissa*. Bhubaneswar: Tribal and Harijan Research-Cum-Training Institute.

M. S. Swaminathan Research Foundation (MSSRF) and World Food Programme (WFP). 2001. *Food Insecurity Atlas of Rural India.* Chennai: MSSRF.

Ministry of Tribal Affairs, Government of India. 2001. *Annual Report*. New Delhi: Ministry of Tribal Affairs.