WPS8015

Policy Research Working Paper 8015

# An Employment Guarantee as Risk Insurance?

## Assessing the Effects of the NREGS on Agricultural Production Decisions

*Esther Gehrke*

WORLD BANK GROUP
Development Economics Vice Presidency
Operations and Strategy Team
March 2017

Policy Research Working Paper 8015


## Abstract

Uninsured risk constrains households in their production decisions in many developing countries. Similarly to crop insurance, employment guarantees can support farmers in managing agricultural production risks. Evidence from representative panel data of Andhra Pradesh, India, suggests that the National Rural Employment Guarantee Scheme (NREGS) reduces households' uncertainty about future income streams because it provides employment opportunities in rural areas independently of weather shocks and crop failure. Because the NREGS makes an ex-post labor supply response to agricultural shocks more efficient, households with access to the NREGS can shift their production towards riskier but also more profitable crops. The observed shifts in agricultural production do considerably raise the profitability of agricultural production and hence the incomes of smallholder farmers. The findings are not driven by changes in the labor or cost intensity of those crops, which supports the idea that the causal mechanism underlying the observed changes is indeed an insurance effect.


This paper is a product of the Operations and Strategy Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at esther.gehrke@unigoettingen.de..

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

Esther Gehrke*

**Keywords:** Uncertainty; Employment Guarantee; Crop choice
**JEL:** I38; O12; Q16

*Esther Gehrke is post-doctoral researcher at the University of Göttingen, e-mail: *esther.gehrke@uni-goettingen.de*. The author is grateful to the editor, Andrew Foster, as well as three anonymous referees for their valuable comments and suggestions. Moreover, she would like to thank Michael Grimm for continuous guidance and advice and Denis Cogneau, Clement Imbert, Markus Loewe, Abhiroop Mukhopadhyay, Dominique Van de Walle and the participants of the German Economic Association AEL Conference (Munich), the AEL PhD seminar (Zürich), the ECINEQ Conference (Bari), the IZA/World Bank Conference on Employment and Development (Bonn), the PEGNet Conference (Copenhagen) and the Brown University Applied Microeconomics lunch for many helpful comments. This work was written while the author was affiliated with German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) and the University of Passau, and supported by the German Federal Ministry for Economic Cooperation and Development [grant number 2011 95 809]. The data used in this publication comes from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh), Peru and Vietnam (www.younglives.org.uk). The views expressed here are those of the author. They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders.

# 1 Introduction

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

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

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

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

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

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

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

The remainder of this article proceeds as follows: Section 2 introduces a theoretical framework for analyzing the effects of an employment guarantee on crop choice. Section 3 presents the data and summary statistics. Section 4 outlines the estimation strategy. Section 5 presents the empirical results, and Section 6 concludes.

## 2 Risk management and households' crop choices: A theoretical framework

Providing additional employment opportunities to a total of 55 million households has brought about considerable changes in the social and economic realities in India. The NREGS affects households in rural areas through various channels.[1] The most obvious and so far most intensely researched effect is the increase in available income and wealth of those households participating in the program. This wealth effect is most pronounced for households with surplus labor - namely households whose labor supply exceeds the labor demand of their farm firm - and in regions where regular labor markets fail to absorb this excess. The increase in income resulting from NREGS participation has been shown to increase consumption levels (Jha, Gaiha, and Pandey, 2012) and to reduce poverty (Klonner and Oldiges, 2014).[2]

Another effect, which is much less well understood, is the insurance effect. It is particularly relevant for households that are highly exposed to covariate shocks such as droughts, floods or large-scale crop diseases. In rural areas of India, casual agricultural

[1] I outline here only the three mechanisms that seem to matter in the context of this study, namely in Andhra Pradesh and within the first two years of program implementation. Other mechanisms though which the NREGS could affect crop choice could be irrigation infrastructure created within the program (Deininger and Liu, 2013), education (Shah and Steinberg, 2015) and reduced conflict (Fetzer, 2014).

[2] Increases in disposable income and wealth might also positively influence the capacity to take risks and investment behavior. This effect is different from the insurance effect, which is the main focus of this article. I discuss how I attempt to isolate the insurance effect in Section 4.

employment is the dominant source of employment, and, as shown by Jayachandran (2006), wages severely fluctuate with covariate shocks. In the case of major weather shocks, farmers have to expect to not find any employment at all (Kaur, 2014). Such wage fluctuations severely limit households' possibilities to cope with shocks through the labor market. By giving households the right to work and making employment opportunities available independently of shocks, the NREGS greatly influences households' ability to smooth income in the case of a shock. In expectation of having access to the NREGS, households could take more risk in their production decisions, and reach higher expected incomes. If a shock then occurs, households can cope with it by working for the NREGS.[3]

Finally, the NREGS was shown to raise wage levels in the private sector through general equilibrium effects in the village economy. Because NREGS wages are higher than the wages paid for casual work, households shift their labor supply from the private sector towards the public works program (Berg et al., 2012; Imbert and Papp, 2015). Increases in wages could also affect production levels or crop choice in agriculture because they raise production costs, particularly for large-scale farmers.[4]

This article focuses specifically on the insurance effect, and how it affects the allocation of inputs to risky crops in a household's farm.[5] The following theoretical model of household decision-making under uncertainty shows more systematically how the introduction of NREGS can affect crop choice via the insurance effect. The model primarily builds on Dercon and Christiaensen (2011). The possibility to smooth consumption over time is therein constrained by two main factors: the lack of adequate risk management strategies and limited access to credit. Crop choice is first modeled in a world without risk but with imperfect credit markets and then extended to a world with uncertainty. This allows for the isolation of the effects of uncertainty and risk aversion on production decisions. Finally, the effects of the NREGS on input allocation decisions are discussed

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

[4]Bhargava (2014) for example shows, that the NREGS induces farmers to shift their production technology towards labor-saving equipment. I show in Section 5 that the results of this article are not driven by differences in the labor intensity of crops.

[5]The focus lies on input allocation because of data constraints: information on land allocation was not consistently collected.

in both scenarios.

## 2.1 General setup

Assume that a household engaging in agricultural production has the choice between two agricultural products $Q^d$ and $Q^s$. Given that both products are well known to the household and have been produced in the region for some time, we can abstract from learning and other sunk costs. These products are produced with two different types of production functions: one is deterministic and the other stochastic.[6] It is also assumed that the risky crop is more productive on average. Both products can be sold at local markets at the same price $p$.

Agricultural production takes place over two periods, the planting and the harvesting seasons. The total yield of both products $Q$ depends on land $a$, labor $l_1$ and input $k$ allocation in period one:[7]

$$Q^d = f^d(a^d, l_1^d, k^d) \tag{1}$$

$$Q^s = \epsilon f^s(a^s, l_1^s, k^s) \qquad E[\epsilon] = 1. \tag{2}$$

Inputs $k$ are defined as a bundle of variable inputs such as seeds, fertilizer and pesticides. The total yield of the risky product additionally depends on the realization of a multiplicative, random, serially uncorrelated shock $\epsilon$ at the end of the first period, thus after input allocation has been decided upon (Fafchamps, 1993; Van Den Berg, 2002). The expected value of this shock is 1; thus in expectation, the yield of the risky crop is just $f^s(a^s, l_1^s, k^s)$. Total yield has to be harvested in the second period, and labor required for harvesting $l_2$ is a linear function of realized yields, e.g. $l_2 = \alpha(Q^d + Q^s)$, where $\alpha$ is a parameter indicating how much labor is needed for harvesting given any realized yield.[8]

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

[7]This model abstracts from fixed capital because the marginal effect of productive capital was found to be close to zero in the data used here.

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

The household maximizes utility from consumption $C$ in both the planting and the harvesting periods. The utility function is additive over both periods and future utility is discounted by the factor $\delta$. The utility function satisfies the usual properties: it is twice differentiable and increases in $C$ but at decreasing rates, $\partial U/\partial C > 0$ and $\partial^2 U/\partial C^2 < 0$. This also implies that the household is risk averse. Tis model abstracts from leisure because it does not change the choice under uncertainty.[9] The household generates income from wage employment on local labor markets and from agricultural production. Building on the full-income approach, the household maximization problem can be described as follows:

$$
\begin{aligned}
max \qquad V =& U_1(C_1) + \delta U_2(C_2) \\
& s.t. \\
& \quad C_1 \leq w_1(T_1 - l_1^d - l_1^s) - g(k^d + k^s) + B \\
& \quad C_2 \leq p(Q^d + Q^s) + w_2(T_2 - l_2) - (1 + r)B \\
& \quad B \leq B^m \\
& \quad a^d + a^s \leq 1.
\end{aligned} \tag{3}
$$

Total time endowment is represented by $T_1$ and $T_2$. In both periods total time can be allocated between working in the labor market and working in own fields. In the first period, the household obtains income from wage work at level $w_1$ and from borrowing $B$. Inputs for agricultural production can be purchased at price $g$. In the second period, the household obtains income from the sale of its own agricultural production $p(Q^d + Q^s)$ and wage work at level $w_2$. Note here that the household has to allocate labor to harvesting in order to generate income from agricultural production. It seems plausible that the household always prioritizes its own harvest over wage employment; therefore the cost of harvesting is valued at reservation wages rather than market wages. The wage cost of harvesting $w_2 l_2$ in the budget constraint can then be replaced with $\alpha w_2^r(Q^d + Q^s)$, where

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

$w_2^r$ is the reservation wage and $\alpha(Q^d + Q^s)$ is the effort necessary for harvesting expressed in units of realized yield.

Incurred debts have to be repaid in the second period at an interest rate of $r$, and $B^m$ describes the maximum amount a household can borrow for productive purposes.[10] In contrast to input credit, consumption credit is much more difficult to obtain and highly expensive. Because households are expected to opt for that source of credit only under extreme circumstances, this model does not allow for any borrowing beyond the harvesting period.

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

Finally, $a^d + a^s = 1$ describes the restrictions on allocable land. The assumption is that there are no functioning land markets and that owned land is used for own agricultural production or left fallow.[11]

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

## 2.2 Deterministic case

First, consider a scenario without uncertainty. In such a world each household maximizes utility by maximizing profits from agricultural production plus income from wage

[10] Input credits are relatively common in rural Andhra Pradesh, although it seems that the amount of credit conceded is limited by a household's wealth. In the sample around 18% of the households that applied for credit reported not receiving the total amount of credit they applied for.

[11] This is obviously a simplifying assumption that does not hold everywhere in India. Nonetheless, observed levels of land renting are relatively low in rural Andhra Pradesh and land sales are virtually absent.

employment.[12] Because both production functions are deterministic in this scenario, optimal land, input and labor allocations are achieved when their marginal products equal respective prices.[13] The decision rule for the allocation of variable inputs to each of the crops is

$$\frac{\partial f^{d,s}}{\partial k^{d,s}} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial U_2}{\partial C_2}}. \tag{4}$$

In the absence of risk, the decision rule is equal for both crops, and optimal allocation implies that the marginal product of inputs in $d$ is equal to the marginal product of inputs in $s$. Because realized yield is harvested in the second period, input allocation does not only depend on input and output prices but also on reservation wages in the harvest season and on the intertemporal marginal rate of substitution in consumption. If credit constraints bind, input allocation to both crops is lower, and the household allocates more time to the labor market.[14]

## 2.3 Introducing uncertainty

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

$$\frac{\partial f^{d}}{\partial k^{d}} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial EU_2}{\partial C_2}} \tag{5}$$

for the deterministic crop, and to

$$\frac{\partial f^{s}}{\partial k^{s}} = \frac{g}{(p - \alpha w_2^r)} \frac{\frac{\partial U_1}{\partial C_1}}{\delta \frac{\partial EU_2}{\partial C_2}} - \frac{cov(\frac{\partial U_2}{\partial C_2}, \epsilon)}{(p - \alpha w_2^r)\delta \frac{\partial EU_2}{\partial C_2}} \tag{6}$$

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

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

[14] C.f. appendix A for a derivation of this result.

for the stochastic crop. Equation (5) looks similar to equation (4), except that the household now maximizes expected utility of consumption in the harvest period. For any expected consumption level $C_2$, expected utility $EU_2(C_2)$ is lower than the utility of the expected value $U_2(E(C_2))$, and marginal expected utility is higher than the marginal utility of the expected value. Thus, under uncertainty, the right-hand side term is lower than in the deterministic case, implying that the household allocates more inputs to the safe crop than it would in the absence of risk. This reflects the greater weight households put on future consumption relative to current consumption when facing uncertainty. Equation (6) shows the effect of uncertainty on input allocation to the risky crop. Here the decision rule changes considerably and the overall effect is less clear. Again, marginal expected utility is higher than marginal utility, thus implying higher input allocation to the risky crop also. However, the covariance between marginal utility of consumption and the random shock $\epsilon$ is strictly negative.[15] This term increases the value of the right-hand side of equation (6), which means that input allocation to the risky crop is lower under uncertainty. Which of the two effects is stronger depends on the degree of risk aversion of the household, expected consumption levels $C_2$ and the amount of covariance between marginal utility and the random shock. Since the covariance is greater with lower wages in period two and with a higher interest rate $r$, the net effect of uncertainty on input allocation can be expected to be negative in this context. Irrespective of total levels of input allocation, it can be clearly seen that under uncertainty, input allocation shifts towards the safe crop $d$ relative to the risky crop $s$. Thus under uncertainty, the share of risky crops in a household's portfolio is always lower than in the deterministic scenario.

## 2.4 The insurance effect of an Employment Guarantee

The insurance effect of an employment guarantee, such as the NREGS, stems from an improvement in the possibilities to offset income losses associated with a failed harvest ex-post. Because employment opportunities are now available also in bad agricultural

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

years, I model the insurance effect of the NREGS as an increase in harvest stage wages in periods with agricultural shocks. This increases expected harvest stage wages and reduces the covariance between harvest stage wage levels and covariant shocks.[16] The comparative statics in this section show that the introduction of NREGS affects optimal input allocation under certainty differently than under uncertainty.

Without uncertainty, an increase in average harvest period wages $w_2$ affects optimal input allocation by increasing consumption levels that can be realized in the second period (c.f. eq. 4). Households that hire labor out (i.e. those whose land is too small to produce at higher levels) increase consumption. One can thus see a decrease in input allocation for net lenders of labor because of increases in $C_2$, which reduces $\partial U_2/\partial C_2$ and increases the second part of the right-hand side of equation (4). The effect of increased wages on agricultural production levels (through consumption) can be understood as a substitution effect. Because working outside the farm becomes more profitable for households with little cultivated land, the allocation of inputs to those lands should decrease from very high levels to more efficient ones.

An entirely different effect can be observed if uncertainty reduces input allocation to risky crops as given by equation (6). If harvest stage wages increase, we can observe the same effects on marginal utility of consumption as in the deterministic case. Under uncertainty, however, the negative covariance term reduces input allocation to the risky crop, and this effect is now partially offset by the introduction of an employment guarantee. As possibilities to generate market income improve, the effect of shocks on harvest period consumption decreases. Because the household knows that it can earn additional income in instances of negative production shocks by spending more time working for the NREGS, it can afford to take a greater amount of risk in his agricultural production. The more the covariance term on the right-hand side of equation (6) approaches zero, the more the ratio of inputs allocated to the risky crop (versus the safe crop) approaches the deterministic scenario. This means that even if total input (or similarly labor) allocation

[16] I subsume wages in the agricultural lean season in harvest wages, since the important criterion is the timing of the shock, and both harvest and the lean season follow the realization of weather outcomes. In a scenario without uncertainty, expected wage levels need to be replaced by average wage levels.

is reduced due to the employment guarantee, the share of total inputs allocated to each of the crops approaches the ratio of the deterministic scenario. Interestingly, this effect holds independently of whether credit constraints reduce total input allocation or not.[17]

# 3 Data

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

Given these considerations, the model specified above is tested using the Young Lives Survey (YLS) data for Andhra Pradesh. Andhra Pradesh is particularly suited to studying the question of interest because it is one of the best performing states in India in terms of the number of workdays generated per household and meeting the demand for work (Dutta et al., 2012). Regarding outreach, only Chhattisgarh, West Bengal, Madya Pradesh and Rajasthan reached higher proportions of rural households in the financial year 2009/10.[18]

The YLS data set covers 3,019 households living in six different districts, 17 subdistricts (blocks) and 87 villages. The selection process of districts for the YLS ensured that all three geographical regions were surveyed, as too were the poor and non-poor districts of each region, such that the YLS is broadly representative of the population of Andhra Pradesh (Galab et al., 2011).[19] Three rounds of interviews have been conducted

[17] C.f. appendix A for a detailed derivation of this result.

[18] At the same time, Andhra Pradesh has been a forerunner in terms of innovative approaches to the implementation of the NREGS. First, it has a lot of experience with performing social audits to increase accountability within the scheme. Second, it was one of the first states to cooperate with IT enterprises to strengthen the efficiency of administrative processes. To increase transparency, entries on muster rolls and the number of workdays generated per job card holder etc. are publicly accessible. Nonetheless, the program continues to be implemented in a top-down manner in Andhra Pradesh. Usually, work is not generated upon demand, rather work applications are only accepted if there is work available (c.f. (Desai and Joshi, 2015; Muralidharan, Niehaus, and Sukhtankar, forthcoming).

[19] This is in reference to the State of Andhra Pradesh in 2013, prior to its division into the states of

so far (2002, 2007 and 2009/10). But for reasons of comparability, only the second (2007) and third (2009/10) rounds are considered in the current analysis. Furthermore, the analysis is restricted to households with non-zero agricultural production in 2007 and 2009/10. This data is complemented by secondary data for the calculation of the dependent variable as well as for a number of controls.

For the empirical analysis, the sample is split in treatment and control group. Treatment indicates that a household has access to the NREGS at the district level at the beginning of the agricultural cycle of the 2009/10 round of interviews. The period of reference for the 2007 round of interviews is the agricultural year 2005/06 (June 2005 to May 2006). NREGS activities started around April and May 2006 in the treatment districts, and assuming that learning about the program took a little bit of time, it seems reasonable to code all households as not having access to the NREGS in the baseline reporting period. The period of reference for the 2009/10 interviews is the agricultural year 2008/09. By that time, NREGS works had started in the districts Anantapur, Cuddapah, Karimnagar and Mahaboobnagar, the treatment districts. In Srikakulam and West Godavari, the control districts, the introduction of the NREGS was in August 2007 and in March 2008 respectively. Since activities started only very slowly in most sub-districts of Srikakulam, this district is used as control district despite the introduction of the NREGS in mid 2007.[20]

Summary statistics suggest that the treatment and control groups are not perfectly comparable (c.f. Table 1). Agricultural production levels as well as the amount spent on variable inputs (such as seeds, fertilizer and pesticides) are not statistically different between treatment group households and control group households. In contrast, the area cultivated, irrigation levels, the probability to apply fertilizer and to use high yielding variety (HYV) seeds are all higher in the treatment group than in the control group.

Andhra Pradesh and Telangana.

[20] Figure D.1 in the appendix shows monthly employment creation at the sub-district level in Srikakulam and West Godavari, as well as the average for all 'treatment' sub-districts. As can be seen, two sub-districts in Srikakulam had substantial NREGS employment creation in the two months before the relevant agricultural year. In order to gauge the robustness of my results to treatment timing, I exclude these two districts from my analysis in Table E.1 in the appendix. As expected the estimated effect of the NREGS increases when excluding one or both of these sub-districts. For a detailed discussion of data sources, treatment timing and the construction of variables refer to appendix B.

Households in the treatment group also experienced a rainfall shock in the agricultural year previous to the baseline period. This shock was less pronounced in the control group. Finally, Table 1 reports the participation status with the NREGS at the time of the baseline data collection. As can be seen, 66% of the households in the treatment districts report having registered with the NREGS in 2007.

For the calculation of the dependent variable - a risk index of each households' crop portfolio - survey data on input allocation to each crop is combined with District-level crop production statistics. The time series of crop production statistics are used to calculate the coefficient of variation of each crop's yield. With this information, a risk index $R_i$ of each household's crop portfolio is constructed given the reported allocation of inputs to each of the crops.[21] The risk index for household $i$ given input allocation $k$ to crop $m$ is defined as $R_i = \sum r_m k_m / \sum k_n$, where $r_m$ is the coefficient of variation of the yield of crop $m$.[22] Note here, that $r_m$ is only available for a subset of all crops $n$ (26 out of 42), such that $m \subseteq n$. Still, $\sum k_m$ represents roughly 90% of the total allocation of inputs in the sample. To reduce potential bias, I drop all observations from the sample which have no crop in their portfolio for which risk information is available, e.g. $\sum k_m = 0$ or $R_i = 0$, in one or both of the survey rounds.[23] As can be seen in Table 1, the risk index at baseline is higher in the treatment group (0.36) than in the control group (0.26). The difference is statistically significant at the 1% level.

There are two main caveats associated with the choice of the dependent variable. First, risk in crop returns not only stems from yield fluctuations but also from the variability in prices. Also, for domestically traded crops it is quite likely that yield and prices are negatively correlated, such that prices are high in years with a bad harvest and low in years with a good harvest. This would reduce the variability in returns. If these crops are also the crops that display the highest variability in yields, I would be grossly misinterpreting

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

[22] The distribution of the risk index as well as of the change in this variable between survey rounds is plotted in Figures D.2 and D.3 respectively in the appendix.

[23] Appendix B provides more information on how the variable is constructed. The robustness of my findings to different methods of aggregating the risk index is shown in Table E.2 in the appendix.

the results of this article. Second, as yields of different crops are not perfectly correlated, households might chose to increase the number of crops in their portfolio as strategy to diversify risk. As the risk index presented above ignores the correlation between crops' yields, it does not adequately predict the amount of risk a household is willing to take in his production decisions. In order to address these concerns, I test if my results are robust to alternative specifications of the dependent variable. I find that the NREGS also increases the price risk in the crop portfolio, as well as the standard deviation of portfolio returns, a variable that accounts for risk in crop returns, as well as the amount of correlation between the returns to different crops. The NREGS also seems to have a positive (albeit not statistically significant) effect on crop concentration, measured by the Herfindahl index, which suggests that farmers are not trading increased yield risk for higher diversification in the presence of the NREGS.[24]

# 4 Estimation strategy

The key prediction of the model described in Section 2 is that the introduction of the NREGS, *ceteris paribus*, increases the share of inputs allocated to risky crops if households were previously constrained in their crop choice by high levels of uncertainty regarding output levels and dysfunctional insurance and labor markets.

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

$$R_{ijt} = \beta_0 + \beta_1 D_{ijt} + \beta_2 X_{it} + \beta_3 Z_{jt} + u_i + \gamma_j + \delta_t + \upsilon_{ijt}. \tag{7}$$

[24]Results are reported in the appendix, Table E.2. Price risk is measured by the coefficient of variation of trend-corrected farm harvest prices. For more information on how the variables are constructed please refer to appendix B.

The dependent variable is the risk index of household $i$'s crop portfolio at time $t$. $D_{ijt}$ represents a household's access to the NREGS. Let $X_{it}$ be a set of time-varying household characteristics that affect preferences and crop choice (such as education, wealth, income and past experience with shocks) and $u_i$ be time-constant unobserved household characteristics (such as risk aversion, farming ability and land quality). $Z_{jt}$ is a set of time-varying village-level characteristics (e.g. weather trends, extension services, prices, etc.), $\gamma_j$ are time-constant village characteristics (such as the land's suitability for certain crops), $\delta_t$ is a time fixed-effect and $\upsilon_{ijt}$ is the error term.

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

$$\triangle R_{ij} = R_{ij,t+1} - R_{ij,t} = \beta_1 \triangle D_{ij} + \beta_2 \triangle X_i + \beta_3 \triangle Z_j + \triangle \delta + \triangle \upsilon_{ij}. \tag{8}$$

For $\beta_1$ to have a causal interpretation, the differences in the change of the risk index between the treatment and control groups must be entirely due to the NREGS. This assumption could be violated for a number of reasons. First, since the access to the NREGS is non-random, treatment could be correlated with potential outcomes of $R_{ijt}$. Second, households in the treatment and control group may not be following parallel trends in their crop choices. The remainder of this section discusses how I address these points.

This article uses four different treatment variables. First, as discussed above, I explore the universal nature of the NREGS by coding as 'treated' those households based in districts where the NREGS was introduced in 2006. Second, I use lagged block-level disbursements under the program as an indicator of the intensity of treatment, arguing that households living in blocks with higher past disbursements expect employment to be more readily available in situations of need. The average lagged disbursement in treatment districts is INR 14.27 Mio. with a standard deviation of 9.64 in 2009/10. Third, following the same logic, I use the lagged annual total of employment person-

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

days generated per job card at the block-level. In 2009/10, the number of person-days generated was 11.15 on average with a standard deviation of 5.58. Fourth, I explore the self-selection of households into the program by comparing the changes in the risk index of households who were registered with the NREGS by 2007 with the rest of the sample.

At the district level, the NREGS should have been introduced in the poorest districts first. This could potentially bias the estimates downwards because poorer districts are less likely to have extension services and marketing structures in place that would enable households to seize the opportunity to plant more profitable cash crops. However, in most states - and in Andhra Pradesh in particular - the prioritization of the poorest districts was not systematically implemented. In this sample, the general economic characteristics of treatment and control districts do not differ greatly.[26] The treatment intensity at the block level should also be exogenous to potential outcomes. Estimates could be biased if funds allocated to blocks responded to rainfall shocks and if these rainfall shocks also affected a household's input allocation decision. However, the amount of funds to be sanctioned per block is defined between December and March for the following financial year (April to March). Since I am using lagged values of disbursed funds, these amounts are fixed 14 to 18 months before household's decide on their input allocation.[27] Lastly, I explore differences in crop choices across households who registered with the NREGS or not. Here, the possibility that unobserved shocks or other time-varying variables affect both the decision to register and a household's crop choice cannot be ruled out. I employ matching techniques to reduce selection bias, but this is admittedly not sufficient to rule out non-random assignment.

The parallel trends assumption could be violated due to differences in crop productivity which cause the share of certain crops in total input allocation to increase independently of the NREGS. Given the small number of districts in the sample, this could significantly bias the results. District-wise time trends in the risk index of crop production are displayed in Figure 1. One of the treatment districts (Mahaboobnagar) displays

[26] See appendix B and Table E.3 for more information.

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

a decreasing trend in the risk index, while all other districts seem to be following the same trend.

Another - more subtle - violation of the parallel trends assumption could emerge from mean reversion in the dependent variable. Why might households with riskier crop portfolios display a negative change in the risk index? The reason could be effects of lagged shocks on current input choices which are rooted in the non-separability of production and consumption decision of agricultural households (Sadoulet and De Janvry, 1995). In a world with imperfect credit markets and risk, past shocks affect current wealth and therefore also current input allocation decisions. If household wealth is perfectly captured by the data, controlling for changes in wealth should eliminate any bias. If wealth is, however, also reflected in soil nutrition, which is affected by weather shocks and not captured in the data, then controlling for wealth is not sufficient (Foster and Rosenzweig, 2010a).

Assume that the risk index of each household's crop portfolio follows a modified AR(1) process, where - in the absence of a shock - the risk index at time $t+1$, $R_{t+1}$, is equal to a linear transformation of the risk index of the previous period plus some random noise, e.g $\rho R_t + \epsilon_{t+1}$.[28] In contrast, if a shock occurs, households with higher risk in their crop portfolio also face higher losses in agricultural production. This forces them to choose a more conservative crop portfolio in the following period. Formally, this process can be described as follows:

$$R_{t+1} = \rho R_t + \delta u_t + g(R_t)u_t + \epsilon_{t+1}. \tag{9}$$

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

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

choice would be the following:

$$\begin{aligned}\triangle R &= R_{t+1} - R_t \\ &= (\rho - 1)R_t + \delta u_t + g(R_t)u_t + \epsilon_{t+1}. \end{aligned} \tag{10}$$

In expectation this change would be $E(\triangle R) = (\rho - 1)R_t$. A placebo treatment effect is zero in expectation only if the process approaches a random walk (e.g. $\rho = 1$) or if the distribution of $R_t$ is equal in treatment and control groups. The placebo treatment effect is also different from zero if the occurrence of lagged shocks $u_t$ is different in both groups. The low number of districts used in this analysis warrants special attention to this phenomenon. As discussed earlier, baseline levels of risk as well as the occurrence of shocks are substantially different between treatment and control groups. I estimate the importance of mean reversion in the control group only and find estimates of $\rho - 1$, $\delta$ and $g(R_t)u_t$ equal to $-0.70$, 0.08 and $-0.44$ respectively.[29]

I account for shock induced mean reversion by adjusting equation (8) in a way that eliminates sources of correlation between $\triangle D_{ij}$ and $(\upsilon_{ij,t+1} - \upsilon_{ij,t})$. Using equation (10) to rewrite eq. (8) yields:

$$\begin{aligned}\triangle R_{ij} = \beta_1 \triangle D_{ij} &+ \beta_2 \triangle X_i + \beta_3 \triangle Z_j + \triangle \delta \\ &+ (\rho - 1)R_{ijt} + \delta u_{jt} + g(R_{ijt})u_{jt} + \triangle \upsilon_{ij}. \end{aligned} \tag{11}$$

Following Chay, McEwan, and Urquiola (2005), I estimate a simplified version, such as:

$$\begin{aligned}\triangle R_{ij} = \beta_1 \triangle D_{ij} &+ \beta_2 \triangle X_i + \beta_3 \triangle Z_j + \triangle \delta \\ &+ \beta_4 R_{ijt} + \beta_5 u_{jt} + \beta_6 R_{ijt} u_{jt} + \triangle \upsilon_{ij}. \end{aligned} \tag{12}$$

Before proceeding, one last empirical challenge needs to be addressed: within cluster correlation in $\triangle \upsilon_{ij}$. Throughout the article, I assume that the first-differenced errors

[29]The standard errors are 0.03, 0.08 and 0.19 respectively. I use the level and the square of $R_t$ as approximation for $g(R_t)$. Detailed results can be found in Table E.4. in the appendix.

are equicorrelated at the block level and estimate all equations in feasible GLS (FGLS) to improve efficiency (Cameron and Miller, 2015).[30] To guard against misspecification of the error structure, I additionally calculate Eicker-White standard errors clustered at the sub-district (block) level or district level depending on the level of aggregation of the regressors.

Since the number of clusters is fairly small, these standard errors are likely to be downward biased. Inference on the main hypothesis is therefore based on p-values obtained from performing a wild cluster-bootstrap with Rademacher weights as suggested by Cameron, Gelbach, and Miller (2008).[31] In a more recent paper, Cameron and Miller (2015) suggest the use of Webb's (2014) weights if the number of clusters is smaller than ten, which seems reasonable when using a district level treatment variable. P-values of the respective treatment variable using both versions of the bootstrap with 4,999 replications are reported in the results section.[32]

# 5 Results

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

## 5.1 Identifying profitable production strategies

To understand inhowfar households' crop choice can improve their income from agricultural production, I estimate an agricultural production function, linking the total value of

[30] Estimation if FGLS is more efficient than OLS as it specifies a model for the error variances.

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

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

agricultural output $Q_{ijt}$ to input allocation $K_{ijt}$, labor $L_{ijt}$, plot size $A_{ijt}$ and risky crop choice $R_{ijt}$. The agricultural production function is assumed to be translog, in which the choice of crops affects output multiplicatively,

$$Q_{ijt} = (K_{ijt}^{\beta_1} L_{ijt}^{\beta_2} A_{ijt}^{\beta_3}) e^{g(R_{ijt})}. \tag{13}$$

Because it seems very likely that increasing the average risk in a crop portfolio is only beneficial to a certain extent, beyond which risk is simply too high to increase output, $g(R_{ijt})$ is a non-linear function. The production function described in equation (13) can be estimated by log-transforming the data and controlling for shocks $Z_{ijt}$, unobserved characteristics $\gamma_{ij}$ and time effects $\delta_t$. Again, $g(R_{ijt})$ is approximated by the level and the square of $R_{ijt}$:

$$\begin{aligned} \ln(Q_{ijt}) = \beta_0 + \beta_1 \ln(K_{ijt}) + \beta_2 \ln(L_{ijt}) + \beta_3 \ln(A_{ijt}) + \beta_4 R_{ijt} + \beta_5 R_{ijt}^2 \\ + \beta_6 Z_{ijt} + \gamma_{ij} + \delta_t + \upsilon_{ijt}. \end{aligned} \tag{14}$$

The production function is estimated in FGLS. The estimates suggest that households could significantly raise the value of their agricultural production if they were to increase the share of inputs allocated to riskier crops (c.f. Table 2). However, this is only true up to a certain level. The square of the risk index is statistically significant at the 5% level in all specifications. Based on the estimates in col. (4), predicted agricultural output reaches its maximum at a risk index of 0.41.[33] Beyond this point, a further increase in risk would reduce total agricultural output. Average risk levels in households' crop portfolios are well below this value; in the survey round of 2007 the average risk index was 0.36 in the treatment group and 0.26 in the control group (c.f. Table 1).

As can be seen from columns (1) and (2) of Table 2, the estimates are not affected by the exclusion of labor from the agricultural production function. The analysis here is restricted to the survey round of 2007.[34] Columns (3) and (4) show estimates from the

[33]C.f. Figure D.4 in the appendix.

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

first-differenced sample. In column (4), the effect of rainfall is allowed to vary with the amount of risk in a household's crop portfolio. The interaction term of rainfall and the risk index is positive and statistically significant at the 5% level. At the optimal risk level of 0.41, the marginal effect of rainfall is as high as 0.22 with a standard error of 0.15. Other variables, such as the amount of inputs allocated, total cultivated area and labor have the expected sign and are all statistically significant.[35]

The robustness of these results can be gauged by comparing returns per hectare of different crops to the variability of these returns.[36] Plotting average returns against the standard deviation of these returns in Andhra Pradesh, reveals a clear positive relationship between average returns and their volatility, indicating again that the riskiness of crops is strongly correlated with returns to producing these crops (c.f. Figure 2).[37]

## 5.2 Does the NREGS support households in coping with shocks?

Next, I estimate to which extent the NREGS helps households in coping with shocks. The theoretical argument presented in Section 2 is based on the premise that work provision within the NREGS sufficiently reacts to increasing demand in the case of a shock. This can be tested by analyzing whether deviations from mean rainfall levels, as well as households' self-reported shocks, drive changes in the number of days households report to have worked for the NREGS. The analysis is restricted to phase one districts; thus only households who had access to the NREGS in both survey rounds are considered.

The results suggest that the number of days worked for the NREGS changes considerably with variation in rainfall levels (c.f. Table 3).[38] The greatest change is observed for

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

[36]These statistics are available at the state-level for major crops and for the years 1996 to 2006 from the Cost of Cultivation Statistics.

[37]Many of these commodities are traded internationally, such that risk-aversion of farmers alone can probably not explain the observed correlation between the riskiness of crops and their returns. Other reasons could be differences in the concentration of supply or demand between crops. Analyzing the reasons for the apparent relationship between risk and returns in crop portfolios is beyond the scope of this article.

[38]In the first two columns, the total number of days worked in the past 12 months is the dependent variable; in the last two columns it is the log of this variable.

lagged rainfall levels - that is, cumulative rainfall in the agricultural year preceding the period of reference. The coefficient of the lagged rainfall variable is negative 66.2, which implies that households worked 6.6 more days for the NREGS if lagged rainfall levels were 10% below average. This supports the assumption that the NREGS helps households in coping with shocks, because households use the program to smooth income ex post - for instance, after harvest and after agricultural products have been sold.[39] Self-reported shocks also seem to increase the number of days worked for the NREGS, although the effect is not statistically significant.[40]

To quantify the contribution of the NREGS to households' risk coping, agricultural losses due to rainfall shortages can be compared with income gains through the NREGS. The agricultural production function estimated in Table 2 (col. 6) suggests that a deviation from average annual rainfall by negative 25%, would reduce agricultural output by 5.6% at the optimal level of the risk index. For the average household, this implies a nominal loss of about INR 1,654 (or US$ 35.7 in constant July 2006 values). The same deviation in lagged rainfall would lead households to work about 16.5 more days for the NREGS, which would generate an additional income of INR 1,067 (US$ 23) at mean wages observed in the sample. The NREGS thus allows households to compensate about 64% of agricultural production losses caused by rainfall shortages. Since rainfall fluctuations are among the most important sources of risk for rural households, these results suggest that the NREGS could indeed have an insurance effect in Andhra Pradesh.

## 5.3 The effects of the NREGS on households' crop choices

Table 4 reports estimates of the effect of the NREGS on households' input allocation decisions. As described in Section 4, all equations are estimated in the first-differenced sample, and controls for initial conditions are included in columns (2), (4) and (6). To isolate the insurance effect, controls include variables that might be affected by the NREGS and might influence a household's crop choice through effects other than the insurance

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

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

effect. These variables are household off-farm income and wealth, as well as key farming characteristics, such as the size of cultivated land, irrigation and total value of variable inputs allocated.[41] All specifications also control for self reported shocks, access to other government programs and rainfall levels (current and lagged). The constant captures state-wide changes in input and output prices, weather trends that are not captured by rainfall data and other changes at the state level that could influence a household's crop choice.

The results show a positive effect of the NREGS on the risk index of a household's crop portfolio. Consistent with the higher prevalence of shocks in the treatment districts and higher initial values of the risk index, controlling for mean reversion increases the estimated effect of the NREGS. Given the low number of clusters, inference should be based on the p-values obtained from the wild-cluster bootstrap. The effect of the introduction of the NREGS at district level is statistically significant at the 10% and 5% level in columns (1) and (2), respectively (using Webb weights). In columns (3) to (6), alternative explanatory variables such as cumulative expenditure and total employment generated per job card under the NREGS are considered. The coefficient on cumulative spending in the sub-district is statistically significant at the 10% and 5% level, depending on the specification considered. Only the amount of employment generated in the sub-district does not yield statistically significant effects when inference is based on the wild-cluster bootstrap.

Results presented in column (2) suggest that the risk index in households' crop portfolios increased by 7.2 percentage points due to the introduction of the NREGS at the district level. Given that the risk index in the treatment group was 0.36 at baseline, the introduction of the NREGS raised the average risk index to 0.43 (absent any shock induced mean reversion), which is remarkably close the optimal risk index of 0.41 identified in Table 2.

[41] Household off-farm income consists, inter alia, of income generated through the NREGS in the past 12 months. Optimally, this should be a lagged value because input allocation decisions are taken at the beginning of the season, while the income variable refers to the time period shortly after these allocative decisions were taken. Unfortunately, the survey does not include this information. Table E.5 in the appendix shows that the results are not influenced by changes in income or changes in total input allocation.

In terms of economic relevance, the results suggest that per additional day of employment generated in the block, each household would increase the risk index by 0.24 percentage points (col. 6). One standard deviation increase in the number of person-days generated per job card (6.9) would increase a household's risk index by 1.66 percentage points and raise net income from agricultural production, *ceteris paribus*, by about INR 640 (or US$ 13.9). This is particularly interesting from a cost-benefit perspective, since these net income gains per household are higher than the wage cost (evaluated at the sample average of observed NREGS wages) of creating 6.9 days of employment under the NREGS, e.g. INR 467 (US$ 10). Of course, wage costs make up for only a part of overall program costs and not all of the NREGS participants own their own land, but nevertheless the magnitude of this effect is striking.

### 5.4 Robustness checks

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

As a first robustness check, I test whether households that registered with the NREGS change their input allocation more strongly than households who are not registered with the NREGS. To account for potential self-selection bias, I match households on their probability to register with the NREGS by using entropy balancing, a method developed by Hainmueller (2012).[42] Table 5 reports the effects of registering with the NREGS on the risk index of households' crop portfolios. I find that households that already registered with the NREGS in 2007 are more likely to grow a higher share of risky crops in the follow-up period. Five different specifications are presented: Columns (1) and (2) show estimates without matching, where column (2) additionally controls for initial conditions. Column (3) excludes all households that did not register with the NREGS

[42] More details on the matching strategy can be found in appendix C. The distribution of the propensity scores is displayed in Figure D.5, and the resulting covariate balance is shown in Table E.6.

by 2009/10.[43] Column (4) shows estimates for the matched sample. As we can see, the effects are only slightly smaller when matching households on their probability to register with the NREGS. Overall, the effects are of a similar size in most specifications though somewhat lower than the estimates presented in Table 4, column (2). Column (5) shows the estimation results for the full sample without matching. Here, being registered by 2009/10 is the main explanatory variable. As we would expect, households that registered with the NREGS only shortly before or even after deciding on their crop portfolio, did not alter their input allocation in a meaningful way.

As mentioned before, the NREGS can affect household decisions via different mechanisms. This set of robustness checks seeks to understand if the observed effect of the NREGS is indeed an insurance effect and not due to alternative mechanisms such as the increase in income of participating households or the change in agricultural wages. If, for example, risky crops are also more capital intensive, then observed outcomes could also be driven by increases in income and wealth or better access to credit of participating households. Likewise, if risky crops are also less labor intensive, then observed outcomes could be driven by wage changes due to the NREGS instead of its insurance effect. I therefore test if the NREGS affects on the labor intensity or cost intensity of households' crop portfolios (c.f. Table 6). Labor intensity per crop is calculated as the share of expenditures on labor in total production costs. Cost intensity is defined as the total production cost that has to be incurred per hectare for each crop.[44] The coefficient on labor intensity is positive and statistically significant at the 10% level in one of the two specifications. This indicates that the NREGS, if anything, increases the labor intensity of crop portfolios (c.f. cols. (1) and (2)). The coefficient on cost intensity is also positive, suggesting that households are able to spend more on their agricultural production. Again, only one out of two specifications is statistically significant at conventional levels (c.f. cols (3) and (4)). I can therefore not rule out the possibility that the observed

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

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

change in the risk index is driven by the fact that the cultivation of riskier crops is also more costly, and households are becoming wealthier due to the NREGS.

As a final set of robustness checks, I explore the extent to which heterogeneity in treatment effects is in line with the predictions of the theoretical model.[45] I find evidence that households who registered with the program in 2007 while experiencing a shock to agricultural production (i.e. a rainfall shock), adjusted their production portfolio towards riskier crops, while households who registered with the NREGS despite experiencing favorable rainfall levels did not alter their production decisions.[46] This suggests that households might register with the NREGS for different reasons. For some households, consumption needs are a much more important reason for registering with the program than the insurance effect. These households would need to work for the NREGS as much as possible to satisfy their consumption needs - even in good years, and are not likely to cultivate higher-risk crops despite working for the NREGS. Other households might rely on the NREGS only to cope with shocks; these would also be the households that are enabled to adjust their input allocation towards more profitable crops through the insurance effect. I also find suggestive evidence that treatment effects are smaller in villages with existing watershed development projects, crop insurance programs and public works schemes, which would be in line with the NREGS having an insurance effect. If households already have access to other insurance or risk mitigation mechanisms, they do not need the NREGS to cope with shocks. However, none of these village-level differences are statistically significant at the 10% level.

# 6 Conclusions

This article presents theoretical and empirical evidence that an employment guarantee, such as the NREGS in India, improves households' ability to cope with shocks in agriculture by guaranteeing income opportunities in areas where and time periods when they previously did not exist. By improving the risk management of households, the NREGS

[45]Results are reported in the appendix, Table E.7.

[46]For better visualization, the marginal effect of registering with the NREGS conditional on lagged rainfall is plotted in Figure D.6 in the appendix.

enables households to switch their production towards riskier but also higher profitability products and to generate higher incomes from agricultural production.

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

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

The findings here contain some lessons for the ongoing debates on the effectiveness of the NREGS and for other countries considering the implementation of such schemes. First, for the insurance effect to unfold, the design of a public works program is crucial.

An employment guarantee that is entitled by law and entails adequate grievance redress mechanisms provides households with the necessary protection against agricultural production risks to enable them to take more risks in their production and investment decisions. Additionally, it is crucial not to severely limit the number of workdays, otherwise such a scheme's potential as a risk-coping instrument cannot be realized. Second, implementation matters. The data analyzed in this article cover only the state of Andhra Pradesh. This is, *inter alia*, because the performance of the NREGS in terms of the number of workdays generated per eligible household varies immensely across states and even across districts in India. Andhra Pradesh is one of the best performing states in the implementation of the NREGS, so it goes without saying that many of the effects captured in this article might not be found in all Indian states. Third, working for a public works scheme is always associated with opportunity costs. In countries or regions with well functioning off-farm labor markets, providing public works schemes might not be necessary. A food-for-work program or cash-for-work program is only effective in areas and time periods where labor is in surplus.

# References

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

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

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

Berg, E., S. Bhattacharyya, R. Durgam, and M. Ramachandra. 2012. *Can rural public* *works affect agricultural wages? Evidence from India*. CSAE Working Paper 2012-05, Oxford: Center for the Study of African Economies.

Bhargava, A.K. 2014. *The impact of India's rural employment guarantee on demand for agricultural technology*. IFPRI Discussion Paper 1381, Washington, DC: International Food Policy Research Institute.

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

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

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

Carter, M.R., A. de Janvry, E. Sadoulet, and A. Sarris. 2014. *Index-based weather insurance for developing countries: A review of evidence and a set of propositions for upscaling*. FERDI Working Paper P111, Clermont-Ferrand: Fondation pour les Etudes et Recherches sur le Developpement International.

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

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

Cole, S., X. Gine, and J. Vickery. 2013. *How does risk management influence production decisions? Evidence from a field experiment*. World Bank Policy Research Working Paper 6546, Washington, DC: World Bank.

Deininger, K., and Y. Liu. 2013. *Welfare and Poverty Impacts of India's National Rural*

*Employment Guarantee Scheme: Evidence from Andhra Pradesh*. World Bank Policy Research Paper 6543, Washington, DC: World Bank.

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

Desai, P., Sonalde amd Vashishtha, and O. Joshi. 2015. *Mahatma Gandhi National Rural Employment Guarantee Act: A Catalyst for Rural Transformation*. New Delhi: National Council of Applied Economic Research.

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

Dutta, P., R. Murgai, M. Ravallion, and D. Van de Walle. 2012. *Does India's Employment Guarantee Scheme Guarantee Employment?*. World Bank Policy Research Working Paper 6003, Washington, DC: World Bank.

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

Fetzer, T. 2014. *Can Workfare Programs Moderate Violence? Evidence from India*. Suntory and Toyota International Centres for Economics and Related Disciplines, LSE, Working Paper 53, London: London School of Economics.

Foster, A.D., and M.R. Rosenzweig. 2010a. *Is There Surplus Labor in Rural India?*. Department of Economics Working Paper 85, New Haven, CT: Yale University.

—. 2010b. "Microeconomics of technology adoption." *Annual Review of Economics* 2:395–424.

Galab, S., S. Vijay Kumar, P. Prudvikhar Reddy, R. Singh, and U. Vennam. 2011. *The Impact of Growth on Childhood Poverty in Andhra Pradesh: Initial Findings from India*. Young Lives Round 3 Survey Report: India, Oxford: University of Oxford.

Hainmueller, J. 2012. “Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies.” *Political Analysis* 20:25–46.

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

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

Jha, R., R. Gaiha, and M.K. Pandey. 2012. “Net transfer benefits under India’s Rural Employment Guarantee Scheme.” *Journal of Policy Modeling* 34:296–311.

Johnson, D. 2009. *Can Workfare Serve as a Substitute for Weather Insurance? The Case of NREGA in Andhra Pradesh*. Centre for Micro Finance Working Paper Series No 32, Chennai: Institute for Financial Management and Research.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry. 2014. “Agricultural Decisions after Relaxing Credit and Risk Constraints.” *Quarterly Journal of Economics* 129:597–652.

Kaur, S. 2014. *Nominal Wage Rigidity in Village Labor Markets*. Working Paper, New York: Columbia University.

Klonner, S., and C. Oldiges. 2014. *Safety Net for India’s Poor or Waste of Public Funds? Poverty and Welfare in the Wake of the World’s Largest Job Guarantee Program*. Department of Economics Working Papers 0564, Heidelberg: University of Heidelberg.

Ministry of Rural Development, Government of India. 2012. *MGNREGA Sameeksha: An Anthology of Research Studies on the Mahatma Gandhi National Rural Employment Guarantee Act, 2005. 2006-2012, edited and compiled by Mihir Shah, Neelakshi Mann and Varad Pande*. New Delhi: Orient BlackSwan.

Mobarak, A.M., and M.R. Rosenzweig. 2013. “Informal Risk Sharing, Index Insurance, and Risk Taking in Developing Countries.” *American Economic Review* 103:375–80.

Muralidharan, K., P. Niehaus, and S. Sukhtankar. forthcoming. “Building state capacity: Evidence from biometric smartcards in India.” *American Economic Review*, pp. .

Sadoulet, E., and A. De Janvry. 1995. *Quantitative development policy analysis*. Baltimore, MD: Johns Hopkins University Press.

Shah, M., and B.M. Steinberg. 2015. *Workfare and Human Capital Investment: Evidence from India*. NBER Working Paper 21543, Cambridge, MA: National Bureau of Economic Research.

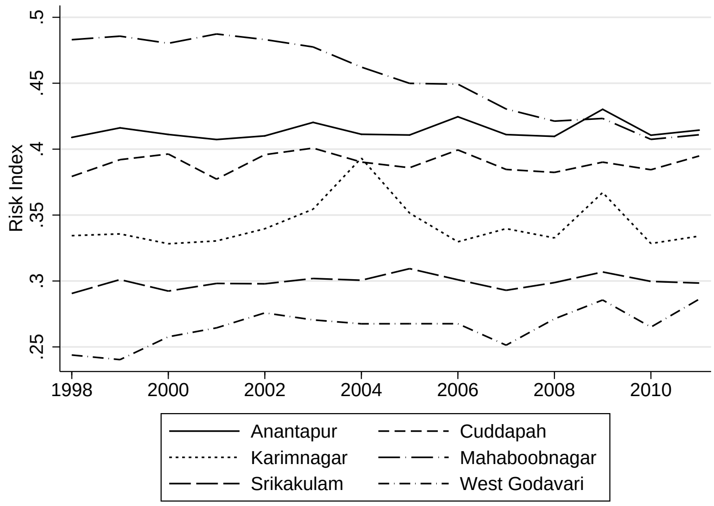
Suri, T. 2011. “Selection and Comparative Advantage in Technology Adoption.” *Econometrica* 79:159–209.

Van Den Berg, M. 2002. “Do public works decrease farmers’ soil degradation? Labour income and the use of fertilisers in India’s semi-arid tropics.” *Environment and Development Economics* 7:487–506.

Webb, M.D. 2014. *Reworking Wild Bootstrap Based Inference for Clustered Errors*. Queen’s Economics Department Working Paper 1315, Ontario: Queen’s University.

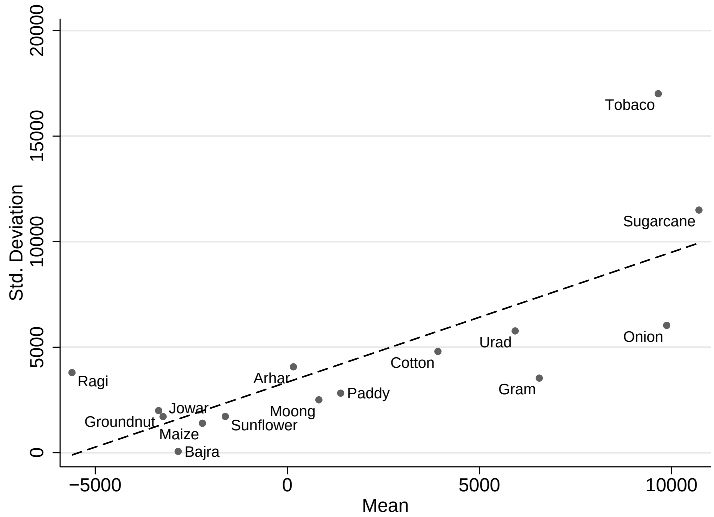
# Figures

Figure 1: District-wise risk-index of land use




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

Figure 2: Returns per hectare of major crops




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

# Tables

Table 1: Baseline characteristics

| | Treatment | | Control | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | p-value |
| *Household characteristics* | | | | | |
| Male household head | 0.96 | 0.20 | 0.97 | 0.18 | 0.41 |
| Age of household head | 41.93 | 12.13 | 41.02 | 11.85 | 0.25 |
| Household head is literate | 0.32 | 0.47 | 0.24 | 0.43 | 0.01 |
| Household size | 6.10 | 2.62 | 5.61 | 2.08 | 0.00 |
| Wealth index | 0.39 | 0.13 | 0.39 | 0.20 | 0.79 |
| Annual income, off-farm activities | 24.70 | 24.82 | 19.82 | 26.17 | 0.00 |
| Hh benefits from credit/training program | 0.62 | 0.49 | 0.58 | 0.49 | 0.17 |
| Any serious debts | 0.63 | 0.48 | 0.47 | 0.50 | 0.00 |
| Able to raise 1000 rupees in one week | 0.61 | 0.49 | 0.33 | 0.47 | 0.00 |
| *Farm characteristics* | | | | | |
| Value of agr. production | 28.49 | 45.76 | 24.43 | 125.14 | 0.56 |
| Value of variable inputs | 14.51 | 21.34 | 14.46 | 69.62 | 0.99 |
| Area cultivated (acres) | 4.15 | 4.57 | 2.74 | 5.47 | 0.00 |
| Time in crop production (hours per year) | 2085 | 2280 | 1369 | 1310 | 0.00 |
| Irrigated area (% of total) | 0.18 | 0.32 | 0.14 | 0.31 | 0.07 |
| Fertilizer (dummy) | 0.98 | 0.15 | 0.87 | 0.33 | 0.00 |
| HYV seeds (dummy) | 0.77 | 0.42 | 0.63 | 0.48 | 0.00 |
| Participated in labor sharing (dummy) | 0.75 | 0.43 | 0.78 | 0.41 | 0.23 |
| Risk index of crop portfolio | 0.36 | 0.12 | 0.26 | 0.08 | 0.00 |
| SD of crop returns | 2472 | 1110 | 3111 | 2129 | 0.00 |
| Risk index (CV of trend corrected yield) | 0.22 | 0.10 | 0.16 | 0.07 | 0.00 |
| Risk index (alt. aggregation) | 0.37 | 0.12 | 0.27 | 0.08 | 0.00 |
| Price risk of crop portfolio | 0.15 | 0.04 | 0.18 | 0.08 | 0.00 |
| SD of portfolio return | 2239 | 1093 | 2889 | 2125 | 0.00 |
| Labor intensity of crop portfolio | 0.27 | 0.07 | 0.28 | 0.07 | 0.00 |
| Cost intensity of crop portfolio | 21166 | 7374 | 26207 | 10097 | 0.00 |
| Herfindahl index of crop portfolio | 0.76 | 0.25 | 0.80 | 0.23 | 0.00 |
| Number of crops | 2.04 | 1.03 | 2.08 | 1.33 | 0.65 |
| *Shocks* | | | | | |
| Rainfall (deviation) | 0.33 | 0.28 | -0.06 | 0.16 | 0.00 |
| Rainfall (deviation, lag) | -0.39 | 0.10 | -0.12 | 0.10 | 0.00 |
| Self-reported shock | 0.81 | 0.39 | 0.52 | 0.50 | 0.00 |
| *NREGS participation* | | | | | |
| Household registered with NREGS | 0.66 | 0.47 | 0.00 | 0.00 | 0.00 |
| Household generated income from NREGS | 0.54 | 0.50 | 0.00 | 0.00 | 0.00 |
| Income, NREGS | 1.24 | 2.39 | 0.00 | 0.00 | 0.00 |
| Observations | 750 | | 337 | | |

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

Source: Author's estimation based on the Young Lives data.

Table 2: Agricultural Production Function

| | 2007 cross-section | | First Differences | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Risk index of crop portfolio | 5.981** | 6.236** | 6.948* | 6.546* |
| | (2.039) | (2.013) | (3.182) | (3.015) |
| Risk index of crop portfolio (squared) | -7.526** | -7.275** | -8.339* | -8.619* |
| | (2.363) | (2.249) | (4.054) | (3.827) |
| Variable inputs (log) | 0.887*** | 0.818*** | 0.746*** | 0.753*** |
| | (0.112) | (0.106) | (0.124) | (0.123) |
| Area cultivated (acres, log) | 0.790*** | 0.622** | 0.667*** | 0.681*** |
| | (0.232) | (0.202) | (0.194) | (0.191) |
| Labour (hours, log) | | 0.230*** | | |
| | | (0.056) | | |
| Rainfall (deviation) | -0.349 | -0.401 | -0.010 | -1.164$^{+}$ |
| | (0.436) | (0.422) | (0.155) | (0.627) |
| Rainfall (deviation) × Risk index | | | | 3.386* |
| | | | | (1.580) |
| Observations | 1087 | 1087 | 1087 | 1087 |
| $R^2$ | 0.299 | 0.323 | 0.127 | 0.130 |

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

Table 3: Number of days worked with NREGS

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | NREGS days | | NREGS days (log) | |
| Rainfall (deviation, lag) | -67.245*** | -66.185*** | -3.128*** | -3.102*** |
| | (16.498) | (16.803) | (0.881) | (0.841) |
| Rainfall (deviation) | -30.766*** | -33.383*** | -1.006+ | -1.096* |
| | (6.795) | (6.883) | (0.576) | (0.523) |
| Self-reported shock | 1.918 | 1.774 | 0.144 | 0.141 |
| | (4.148) | (4.500) | (0.093) | (0.102) |
| Additional controls | No | Yes | No | Yes |
| Observations | 740 | 740 | 740 | 740 |
| $R^2$ | 0.055 | 0.070 | 0.087 | 0.100 |

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

Table 4: Effect of the NREGS on risk index of crop portfolio

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| NREGS introduced in district | 0.038$^{*}$ | 0.072$^{**}$ | | | | |
| | (0.011) | (0.017) | | | | |
| Cumulative expend., NREGS (log, lag) | | | 0.007$^{**}$ | 0.015$^{***}$ | | |
| | | | (0.002) | (0.003) | | |
| Employment per JC generated, NREGS (lag) | | | | | 0.002$^{+}$ | 0.002 |
| | | | | | (0.001) | (0.002) |
| Rainfall (deviation) at baseline | | 0.120$^{*}$ | | 0.102 | | 0.130$^{+}$ |
| | | (0.051) | | (0.073) | | (0.075) |
| Risk index at baseline | | -0.602$^{***}$ | | -0.659$^{***}$ | | -0.659$^{***}$ |
| | | (0.061) | | (0.038) | | (0.038) |
| Risk index × Rainfall (deviation) | | -0.110 | | -0.055 | | -0.069 |
| | | (0.142) | | (0.097) | | (0.099) |
| Bootstrap p-value of main treatment variable | | | | | | |
| Rademacher weights: | 0.107 | 0.047 | 0.072 | 0.015 | 0.326 | 0.388 |
| Webb weights: | 0.099 | 0.045 | 0.062 | 0.013 | 0.315 | 0.391 |
| Observations | 1087 | 1087 | 1087 | 1087 | 1087 | 1087 |
| $R^2$ | 0.068 | 0.443 | 0.066 | 0.431 | 0.058 | 0.389 |

Notes: Estimation in FGLS. Dep. var.: Risk index of a household's crop portfolio. Additional controls are variable inputs (log), area cultivated (log), share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, annual income from off-farm activities (log), if household benefits from credit/training program, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Standard errors (clustered at the district in cols. (1) and (2) and at the sub-district in cols. (3) to (6)) in parentheses. P-values are obtained by performing a wild cluster-t bootstrap with 4,999 replications and two alternative weights. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Table 5: Effect of registration with the NREGS on risk index of crop portfolio

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| NREGS registered (2007) | 0.019$^{+}$ | 0.022$^{***}$ | 0.034$^{**}$ | 0.026$^{*}$ | |
| | (0.010) | (0.006) | (0.010) | (0.010) | |
| NREGS registered (2009/10) | | | | | 0.007 |
| | | | | | (0.006) |
| Rainfall (deviation) at baseline | | 0.175$^{**}$ | 0.194$^{***}$ | 0.205$^{**}$ | 0.227$^{***}$ |
| | | (0.054) | (0.055) | (0.061) | (0.056) |
| Risk index at baseline | | -0.635$^{***}$ | -0.572$^{***}$ | -0.477$^{***}$ | -0.500$^{***}$ |
| | | (0.040) | (0.058) | (0.073) | (0.061) |
| Risk index × Rainfall (deviation) | | -0.134 | -0.276$^{*}$ | -0.373$^{*}$ | -0.342$^{*}$ |
| | | (0.096) | (0.128) | (0.146) | (0.136) |
| Observations | 1087 | 1087 | 838 | 1087 | 1087 |
| $R^2$ | 0.058 | 0.402 | 0.459 | 0.387 | 0.395 |

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

Table 6: Effect of NREGS on labor and cost intensity

| | Labor intensity | | Cost intensity | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| NREGS introduced in district | 0.011* | | 1727.235*** | |
| | (0.004) | | (382.798) | |
| NREGS registered (2007) | | 0.003 | | 743.686 |
| | | (0.005) | | (498.628) |
| Observations | 1010 | 1010 | 1010 | 1010 |
| $R^2$ | 0.03 | 0.03 | 0.03 | 0.02 |

Notes: Estimation in FGLS. Dependent variable in cols. (1) & (2) is labor intensity of crop portfolio, in cols. (3) & (4) cost intensity of crop portfolio. Additional controls are area cultivated (log), share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, if household benefits from credit/training program, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Variable definitions and sources are described in appendix B. Standard errors (clustered at the district in cols. (1) and (3), and clustered at the sub-district in cols. (2) and (4)) in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

# APPENDIX

# A Mathematical Appendix

## Deterministic Case

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

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

Differentiating the Lagrange with respect to the choice variables, leads to the following first order conditions:[47]

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

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

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

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

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

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

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

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

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

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

Rearranging the first order conditions (S.1.1) and (S.1.2) gives:

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

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

And including (S.1.10) and (S.1.11) into (S.1.3)-(S.1.9) gives our decision rules:

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

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

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

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

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

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

Equation (S.1.17) can be rewritten to describe the optimal consumption rule over both periods given credit constraints:

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

If the credit constraint is binding, $\varphi$ is greater than zero and the marginal utility from consumption in the planting period greater than the discounted marginal utility from consumption in the harvesting period. This means that consumption in the planting stage is lower than what could be achieved if the credit constraints were not binding. Including equation (S.1.18) into equation (S.1.14) also reveals the effect of the credit constraint on input allocation:

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

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

## Stochastic Case

When introducing uncertainty, the Lagrange becomes the following:

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

Note here that the household forms expectations not only about the utility he derives from consumption in period 2, but also about the level of consumption that can be achieved. Differentiating the Lagrange with respect to the choice variables, leads to the following first order conditions:[48]

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

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

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

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

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

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

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

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

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

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

Rearranging (S.1.20) and (S.1.21) gives:

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

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

And the optimal consumption rule becomes:

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

The consumption rule - equation (S.1.31) - changes slightly when introducing uncertainty because for any expected consumption level $C_2$, expected utility $EU_2(C_2)$ is lower than the utility of the expected value $U_2(E(C_2))$, and marginal expected utility is higher than the marginal utility of the expected value. Since all other variables remain constant, $C_2$ has to be higher relative to $C_1$ under uncertainty for the identity to hold. This is equivalent with the well-known argument that risk decreases current consumption levels and enhances savings.

Including (S.1.29) and (S.1.30) into (S.1.22)-(S.1.27) gives our decision rules for $l_1^d$,

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

for $l_1^s$,

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

for $i^d$,

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

for $i^s$,

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

for $a^d$,

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

and $a^s$,

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

resulting in:

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

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

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

We can see from equation (S.1.37) that both risk and credit constraints go in the same direction and reduce the input allocation to the risky crop. More importantly, it also shows that uncertainty reduces input allocation to the risky crop relative to the deterministic crop even if credit constraints are not binding.

# B Data Description

## Young Lives Survey

- Reference periods: In most questions the references period of the YLS are the 12 months prior to the date of interview. However, for all questions on agricultural production, the period of reference is a particular agricultural year. In 2007, the reference period for agricultural production was June 2005 to May 2006. In 2009/10, the reference period was June 2008 to May 2009. In order to be consistent with the survey data, I use this definition of the agricultural year throughout the paper.

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

## Crop production

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

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

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

In order to gauge the robustness of my results, I recalculate the main results using a range of alternative risk measures, see Table E.2. In column (1), I use the coefficient of variation of the trend-corrected yield as risk measure for each crop. In column (2), I compute a risk measure that takes into account only those crops for which information is available, e.g. $R_i^{alt} = \sum r_m k_m / \sum k_m$. Here $r_m$ is again the coefficient of variation of the yield of crop $m$. And finally, column (3) uses the CV of trend-corrected prices as dependent variable, col. (4) the SD of portfolio returns and in col. (5) the Herfindahl index of the crop portfolio. A description of how price risk and the SD of portfolio returns is calculated can be found below.

To calculate the risk-index in district-level land use (Figure 1), I merge this information with the district wise land use statistics, which are also available from the Directorate of Economics and Statistics, Ministry of Agriculture, GoI. The risk index is calculated as follows: $R_{jt} = \sum r_m a_{mjt} / \sum a_{mjt}$, where $a_{mjt}$ is the land allocated to crop $m$ in district $j$ at time $t$ and $r_m$ is the coefficient of variation of crop $m$.

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

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

- Portfolio risk: Portfolio risk is calculated as the square root of the variance in portfolio returns. Consistent with modern portfolio theory, the variance of portfolio return is defined as $\sigma^2 = \sum w_i \sigma^2 = \sum_i \sum_{i \neq j} w_i w_j \sigma_i \sigma_j \rho_{ij}$, where $w_i$ is the share of crop $i$ in the portfolio, $\sigma_i$ is the standard deviation of the return of crop $i$ and $\rho_{ij}$ is the correlation coefficient of returns of crop $i$ and $j$. The variance of portfolio returns is calculated using the crop-wise returns per hectare in the cost of cultivation statistics for Andhra Pradesh from 1995/96 to 2009/10. The data were obtained from the Directorate of Economics and Statistics, Ministry of Agriculture, GoI, and are available online: $http://eands.dacnet.nic.in$.

- Price risk: Price risk is calculated as the weighted average of the coefficient of variation of trend-corrected Farm Harvest Prices in Andhra Prasdesh. District-level Farm Harvest Prices are available for 13 crops and the years 1998/99 to 2009/10. The data were obtained from the Directorate of Economics and Statistics, Ministry of Agriculture, GoI, and are available online: $http://eands.dacnet.nic.in$.

## Rainfall data

The rainfall data used in this paper were compiled by the Directorate of Economics and Statistics, Government of Andhra Pradesh. Rainfall data are available at the sub-district (i.e. block) level for the years 2002/03 to 2011/12. Rainfall deviation and rainfall deviation (lag) describe the relative deviation of cumulative rainfall over the agricultural year (June - May) from the long-term average, e.g. $devrain^{05/06} = (rf^{05/06} - \overline{rf})/\overline{rf}$. For the 2007 round of interviews, current rainfall uses the 2005/06 rainfall, and lagged rainfall uses rainfall in the agricultural year 2004/05. For the 2009/10 round of interviews, current rainfall uses the rainfall in the agricultural year 2008/09, and lagged rainfall uses data from the agricultural year 2007/08.

## NREGS data

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

- NREGS introduced in District: This variable is an indicator which equals 1 if a household has access to the NREGS at the district level at the beginning of the agricultural cycle (i.e. when most input allocation decisions have to be taken). The YLS was collected in six districts, four of which are treatment districts: Anantapur, Cuddapah, Karimnagar and Mahaboobnagar, and two are control districts Srikakulam and West Godavari.

  The period of reference for the 2007 round of interviews is the agricultural year 2005/06 (June 2005 to May 2006). According to the Employment Generation Report of the Government of Andhra Pradesh, NREGS activities started in April 2006 in half of the 'treatment' sub-districts and in May 2006 in the other half of the 'treatment' sub-districts. I therefore set, $D_{ijt}$ equal to 0 for all households at baseline.[49]

The period of reference for the 2009/10 interviews is the agricultural year 2008/09. By that time, NREGS works had started in the districts Anantapur, Cuddapah, Karimnagar and Mahaboobnagar. In Srikakulam and West Godavari the introduction of the NREGS was in August 2007 and in March 2008 respectively. Since activities started only slowly in most sub-districts of Srikakulam, I treat this district as control district despite the introduction of the NREGS mid 2007. Figure D.1 in the appendix shows monthly employment creation at the sub-district level in Srikakulam and West Godavari, as well as the average for all 'treatment' sub-districts. As can be seen, two sub-districts in Srikakulam had substantial NREGS employment creation in the two months before the relevant agricultural year. In order to gauge the robustness of my results to treatment timing, I exclude these two districts from my analysis in Table E.1 in the appendix. As expected the estimated effect of the NREGS increases when excluding one or both of these sub-districts.

- Treatment intensity, NREGS: Cumulative expenditure and number of person-days of employment generated at the block level are used to capture the treatment intensity of the NREGS. Data are retrieved from Government of Andhra Pradesh, Department for Rural Development, $http://www.nrega.ap.gov.in$, and cover the respective financial year (April to May).

  The amount sanctioned per village depends on a village's list of projects, which has to be approved by the block (i.e. sub-district) program officer. The block program officer has to estimate employment demand for the following financial year and consolidate all village lists before submitting the Block Employment Guarantee Plan to the district program coordinator. The district council (zilla parishad) has

[49]Data were retrieved from the Department of Rural Development, Government of Andhra Pradesh, and are available online: $http://www.nrega.ap.gov.in$.

to approve all plans before transferring them to the state government.

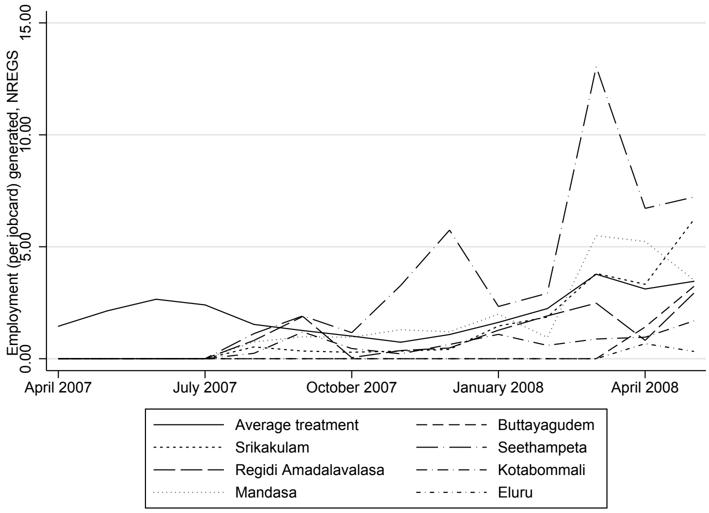
# C Matching strategy

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

I match households on the mean and the variance of variables that determine a household's registration with the NREGS and potentially influence post-treatment outcomes, such as cost incurred in agricultural production, total cultivated area, percentage of area irrigated, a dummy indicating whether a household participates in labor sharing in agriculture, wealth levels and off-farm income, and household characteristics, e.g. education, age and sex of the household head, indebtedness, and the ability to raise INR 1,000 (US$ 21.6) in one week. The resulting covariate balance is shown in Table E.6. This method focusses on the covariate balance and less on the common support among the treatment and control group. In order to understand how comparable both groups are in terms of the selected variables, I estimate the propensity score for each household based on the selection variables described above, and plot its distribution in Figure D.5. As can be seen, there is substantial overlap in the estimated propensity scores.

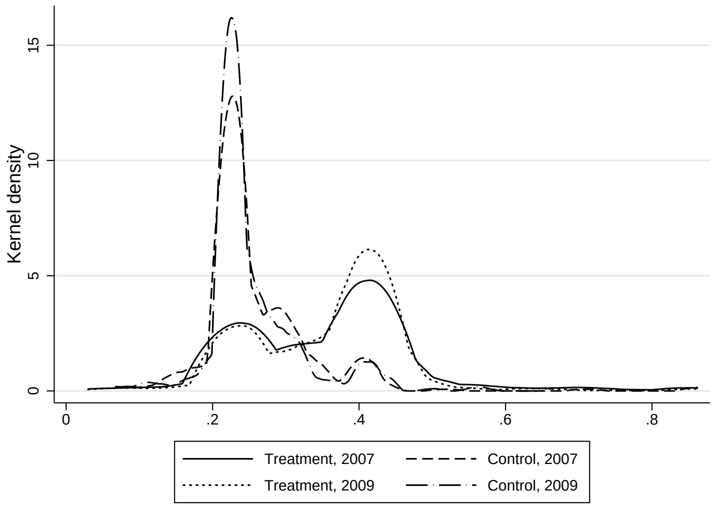
# D Supplementary Figures

Figure D.1: Block-wise NREGS employment per Jobcard (monthly)



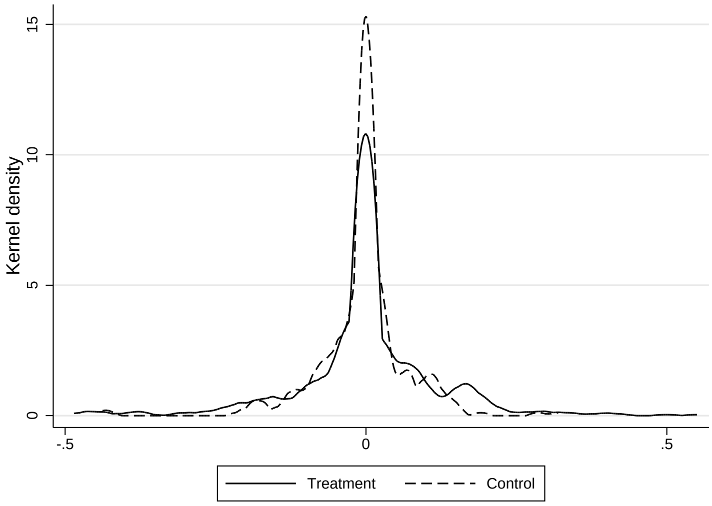

Source: Author's estimation based on Employment Generation Report, Department of Rural Development, Government of Andhra Pradesh.

Figure D.2: Distribution of risk-index



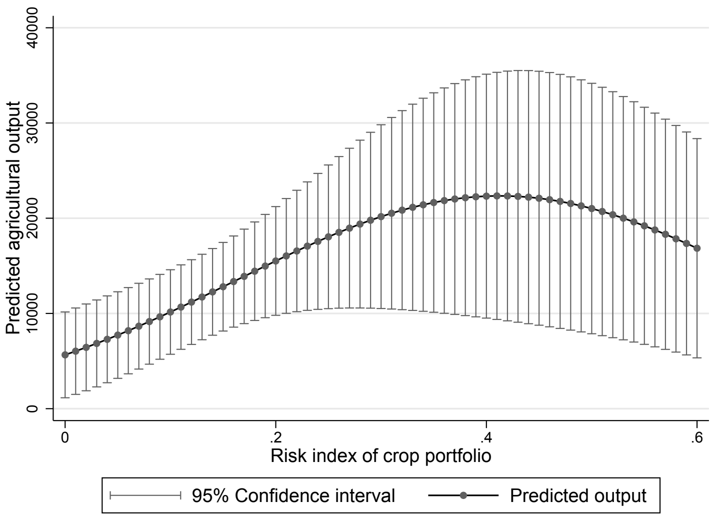

Source: Author's estimation based on District-wise Crop Production Statistics, Ministry of Agriculture, GoI, and Young Lives data.

Figure D.3: Distribution of change in risk index



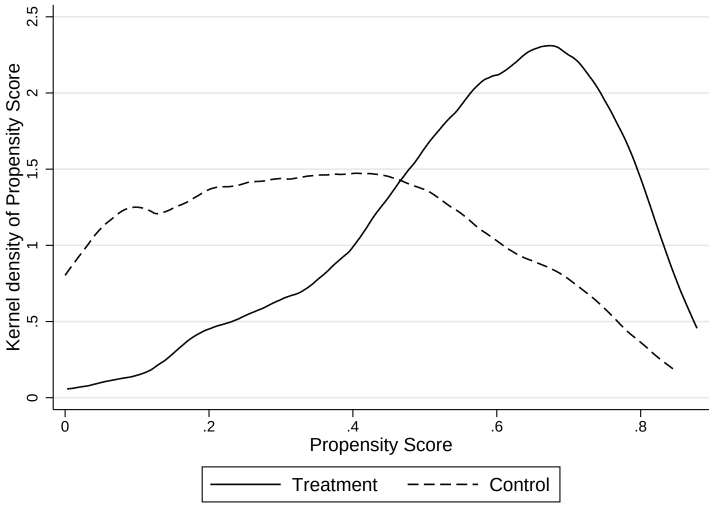

Source: Author's estimation based on District-wise Crop Production Statistics, Ministry of Agriculture, GoI, and Young Lives data.

Figure D.4: Agricultural output as function of the riskiness of crops



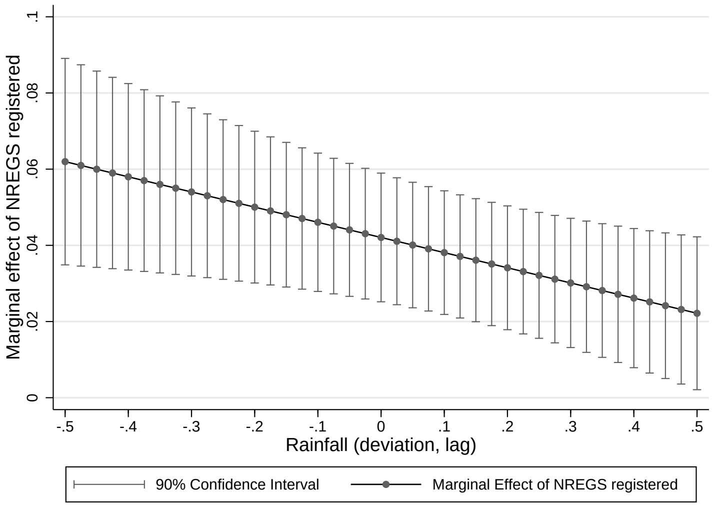

Source: Author's estimation based on the Young Lives data.

Figure D.5: Distribution of the propensity score




Source: Author's estimation based on Young Lives data.

Figure D.6: Effect of the NREGS on risk index conditional on lagged rainfall




Source: Author's estimation based on the Young Lives data.

# E Supplementary Tables

Table E.1: Sensitivity of results to treatment timing

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| NREGS introduced in district | 0.038*** | 0.041** | 0.040** | 0.045** |
| | (0.011) | (0.014) | (0.013) | (0.015) |
| Observations | 1087 | 1003 | 1003 | 941 |
| $R^2$ | 0.068 | 0.070 | 0.070 | 0.072 |

Notes: Estimation in FGLS. Column (1) reports main results, col. (2) excludes Seethampeta sub-district, col. (3) excludes Mandasa sub-district, and col. (4) excludes both sub-districts. Dep. var.: Risk index of a household's crop portfolio. Additional controls are variable inputs (log), area cultivated (log), share of area under irrigation, fertilizer application, HYV seeds application, labor sharing, annual income from off-farm activities (log), if household benefits from credit/training program, rainfall (deviation), rainfall (deviation, lag), self-reported shocks, and time trend. Standard errors (clustered at the district) in parentheses. $^+$ $p < 0.10$, $^*$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Source: Author's estimation based on the Young Lives data.

## Table E.2: Sensitivity of results to alternative dependent variables

| | CV detrended yield | $Riskindex^{alt}$ | CV detrended prices | SD of portfolio returns | Concentration |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| NREGS introduced in district | 0.023* | 0.024+ | 0.014+ | 338.413* | 0.024 |
| | (0.010) | (0.013) | (0.008) | (155.036) | (0.027) |
| Area cultivated (acres, log) | 0.006 | 0.010* | -0.007* | -187.479 | -0.133*** |
| | (0.004) | (0.005) | (0.003) | (133.418) | (0.023) |
| Hh benefits from credit/training program | -0.006 | -0.005 | 0.000 | -42.957 | 0.004 |
| | (0.005) | (0.008) | (0.002) | (81.503) | (0.008) |
| Rainfall (deviation) | 0.010+ | 0.012 | -0.002 | 8.942 | 0.011 |
| | (0.006) | (0.008) | (0.009) | (112.598) | (0.028) |
| Rainfall (deviation, lag) | -0.030* | -0.034* | -0.018 | -293.702 | -0.032 |
| | (0.012) | (0.015) | (0.015) | (234.946) | (0.031) |
| Self-reported shock | 0.008* | 0.004 | -0.001 | 137.167 | -0.008 |
| | (0.004) | (0.005) | (0.004) | (110.967) | (0.007) |
| Irrigated area (% of total) | -0.025 | -0.024 | 0.001 | -115.958 | -0.108*** |
| | (0.017) | (0.017) | (0.007) | (137.550) | (0.027) |
| Fertilizer (dummy) | -0.044 | -0.046 | -0.007 | 122.150 | 0.013 |
| | (0.030) | (0.034) | (0.012) | (253.983) | (0.030) |
| HYV seeds (dummy) | 0.007 | 0.003 | -0.005+ | 193.225+ | -0.032* |
| | (0.005) | (0.008) | (0.003) | (114.449) | (0.013) |
| Participated in labor sharing (dummy) | -0.005** | -0.001 | -0.002 | -134.633** | 0.004 |
| | (0.002) | (0.002) | (0.003) | (41.143) | (0.015) |
| Observations | 1087 | 1087 | 1050 | 1010 | 1087 |
| $R^2$ | 0.05 | 0.04 | 0.02 | 0.03 | 0.10 |

Notes: Estimation in FGLS. The dependent variable in column (1) is the coefficient of variation of the trend-corrected yield, in column (2) the risk index but now taking into account only those crops for which information is available, e.g.$R_i^{alt} = \sum r_m k_m / \sum k_m$. The dependent variable in column (3) is the CV of trend-corrected prices, in col. (4) the SD of portfolio returns and in col. (5) the Herfindahl index of the crop portfolio. For more information, refer to appendix B. Standard errors (clustered at the district) in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Source: Author's estimation based on the Young Lives data.

Table E.3: District-level statistics

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

Notes: Nominal values in current INR.

Source: Author's estimation based on the Districts at a Glance statistics, Directorate of Economics & Statistics, Govt. of Andhra Pradesh.

Table E.4: Evidence on mean reversion

| | (1) | (2) |
|---|---|---|
| Risk index of crop portfolio | -0.703*** | -0.476** |
| | (0.028) | (0.172) |
| Rainfall (deviation) | 0.083 | -0.097 |
| | (0.082) | (0.279) |
| Risk index of crop portfolio × Rainfall (deviation) | -0.439* | 0.818 |
| | (0.188) | (2.052) |
| Risk index of crop portfolio (squared) | | -0.309 |
| | | (0.211) |
| Risk index of crop portfolio (squared) × Rainfall (deviation) | | -2.094 |
| | | (3.639) |
| Observations | 337 | 337 |
| $R^2$ | 0.405 | 0.420 |

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

Source: Author's estimation based on the Young Lives data.

Table E.5: Robustness to inclusion of controls

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| NREGS introduced in district | 0.007 | 0.027* | 0.028* | 0.038*** | 0.059+ | 0.069*** | 0.070*** | 0.072*** |
| | (0.010) | (0.012) | (0.011) | (0.011) | (0.030) | (0.017) | (0.016) | (0.017) |
| Rainfall (deviation) | | 0.026*** | 0.027*** | 0.025*** | | 0.058 | 0.061+ | 0.050* |
| | | (0.007) | (0.007) | (0.006) | | (0.036) | (0.035) | (0.025) |
| Rainfall (deviation, lag) | | -0.036* | -0.036* | -0.044*** | | -0.048+ | -0.050+ | -0.041* |
| | | (0.015) | (0.016) | (0.011) | | (0.029) | (0.028) | (0.019) |
| Self-reported shock | | 0.007+ | 0.007+ | 0.008* | | -0.001 | -0.001 | 0.000 |
| | | (0.004) | (0.004) | (0.003) | | (0.003) | (0.003) | (0.004) |
| Annual income, off-farm activities (log) | | | 0.001 | -0.000 | | | -0.001 | -0.003* |
| | | | (0.002) | (0.002) | | | (0.001) | (0.001) |
| Hh benefits from credit/training program | | | -0.009 | -0.011 | | | -0.010** | -0.013*** |
| | | | (0.007) | (0.008) | | | (0.004) | (0.004) |
| Variable inputs (log) | | | | -0.016* | | | | -0.005 |
| | | | | (0.006) | | | | (0.005) |
| Area cultivated (acres, log) | | | | 0.018*** | | | | 0.011*** |
| | | | | (0.002) | | | | (0.002) |
| Irrigated area (% of total) | | | | -0.033+ | | | | -0.024 |
| | | | | (0.019) | | | | (0.015) |
| Fertilizer (dummy) | | | | -0.043 | | | | -0.028+ |
| | | | | (0.032) | | | | (0.016) |
| HYV seeds (dummy) | | | | 0.013* | | | | 0.025*** |
| | | | | (0.005) | | | | (0.006) |
| Participated in labor sharing (dummy) | | | | -0.003 | | | | -0.002 |
| | | | | (0.002) | | | | (0.002) |
| Controls: Rainfall and risk index at baseline | No | No | No | No | Yes | Yes | Yes | Yes |
| Observations | 1087 | 1087 | 1087 | 1087 | 1087 | 1087 | 1087 | 1087 |
| $R^2$ | 0.001 | 0.019 | 0.021 | 0.068 | 0.391 | 0.407 | 0.416 | 0.443 |

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

Source: Author's estimation based on the Young Lives data.

Table E.6: Weighted summary statistics

| | Treatment | | Control | | | |
|---|---|---|---|---|---|---|
| | | | (not matched) | | (matched) | |
| | Mean | SD | Mean | SD | Mean | SD |
| Value of variable inputs | 12.81 | (16.41) | 15.91 | (55.75) | 14.10 | (27.61) |
| Area cultivated (acres) | 3.96 | (4.53) | 3.50 | (5.20) | 3.88 | (3.85) |
| Irrigated area (% of total) | 0.14 | (0.29) | 0.19 | (0.33) | 0.14 | (0.28) |
| Participated in labor sharing (dummy) | 0.79 | (0.41) | 0.73 | (0.44) | 0.79 | (0.41) |
| Annual income, off-farm activities | 23.68 | (21.48) | 22.77 | (28.17) | 24.45 | (26.72) |
| Male household head | 0.96 | (0.19) | 0.96 | (0.20) | 0.96 | (0.19) |
| Age of household head | 41.20 | (12.07) | 42.02 | (12.02) | 41.20 | (11.79) |
| Household head is literate | 0.32 | (0.47) | 0.28 | (0.45) | 0.32 | (0.47) |
| Wealth index | 0.37 | (0.11) | 0.40 | (0.19) | 0.37 | (0.13) |
| Household size | 6.02 | (2.56) | 5.88 | (2.40) | 6.02 | (2.51) |
| Able to raise 1000 rupees in one week | 0.56 | (0.50) | 0.49 | (0.50) | 0.56 | (0.50) |
| Any serious debts | 0.67 | (0.47) | 0.50 | (0.50) | 0.67 | (0.47) |
| Observations | 496 | | 591 | | 591 | |

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

Source: Author's estimation based on the Young Lives data.

Table E.7: Interaction with rainfall and previously existing programs

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| NREGS registered (2007) | 0.032** | | | |
| | (0.011) | | | |
| NREGS × Rainfall (deviation, 2005-06) | -0.070** | | | |
| | (0.026) | | | |
| NREGS introduced in district | | 0.080*** | 0.077*** | 0.085*** |
| | | (0.015) | (0.017) | (0.019) |
| NREGS × Crop insurance | | -0.019 | | |
| | | (0.028) | | |
| Crop insurance | | -0.018 | | |
| | | (0.023) | | |
| NREGS × Watershed dev. | | | -0.025 | |
| | | | (0.019) | |
| Watershed dev. | | | 0.008+ | |
| | | | (0.004) | |
| NREGS × Public works | | | | -0.021 |
| | | | | (0.014) |
| Public works | | | | 0.010 |
| | | | | (0.007) |
| Controls: Rainfall and risk index at baseline | No | Yes | Yes | Yes |
| Observations | 1087 | 1083 | 1083 | 1083 |
| $R^2$ | 0.043 | 0.438 | 0.420 | 0.417 |

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

Source: Author's estimation based on the Young Lives data.