Moral Hazard:

Experimental Evidence from Tenancy Contracts*

Konrad Burchardi

Selim Gulesci

Benedetta Lerva

Munshi Sulaiman

Abstract

Agricultural productivity is particularly low in developing countries. Output sharing rules that make farmers less-than-full residual claimants are seen as a potentially important driver of low agricultural productivity. We report results from a field experiment designed to *estimate* and *understand* the effects of sharecropping contracts on agricultural input choices, risk-taking, and output. The experiment induced variation in the terms of sharecropping contracts. After agreeing to pay 50% of their output to the landlord, tenants were randomized into three groups: (i) some kept 50% of their output; (ii) others kept 75%; (iii) others kept 50% of output and received a lump sum payment at the end of their contract, either fixed or stochastic. We find that tenants with higher output shares utilized more inputs, cultivated riskier crops, and produced 60% more output relative to control. Income or risk exposure have at most a small effect on farm output; the increase in output should be interpreted as an incentive effect of the output sharing rule.

JEL Classification: O12, Q12, Q15.

*We thank the Riksbanken Jubileumsfond for financial support through grant P13-0161; staff of BRAC Uganda, in particular Kashem Mozumder and Esau Tugume, and at Metajua Ltd for their collaborative efforts throughout the project; seminar audiences at Bocconi University, Bristol University, CEPR/PODER Development Economics Conference (LSE), CEMFI (Madrid), CREI (Barcelona), Development Study Group (Stockholm), European University Institute, Frankfurt University, HECER (Helsinki), Manchester University, MPI Bonn, NBER SI 2017, Paris School of Economics, SEEDEC 2017, Stockholm School of Economics, Swedish Ministry of Finance, Trinity College Dublin, Universitat Autònoma de Barcelona, University of Groeningen, UNU-WIDER for providing useful comments; Oriana Bandiera, Tim Besley, Timo Boppart, Esther Duflo, Pascaline Dupas, Martin Hellwig, Seema Jayachandran, Eliana La Ferrara, Sara Lowes, Andreas Madestam, Eduardo Montero, Arash Nekoei, Torsten Persson, Mireia Raluy i Callado, Jakob Svensson, Diego Ubfal, and Chris Udry for their helpful feedback; and the editor Andrei Shleifer and three anonymous referees for their very swift handling of the paper and excellent suggestions which considerably improved the paper. Martin Hellwig and Torsten Persson provided feedback that substantially impacted the experimental design. Cristina Clerici, Charlotte Swart, Fabrizio Santoro, Silvia Gianfredi, and Emilio Dal Re provided excellent research assistance. Lerva and Sulaiman were employed by our partner NGO, BRAC Uganda, at the time of the experiment. All errors and omissions are the responsibility of the authors alone.

"For, when the cultivator has to give to his landlord half of the returns to each dose of capital and labor that he applies to the land, it will not be to his interest to apply any doses the total return to which is less than twice enough to reward him." (Marshall, 1890, Book VI, Chapter X.14)

I Introduction

Agriculture is the main source of income for a majority of the rural poor in developing countries; yet agricultural productivity remains notoriously low (Gollin, Lagakos, and Waughn, 2014). A commonly cited explanation for low agricultural output in developing countries is the prevalence of output sharing rules that make farmers less-than-full residual claimants.[1] Such output sharing rules may take the form of sharecropping contracts, whereby a tenant farmer pays a share of her output to the landowner (Banerjee, Gertler, and Ghatak, 2002), formal taxes or informal taxes such as kinship taxation (Lewis, 1955; Jakiela and Ozier, 2016), or imperfectly defined and secured property rights (Besley, 1995; Shleifer and Vishny, 1998; Acemoglu, Johnson, and Robinson, 2001). It is a central idea of modern microeconomics that such output sharing rules induce inefficient behavior by the agent as long as she is not the full residual claimant. This powerful idea dates back to the classical authors Adam Smith and, in particular, Alfred Marshall, who stated it succinctly, precisely to highlight sharecropping contracts as a potential source of low agricultural output.

How important is the degree of output sharing in explaining low agricultural output? How would tenant farmers adjust their behavior in response to a higher share? How much of that effect is due to the incentive effect conjectured by Alfred Marshall? These questions are empirical in nature. Over the past couple of decades research has attempted to answer these questions using observational data (Rao, 1971; Bell, 1977; Shaban, 1987). This paper reports novel results from the first field experiment designed to estimate and understand the effects of sharecropping contracts on tenant farmers' input choices, risk-taking behavior and output. These estimates provide answers to the three questions set out above.

Quantifying the incentive effects of contracts on production decisions generally poses at least two challenges. First, the outcomes of interest as well as the contractual terms are likely to be determined jointly by unobservable factors. In tenancy contracts, technology adoption and investment choices are likely to be a function of factors such as unobserved productivity, farmer ability or outside options, and contractual terms are chosen endogenously as a function of the

[1] According to a household panel survey by Uganda Bureau of Statistics, 38% of the rural households producing crops were engaged in sharecropping arrangement in 2009-10 (Khandker and Koolwal, 2014). According to a nationally representative survey of rural areas in Bangladesh in 2007, 26% of the cultivable land were under sharecropping compared to 9% with rental arrangement (Hossain and Bayes, 2015).

same factors. In fact, an extensive theoretical literature discusses the potential determinants of agricultural tenancy contracts.[2] This body of work implies that a positive correlation between the tenant's share in output and the level of total output might be the consequence of unobservable factors driving both the adoption of certain contractual terms and agricultural output, rather than evidence of incentive effects. Secondly, even when plausibly exogenous variation in a tenant's share of the output exists, it cannot solely be interpreted as an incentive effect: a higher output share has an incentive effect, but additionally induces higher income and higher exposure to risk, both of which might influence farmers' choices independently.

To overcome these challenges, we collaborated with the NGO BRAC in Uganda to implement a randomized controlled trial that induces variation in real-life tenancy contracts. As part of their operations, BRAC leased plots of land to women from low socio-economic backgrounds who were interested in becoming farmers (henceforth 'tenants') and provided them with agricultural training and a package of seeds for cultivation – effectively acting as the 'landlord'. The experiment was conducted with 304 tenants located in 237 villages, and at most two tenants per village. In all villages, tenants were contracted for one season under a sharecropping contract that gave them a 50% stake in the output. After signing the contract, villages were randomized into three groups.[3] In the first group (C), the contract was maintained – i.e. tenants received 50% of output. In the second group (T1), tenants were offered to keep 75% of the output. Tenants in a third group (T2) kept the same output share as in control (50%) but received an additional fixed payment which was independent of their output level, paid at harvest and announced at the same time as T1 received news of the higher share.[4] Within this third group, half of the tenants (T2A) received it as a risk-free cash transfer while the other half received part of their additional payment as a lottery (T2B), the expected payment in T2A and T2B being the same. The plots were visited pre-harvest to measure output levels and crop choice; and all tenants were surveyed shortly after the harvest to record their input use, such as labor, fertilizer, and tools.

The experimental design entails six key elements that allow us to estimate and understand the

[2] Sharecropping contracts can be understood as trading off incentive and risk-sharing motives (Stiglitz, 1974), as incentivizing the landlord's inputs, some of which may be unobservable and therefore non-contractible (Eswaran and Kotwal, 1985), as trading off moral hazard in effort and risk-taking (Ghatak and Pandey, 2000), as screening tenants of different abilities (Hallagan, 1978; Newbery and Stiglitz, 1979) and as the optimal contract under financial constraints (Shetty, 1988; Laffont and Matoussi, 1995; Banerjee et al., 2002). See Binswanger and Rosenzweig (1982) and Otsuka and Hayami (1988) for reviews of the literature on contract choice and the co-existence of different types of tenancy contracts.

[3] The village level randomization guarantees that if there were two tenants in a village, both were exposed to the same treatment condition.

[4] Note that the treatment is designed to emulate the nature of the income effect of treatment T1; it should not be thought of as realized unconditional cash transfer (Haushofer and Shapiro, 2016).

effects of output sharing rules on farmers' decisions. First, by randomly assigning tenants to contracts with varying terms, we ensure that tenants in different groups are not systematically different in their (unobservable) characteristics, such as their abilities, time preferences or risk attitudes. Second, the same contract was advertised in all groups to rule out ex-ante selection effects.[5] Further, tenants in the treatment groups were offered a change in contract that was unambiguously beneficial to avoid design-induced attrition. Third, we changed the terms of the tenancy agreements in T1 to generate exogenous variation in the tenant's share of output. This variation is key for estimating the incentive effect of the sharecropping contract. However, tenants entitled to 75% of their output are not only exposed to stronger incentives relative to those who receive 50%; they also have higher expected income, and they are exposed to additional risk. Fourth, the additional income may influence tenants' input choice and risk-taking through various mechanisms, rendering the direction and the magnitude of the effect unclear.[6] For that reason we implemented T2. The comparison of T2 with C allows us to test for the presence of an income effect on agricultural productivity. Fifth, to test whether tenants' exposure to risk alters their agricultural choices, some tenants within T2 received a risky income transfer while others received a safe one. The comparison of T2B with T2A allows us to test for the presence of a risk exposure effect. Sixth, tenants might have an incentive to misreport the agricultural output when a share of the output has to be given to the landlord. We therefore conducted pre-harvest plot-surveys to obtain an objective measure of output.

We present a model that specifies how incentives, income and risk exposure will impact tenants' input choices and risk-taking behavior, and consequently output. We model tenants as expected-utility-maximizing risk-averse agents who must decide the level and risk profile of inputs to be used on a plot. In particular, a tenant can choose between a risk-free cultivation technique or a risky but, in expectation, more productive one. Her compensation is in the form of a share s of the realized output and a fixed payment w, which could be positive (a wage), negative (a rent) or zero. The model predicts that an increase in s leads to an increase in the level of inputs the tenant chooses to employ in cultivation (the 'Marshallian inefficiency' effect); but has an ambiguous effect on her risk-taking, the direction of which depends on the

[5] Ackerberg and Botticini (2002) show that tenants are matched endogenously to contracts (and plots/crops). Randomization also ensures that there are no systematic differences in terms of plot or crop characteristics ex-ante across the different treatment groups. There may still be ex-post differences in tenant characteristics due to differential attrition, which we test for.

[6] Higher expected income may lower an individual's labor supply through a standard income effect. It may also affect incentives for risk-taking, as we demonstrate in Section II. Moreover, since tenants in T1 receive a better contract than what they had initially agreed on, they may increase their effort due to the presence of an efficiency wage. Finally, higher expected income may increase a tenant's access to credit which may enable him to increase the supply of inputs.

shape of her utility function.[7] On the other hand, an increase in w should have no effect on the level of her investment in inputs, independent of the risk profile of w. A safe increase in w will have a non-negative effect on her risk-taking (positive if the tenant's absolute risk aversion is decreasing with income). Additional exposure to uncorrelated risk will lead to less risk-taking. In terms of output, the effect of increasing s is positive, as long as the effect on risk-taking does not offset the effect on increasing the level of inputs. The effect on output of increasing w depends on how w affects tenants' risk-taking: if higher w leads to greater risk-taking by the tenant, it will lead to greater expected output as well. The experiment allows us to test these predictions.

We find that the fields of tenants with 75% output share generated on average 60% higher agricultural output compared to tenants who were allowed to keep 50% of output (T1 vs. C). We do not find that tenants who received a higher income produced significantly more (or less) output than tenants in the control group (T2 vs. C). We do observe a small, negative and imprecisely-estimated effect of risk exposure on the output level of tenants (T2B vs. T2A).

Next we show how tenants respond in terms of input levels and risk-taking. For input levels, we find that the tenants who retained a higher share of their output (T1) invested more in capital inputs to cultivate their plots. In particular, they used more fertilizer (120% more than the control group) and they acquired more agricultural tools (29% more relative to the control). We also find an increase in their use of unpaid labor (by 64% relative to control), but the effect on total labor hours is imprecisely estimated. In contrast, tenants who received higher income (T2) did not invest more in capital or labor inputs relative to the control group.

We assess changes in the tenants' risk-taking in three ways. The most direct approach is to study the crop-mix the tenants chose to cultivate on their plots. In effect, crops are differently risky assets between which the tenant chooses, conditional on a level of investment. In order to determine the relative riskiness of the different crops, we assess their sensitivity to rainfall and the volatility of their yield. We then study the differential crop choice of tenants across treatment groups.[8] Secondly, we analyze the dispersion of output across treatment groups. Third, we estimate the responsiveness of output to rainfall variation across treatment groups. Across these approaches we consistently find evidence of significantly higher risk-taking amongst ten-

[7] The latter is a standard result in public finance literature that studies the effect of taxation on entrepreneurial risk-taking (Domar and Musgrave, 1944; Mossin, 1968; Stiglitz, 1969; Feldstein, 1969).

[8] In particular, we study the sensitivity of each crop's yield to rainfall in two ways: first, by exploiting rainfall variation across plots cultivated by the tenants in the control group; second, in a panel data of crop yields in Sub-Saharan African countries from FAOStat. Both methods show that beans are less sensitive to rainfall compared to maize, tomatoes or peanuts. Moreover, the yield of beans has a lower coefficient of variation in the country level panel data. We find that tenants in T1 cultivated more of the riskier crops (maize, tomatoes and peanuts) while there was no significant effect on their cultivation of the safer crop (beans) relative to the control group.

ants with a higher output share (T1), mildly higher risk-taking amongst tenants who receive a risk-free income (T2A), and mildly lower risk-taking amongst tenants who received the risky income transfer (T2B), all relative to control (C). It should be noted that our approach does not allow to measure the returns to risk taking. Standard asset pricing theory and empirical work suggests that they are positive.

We do not find that the increase in output for tenants with higher output share had other adverse effects. In theory, tenants in T1 may have diverted their investments from other plots or reduced their involvement in other income-generating activities to generate the high output we observe on the experimental plots. We find no evidence of such adverse effects: total household income is significantly higher among the tenants in T1 and we find no crowding out of other income generating activities at the household level. Another concern with high-powered incentives is they may lead to over-investment in technologies that maximize short-term output at the expense of long-term soil quality. We collected soil samples from the experimental plots and tested for any impact on indicators of soil quality.[9] We do not find any evidence that the high-incentive tenancy contracts had led to soil degradation by the end of the experiment.

In Section V we discuss the interpretation of the results. First, we explain that the empirical findings are in line with the predictions of our theoretical framework. Second, we demonstrate that the output increase of tenants with a high share can also be quantitatively accounted for by observed changes in the input levels and risk-taking behavior of tenants, with each contributing about half of the full effect. Third, we simulate the welfare consequences of a higher crop share and find that these are large for reasonable levels of risk-aversion. This is unsurprising given that the gross income of tenants with high output share increases by 140% relative to control. Last but not least we discuss a number of important limitations of our approach: we work with an implicitly selected sample of farmers; we estimate output responses using data from two agricultural seasons which necessarily constitute a particular realization of weather and other risks; the experiment was conducted in a setting where formal sharecropping contracts are uncommon; and the experimental design does not allow to capture potential externalities. We discuss at length whether and how these features of the experiment limit the ability to extrapolate from our findings.

Our paper contributes foremost to the empirical literature on the incentive effects of agricultural tenancy contracts. Rao (1971) shows that output is higher in owner-operated relative to sharecropped farms in India, but a large share of the difference can be attributed to differences in land size. Controlling for farm size changes the sign of the correlation between

[9]In particular, we test for the levels of nitrogen, potassium, phosphorous, organic matter and the Ph-level.

ownership and output. An important methodological contribution is made by Bell (1977) and Shaban (1987) who use plot level data, and compare output and input levels across plots with different tenancy statuses within the same household, thus controlling for many unobservable household level characteristics. Nevertheless, the endogeneity of contract choice and the presence of unobserved plot level characteristics are potential sources of bias in their findings (Arcand et al. (2007); Braido (2008); Jacoby and Mansuri (2009)). Banerjee et al. (2002) show that a tenancy reform which simultaneously changed legal output share of registered tenants and reduced their likelihood to be evicted by the landlord increased agricultural output in West Bengal. However, it is not clear to what extent this effect was driven by the change in tenants' legal crop share or their security of tenure.[10] As far as we are aware, the current paper is the first to provide experimental evidence on the incentive effects of tenancy contracts.

More broadly this paper contributes to the growing literature that seeks to understand the agricultural productivity shortfall in developing countries and identifies policies that increase agricultural productivity and output. Notable contributions are the work by Karlan, Osei, Osei-Akoto, and Udry (2014), who find that farmers in Ghana make riskier production choices when provided with insurance; Duflo, Kremer, and Robinson (2008), who show that subsidies can be carefully designed to increase the adoption of profitable technologies in the presence of hyperbolic discounting; and research by Adamopoulos and Restuccia (2014) and Restuccia and Santaeulalia-Llopis (2017), who show that the reallocation of agricultural land across heterogenous farmers might have large output and welfare gains. We show that policies that effectively strengthen the cultivators' position as residual claimant also have the potential to substantially increase agricultural output.

The paper is also related to recent empirical studies that have demonstrated the role of agents' incentives in other contexts; see, for example, Prendergast (1999) and Bandiera et al. (2011) for the role of contracts in incentivizing workers within a firm. Tenant farmers, compared to typical wage workers, have a wider set of decisions to make, often trading off expected returns with the riskiness of production (Ghatak and Pandey, 2000). In this respect, the decisions of tenant farmers are conceptually closer to those of entrepreneurs or corporate executives, analyzed in public economics (Domar and Musgrave, 1944; Mossin, 1968) and corporate finance (Jensen and Meckling, 1976). This literature highlights the role of output sharing rules for risk-taking and shows that the effect of taxation on risk-taking is ambiguous in a general setup. The sign of the effect depends on the exact shape of the tax schedule as well as the agent's utility

[10]Related to the tenure security effect of the reform, an eminent literature demonstrates the role of property rights in driving agricultural decisions and productivity (Besley, 1995; Braselle et al., 2002; Jacoby et al., 2002; Goldstein and Udry, 2008; Hornbeck, 2010; Montero, 2018; Iwanowsky, 2018).

function (Domar and Musgrave, 1944; Mossin, 1968; Stiglitz, 1969; Feldstein, 1969). Empirical tests of the theory have been limited due to the endogeneity of taxes to income and wealth (Feldstein, 1976). While some papers (see e.g. Poterba and Samwick, 2003) have exploited changes in tax regimes to study household portfolio choice, the evidence on the effect of taxation on entrepreneurial risk-taking is limited. We contribute to this literature by providing evidence that a lower tax (higher output share) increases risk-taking among farmers.

II Theory

Set-Up Suppose that a tenant's preferences can be represented by expected utility maximization and a Bernoulli utility function $u(c)$, defined over a consumption good c, with $u : \mathbb{R}^+ \to \mathbb{R}$ being increasing, concave and twice differentiable. When assessing welfare consequences, we assume specifically $u(c) = \frac{c^{1-\eta}-1}{1-\eta}$, where η is the (constant) coefficient of relative risk aversion.

The tenant faces two choices: she purchases a bundle of inputs x at unit price p; and she determines the risk profile of returns to her investments. The latter choice represents both which input mix the tenant purchases, and how she chooses to use these inputs. We parametrize this notion by assuming that a tenant's output can be written as

$$y = a\theta f(x) + (1-a)f(x),$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an increasing, concave and twice differentiable production function, θ is a random variable with positive support, and $a \in [0, 1]$ captures the extent to which tenants take on risk. For $a = 0$ the tenant chooses not to be exposed to risk; for $a = 1$ she chooses the maximal level of risk; intermediate choices of a represent a convex combination of the return profiles of these polar cases. We implicitly normalize the return of the risk-free investment to 1. Let the c.d.f. of the distribution of θ be denoted by $G(\theta)$, with support $[\underline{\theta}, \overline{\theta}]$. We assume $\underline{\theta} \in [0, 1]$ and $\mathbb{E}_\theta [\theta] > 1$; those are necessary conditions for an interior solution for a. The formulation also implicitly assumes that tenants take output prices as given.

A linear sharecropping contract specifies that the tenant pays a share $(1-s)$ of gross output to the landlord, in addition to a fixed payment. The fixed payment to the tenant can be positive (a wage) or negative (a fixed rent). The tenant may also have additional income. We denote with w the sum of additional income and any payment to the tenant agreed with the landlord.

The tenant's consumption is then $c = s[a\theta f(x) + (1-a)f(x)] - px + w$. She will choose the input bundle x and the risk-profile of investment a to maximize

$$\mathbb{E}_\theta[u(c)] = \int u\left(s[a\theta f(x) + (1-a)f(x)] - px + w\right)\, dG(\theta). \tag{1}$$

This framework captures a number of aspects of a tenant's choice that we consider realistic and potentially important. Firstly, agricultural output is typically subject to aggregate risks that are difficult to insure locally, such as output risks resulting from rainfall and temperature variation or pest outbreaks. Secondly, we model tenants' risk aversion. There is both empirical evidence suggesting that tenants are risk averse, and theoretical reasons to believe that an agent's risk aversion might be important for her productive choices.[11] Third, we restrict attention to linear incentive contracts. This aspect of the model lacks theoretical generality, but not realism: surveys of tenancy contracts show that a large majority of observed sharecropping contracts take a linear form.[12] Fourth, and most importantly, we think of the tenant's problem as choosing both the level of investment and the risk profile of investments. We believe this to be a realistic representation of a tenant's choice. Agricultural tenants typically choose the level of inputs such as their own or hired labor, the intensity of their own labor (often referred to as "effort"), total expenditures on seeds, fertilizer, pesticides and irrigation, amongst others. However, in choosing the specific mix of these inputs, such as the composition of seeds, or how to apply them, they also effectively choose between investments with different risk profiles. Our set-up allows us to study both choices jointly: A change in the terms of the sharecropping arrangement – or, under an alternative interpretation, the effective tax schedule – will potentially lead to a change in the tenant's level of input purchases. A change in the sharecropping arrangement might also change the incentives for risk-taking. Importantly, both of these decisions might interact, and understanding them in isolation might not be possible. The framework outlined here allows us to study the joint determination of the level of investment and its risk profile. It will guide how we interpret the reduced form effects of variation in sharecropping arrangements on outcomes of interest.

This formulation is special in at least two ways. First, a set-up where $f(x)$ is linear in x would be closer to the problem analyzed in the theory of portfolio choice, where typically the level of asset holdings does not alter the distribution of marginal returns of each asset. Second, we assume, given a level of investment x, a particular relationship between the mean gross return

[11] Smallholder farmers have been shown to exhibit substantial risk aversion in both survey and lottery based measures of risk aversion (Binswanger, 1980) and farmers' behavior (Karlan et al., 2014). Risk aversion is central to standard explanations for the existence of partial incentive contracts, pioneered by Stiglitz (1974).

[12] Holmstrom and Milgrom (1987) present sufficient conditions for linear contracts to be optimal theoretically.

of an investment and the associated dispersion around the mean. In a general framework the tenant would choose between a set of investments with unrestricted distributions of returns.[13]

Understanding Tenants' Choices Assuming an interior solution, a tenant's optimal choice of (x, a) is characterized by the following first order conditions:

$$\int u_c \cdot [s[a\theta f_x(x) + (1-a)f_x(x)] - p] \, dG(\theta) = 0 \tag{2}$$

$$\int u_c \cdot [s\theta f(x) - sf(x)] \, dG(\theta) = 0, \tag{3}$$

where $u_c \equiv \frac{\partial u(c)}{\partial c}$. We will denote the elements of the associated Hessian as $D_{ij} = \frac{\partial^2 \mathbb{E}_\theta[u(c)]}{\partial i \partial j}$.

To understand the tenant's input and risk-taking choices it is instructive to first consider (3), the first order condition with respect to a. It captures the trade-off between higher mean returns and additional risk. It states that the tenant will take on risk until the marginal expected utility from additional risk is equal to 0. Now consider (2) and note that it can be rearranged in two parts as $\int u_c \cdot [sf_x(x) - p] \, dG(\theta) + \int u_c \cdot [saf_x(x)(\theta - 1)] \, dG(\theta) = 0$. The first part captures the increase in the expected marginal utility from increasing the level of returns across investments. The second part captures that a higher x also increases the absolute dispersion of returns, just like risk-taking does. We can derive the following prediction. (All proofs are in Online Appendix I.)

Prediction 1. *(Input Effects)*

i. An increase of the tenant's share in output, s, increases level of investment, $\frac{dx}{ds} = -\frac{f_x(x)}{sf_{xx}(x)} > 0$.

ii. An increase of the tenant's income level, w, leaves the level of investment unchanged, $\frac{dx}{dw} = 0$. (This result is independent of the stochastic profile of w.)

The first part of the result captures the intuition that Alfred Marshall had in mind: a higher share of the agent increases the marginal return to investments keeping the costs constant, which increases the level of investments. This result would be straight-forward to demonstrate in a framework where the agent is risk neutral. Prediction 1 demonstrates that it also holds for risk averse tenants, as long as the tenant can adjust the level of risk-taking: the endogenous adjustment of risk-taking allows to offset the risk exposure effect of higher investment levels.[14]

[13] Conditional on any mean return a preferred investment portfolio will always exist. However, the dispersion of returns around the mean of that portfolio might have a general form. In contrast, our formulation implies a particular relationship: at the mean return $[a\mathbb{E}_\theta[\theta] + (1-a)]f(x)$ gross returns have one specific distribution, with variance $a^2\mathbb{E}_\theta[\theta - \mathbb{E}_\theta[\theta]]^2 (f(x))^2$. A feature of this relationship is that higher mean returns require a tenant to take on additional dispersion of returns.

[14] Note that this also implies that the second order conditions are satisfied.

Therefore the only effect determining the level of investment is the standard trade-off between expected marginal utility gains and costs. If a risk-averse agent cannot adjust the level of risk-taking, an increase in s would, in addition to the standard incentive effect, also have a risk exposure and wealth effect. These effects might work in opposite direction, which is a well-known result since Pratt (1964) and Arrow (1971), and the sign of the sum of them would be ambiguous. (See Online Appendix I.A.)

It is worth noting that the effect of s on x will be larger when a adjusts endogenously than when a is kept fixed.[15] The intuition for this result is that the tenant does not take into account any effect of x on risk exposure when choosing its optimal level, since risk exposure can be undone by adjusting the level of risk-taking conditional on x – an instance of Le Chatelier's principle. The result is important for the interpretation of our results. As we will show, tenants do adjust the risk level in our setting. If however in some other setting tenants cannot adjust a – for technological, institutional or behavioral reasons – we would expect to see smaller effects of changes in the tenants' share on investment levels.

A useful corollary of Prediction 1 is that $\frac{-f_x(x)}{x f_{xx}(x)}$ is a sufficient statistic for the elasticity of investments with respect to the tenant's share s. In particular, no knowledge of the specific utility function is required to predict changes in the investment level when changing s. This implies that estimates of $\frac{dx}{ds}$ have external validity as long as production choices are common – even though tenants might have heterogeneous utility functions.

Lastly, an increase in w is predicted to leave the choice of x unchanged. This result also holds when the increase in w is stochastic, independent of the type of correlation structure between θ and w. Both additional income and risk exposure do not affect the choice of x.

Next, we turn to the effects of the contractual terms on the tenant's risk-taking behavior.

Prediction 2. *(Risk-Taking)*

i. *An increase in s has an ambiguous effect on risk-taking, a when $u(\cdot)$ exhibits DARA.*

ii. *Consider an increase in safe w. Then the level of risk-taking, a, increases when $u(\cdot)$ exhibits DARA, $\frac{da}{dw} > 0$.*

iii. *Consider an increase in stochastic w, with w independent of θ. An increase in w has an ambiguous effect on the level of risk-taking, a, when $u(\cdot)$ exhibits DARA.*

[15] If the level of risk-taking adjusts endogenously, we can write $\frac{dx}{ds}$ as

$$\Psi \times \frac{-1}{D_{xx}} \int u_c \cdot [a\theta f_x(x) + (1-a) f_x(x)] \, dG(\theta),$$

with $\Psi := \frac{D_{xx}}{\int u_c[sa\theta f_{xx}(x) + s(1-a) f_{xx}(x)]} > 1$. Compare this to the incentive effect in Online Appendix I.A.

A large literature in public finance studies the theoretical effect of taxation on risk-taking, especially entrepreneurial risk-taking. It analyzes the risk-taking effects of taxation in isolation of effects on investment levels. It finds the sign of the effect of taxation on risk-taking to be indeterminate in a general setup; predictions depend on the shape of the tax schedule and the utility function (Domar and Musgrave, 1944; Mossin, 1968; Stiglitz, 1969; Feldstein, 1969).

The first part of Prediction 2 shows that this conclusion carries over to our framework. An increase in s implies a higher exposure to risk – both mechanically and because x increases – as well as higher wealth. The income effect is described by part (ii): absolute risk-taking increases for an agent characterized by a DARA utility function. This is nothing more than the name-giving property of such utility functions. Part (iii.) in combination with part (ii) highlights that additional exposure to income risk dampens the tenants willingness to take on additional risk through a. This explains part (i): Under DARA further assumptions are needed to sign the effect of s on risk-taking.[16] Note that DARA is likely a plausible assumption. Therefore this result also highlights how understanding the effect of the tenant's share on risk-taking is an inherently empirical question.

In summary, the theory predicts that an increase in the tenant's share s increases the level of inputs purchased. This is an incentive effect, both income and risk-exposure have no effect on the level of inputs. And an increase in s has an ambiguous total effect on risk-taking. The income effect of the increase in s has a positive effect on risk-taking and the risk-exposure effect of the increase in s has a negative effect on risk-taking.

Lastly, much of the interest in sharecropping contracts is concerned with designing contracts and regulation to increase agricultural output. Predictions 1 and 2 do translate into implications for expected output.

Prediction 3. *(Output Effects)*

i. *The tenant's expected output increases with s, as long as $\frac{da}{ds}$ exceeds some negative bound.*

ii. *The tenant's expected output increases with w if and only if $\frac{da}{dw} > 0$.*

This result highlights how an increase in the tenant's share does not necessarily need to translate into higher expected output. The reason is that the increase in output implied by the Marshallian incentive effect might be offset by the tenant taking on less risk. However, moderate levels of risk reduction still imply increases in expected output, and increases in the level of

[16] Under CARA an increase in safe w leaves risk-taking unchanged, which is again the name-given property of such utility functions, and an increase in stochastic w decreases risk-taking.

risk-taking amplify the effect of the tenant's share on output. Increases in the tenant's income w do not effect the input choice, therefore any effect on expected output from changes in w is coming from changes in the level of risk-taking.

The predictions are summarized in Supplementary Table I in the Online Appendix.

III Methods

III.A Setting

In order to test the theoretical predictions above, we implemented a field experiment in collaboration with BRAC. Uganda has one of the youngest populations in the world. In 2014, 48% of Uganda's population of 35 million was aged 15 or below, while – as a point of comparison – the figure is 21.2% in the US. Among the youth, young women are particularly at risk as they are more likely to drop out of school at an early age and face social and economic constraints in entering the labor market. BRAC operates a program called Empowerment and Livelihood for Adolescents (ELA) with the objective to empower young women in Uganda. At the core of this program is to open, finance and operate youth "clubs" for girls. In rural areas, each club is assigned to a village. BRAC provides vocational and life skills training, as well as various social activities through these clubs.[17] As part of these efforts, BRAC decided to lease plots of land to women who were interested in becoming farmers. Women in Uganda head 26% of rural households and grow 70% – 80% of food crops, yet own less than 8% of the land (Nafula, 2008). Moreover, even on plots of land controlled by women, productivity is likely to be lower due to differential access to factors of production (Udry, 1996). In order to assist young women who wanted to become farmers but faced difficulties in setting up their farm activities, BRAC started implementing the intervention that forms the setting of our experiment. During the design phase, focus group discussions with club members revealed that due to credit constraints and concerns about the riskiness of cultivation, most potential tenants would not find a fixed-rent contract suitable. As such, BRAC decided to implement the intervention under a sharecropping arrangement.

III.B Timeline

Season 0. In July 2013, BRAC selected 300 clubs in Eastern, Western and Central regions of Uganda to implement the intervention.[18] At most one club was selected per village; for the purpose of our experiment the club and village level are hence the same. BRAC then attempted

[17]See Bandiera et al. (2018) for further details of the ELA program.

[18]Uganda has four main regions: Eastern, Western, Southern and Northern. The Northern region differs significantly from the other three in terms of geography, climate and socio-economic organization.

to rent a plot of agricultural land of roughly 0.5 acre close to the club, and searched for up to three club members willing to rent the plot under a $s = 0.5$ sharecropping contract, with no fixed payment component, for one season. In 285 clubs both land and potentially interested tenants were found. Figure I shows the location of these clubs. The interested girls were then offered the land, in an order randomized by the authors, until one of them decided to take up the offer and become a tenant. Both a plot and a farmer who actually signed up as tenant of the plot were found in 259 clubs. The tenants cultivated the plot for the following agricultural season, from September 2013 to January 2014 (henceforth 'Season 0'), which served as a pilot season.

Seasons 1 and 2. We collaborated with BRAC to implement the experiment in two agricultural seasons of 2014, spanning from March to July ('Season 1') and September 2014 to January 2015 ('Season 2'). In Season 1, the plots were advertised to be available for tenants under a 50% sharecropping contract with no fixed component. Tenants who had cultivated the plots in Season 0 were given priority. Roughly half of the Season 0 tenants decided to continue in Season 1. In the remaining cases new tenants signed up. Additionally BRAC decided to scale up the program for Season 1, both by renting an additional plot in clubs where a plot was rented in Season 0, and also by re-attempting to rent plots close to clubs for which no plots were found in Season 0. As a result of these changes 304 tenants signed a 50% sharecropping contract at the beginning of Season 1. In preparation of Season 2, the plots were again offered under a 50% sharecropping contract with no fixed component, with priority given to Season 1 tenants.

Within-Season Procedures In each agricultural season BRAC provided the tenants with agricultural training. The training taught best-practice recommendations on (a) how to prepare the land and plant, (b) grow, and (c) harvest crops. The first training session took place before planting, the last training session took place before harvesting.[19] During the first of these training sessions, BRAC also provided the tenants with a bundle of high yield variety seeds. In Season 0 tenants were given maize, beans, cabbages and tomato seeds, for a total seed bundle value of 12 PPP USD; in Seasons 1 and 2 tenants were given maize, beans, and peanut seeds for a total seed bundle value of 32 PPP USD.[20] The training focused on techniques related to these crops, respectively. During the first training session the tenants signed the 50% sharecropping contract, valid for one season, in the presence of the BRAC program assistant as well

[19] In Seasons 1 and 2 there were only two training sessions, and topics (a) and (b) were both taught during the first training session. In Season 0, topic (b) was taught in a separate mid-season training session.

[20] In two areas potato seedlings were provided instead of peanuts. In that case the seed bundle value was 28 PPP USD. BRAC decided to change the seed mix provided to farmers between Season 0 and the following seasons after program assistants reported that farmers preferred peanuts or potatoes to tomatoes and cabbages.

as another witness.

III.C Experiment

Treatments. The experiment was implemented in Seasons 1 and 2.[21] At the start of both Season 1 and Season 2 the plots were advertised with a 50% sharecropping contract. Tenants who agreed to rent the plot under that contract signed it during the first training session. The contracts explicitly stated and the tenants were clearly told that the arrangement would last for one season, and there was no guarantee that their tenancy would be renewed in the future.

Tenants were assigned to one of four treatment conditions (see Figure II):

Control (C): Tenants keep the $s = 0.5$ contract.

High s (T1): Tenants are offered a contract with $s = 0.75$.

High w, safe (T2A): Tenants keep $s = 0.5$ and are offered a fixed payment w, with w being set to 25% of Season 0's median harvest value, to be paid at the time of the next harvest.[22]

High w, risky (T2B): Tenants keep $s = 0.5$ and are offered a payment w, with w being 20% of Season 0's median harvest value with probability 0.5, and 30% of Season 0's median harvest value with probability 0.5, to be determined and paid out at harvest time.

We refer to the union of T2A and T2B as T2.

While tenants were assigned to the same treatment condition across seasons, in both Season 1 and 2 the tenants initially signed a 50% sharecropping contract, and only heard about the update to their contract when they were contacted by BRAC program staff after the training of the respective season. During these calls tenants were reminded that they have signed a $s = 0.5$ sharecropping agreement, and comprehension checks were performed and repeated until passed satisfactorily. Tenants in treatment groups T1 and T2 were informed about the change in the terms of their contract, and comprehension checks were performed. The tenants in group T1 and T2 were told that they had been selected for the more favorable contract by a lottery. The terms of the new contract were explained to them in detail and were stated as applying to the upcoming season. Tenants in T2 were informed of the amount of cash transfer they would receive at the end of the season, those in T2B were explained the details of the

[21] In the study area there are two agricultural seasons per year. The first one extends from March to August, the second from September to February. Rains in the first season are usually heavier, and the chance of crop failure is lower.

[22] The level of the transfer was calculated at the BRAC branch office level, using data on the harvest value of all Season 0 farmers. Note that Season 0 is the baseline season; no experimental variation in contracts had been induced or announced in Season 0.

lottery (i.e. the risky cash transfer) they would participate in. After the phone calls the BRAC program assistant delivered a letter to the tenant specifying the updated contract. Additionally all tenants received this information in a text message.

Rationale. The objective of the research project was to understand the nature and magnitude of a number of specific effects of agricultural land tenure systems on the behavior of the tenants on input choices, risk-taking and agricultural output. The experimental design allows us to test the Marshallian hypothesis and identify the mechanisms behind it.

Firstly, BRAC advertised the same contract (with s=50%) in all treatment groups. This is a version of the seminal experimental design in Karlan and Zinman (2009) and controls for selection effects. As such, there is no reason to believe that tenants who sign up are systematically different on any unobservable characteristics across the different treatment groups.

Secondly, after the tenancy contracts were signed, tenants in T1 were offered s=75%, in order to generate variation in the tenant's share in output. We chose to implement a change to the tenancy contracts in T1 which we surely knew was dominating the original contract from the perspective of the tenant, in order to avoid design-induced attrition. The exogenous variation in output share induced in T1 is key to test the incentive effects of sharecropping contracts.

Third, the comparison of input intensities and output levels between C and T1 does not necessarily allow us to estimate the incentive effect of a higher share in the output. Increasing a tenant's share of the output does not only have an effect on the marginal revenue of the tenant, but might also have an income effect. A classic income effect driven by the tenant's labor-leisure choice would suggest that individuals at higher expected income levels may choose to work less. Higher expected income may also increase the tenant's access to credit which may enable her to increase the supply of inputs. In order to test for the collection of these effects, we introduce T2. In this group, tenants are offered the same output share (s=50%) as in C, but receive a fixed payment. This allows us to estimate the size of the income effect. If this estimate is 0, the comparison of C and T1 estimates the incentive effect.[23]

Finally, within T2, half of the tenants were offered a risk-free cash transfer (T2A) while for half of them, part of the payment was based on a lottery (T2B). The expected transfer amount is the same across the two groups. To the extent that any income effect exists in T1, this is the effect of a *risky income*, since agricultural output is necessarily stochastic from the point of view of the tenant. Any income effect likely varies with the risk profile of the additional income, either be-

[23] If the estimate of the income effect is significantly different from 0, we can estimate a structural model of labor supply which features two structural parameters, one governing the income effect, and one governing the incentive effect.

cause the tenant is not risk neutral or because credit access is affected by the stochastic nature of the additional income. The treatment T2B allows us, by comparison with T2A, to test whether indeed the risk profile of income is important to understand tenants' behavior. We designed T2B such that the first and second moments of the distribution of additional income in T2B roughly match the first and second moments of the distribution of additional income induced by treatment T1. However, higher moments of these two distributions are likely different; further the additional risk in T2B is perfectly uncorrelated with agricultural yields, whereas the additional income risk induced by T1 is perfectly correlated with agricultural yields. This should be kept in mind when interpreting the results.

Implementation Challenges. In implementing the experimental design we faced two challenges. First, the amount of additional income provided in T2 was determined as 25% of the BRAC branch level median output of Season 0. This might incorrectly reflect the (expected) income effect of treatment condition T1, which it would ideally match. We will address this when discussing the main effect of treatment condition T2 relative to treatment condition T1 in Section IV.B. Second, the information about the updated contract was to be provided shortly after the first training session, prior to the start of the agricultural season. This feature was implemented as such in Season 1. However, in Season 2, due to administrative constraints on the ground, the information about the updated contracts was provided to the tenants only in January 2015, three months late into the agricultural season. This needs to be kept in mind when interpreting the findings.

Randomization. We grouped the 300 clubs originally designated as potential study sites into clusters of three clubs (henceforth referred to as 'blocks'), with the heuristic objective to minimize within-block geographic distance. The study clubs were typically geographically bunched – see Figure I for a visualization of this. We grouped clubs into these large clusters (henceforth referred to as 'strata'). Assignment to treatment was randomized at the club and hence village level at the beginning of Season 0. We assigned equal fractions of the 300 potential study clubs to C, T1 and T2, stratified by blocks. Within T2 clubs we assigned 50 clubs to T2A and T2B, stratified by strata.

III.D Measurement

We collected data through two types of survey instruments, a tenant level survey ('Tenant Survey') and a plot level survey designed to estimate outputs on the field ('Crop Assessment').

The Tenant Survey collected information on the tenants' and their households' demographic and socioeconomic characteristics. We recorded their educational history, health status, labor supply and employment characteristics, household structure, detailed agricultural practices and output on each of the household's cultivated plots, including the plot rented from BRAC, ownerships status of plots, the household's asset holdings, and consumption expenditures, the tenant's savings and loans. The survey was administered by enumerators who were hired by BRAC and managed by the research team. The survey was administered to all potential tenants before each season of cultivation. It was also administered to all tenants about one month after the end of the season. It provides baseline information on the tenants in our sample (collected at the end of Season 0), as well as followup information at the end of Seasons 1 and 2.

A central challenge was to measure agricultural output in a way that is immune to manipulation by the tenant. Neither self-reported yields, nor crop-cutting and whole-plot harvesting techniques – commonly used to measure agricultural output[24] – satisfy this criterion.[25] Instead from Season 1 onwards we conducted a plot level survey of yields shortly before maturity of the crops ('Crop Assessment').[26] For this survey we hired students of agriculture as enumerators. They measured the size of plots and its parcels using GPS trackers; collected exhaustive data on the plot, including agricultural practices applied; took soil samples and tested levels of nitrogen, phosphorus, potassium, organic matter and soil pH. Importantly, to assess the output, they placed 1.5m $\times$ 1.5m quadrants on representative sections of the plot's parcels (8 quadrants per acre), and recorded detailed plant characteristics within each quadrant. Further they were trained to assess the expected output at harvest time for every plant in every quadrant. In order to value the output of a given crop, we conducted a survey of crop prices at the nearest local markets at harvest time. While in theory it is possible that local prices may be affected by the treatment assignment, in practice it is unlikely as the plots are small (0.5 acre on average) and therefore the crops harvested from the experimental plots make up only a very small fraction of the total output in each village. Hence, any general equilibrium effects on local prices are unlikely. Starting from Season 1 these estimates were used to determine the tenants' due payment, which was collected by BRAC field officers.

[24]For a comprehensive list of available techniques, see Fermont and Benson (2011).

[25]Notice that tenants across treatment groups have a differential incentive to misreport their yields. Further, farmers might harvest mature crops at any time before the arrival of the surveying team and again the incentives to do so are differential across treatment groups.

[26]In Season 0 the crop assessment was conducted by BRAC: Two BRAC program assistants, the tenant, and an enumerator visited the plot at harvest time and surveyed plant density, quality and other characteristics for maize, beans, tomatoes and cabbage, and estimated the plot size. In addition the tenants were asked to report the recalled amount and value of crops that had already been harvested, both for sale or own consumption. This procedure turned out to have a number of drawbacks. One drawback is it was conducted shortly before the harvest time of maize. The harvest time of other crops, such as beans and tomatoes for example, would likely have been earlier.

III.E Sample, Attrition and Seasons

Sample. Subsequently we report results using data from the Tenant and Crop Assessment surveys in Seasons 1 and 2. All analysis is based on the sample of tenants who signed the tenancy contract in the beginning of Season 1 and the plots of those tenants. We do not report results for tenants who only started renting a plot in Season 2. Figure II provides a visual summary of the experiment's setup, timeline and sample sizes in each treatment group.

Attrition. Of the 304 tenants who signed a tenancy contract in the beginning of Season 1, we successfully surveyed 253 tenants during Tenant Survey 1, and we surveyed the plots of 228 tenants in Crop Assessment 1.[27] Supplementary Table XI tests for differential attrition during Season 1. In the control group, 24% of tenants did not have a Crop Assessment in Season 1 and 20% of tenants could not be surveyed in the Tenant survey. The attrition rates in the treatment groups were similar to the control and to each other. The table shows that any differences in attrition rates across the different groups are not statistically significant.

As described in Section III.B, tenants who participated in the first season of the experiment were invited to sign a new contract in the second season.[28] In Season 2 we surveyed 179 of the Season 1 tenants in Tenant Survey 2, and we surveyed the plots of 192 of the Season 1 tenants in Crop Assessment 2.[29] In Supplementary Table XII, we test if the attrition rate in Season 2 – defined as a successful Crop Assessment or Tenant survey – was differential across the treatment groups. Differences in the rate of attrition are not significant throughout. They are also small in quantitative terms for the Crop Assessment 2 survey; however, the attrition rate in Tenant Survey 2 is around 11 percentage points higher amongst treatment tenants.

While it is comforting that we do not observe differential attrition across treatment groups, the average attrition rate in our experiment is high. This is likely because the tenants were young women who at the start of the experiment were living in their parents' household and were not married, and geographic mobility amongst this group is relatively high. Throughout we will probe the robustness of our findings to different bounding exercises (Lee, 2009; Fairlie, Karlan, and Zinman, 2015) where the bounds assume the tracked sample is either negatively

[27] This excludes plots on which the measured output was above the 99th percentile of the distribution of measured outputs, which we trimmed. Of those 228 tenants, 195 had rented one plot, 16 had rented two plots and 1 tenant had received 3 plots. There are therefore 262 plots from Season 1 in our dataset.

[28] In most cases where the tenants from Season 1 did not want to carry on cultivating the plot in Season 2, BRAC found replacement tenants. However, since this round of recruitment was carried on after the random assignment into treatment and control groups, we exclude these replacement tenants from the analysis in order to control for any selection effects.

[29] This excludes plots on which the measured output was above the 99th percentile of the distribution of measured outputs, which we trimmed. Of those 192 tenants, 173 had rented one plot, and 19 had rented two plots. There are therefore 211 plots from Season 2 in our dataset.

or positively selected. These are described in detail in Section IV.A below. The key results of the paper are robust to making extreme assumptions about the selection of attritors.

Balance. Table I provides balance tests for the baseline characteristics of the tenants and plots. Panel A present tenant characteristics such as their age, schooling, marital status, household demographics and socioeconomic status. The data was collected at the end of Season 0, prior to the start of Season 1. The average tenant in the sample is 21 years old, has 8 years of schooling, has 2 children and lives in a household with 5.4 people; 51% of the tenants are married. These observable characteristics are balanced across treatment groups. Out of 45 pairwise tests comparing C, T1 and T2 for each characteristic, we find that only one is significantly different at the 10% level based on randomization inference p-values: tenants in T1 had higher consumption expenditure than those in T2. These differences are unlikely to be important for the interpretation of our results.

Panel B presents plot level characteristics collected in Crop Assessment 1. That survey was conducted towards the end of Season 1. Therefore we consider immutable plot characteristics only for the balance checks. In addition summary statistics for total rainfall (pooled sample of both seasons) are reported. Across characteristics we find no economically meaningful and statistically significant differences between plots characteristics across treatment arms.[30]

Experimental Seasons. In Supplementary Figure I we describe the weather conditions on experimental plots during the experimental seasons relative to the typical weather conditions. We calculate the ratio of the estimated rainfall on experimental plots during each month of Season 0, 1 and 2, relative to the historic average rainfall in the same calendar month in the same area. We depict the distribution of that ratio across experimental plots, for each month separately. The figure shows that, on average, the weather conditions during the experimental seasons were similar to typical weather conditions. This suggests that none of our results will likely be driven by unusual weather events. The figure also shows that across plots there is substantial heterogeneity: some experimental plots experienced much lower or higher rainfall than is typical in the area during a calendar month. We will exploit that cross-sectional variation when estimating the responsiveness of yields to weather conditions in Section V.B.

[30]The only statistically significant differences are in terms of longitude and latitude: plots in treatment groups T1 and T2 tend to be located slightly to the North and plots in treatment group T2 tend to be located slightly more westernly. However, the magnitude of these differences is small: a difference of 0.013 degrees in latitude or longitude corresponds to roughly 1.5km around the equator. This difference is highly unlikely to explain differences in agricultural yields.

IV Results

IV.A Estimation

To identify the treatment effects of different contractual variations, we estimate:

$$y_{ict} = \sum_{k=1}^{2} \lambda_k T_{ik} + \delta_s + \epsilon_{ict}, \tag{4}$$

where y_{ict} is the outcome of interest for tenant i from club c in season t; T_{ik} is an indicator variable equal to 1 if tenant i belonged to a club of treatment group k and 0 otherwise, and δ_s are strata fixed effects. The sample includes tenants who were contracted at the beginning of season 1, prior to randomization. We use observations from both seasons 1 and 2 in order to improve statistical power.

The key parameters of interest are λ_k, the difference between outcomes of tenants who were assigned to treatment k and the control group. Under the identifying assumption that the control group represents a valid counterfactual, λ_k identifies the causal effect of the change in tenant i's contract on y_{ict}. In all regressions we report standard errors, clustered at the village level (the unit of randomization).

Throughout the paper, the p-values associated with hypothesis tests are calculated using randomization inference (Fisher's exact test). We estimate the coefficient of interest in 1000 alternative random assignments.[31] In each iteration we cluster standard errors at the village level, and record the distribution of the F-statistic associated with the hypothesis of interest. The randomization inference p-values report the percentile of the F-statistic found under the actual treatment assignment in the distribution of F-statistics found under alternative treatment assignments.

In order to assess the sensitivity of our findings to differential attrition (see Section III.E), we calculate bounds that adjust for differential attrition across the treatment and control groups under different assumptions regarding the positioning of the attritors within the distribution. 'Lee bounds' (Lee, 2009) trim observations from above (below) in the group(s) with lower attrition, to equalize the response rates across the treatment and control groups.[32] We then re-

[31] A treatment assignment is an N-component column vector, denoted $\mathbf{A}$, with i^{th} element $A_i \in \{C, T1, T2A, T2B\}$. Further denote the set of all potential treatment assignments, given our randomization procedure, as $\mathcal{A}$. We sample the 1000 alternative treatment assignments from $\mathcal{A}$ by running the same code which generates the actual treatment assignment another 1000 times to generate alternative treatment assignments.

[32] In particular, we find – by season – the group with highest attrition, and then delete – by season – observations with the highest (or lowest for the upper bounds) values in the other treatment groups until we have the same attrition rate as in the group with the highest attrition.

estimate the treatment effects in the trimmed sample to deliver the lower (upper) bounds for the true treatment effects. We also calculate alternative bounds, following Fairlie et al. (2015). For non-responders we impute – within treatment groups – the mean minus (plus) a specified standard deviation multiple of the observed distribution of outcomes in that treatment group. We then re-estimate the treatment effects in the sample including imputed data to find their lower (upper) bounds.

IV.B Effects on Output

We start by discussing the effects of a higher output share, income and risk exposure on output levels and yields.

Output Level. Table II presents the treatment effects on the total output (of all crops) that was observed on the plots during the pre-harvest crop assessment surveys. Column 1 shows that the average tenant in the control group had an output of USD 95 (at PPP). Relative to that, tenants in T1 had USD 56 more output on their plots. This implies that the 50% increase in their output share (from 50% to 75%) increased their output by 60%. On the other hand, tenants in group T2 had USD 5 more output relative to C, but this is imprecisely estimated. Moreover, the difference between T1 and T2 is significant (p-value=0.023). Overall, these findings imply that the tenants who were given a higher output share were more productive, and this was driven by the incentive effect rather than an income effect.[33]

Column 2 shows the effects for groups T2A and T2B separately. There is no significant difference between the coefficients of T2A and T2B. This implies that the risk profile of additional income does not play a significant role as T2A and T2B had similar effects on output. This reinforces the idea that the effect of treatment status T2 does capture any income effect induced by treatment condition T1. Nevertheless, it is important to note that the point estimates of T2A and T2B have different signs. Moreover, the difference between T1 and T2A is large (the magnitude of the point estimate for T1 is more than thrice as large as that of T2A), but not statistically significant. This suggests that some tenants in T2A, who were promised a safe income transfer at the end of the season, may have generated higher output than the control tenants while for tenants with the risky income transfer (T2B) this was not the case.[34]

[33]The finding that T2 tenants did not generate more output while T1 tenants did suggests that an efficiency wage story or a behavioral mechanism based on reciprocity are unlikely to be driving the effect of T1. If tenants in T1 were more productive because they received a better deal than they expected and wanted to work hard to reciprocate this favor (or to maintain it in the future), then we should see a similar effect on tenants who were given a cash transfer.

[34]Supplementary Table XIII in the Online Appendix shows the effects on self-reported output. The level of output is lower in all groups and while the signs of the point estimates are similar, the magnitudes are much smaller. This highlights the importance of using observed as opposed to self-reported information on output for

Figure III shows the cumulative distribution function (CDF) for output in each treatment group. One can see that the CDF of output for tenants who were assigned the high-incentive contract (T1) lies to the right of the CDF for tenants with the standard contract (control group). This implies that the differences in average output levels reported above are not driven by a particular group of tenants responding to the high-incentive contract, but rather by an effect throughout the distribution, in particular from the median upwards. The figure also shows that tenants in T1 performed better than the tenants who were given a cash transfer (T2), which demonstrates that the effects are not driven by the increase in expected earnings. A summarized version of these findings is presented as a box plot in Supplementary Figure II.

Yield. The rest of Table II presents the effects on yield, measured as output per square meter. We find that an increase in the tenant's share of output from 50% to 75% increases her yield by 0.074 USD per m^2 (p-value=0.024). We find no income effect in the specification of column 3 where we do not differentiate between T2A and T2B tenants (p-value=0.993). When estimating the effect of T2A and T2B separately in column 4 we find a small positive effect of treatment condition T2A and a negative effect of T2B. None of the effects are significant at conventional levels. These effects are qualitatively similar to those on total output.

Robustness. Supplementary Table XXIV provides attrition bounds for the effects on output. Overall, the estimates are robust to different adjustments for differential attrition.

In Table II, output value is trimmed at the top so that the top 99% of each treatment group is coded to missing. Effects without trimming are reported in Supplementary Table XIV where we find an even larger effect for being assigned to T1, and no significant effect of being assigned to T2. The larger effect of T1 in the non-trimmed results are driven by a handful of highly productive tenants in T1. Therefore, we rely on the trimmed observations as the main results.

In Section III.C we discussed that the income transfer in T2 might have been different from the (pre-season expected) income effect of treatment condition T1.[35] Since the income transfer in T2 was determined at the branch level, there is branch level variation in the ratio of the income transfer we implemented over the income transfer we should have implemented. In Supplementary Table XV we exploit this variation to assess whether a mis-calibration of T2 could explain why we do not find any significant income effect. In particular, this table presents results of regressions analogous to those in Table II, with the only exception that T2 is

our methodology.

[35] An alternative experimental design would have been to link the income transfer in T2 to the season's realized output in geographically close control clubs. This would have circumvented the challenges we faced in calibrating T2. That design requires to inform participants of the existence of other treatment conditions, which our design does not require. Whether this is an important advantage will depend on the specific setting.

a continuous variable measuring the aforementioned ratio. We proxy the pre-season expected income effect of T1 by half the realized output of tenants in the control group in the respective season, calculated at the branch level. To the extent that this is a suitable proxy, the ratio will be 1 in branches and seasons where the actual income transfer in treatment group T2 matches what we should have implemented. And it is proportionally higher (lower) in branches where the income transfer in treatment group T2 is higher (lower) than what we should have implemented. Under the assumption that the marginal income effect is constant, the coefficient on T2 will then estimate the true income effect of T1. The analogous statement holds for T2A and T2B. The results in Supplementary Table XV indicate that our previous conclusions in Table II do still hold with this alternative definition of the treatment variable. In particular, we continue to find very similar, quantitatively large, significant effects of treatment condition T1 on output, even though the level of significance decreases somewhat. In contrast, we do not find any significant income effect on output levels.

A concern with measuring output in our setting is that farmers have differential incentives to hide output. For that reason we do not rely on farmer-reported output measures or a crop-cutting survey conducted at harvest, but instead the pre-harvest Crop Assessment as discussed in Section III.D. To address concerns whether these efforts were sufficient to avoid differential output hiding across treatment groups, we examine the heterogeneity of the treatment effects on output, with respect to the timing of the crop assessment survey in each season.[36] Supplementary Table XVI repeats the analysis of Table II, with the exception that we include the interactions of the treatment indicators with *Survey Day* measuring the number of days between the crop assessment conducted on a tenant's plot and the first day of the crop assessment survey. There is no sizable or statistically significant evidence of differential harvesting between T1 and T2/C, both when considering the interaction terms individually and when testing the null of no differential treatment effect by survey days across all treatment arms jointly.[37] Online Appendix V.C. presents additional robustness checks.[38] Taken together, these robustness

[36]The intuition is that on early days of the crop assessment differential harvesting is less plausible; if we observe larger output differences between T1 and T2/C on later harvesting days, this would be worrisome, since it might indicate that farmers in groups T2 and C try to harvest crops before the arrival of the crop assessment survey team.

[37]If anything we observe larger differences between groups T1 and T2/C on plots that were surveyed early, though none of these differences are significant.

[38]During the crop assessment farmers are asked to self-report whether and what quantity of crops they harvested earlier in the season. Supplementary Table XVIII analyzes whether treatment status had any effect on early harvesting. We do not find any significant or sizable effect of treatment status on early harvest behavior at the extensive or on the intensive margin. Further the level of crops reported as being harvested prior to the crop assessment is low. However, in interpreting this robustness check it should be kept in mind that the early harvest data is self-reported. Another robustness check is to assess the heterogeneity of the treatment effects with respect to distance between the plots and the nearest market. The intuition is it may be easier for the tenants to sell their crops the closer their plots are to the market, and tenants in the control group may be particularly prone to do so,

tests build confidence that differential incentives to hide output are unlikely to be driving the effects on output.

A separate concern is whether the pre-harvest Crop Assessment is closely related to realized harvests. Burchardi et al. (2018) validate the pre-harvest Crop Assessment methodology in a setting where farmers have no incentive to hide output. They conduct, also amongst Ugandan farmers, a crop assessment prior to harvest and maturity of maize and conduct a crop-cutting survey at harvest time. They find that output measured through the pre-harvest Crop Assessment is strongly proportionately related to output measured through a crop-cutting survey.[39]

IV.C Effects on Input Use

Next we seek to understand how tenants adjust their behavior to generate the output effects found above. First we present the results on changes in input choices. Prediction 1 in Section II says that the increase in output share (s) for tenants in T1 should induce them to use more inputs while the increase in the output-independent income (w) of tenants in T2 should have no impact on their input use. We test these predictions using data from the tenant surveys conducted at the end of each season and recorded tenants' use of labor and capital inputs.

Capital Inputs. The tenants were asked to report the amount (if any) of any type of fertilizer and insecticide they used; and whether they acquired any agricultural tools during the past season.[40] Table III presents the effects of the treatment(s) on indicators of tenants' investments in capital inputs. Panel A of the table shows the effects on the extensive margin, while panel B presents the effect on the intensive margin (monetary value) of each input used.[41] In the first column, the outcome is any type of fertilizer used (either chemical or organic) by the tenants. Consistent with evidence from other East African settings (Duflo et al., 2011), fertilizer use was low among tenants in our sample. Only 28% of the tenants in the control group reported using any fertilizer on their plots. As a result of the higher output share, tenants in T1 were 9.4 percentage points (ppt) more likely to use any type of fertilizer. This corresponds to a 34% increase relative to the control group. While this effect is large, it is not precisely estimated at

since they have a greater incentive to sell their produce before the arrival of the survey teams. Supplementary Table XVII reports the results. We do not find significant heterogeneity in the output effect with respect to distance to nearest market. We thank an anonymous referee for suggesting these tests and the heterogeneity analysis by survey day.

[39] Crop-cutting surveys are considered the gold standard of agricultural output measurement, but infeasible for the purpose of this project given potential output hiding prior to harvest.

[40] All tenants were provided seeds by BRAC and, while they were free to use other seeds, only 13% of tenants reported using any seeds from another source, and this rate was not different across the treatment and control groups.

[41] For fertilizer and insecticide used, the monetary value corresponds to the amount spent on the relevant input used for the experimental plot; while for tools the monetary value corresponds to the total value of agricultural tools that the tenant owned at the time of the survey.

conventional levels (p-value=0.176). Panel B shows that the intensive margin effect on fertilizer use is even larger (in percentage terms) and precisely estimated. Tenants in T1 used on average USD 1.13 more fertilizer, which is 118% more compared to the average tenant in the control group. The corresponding effects for T2 are imprecisely estimated, although the point estimates are positive and not statistically different from the effects of T1. The test of equality between the treatment effects of T1 and T2 results in a p-value of 0.310 (0.350) for the extensive (intensive) margin of fertilizer use – reported at the lower section of each panel.

The second column of Table III displays the effects on insecticide use. In the control group, 28% of tenants reported using insecticide and spent on average USD 1.8 on it. Relative to the control, insecticide use was not significantly different among tenants in T1 or T2, neither on the extensive nor on the intensive margin. However, tenants in T1 spent significantly more on insecticide relative to tenants in T2 (p-value=0.046). The third column of the table shows that tenants in T1 were 9 ppt more likely to have purchased or acquired tools, and at the end of the season, the value of agricultural tools owned by the respondent was higher by USD 11 in T1 (30% relative to C). This latter effect is precisely estimated. We find no such effect for tenants in T2 and the difference between the coefficients of T1 and T2 is also statistically significant (p-value=0.059).

We have discussed the results of the treatment effect on a number of sub-categories of capital inputs. Testing multiple hypotheses poses well-known challenges to the interpretation of p-values. We present results of two approaches to deal with these challenges. First, in the final column of Table III we use an aggregate index that combines the three indicators presented in the table. To construct this index, we first standardize each outcome into a z-score, by subtracting the control group mean at the corresponding survey round and dividing by the control group standard deviation. We then take the unweighted average of all the z-scores, and again standardize to the control group. The results show that while there were no significant differences on the extensive margin, the tenants in T1 spent on average 0.2 standard deviations more on capital inputs compared to tenants in the control group. The corresponding effect for T2 tenants is -0.05 standard deviations and imprecisely estimated (the difference between T1 and T2 is significant with a p-value of 0.080). Second, we estimate the equations in columns 1 through 4 jointly, and then test the null hypothesis that a specified restriction holds in all estimating equations across columns. The results of these tests are consistent with what we found before when constructing an index: There is no robust evidence for an extensive margin effect. On the other hand, there is robust evidence that tenants in T1 have more intensive use of capital inputs (p-value=0.039). The corresponding effect is insignificant for tenants in T2 (p-

value=0.274). And the effect of treatment condition T1 on the intensive margin of capital use is significantly different from the effect of treatment condition T2 at the 5% level (p-value=0.044).

Supplementary Table XXV in the Online Appendix reports bounds that adjust for differential attrition across the treatment groups. The results show that the effects on the intensive margin (of fertilizer and tools) are robust if we impute – within treatment groups – the mean minus (plus) up to 10% of a standard deviation multiple of the observed distribution of outcomes in that treatment group. However, they are not robust if we conduct the imputation with 20% of a standard deviation, or if we trim observations at the top of the distribution to equalize the attrition rates across the groups (i.e. the lower Lee bound). They should be interpreted with this caveat in mind.

Labor Inputs. Tenants reported their own labor hours as well as any outside labor that they may have used on the plot, broken down into paid versus unpaid labor. Table IV reports the results of estimating specification 4 where the outcomes are variables pertaining to labor inputs used on the plot. Column 1 shows that tenants in T1 and T2 did not spend more hours working on their plots relative to tenants in the control group nor relative to each other. Similarly, in column 2, we do not find any significant differences in terms of paid labor across the treatment groups. On the other hand, column 3 shows that tenants in T1 had more "unpaid workers" working on their plots. In particular, they used 8 more days of unpaid labor during the season.[42] Relative to the mean in the control (12.5 days/season) this corresponds to a 64% greater use of unpaid labor on the plot. The difference between T1 and T2 in terms of unpaid labor is also large (approximately 6 days) but statistically not significant at conventional levels (p-value=0.173).

To address concerns related to multiple hypotheses testing, we again follow the two approaches discussed above. In the final column of the table we use an aggregate index that combines the three types of labor (own, paid and unpaid). The results show that the effect of T1 on this aggregate index is 0.2 standard deviation but imprecisely estimated at conventional levels (p-value=0.157) and the effect of T2 is 0.05 standard deviations, also imprecisely estimated (p-value=0.721). The difference between the two indices is insignificant (p-value=0.280). The same result is obtained when testing the corresponding cross-equation hypothesis.

Supplementary Table XXVI in the Online Appendix shows that these effects are not likely to be driven by differential attrition – the magnitudes of both the lower and upper bounds under alternative assumptions about the attritors are similar to the unadjusted estimates.

[42] A further breakdown of labor shows that the effect is driven by a combination of family and friends helping with cultivation, results available from the authors upon request.

Summary. Figure IV provides a visual summary of the effects on input use. It plots the standardized effect size and the 90% confidence interval around the treatment effects for labor and capital inputs. The solid squares correspond to the effects of T1, while the hollow ones show the effect of T2 relative to control. Overall, the results show that the tenants in T1 have responded to the increase in their output share by increasing their use of inputs – in particular fertilizer, tools and unpaid laborers – while the increase in the income of tenants in T2 had no such impact. These effects are perfectly in line with Prediction 1 of the framework: higher s increases input use, while higher w does not.

IV.D Effects on Risk-Taking

Prediction 2 says that the increase in s and w or risk exposure may also affect tenants' level of risk-taking. The direction of the effect is in general ambiguous, as it depends on the shape of the tenants' utility function. Only the prediction on background risk is unambiguous, it should decrease risk-taking. Typically it is difficult to test this prediction as often the researcher does not observe the risk associated with different input combinations. We provide three distinct pieces of evidence of changes in risk-taking. Note that we do not quantify the returns to risk-taking. Theoretically, standard asset pricing models suggest that the supply of riskier crops is such that their price and hence return compensates for risk-taking, i.e. is higher for riskier crops. Existing evidence on the risk-return trade-off in crop choice faced by farmers in developing countries is consistent with this (Cole et al., 2017).

Approach 1: Crop Choice. First, in our context the crops the tenant chooses to cultivate are a close proxy of risk-taking. The crops that BRAC offered seeds for and which were frequently cultivated by tenants imply different levels of risk exposure for the tenant. In particular, peanuts, tomatoes and maize are very sensitive to rainfall variation and exhibit high output volatility, while beans are relatively insensitive and exhibit less output volatility.[43] In Supplementary Table XX, we present two different approaches to demonstrate this. In Panel A, we exploit geographical variation among the plots cultivated by the control group of tenants to estimate the effect of rainfall throughout the season on the yield of each crop. In Panel B, we use data from FAOStat on crop yields of countries across time in Sub-Saharan Africa.[44] Both approaches show that maize and peanut yields are particularly sensitive to rainfall, while

[43] This may not hold in other contexts. The FAO publication *Irrigation and Drainage Paper No. 33* relates yield to water intake using evapotranspiration as a main parameter, rather than rainfall. It reports maize and beans as sensitive to water deficit, while groundnuts are described as tolerant to water deficit. While these findings are different from ours with respect to beans and groundnuts, one should notice that they are not specific to East African cultivars and local crop management practices.

[44] Available from: `http://www.fao.org/faostat/en`.

beans are less sensitive. We cannot use the first approach for tomatoes or potatoes, since no tenant in the control group chose to cultivate these two crops, but the results from the second approach demonstrate that tomatoes are as sensitive to rainfall as peanuts.[45] To the extent that rainfall is a good proxy for aggregate income shocks and that farmers can effectively not insure against it, this implies that the return to maize, peanuts and tomatoes has a high risk component, while for beans this is not the case.[46]

In order to test Prediction 2, we show how an increase of s in T1 or of w in T2 affects the tenants' decision to grow certain crops more than others. Table V presents the results of estimating specification (4) for outcomes quantifying the extensive and intensive margin of tenants' crop choice. In Panel A the outcome variables are indicators for whether a given type of crop was on the plot at the time of the crop assessment survey (extensive margin); in Panel B the output variable is the number of plants of the respective crop, irrespective of the plants' yield (intensive margin); and in Panel C the outcome variable is the value of the output of the respective crop, taking into account both the number of plants and the number of crops observed on the plants. The first row of Panel A shows that the tenants in T1 were significantly more likely to have maize and tomatoes on their plots compared to tenants in C. While the coefficients for beans and peanuts are also positive, they are not precisely estimated. When we compare the effect of T1 with T2, we find that the only crop that is significantly more likely to be present on T1 plots compared to T2 plots was tomatoes. Panel B shows that on the intensive margin, tenants in T1 grow more tomatoes, maize and peanuts compared to tenants in C, although the former two effects are not statistically significant at conventional levels. They do not grow any more beans. These results are highly consistent with additional risk-taking amongst tenants in treatment group T1. No such pattern exists for tenants in treatment group T2. Panel C shows that a similar conclusion is drawn when measuring the intensive margin in terms of value of output. Tenants in T1 produced more peanuts as well as tomatoes compared to tenants in C and T2. In particular, their expected output was USD 33 more for peanuts and USD 8 more for tomatoes, and these effects are significantly different from the corresponding effects of T2

[45] As an alternative way to quantify the riskiness of these crops, we used the FAOStat data to calculate the coefficients of variation in the yields (output per area cultivated) of maize, beans, peanuts, tomatoes and potatoes. We did so using cross-country variation, as well as time variation within countries, and finally using both cross-country and time variation in the panel data. Supplementary Table XXI shows that the coefficients of variation for maize, peanuts and tomatoes are greater than those for beans.

[46] Another dimension of risk that may affect crop choice is uncertainty of prices, as different crops are likely to have different levels of price variability. Tenants who choose to plant crops with greater price variability would be taking more risk. In the absence of time-series data on local prices in our study area, we use the cross-country panel data on crop prices provided by FAOStat to calculate the coefficient of variation of local average crop prices in Sub-Saharan African (SSA) countries. The lower panel of Supplementary Table XXI shows that the average (across SSA countries) coefficient of variation for price of beans is lower than those for maize, peanuts, tomatoes or potatoes. This suggests that beans are a safer alternative also in this respect.

(p-values of 0.065 and 0.074 respectively). This set of results is highly consistent with the interpretation that the increase in s led to greater risk-taking by tenants in T1, by inducing them to increase their cultivation of riskier crops (maize, peanuts and tomatoes) compared the the safer option (beans).[47] [48]

Our theoretical framework predicts that the increase in w in T2 may also influence the tenants' risk-taking. In particular, a safe increase in w (as in T2A) can lead to more or less risk-taking (Prediction 2.ii) while a stochastic increase in w (as in T2B) is likely to reduce risk-taking (Prediction 2.iii). In order to test for these predictions, we estimate the effect of T2A and T2B separately on crop choice. Supplementary Table XXIII in the Online Appendix shows that the tenants in T2A produced more peanuts as opposed to the other crops. This suggests that some tenants in T2A may have increased their risk-taking, in line with Prediction 2. We do not find discernable effects of T2B on crop choice.

Approach 2: Distribution of Output. Second, risk-taking will affect the distribution of output across plots within treatment groups. One way to detect risk-taking from the distribution of output is to note that the coefficient of variation of output does not vary with choices that scale up production by a constant factor, but it does vary with changes in risk-taking. In particular, the theory discussed in Section II suggests a coefficient of variation of $\frac{a\sqrt{\int(\theta-\mathbb{E}_\theta[\theta])^2 d\theta}}{a\mathbb{E}_\theta[\theta]+(1-a)}$, which is independent of $f(x)$ and increasing in a. The coefficient of variation of output across plots takes values of 1.37, 1.57, 1.66, and 1.28 in treatment arms C, T1, T2A, and T2B, respectively. Consistent with the earlier results, this approach suggests additional risk-taking by farmers in the treatment arm that provides a high output share s (T1) relative to control (C). It also uncovers additional risk-taking when farmers experience additional safe income (T2A) relative to control (C); and less risk-taking when farmers experience additional risky income (T2B) both relative to control (C) and to additional safe income (T2A).

Another way to detect risk-taking from the distribution of output is to estimate quantile treat-

[47] An alternative explanation could be that tenants in T2 diversify their crop portfolio in order to lower output variability. This would be the case if different crops had negatively correlated expected outputs, then the tenants could lower their risk exposure by intercropping them. Supplementary Table XXII shows that, among the control group, outputs of maize, beans and peanuts are not negatively correlated. If anything, the covariances are positive (imprecisely estimated). Moreover, as we show in the following section, tenants in T2 ended up having higher output variability relative to the control group. As such, a diversification strategy to insure against risks is unlikely to be driving the effects we observe on crop choice.

[48] Supplementary Table XXVII provides attrition bounds for the effects on crop choice. While most bounds are similar to the main estimates, there are a few notable differences. The Lee lower bound for the intensive margin of peanuts is close to zero and imprecisely estimated; while for tomatoes the Lee lower bound for both the intensive and the extensive margins are zero. This is because we have a small sample, and most tenants do not grow any tomatoes and few grow peanuts. Therefore when we trim the observations on top of the distribution for both of these crops, we lose all or almost all of the positive observations.

ment effects (QTE). We do this using the following specification:

$$Quant_\tau(y_{ict}) = \sum_{k=1}^{2} \beta_\tau^k T_{ik} + \phi_\tau \delta_s, \tag{5}$$

where y_{ict} is the output level of tenant i from club c in season t; T_{ik} is an indicator variable equal to 1 if tenant i belonged to a club of treatment group k and 0 otherwise and δ_s are strata fixed effects. One caveat to bear in mind is that, due to the small sample size, we have low power in estimating the treatment effects across the distribution.

Figure V displays the results. The QTE estimates reveal that there is considerable heterogeneity in the effects of incentives on the realized output levels: the effect on the 90th centile of output is 4 times as large as the effect on the 50th centile. Moreover, while we observe no negative effect on output at any centile, the treatment effect at the lower centiles are indistinguishable from zero. These effects are again consistent with additional risk-taking by tenants in T1. On the other hand, the lower panel of Figure V reveals that tenants in the high-income group (T2) do not generate more output than the control group, at any decile.

Supplementary Figure III displays QTEs for the sub-group of tenants who received safe versus risky w (T2A vs. T2B) cash transfers. For the group of tenants with additional safe income (T2A) we observe positive point estimates of the treatment effect in the highest deciles. This is consistent with the idea that tenants in T2A take on more risk, as predicted in part (ii.) of Prediction 2. Receiving additional stochastic income (T2B) seems to have the opposite effect. Again this is consistent with the prediction of part (iii.) of Prediction 2: relative to safe income w, additional stochastic income will induce less risk-taking and might have a negative effect on risk-taking. Note that these quantile treatment effects are estimated imprecisely, given the small sample size.

Both approaches to detect risk-taking from the distribution of output should be interpreted with caution. While the results are consistent with risk-taking, they are also consistent with other explanations. Tenants might differ in their innate abilities, and more able tenants in T1 might respond more strongly to the high-incentive contract (Lazaer, 2000), for example by working harder.[49]

Approach 3: Responsiveness to θ. A third approach to uncover risk-taking is suggested by the theory: one can estimate the responsiveness of output to θ across treatment groups. The coefficient estimate will identify the treatment-group-specific $a \cdot f(x)$. The approach can be

[49] We did not find a significant difference in terms of hours worked by T1 tenants relative to the control group (Section IV.C), but they may have exerted more effort during those hours.

operationalized by using weather data to proxy for variation in θ. This allows us to draw inference on risk-taking, a, given information on treatment-group-specific changes in $f(x)$. We explain this approach and how information on treatment-group-specific $f(x)$ can be obtained in detail in Section V.B when discussing the quantitative contributions of input levels and risk-taking to the output effects. As an upshot, this approach also suggests significant additional risk-taking in the treatment arm that provides a high output share s (T1) relative to control (C).

Summary. The collection of evidence in this section shows that tenants with a higher share of output (T1 vs. C) made riskier input choices. Additional safe income w (T2A vs. C) leads to somewhat more risk taking, while additional exposure to uncorrelated background risk (T2B vs. T2A) induces less risk-taking.

IV.E Effects on Other Outcomes

The results thus far showed that tenants in the high-incentive group (T1) invested more in cultivating their rented plots, took on additional risk and generated more revenue from them. A natural question is whether these are achieved at the expense of other detrimental effects for them, their households or the plots. In particular, since we observe an increase in unpaid labor, in part driven by family labor, this raises the question of whether the increased labor activity on the plot crowded out other income-generating activities and reduced household earnings. To shed light on this, we estimate the impacts on respondent's and her household's economic wellbeing. Table VI presents the results. The table shows that tenants in T1 did not have lower labor income, consumption, cash savings, household income or assets at the end of the experiment. If anything, column 4 shows that they had higher household income and column 5 shows that they had more households assets (both marginally significant at 10% level) relative to C.[50] These findings imply that the high incentive contract did not crowd out any other productive activities. If anything, the evidence is in line with it increasing household income.[51]

While high tenant incentives may increase output and their households' economic well-being, they may have negative consequences for the environment. In particular, short-term, high-incentive contracts (such as those we study here) may lead the tenant to overwork the land (e.g. by overusing fertilizers) which may lead to environmental degradation. To test for such an effect, at the end of the experiment (i.e. at the end of the second experimental season) we

[50] Findings in Table III showed that tenants in T1 were more likely to invest in tools for their plots. This may generate positive spillover effects on their households' other plots, which may explain the larger effect on their household income relative to their personal labor income.

[51] Supplementary Table XXVIII displays bounds for differential attrition for the effects reported in Table VI.

collected soil samples from the plots that were part of the experiment, and tested their chemical composition. In particular, we measured the amount of Nitrogen, Phosphorous, Potassium, Organic matter, and the Ph-level of the sample. Table VII shows the results of estimating the effects of the treatment(s) on these soil quality indicators. We do not find any significant differences in terms of soil quality of the plots in different treatment arms. While this suggests that the high incentive contract did not come at a cost to the soil quality in the short run, it does not rule out long-run negative effects or changes in unobservable dimensions of soil quality.

V Discussion

In this section we interpret the experimental findings, discuss their welfare implications and make note of possible limitations.

V.A Understanding Tenants' Choices

The theory in Section II highlights three drivers of the tenants' output and risk-taking choices: incentives, risk exposure, and income. We next revisit these predictions and show that the empirical results are highly consistent with them.

Prediction 1 says that a higher tenant share s leads to higher input levels, x, while income and risk exposure have no such effects. In Section IV.C we show that tenants with a higher output share use more inputs, both capital and labor. Additional income or risk exposure does not lead to substantial changes in input levels.

Prediction 2 says that risk-taking increases with income w under DARA and decreases with risk exposure. An increase in the tenant's share s has both of those effects on risk-taking, in addition to an incentive effect; the sign of the total effect on risk-taking and hence output is theoretically indeterminate. In Section IV.D we show that tenants with a higher output share also take on additional risk. Additional income leads, if anything, to a small increase in risk-taking, whereas background risk discourages risk-taking.

Prediction 3 says that whenever an increase in s induces risk-taking, the aggregate impact on output should be positive; and the effect of income and risk exposure on output has the sign of their effect on risk-taking. In Section IV.B we present output results that are highly consistent with these predictions. We find that an increase in the tenant's share by 25% leads to 60% higher output. Additional income leads, if anything, to a small increase in output; while additional risk exposure leads to a small decrease in output.

Note that the combined income and risk exposure induced by treatment T2B would, in theory, discourage risk-taking less than the combined income and risk exposure induced by treatment T1: treatment T2B induces uncorrelated background risk, while T1 induces additional exposure to risk perfectly correlated with yields; and additional risk in T1 includes the possibility of crop failure which T2B does not allow for and which farmers might be particularly averse to.[52] Therefore the positive effect on risk-taking associated with treatment T1 is likely to be a lower bound on the incentive effect of a higher share s on risk-taking.

Our theory does not allow for income effects resulting from a labor-leisure trade-off (often highlighted in labor economics). Further, the additional income might relax credit constraints (often highlighted in development economics), even though it was to be realized in the future at the time of the agricultural decisions in both T1 and T2. To the extent that such effects are present in the setting under study, our empirical results suggest the sum of these effects is at most small.[53]

V.B Accounting for Output Effects

Next we discuss whether and under what assumptions the output effects can also be accounted for quantitatively by the observed changes in input use and risk-taking.

Taking logarithms of equation (1) gives:

$$\log y = \log[a(\theta - 1) + 1] + \log f(x). \tag{6}$$

Equation (6) suggests that the change in log-output can be decomposed into the additive effects of changes in risk-taking and changes in inputs.

Effects via Inputs. First, let us quantify the change in $[\log f(x)]$ resulting from altered input choices we observe. To that end, we assume a parametric form for $f(x)$. In particular, let $x = (k, l, z)$ and $f(x) = \psi k^{\alpha} l^{\beta} z^{\gamma}$, where k denotes capital, l labor, z is land and ψ is the farm TFP. Substituting into equation (6) yields:

$$\log y = \zeta + \alpha \log k + \beta \log l + \gamma \log z \tag{7}$$

[52] Both these differences between the risky income induced by T1 and in T2B also might imply relatively better credit access in T2B; again this would imply that the positive effect on risk-taking associated with treatment T1 is likely to be a lower bound on the incentive effect.

[53] In the Online Appendix V.F., we use the tenant survey data to assess whether the contractual changes had any impact on tenants' access to credit. In particular, we examine whether the tenants in T1 or T2 had more outstanding loans (both on the extensive and intensive margins) and whether they believe they would be able to borrow 25.000 UGX and 300.000 UGX, respectively, for 6 months. Results presented in Supplementary Table XXIX show no effect along any of these dimensions, nor on the aggregate index. We also examined whether the treatment(s) had any effects on the sources of borrowing (whether the tenant had borrowed from friends, MFIs or NGOs, cooperatives or moneylenders). Supplementary Table XXX shows that the treatment had no effect on the source of borrowing either. Our interpretation is that additional income realized by the farmers in T1 and T2 does not appear to substantially improve credit access.

where $\zeta := \log[a(\theta - 1) + 1] + \log \psi$. This formulation is consistent with the literature estimating factor shares in agriculture. To assess the contribution of changes in input levels to the output effects, we require estimates of the treatment effects on k, l and z, as well as estimates of the factor shares. Table VIII presents the results of estimating the treatment effects on the log values of total output (y), capital (k), labor hours (l) and size of the plot area cultivated (z).[54] In column (4) we estimate log output to increase by 0.38 log-points for tenants in T1 relative to tenants in control. Columns (1) to (3) show that tenants in T1 increase their investments in k by 0.20, l by 0.10 and z by 0.29 log-points, respectively. Factors shares have been estimated, amongst others, by Valentiyi and Herrendorf (2008) and Restuccia and Santaeulalia-Llopis (2017). Using the results for the U.S. agricultural sector in Valentiyi and Herrendorf (2008) (which are $\alpha = 0.36$, $\beta = 0.46$, $\gamma = 0.18$) implies that the observed changes in input levels explain an increase in output of 0.17 in log-points; using Restuccia and Santaeulalia-Llopis (2017)'s estimates of factor shares in Malawi ($\alpha = 0.19$, $\beta = 0.42$, $\gamma = 0.39$), the predicted output increase is 0.19 log-points. Therefore, the observed changes in input levels explain approximately half of the output effect we observe. This also implies that $\frac{f(x_{T1})}{f(x_C)} = e^{0.19} \approx 1.21$.

Effects via Risk-Taking. Quantifying the contribution of risk-taking to output increases requires information about both the level of risk-taking a in each treatment group, and the returns to risk-taking, $\mathbb{E}[\theta]$. We first discuss how to quantify the relative level of risk-taking across treatment arms. The theory suggests that the slope coefficient of a regression of output y on θ identifies $a \cdot f(x)$. For farmers in developing countries, an important subset of variation in θ is weather risk. Therefore the relative responsiveness of output to weather shocks in T1 relative to C is informative about the ratio $\frac{a_{T1} f(x_{T1})}{a_C f(x_C)}$. We obtain an estimate of this ratio through three steps, the details of which are explained in Online Appendix IV. We first obtain satellite-imagery based rainfall data for each month of the agricultural season and match it to the geolocation of the experimental plots. Second, using data from T2 we find a predictive model of how the multidimensional rainfall data maps into a unidimensional measure of weather conditions, scaling proportionately with output. Applying this model we calculate a measure of weather conditions for plots in C and T1. Third, we estimate how strongly output on plots in treatment arms C and T1, respectively, responds to the measure of weather condi-

[54] Since we do not observe the quality of labor hours (i.e. effort) or land cultivated, our measures are, at best, imperfect proxies for the true input levels. To calculate aggregate capital used, we sum the values of fertilizer, insecticide and households tools. When aggregating the labor hours, we need to combine own labor hours (reported for a typical week during the season) and numbers of days of hired labor used during the season. To do so, we assume that each worker-day corresponds to 8 hours; and each season lasted for 3 months. While the size of the plot allocated to tenants in different treatment arms was identical on average (due to the randomization), the tenants could decide to cultivate any fraction of the plot. The land size variable corresponds to the cultivated area as observed during the crop assessment survey.

tions. Denote the estimated coefficients as $\hat{\rho}_k$, respectively, where k_i indicates that plot i is in treatment arm $k \in \{C, T1\}$. The ratio $\frac{\hat{\rho}_{T1}}{\hat{\rho}_C}$ is then a consistent estimate of $\frac{a_{T1} f(x_{T1})}{a_C f(x_C)}$. For tenants in control C the responsiveness of output to weather conditions is estimated to be 0.614 (p-value $= 0.008$), and in treatment group T1 it is estimated to be 1.393 (p-value $= 0.002$).[55] These point estimates suggest a ratio $\frac{a_{T1} f(x_{T1})}{a_C f(x_C)}$ of 2.27.[56] Above we found that $\frac{f(x_{T1})}{f(x_C)} \approx 1.21$. Together these results imply that $\frac{a_{T1}}{a_C} \approx 1.88$.[57]

Lastly, we need to quantify the returns to risk-taking. We can not quantity these, as our experiment does not allow to estimate the distribution of θ separately from the level of a_C. However, existing evidence suggests that the gross returns to risky agricultural techniques are large; and their adoption rates are low in many developing countries and especially in Africa. For example, Duflo et al. (2008) summarize related literature as finding that fertilizer and hybrid seeds increase yield from 40% to 100%. Both hybrid seeds and the fertilizers studied are risky investments, since they are highly complementary with rainfall. The same authors report adoption rates for fertilizer of 35% to 40% for farmers participating in their study in Kenya. If we take adoption rates as rough measure of a, and further assume that adoption of hybrid seeds and fertilizer is akin to moving from the safest input mix to the most risky input mix, these numbers are informative about the extent to which the additional risk-taking in T1 translates into additional output. At the midpoints of the given ranges we have $a_C = 0.375$ and $\mathbb{E}[\theta - 1] = 0.7$, which suggests that the additional risk-taking of tenants in T1 explains 0.17 log points of the 0.19 log points in output difference unexplained by input choices.[58]

These results suggest the estimated output effects can be almost fully explained by additional input use and risk-taking of tenants, each contributing about half of the total output effect. It should be kept in mind that the exact quantitative decomposition depends on assumptions about functional forms as well as effect sizes. The main take-away of these calculations is not to provide an exact decomposition, but rather that a set of reasonable assumptions exists under which the total output response can be rationalized as being the effect of the input and

[55] Reassuringly, this suggests that our measure of weather conditions as constructed in the first and second step is indeed meaningful.

[56] The responsiveness to weather shocks of plots in C and T1 should *not* be compared to the estimated responsiveness to weather shocks in T2, which by construction is 1. The measure of weather conditions is constructed using T2 data, implying that in the second step we likely overfit the predictive model towards output of T2 plots.

[57] An alternative approach is to note that $\frac{SD(y|T1)}{SD(y|C)} = \frac{a_{T1} f(x_{T1})}{a_C f(x_C)}$, where $SD(y|k)$ is the standard deviation of output in treatment arm k. Our results suggest a ratio of standard deviations of 1.71. This approach is simpler, but its results are less straight-forward to interpret. While differential variation in output across treatment arms is consistent with differential risk-taking and input levels, it may also reflect heterogeneity in the tenants' responses to incentives. Nonetheless, it is comforting that the results of those two unrelated approaches are in the same ballpark.

[58] This results is obtained as $\log[(0.375 \times 1.88 \times 0.7 + 1)/(0.375 \times 0.7 + 1)] \approx 0.17$.

risk-taking choices we observe.

Moral hazard models are typically phrased in terms of the agent's 'effort'. Effort is then often interpreted as a metaphor for factors that are non-observable and therefore non-verifiable. Such factors that are truly unobservable by the landlord might exist. They will also not be observable to us as researchers, and such factors would contribute to the small unexplained output increase in T1 relative to control. However, taking the decomposition exercise at face-value, a more suitable interpretation of the standard moral hazard model is to think of both input choices and risk-taking as 'effort'. While these factors are in fact partly observable, they are non-verifiable. Contracts are typically not written contingent on these choices, presumably because the informational costs are prohibitively high and such contingent contracts might be particularly difficult to enforce given the state of courts and other enforcement mechanisms. (In the end, at some cost many if not all important dimensions of agricultural practice are observable. But observing them is costly. As researchers, a large fraction of our research budget was spent on conducting high-intensity pre-harvest land measurements, crop assessment surveys and soil tests.)

V.C Welfare Implications

Tenants with a 75% output share generate 60% higher output than tenants with a 50% output share. Consequently income increases by 140%. However, these tenants are also exposed to a higher variance of income, both mechanically and because they increase their input levels and risk taking. If tenants are risk-averse, this begs the question whether and by how much welfare increases when tenants' are allowed to keep a higher share of outcome.

In order to make any welfare statement, we need to gauge the distribution of income that tenants are facing in T1 and C.[59] We obtain an estimate of the distribution of income over time on each plot. To that end, we use satellite-imagery based data on monthly rainfall in 0.1 degree grid cells for the 16 seasons preceding our experiment as well as the seasons of our experiment, and match it to the geolocation of the experimental plots.[60] Then we use the treatment arm specific estimates of the responsiveness of output to weather conditions to calculate the predicted value of output for every plot i in season t (see Section V.B and Online Appendix IV. for details). To this plot specific vector of agricultural gross income we add the average income obtained from other sources net of costs of inputs by farmers in C and T1, respectively. This procedure yields an estimate of the distribution of income across time t for

[59]Note that the cross-sectional variation in income is not a suitable approximation, since it also reflects unobserved but fixed productivity differences across tenants and space.

[60]In Uganda there are two agricultural seasons per calendar year, and we use rainfall data from 2006 through 2013.

each plot i, assuming that output reacts to weather in the time series the same way as it does in-sample, and that all variation in output is driven by weather shocks.

We then calculate the certainty equivalent of the income stream of T1 for agents with the baseline risk exposure of tenants in C for a range of levels of risk aversion. Figure VI plots the results. There is limited agreement on what level of risk aversion characterizes choices under uncertainty, and estimates of risk aversion yield wildly different results across different methodologies and settings (Rabin, 2000). However, for levels of risk aversion that appear to characterize well choices over larger stakes, such as $\eta \in [1,2]$, we find substantial welfare gains for tenants who are given a higher share s of output (T1) relative to control (C). Tenants in control, who operate under a 50-50 sharecropping contract, would need to be given 45 to 55 USD (PPP) for sure to be as well off as tenants who are residual claimants on 75% of output.

These large benefits for tenants of providing incentives need to be weighed against a moderate loss for landlords. For landlords the high-incentive contract implies a fall of expected income by 20%, or roughly 10 USD (PPP).

V.D Limitations

When extrapolating from the findings reported in this paper it is important to keep in mind at least four limitations of our study.

Selected Sample. First, the sample of farmers was not chosen explicitly to be representative of a general population of farmers. On the contrary, being female and of young age were implicit inclusion criteria. We assess how the experimental sample compares to a general population of farmers, by comparing farmer characteristics to the 2013/2014 wave of the Uganda National Panel Survey, a survey that is part of the World Bank's "Living Standards Measurement Study" (LSMS) program and designed to be nationally representative. Supplementary Table III presents summary statistics on key characteristics of farmers in the control group of our experimental sample (column 1) and all tenant farmers in the UNPS sample (column 2).[61] It also reports normalized differences of means between the UNPS sample and the control group of the experimental sample. We adopt the Imbens and Wooldridge (2009) criterion and judge sample differences as large when the normalized difference exceeds 0.25. Relative to the general population of Ugandan tenant farmers (column 2), tenants in our experiment are younger, less likely to be married and have one and a half more years of education. Both of

[61] The UNPS survey is based on self-reported output measures. Since self-reported and plot based measures of output tend to yield substantially different results Lobell et al. (2018), we compare responses to the UNPS survey to self-reported responses from the experimental sample of farmers. See Online Appendix III. for details.

the latter differences are partially the consequence of gender and cohort effects.[62] In terms of their agricultural production though the farmers in the experimental sample are not dissimilar from the general population of Ugandan farmers: The farmers in the experiment have very similar levels of yields even though the plots we rented out are somewhat smaller than typical plots; the experimental sample of farmers use similar levels of tools and similar amounts of total labor, but less fertilizer; and they grow more of the crops for which they were given seeds, maize, beans and peanuts. With the exception of plot size and the value of peanut output, these differences are not substantial in terms of the normalized differences.[63]

An important question is whether the specific characteristics of the experimental sample are related to particular responses to contractual terms. We provide suggestive evidence by exploiting within sample variation in terms of marital status, age, schooling and plot size to gauge whether responses are heterogeneous along these dimensions. Supplementary Tables V through VIII provide results from this exercise. We do not find significant heterogeneity in the response to the experimental treatments along age, marital status or plot size. We do observe that farmers with more schooling respond more strongly to both the incentive treatment as well as to the safe cash transfer. This suggests that the treatment effect of T1 would have been lower on a less educated sample of farmers.[64] In terms of magnitudes, the coefficient of the interaction term "T1 $\times$ Schooling (years)" in Supplementary Table VII is 13.7, while the difference in schooling levels between the tenant farmers in our sample and the average tenant farmer in Uganda is 1.7 years (see Supplementary Table III). None of the heterogeneity analysis is highly powered, but taking the coefficient estimates at face value this suggests the treatment effect of T1 (56.3) might be lower by roughly 40% for a farmer with average education level.

Seasonal Effects. Rosenzweig and Udry (2018) recently made the important observation that estimating the average effects of interventions on agricultural output is difficult both for farmers and researchers given variation in weather conditions across seasons.

One important and predictable source of such variation is that the two agricultural seasons observed in the tropics typically have different yield potential. For that reason we ran the experiment both during the long (Season 1) and short rains (Season 2). Supplementary Table XIX presents the main output results of Table II disaggregated by season. In line with expectations

[62] Differences in marital status and schooling are somewhat smaller when restricting the UNPS sample to female and below-median age farmers ('UNPS - restricted', column 3). The differences may also reflect effects of the ELA program, which is designed to empower young women (Bandiera et al., 2018).

[63] Conclusions are similar when comparing the UNPS sample to the full experimental sample rather than just farmers in the control group, see Supplementary Table IV.

[64] At the same time the effect of T2 is also increasing in the tenant's schooling level, and the differences between the inventive and income effects (i.e. T1 v.s. T2) is if anything larger on a less educated set of farmers.

the agricultural output in Season 2 is substantially lower than in Season 1, roughly half for the control group. We also find a significantly stronger output response of treatment T1 in the first season: while in Season 1 farmers in treatment group T1 produce 69% more output than farmers in control, in Season 2 the corresponding increase is 27%. Note that any such heterogeneity test has low power. We therefore report the average results throughout the paper, which should be interpreted as weighted average across both seasons (with higher weight on Season 1 given Season 2 attrition).

On top of the variation in the output potential across the two types of agricultural season, there is variation within the same type of season across years. This is particularly important in our setting since risk-taking by farmers in treatment T1 implies even larger than usual variance in output, and results from any given year or season correspond to a particular realization of that risk. In Section V.C we describe a structured approach to extrapolate from the experimental setting: we exploit variation in the weather conditions across experimental plots and seasons to estimate the responsiveness of farmers in each treatment group to weather conditions; we then use the distribution of weather conditions across several past seasons to proxy for the distribution of weather conditions faced by farmers; combining these allows to estimate the average output response across plots in a treatment arm in a typical season. This exercise suggests that on average across seasons farmers in treatment group T1 would produce 57% more output than farmers in control.[65] These results do not suggest strong reasons to think the experimental results are specific to the agricultural seasons during which the experiment was conducted.

The same exercise also predicts output of 188.2 USD and 116.4 USD in T1 and C, respectively, in Season 1 and 83.2 USD and 68.7 USD, respectively, in Season 2. Therefore treatment fixed effects and average responsiveness to weather conditions across season in T1 and C – together with realized weather conditions in Season 1 and Season 2 – explain an output increase in T1 relative to C of 61.7% in Season 1 and of 21.0% in Season 2, very close to what we find in Supplementary Table XIX. The differently sized responses across seasons might also reflect that treatment effects decrease with experience, or that output sharing rules have other dynamic effects, or the implementation challenges we describe in Section III.C. We cannot ultimately reject those hypotheses.[66] But these results suggest that the differences in responses might plausibly be driven by the combination of different levels of inputs and risk-taking across

[65]A less structured approach is to observe that the rainfall patters during the two experimental season were no particular outlier relative to historic rainfall patterns, see Supplementary Figure I). However, this masks that any given plot might have been exposed to an extraordinary rainfall pattern.

[66]We think the dynamic effects of the terms of output sharing rules are an interesting area for future research. Our results cannot contribute to that debate.

treatment groups and the particular weather realizations of Season 1 and 2.

Prevalence of Sharecropping. Third, formal sharecropping contracts are not as common in rural Uganda as in other places, in particular Southeast Asia.[67] To the extent that this implies that the tenants are imperfectly aware of the functioning of sharecropping contracts, this would again imply a muted response toward contractual changes relative to a situation where sharecropping contracts are well understood. However, the fact that sharecropping contracts are largely absent in Uganda might also be the consequence of underlying differences between rural Uganda and other areas where sharecropping contracts are more prevalent. If such differences are related to the elasticity of tenant responses towards changes in s – as would for example be the case if the underlying agricultural production function is different – our findings are unlikely to be externally valid.

Externalities. Finally, we find that tenants respond to higher incentive contracts both by acquiring more inputs, and by taking on more risk. To the extent that either of these responses is having externalities, such responses may not be socially optimal. For example, the tenants could be depleting their land of nutrients such that land quality is substantially reduced in the long run. We do not find any such evidence, but we cannot exclude that unmeasured negative effects do exist. Also, tenant choices might have pecuniary externalities on crop prices; if insurance markets are incomplete, the optimal level of private risk-taking might be different from the socially optimal level of risk-taking.

VI Conclusion

The question of how output sharing rules affect economic agents' incentives for investment and risk-taking is central to development economics, contract theory, and public economics. In the context of agricultural tenancy contracts, the idea that a tenant who has to share a large part of her output with the landowner will have little incentive to invest in cultivation has been long established. Yet, the empirical evidence on this is scant. We find that an increase in the output share from 50% to 75% leads tenants to invest more in inputs, especially capital (fertilizer and tools) and take on more risk. As a result of these changes, they produce 60% more output.

We find the effects of high-incentive output sharing rules on agricultural input choices and output are largely to be interpreted as an incentive effect. Taken at face value, our results

[67] However, when tenancy contracts exist, a 50% output sharing rule is very common around the world, see Otsuka et al. (1992) or Banerjee et al. (2002).

suggest that increasing the tenants' income is unlikely to trigger the same type of output response. However, this interpretation ought to be cautioned. The income treatment in this paper promised future income to tenants – to mirror the income effect of the high output share and gauge its size. This should not be compared with policies such as the unconditional cash transfers studied by Haushofer and Shapiro (2016) which might have a stronger effect on relaxing liquidity constraints, inducing changes in labor supply and consumption. Their evaluation also considers cash transfers which were at least an order of magnitude larger than our income treatment.

Moreover, we find that one effect of strengthening the cultivator's position as residual claimant is increased uptake of profitable but risky agricultural techniques. This finding speaks to the large theoretical literature in public finance that studies the effect of taxation on entrepreneurial risk-taking (Domar and Musgrave, 1944; Mossin, 1968; Stiglitz, 1969; Feldstein, 1969). This literature highlights that even the sign of the effect is theoretically indeterminate in the absence of strong assumptions. Our findings suggest that – in our context – output taxation discourages risk-taking.[68]

Our findings are also consistent with the recent work by Karlan, Osei, Osei-Akoto, and Udry (2014) who find that farmers in Ghana make riskier and presumably profitable production choices when provided with insurance. The socially inefficient production choices induced by incomplete insurance markets will best be addressed by effectively providing insurance. However, in the absence of perfectly functioning insurance markets, our results suggest that increasing the tenant's share in output, may also encourage profitable risk-taking, in addition to the incentive effects on input levels.

IIES, Stockholm University; BREAD; and CEPR

Department of Economics, IGIER and LEAP, Bocconi University; and CEPR

IIES, Stockholm University

BRAC Institute of Governance and Development

[68]Recent experimental studies highlight the importance of kinship taxes in the African context. This literature suggests that demands from individuals' social networks to share output may lower individuals' incentives to invest in high-return projects (Jakiela and Ozier, 2016) and lower enterprise growth (Squires, 2017). Studying the interaction of kinship taxes with formal output sharing rules (such as sharecropping contracts) can be valuable for future research.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001, December). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review 91*(5), 1369–1401.

Ackerberg, D. A. and M. Botticini (2002). Endogenous Matching and the Empirical Determinants of Contract Form. *Journal of Political Economy 110*(3), 564–591.

Adamopoulos, T. and D. Restuccia (2014, June). The Size Distribution of Farms and International Productivity Differences. *American Economic Review 104*(6), 1667–97.

Arcand, J.-L., C. Ai, and F. Ethier (2007). Moral Hazard and Marshallian Inefficiency: Evidence from Tunisia. *Journal of Development Economics 83*, 411–445.

Arrow, K. J. (1971). The Theory of Risk Aversion. In K. J. Arrow (Ed.), *Essays in the Theory of Risk Bearing*. Amsterdam: North-Holland.

Bandiera, O., I. Barankay, and I. Rasul (2011). Field Experiments with Firms. *Journal of Economic Perspectives 25*(3), 63–82.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2018). Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. Working paper.

Banerjee, A., P. Gertler, and M. Ghatak (2002). Empowerment and Efficiency: Tenancy Reform in West Bengal. *Journal of Political Economy 110*(2), 239–280.

Bell, C. (1977). Alternative Theories of Sharecropping: Some Tests Using Evidence from Northeast India. *Journal of Development Studies 13*, 317–346.

Besley, T. (1995). Property Rights and Investment Incentives: Theory and Evidence from Ghana. *Journal of Political Economy 103*(5), 903–937.

Binswanger, H. P. (1980). Attitudes Toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics 62*(3), 395.

Binswanger, H. P. and M. R. Rosenzweig (1982). *Rural Labor Markets in Asia: Contractural Arrangements, Employment and Wages*, Chapter Contractual Arrangements, Employment and Wages in Rural Labor Markets: A Critical Review. New Haven: Yale University Press.

Braido, L. H. B. (2008). Evidence on the Incentive Properties of Share Contracts. *Journal of Law and Economics 51*(2), 327–349.

Braselle, A.-S., F. Gaspart, and J.-P. Platteau (2002). Land Tenure Security and Investment incentives: Puzzling Evidence from Burkina Faso. *Journal of Development Economics 67*(2), 373–418.

Burchardi, K., J. de Quidt, B. Lerva, and S. Tripodi (2018). Pre-Harvest Measurement of Agricultural Output. IIES working paper.

Cole, S., X. Giné, and J. Vickery (2017). How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment. *The Review of Financial Studies 30*(6), 1935–1970.

Domar, E. D. and R. A. Musgrave (1944). Proportional Income Taxation and Risk-Taking. *The Quarterly Journal of Economics 58*(3), 388.

Duflo, E., M. Kremer, and J. Robinson (2008, May). How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya. *American Economic Review 98*(2), 482–88.

Duflo, E., M. Kremer, and J. Robinson (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *The American Economic Review 101*(October), 2350–2390.

Eswaran, M. and A. Kotwal (1985). A Theory of Contractual Structure in Agriculture. *The American Economic Review 75*(3), 352–367.

Fairlie, R. W., D. Karlan, and J. Zinman (2015). Behind the GATE Experiment: Evidence on Effects of Rationales for Subsidized Entrepreneurship Training. *American Economic Journal: Economic Policy 7*(2), 125–161.

Feldstein, M. S. (1969). The Effects of Taxation on Risk Taking. *Journal of Political Economy 77*(5), 755–764.

Feldstein, M. S. (1976). Personal Taxation and Portfolio Composition: An Econometric Analysis. *Econometrica 44*(4), 631–650.

Fermont, A. and T. Benson (2011). Estimating Yield of Food Crops Grown by Smallholder Farmers. IFPRI Discussion Paper 01097.

Ghatak, M. and P. Pandey (2000). Contract Choice in Agriculture with Joint Moral Hazard in Effort and Risk. *Journal of Development Economics 63*(2), 303–326.

Goldstein, M. and C. Udry (2008). Property Rights and Investment Incentives: Theory and Evidence from Ghana. *Journal of Political Economy 116*(6), 981–1022.

Gollin, D., D. Lagakos, and M. E. Waughn (2014). The Agricultural Productivity Gap. *Quarterly Journal of Economics 129*(2), 939–993.

Hallagan, W. (1978). Self-Selection by Contractual Choice and the Theory of Sharecropping. *Bell Journal of Economics 9*, 344–354.

Haushofer, J. and J. Shapiro (2016). The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics 131*(4), 1973–2042.

Holmstrom, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica 55*(2), 303–328.

Hornbeck, R. (2010). Barbed Wire: Property Rights and Agricultural Development. *Quarterly Journal of Economics 125*(2), 767–810.

Hossain, M. and A. Bayes (2015). Leading Issues in Rural Development, Bangladesh Perspective.

Imbens, G. W. and J. M. Wooldridge (2009, March). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature 47*(1), 5–86.

Iwanowsky, M. (2018). Property Rights, Resources and Wealth: Evidence from a Land Reform in the United States. IIES Working paper.

Jacoby, H. G., G. Li, and S. Rozelle (2002). Hazards of Expropriation: Tenure Security and Investment in Rural China. *American Economic Review 92*(5), 1420–1447.

Jacoby, H. G. and G. Mansuri (2009). Incentives, Supervision, and Sharecropper Productivity. *Journal of Development Economics 88*, 232–241.

Jakiela, P. and O. Ozier (2016). Does africa need a rotten kin theorem? experimental evidence from village economies. *The Review of Economic Studies 83*(1), 231–268.

Jensen, M. C. and W. H. Meckling (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics 3*(4), 305–360.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014). Agricultural Decisions after Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics 129*(2), 597.

Karlan, D. and J. Zinman (2009). Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. *Econometrica 77*(6), 1993–2008.

Khandker, S. R. and G. B. Koolwal (2014). Does Institutional Finance Matter for Agriculture? Evidence Using Panel Data from Uganda. Policy Research Working Paper 6942.

Laffont, J.-J. and M. Matoussi (1995). Moral Hazard, Financial Contraints and Sharecropping in El Oulja. *The Review of Economic Studies 62*(3), 381–399.

Lazaer, E. P. (2000). Performance Pay and Productivity. *The American Economic Review 90*(5), 1346–61.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies 76*(3), 1071–1102.

Lewis, W. A. (1955). *The Theory of Economic Growth*. Homewood, Illinois: RD Irwin.

Lobell, D. B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, and T. K. S. Murray (2018). Eyes in the Sky, Boots on the Ground. Policy Research Working Paper 8374.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Montero, E. (2018). Cooperative Property Rights and Development: Evidence from Land Reform in El Salvador. Working Paper, Harvard.

Mossin, J. (1968). Taxation and Risk-taking: An Expected Utility Approach. *Economica 35*, 74–82.

Nafula, J. (2008, October). Uganda: No Land for Women; the Attitude Behind Hunger.

Newbery, D. M. G. and J. E. Stiglitz (1979). *Risk, Uncertainty and Agricultural Development*, Chapter Sharecropping, Risk Sharing and the Importance of Imperfect Information. New York: Agricultural Development Council.

Otsuka, K., H. Chuma, and Y. Hayami (1992). Land and Labor Contracts in Agrarian Economies: Theories and Facts. *Journal of Economic Literature 30*(4), 1965–2018.

Otsuka, K. and Y. Hayami (1988). Theories of Share Tenancy: a Critical Survey. *Economic Development and Cultural Change 37*(1), 31–68.

Poterba, J. M. and A. A. Samwick (2003). Taxation and Household Portfolio Composition: US Evidence from the 1980s and 1990s. *Journal of Public Economics 87*(1), 5–38.

Pratt, J. W. (1964). Risk Aversion in the Small in the Large. *Econometrica 32*, 122–136.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature 37*(1), 7–63.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica 68*(5), 1281–1292.

Rao, C. H. H. (1971). Uncertainty, Entrepreneurship, and Sharecropping in India. *Journal of Political Economy 79*(3), 578–595.

Restuccia, D. and R. Santaeulalia-Llopis (2017). Land Misallocation and Productivity. NBER Working Paper 23128.

Rosenzweig, M. and C. Udry (2018). External Validity in a Stochastic World: Evidence from Low-Income Countries . Working paper.

Shaban, R. A. (1987). Testing between Competing Models of Sharecropping. *Journal of Political Economy 95*(5), 893–920.

Shetty, S. (1988). Limited Liability, Wealth Differences and Tenancy Contracts in Agrarian Economies. *Journal of Development Economics 29*(1), 1–22.

Shleifer, A. and R. Vishny (1998). *The Grabbing Hand*. Harvard University Press.

Squires, M. (2017). Kinship Taxation as a Constraint to Microenterprise Growth: Experimental Evidence from Kenya. Working paper.

Stiglitz, J. (1974). Incentives and Risk Sharing in Sharecropping. *Review of Economic Studies 41*(2), 219–255.

Stiglitz, J. E. (1969). The Effects of Income, Wealth, and Capital Gains Taxation on Risk-Taking. *The Quarterly Journal of Economics 83*(2), 263.

Udry, C. (1996). Gender, Agricultural Production, and the Theory of the Household. *Journal of Political Economy 104*(5), 1010–1046.

Valentiyi, A. and B. Herrendorf (2008). Measuring Factor Income Shares at the Sector Level. *Review of Economic Dynamics 11*(4), 820–835.

Tables

TABLE I: DESCRIPTIVES CHARACTERISTICS AND BALANCE TESTS

	(1)	(2)	(3)	(4)	(5)
		Difference between			
	C	T1-C	T2-C	T1 - T2	N
Panel A:			*Farmer Characteristics*		
Young (Age $\leq$ 21)	0.557 (0.500)	-0.044 [0.552]	0.027 [0.721]	0.071 [0.313]	262
Low Schooling ($\leq$ 8 years)	0.550 (0.501)	-0.028 [0.731]	0.005 [0.946]	0.033 [0.671]	265
School enrolment	0.089 (0.286)	-0.010 [0.823]	-0.038 [0.317]	-0.028 [0.519]	264
Raven test score (0-100)	51.54 (24.95)	2.88 [0.419]	5.02 [0.159]	2.13 [0.527]	269
Health status (0-10)	8.111 (1.643)	0.190 [0.391]	0.044 [0.865]	-0.146 [0.418]	269
Married	0.512 (0.503)	-0.004 [0.962]	-0.029 [0.718]	-0.026 [0.761]	268
Number of children	1.750 (1.710)	-0.197 [0.426]	-0.026 [0.906]	0.171 [0.484]	268
Labor income (USD)	29.3 (38.1)	3.8 [0.574]	-5.7 [0.270]	-9.5 [0.123]	264
Cash savings (USD)	122.2 (145.5)	-13.3 [0.523]	-7.9 [0.725]	5.4 [0.826]	266
Consumption (USD)	142.6 (91.2)	11.0 [0.495]	-17.3 [0.187]	-28.3 [0.062]*	261
Household size	5.346 (2.001)	-0.213 [0.488]	0.010 [0.970]	0.223 [0.431]	269
Household sex ratio	0.425 (0.208)	-0.041 [0.174]	-0.002 [0.957]	0.039 [0.212]	269
Household income (USD)	194.6 (171.9)	10.8 [0.666]	-3.6 [0.872]	-14.4 [0.550]	235
Household assets (USD)	1,506.9 (2,714.3)	-273.9 [0.480]	-518.4 [0.135]	-244.5 [0.409]	265
Agricultural tools (USD)	49.12 (33.04)	-6.76 [0.178]	-3.49 [0.422]	3.27 [0.499]	265
Panel B:			*Plot Characteristics*		
Distance to Market (km)	2.712 (4.629)	-0.452 [0.329]	-0.347 [0.477]	0.105 [0.756]	270
Soil type: loam	0.386 (0.490)	-0.018 [0.860]	0.083 [0.297]	0.101 [0.296]	270
Soil type: clay	0.084 (0.280)	-0.003 [0.937]	-0.038 [0.427]	-0.035 [0.388]	270
Soil type: sandy	0.108 (0.313)	-0.006 [0.897]	0.000 [0.990]	0.006 [0.875]	270
Soil type: rocky	0.048 (0.215)	-0.010 [0.754]	-0.031 [0.301]	-0.021 [0.474]	270
Slope: steep	0.084 (0.280)	-0.001 [0.983]	0.014 [0.752]	0.015 [0.642]	270
Slope: gentle	0.602 (0.492)	0.016 [0.816]	-0.037 [0.550]	-0.054 [0.390]	270
Slope: valley	0.072 (0.261)	-0.041 [0.352]	0.022 [0.649]	0.064 [0.137]	270
Slope: flat	0.241 (0.430)	0.026 [0.667]	0.001 [0.992]	-0.025 [0.674]	270
Latitude (degrees N)	0.361 (0.625)	0.014 [0.026]**	0.012 [0.061]*	-0.003 [0.692]	270
Longitude (degrees E)	32.168 (1.403)	0.000 [0.980]	-0.012 [0.100]	-0.012 [0.056]*	270
Total rainfall (dm)	5.096 (1.670)	0.040 [0.427]	0.020 [0.729]	-0.020 [0.743]	479

Notes: Column 1 shows the mean (and standard deviation in brackets) of each baseline characteristics in the control group. Columns 2 through 4 show differences in characteristics assigned to treatment and control groups. These are calculated from a regression a regression of the characteristic on dummy variables for treatment status, controlling for strata fixed effects. In square brackets we provide the randomization inference p-value of a test of the null hypothesis that C, $T1-C$, $T2-C$ and $T1-T2$ is equal to 0, respectively. Panel A presents tenant characteristics measured before the start of Season 1. All monetary values are in PPP USD. Panel B presents plot level characteristics which are unaffected by treatment, measured at the end of Season 1. Detailed variable definitions are provided in Section .

TABLE II: EFFECTS ON OUTPUT

	Output, y		Yield, y/m^2	
	(1)	(2)	(3)	(4)
High s (T1)	56.28***	56.07***	0.074**	0.073**
	(18.52)	(18.58)	(0.031)	(0.031)
	[0.004]	[0.004]	[0.024]	[0.027]
High w (T2)	5.36		-0.000	
	(17.17)		(0.030)	
	[0.765]		[0.995]	
High w, safe (T2A)		18.29		0.043
		(25.84)		(0.048)
		[0.543]		[0.403]
High w, risky (T2B)		-7.25		-0.043
		(15.82)		(0.032)
		[0.641]		[0.206]
H_0: T1 = T2	0.023		0.046	
H_0: T1 = T2A		0.218		0.590
H_0: T1 = T2B		0.001		0.002
H_0: T2A = T2B		0.343		0.120
Mean Outcome (C)	95.13	95.13	0.174	0.174
Observations	473	473	473	473

Notes: The table reports ordinary least square estimates based on specification (4) at the plot level, for both Season 1 and Season 2. *Output, y* is the expected output of the plot measured through the pre-harvest crop assessment survey. It is calculated by multiplying the expected quantity of output of each crop with the price of the relevant crop measured on local markets, and summing over crops. *Yield,* y/m^2 is the expected output of the plot divided by the area (in square meters) cultivated. Values are in PPP USD. T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. T2A and T2B indicate subgroups of treatment group 2 (T2). T2A received a fixed income transfer, and T2B received a stochastic income transfer, with mean equal to T2A. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. In square brackets randomization inference p-values of the null hypothesis of no effect are provided; *** (**) (*) indicates significance of that test at the 1% (5%) (10%) level. Additionally randomization inference p-values for the specified compound hypotheses are reported.

TABLE III: EFFECTS ON CAPITAL INPUTS

	Fertilizer	Insecticide	Tools	Index
	(1)	(2)	(3)	(4)
Panel A:	*Extensive Margin*			
High s (T1)	0.094	-0.010	0.086	0.201
	(0.061)	(0.053)	(0.055)	(0.133)
	[0.176]	[0.860]	[0.123]	[0.162]
High y (T2)	0.027	-0.064	0.007	-0.049
	(0.060)	(0.055)	(0.053)	(0.140)
	[0.690]	[0.261]	[0.901]	[0.739]
Within-Equation Test				
H_0: T1 = T2	0.310	0.320	0.142	0.080
Cross-Equations Test				
H_0: T1 = 0		0.283		-
H_0: T2 = 0		0.594		-
H_0: T1 = T2		0.375		-
Mean Outcome (C)	0.277	0.276	0.500	0.000
Observations	432	423	432	423
Panel B:	*Intensive Margin (USD)*			
High s (T1)	1.13*	0.43	11.36**	0.436***
	(0.55)	(0.51)	(5.04)	(0.153)
	[0.056]	[0.416]	[0.039]	[0.008]
High y (T2)	0.59	-0.50	1.59	0.029
	(0.43)	(0.47)	(4.32)	(0.126)
	[0.205]	[0.282]	[0.727]	[0.808]
Within-Equation Test				
H_0: T1 = T2	0.350	0.046	0.059	0.008
Cross-Equations Test				
H_0: T1 = 0		0.039		-
H_0: T2 = 0		0.274		-
H_0: T1 = T2		0.044		-
Mean Outcome (C)	0.96	1.81	37.81	0.000
Observations	419	413	427	402

Notes: The table reports ordinary least square estimates based on specification (4). T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. In square brackets randomization inference p-values of the null hypothesis of no effect are provided; *** (**) (*) indicates significance of that test at the 1% (5%) (10%) level. Cross-Equations Tests report the randomization inference p-value for a test of the hypothesis that the specified restriction holds in all estimating equations across columns. Within-Equation Tests report the randomization inference p-value for a test of the specified compound hypothesis. In Panel A, "Fertilizer (Insecticide)" is a dummy variable equal to 1 if the tenant said she used fertilizer (insecticide) on her plot during the past season; "Tools" is a dummy variable equal to 1 if the tenant said she bought agricultural tools to cultivate her plot. In Panel B, the dependent variable is the monetary value of the input used in PPP USD terms. For agricultural tools, the intensive margin is the value of agricultural tools owned by the respondent's household at the time of the survey. The "Index" combines the four indicators by first standardizing each outcome into a z-score (by subtracting the control group mean at the corresponding survey round and dividing by the control group standard deviation), then takes the average of the z-scores, and again standardizes to the control group.

TABLE IV: EFFECTS ON LABOR INPUTS

	Own labor	Paid	Unpaid	Index
	(hours/week)	*(days/season)*		
	(1)	(2)	(3)	(4)
High s (T1)	0.34	-0.05	8.02*	0.20
	(1.28)	(1.98)	(4.03)	(0.12)
	[0.781]	[0.982]	[0.065]	[0.157]
High y (T2)	-0.03	1.06	1.79	0.05
	(1.22)	(2.08)	(3.31)	(0.12)
	[0.984]	[0.628]	[0.626]	[0.721]
Within-Equation Test				
H_0: T1 = T2	0.783	0.550	0.173	0.280
Cross-Equations Test				
H_0: T1 = 0		0.277		-
H_0: T2 = 0		0.909		-
H_0: T1 = T2		0.575		-
Mean Outcome (C)	17.13	4.28	12.54	-0.00
Observations	417	432	432	417

Notes: The table reports ordinary least square estimates based on specification (4). T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. In square brackets randomization inference p-values of the null hypothesis of no effect are provided; *** (**) (*) indicates significance of that test at the 1% (5%) (10%) level. Cross-Equations Tests report the randomization inference p-value for a test of the hypothesis that the specified restriction holds in all estimating equations across columns. Within-Equation Tests report the randomization inference p-value for a test of the specified compound hypothesis. "Own labor" is the number of hours that the tenant said she worked on the plot in a typical week during the past season. The dependent variables in columns 2 and 3 are the number of worker-days of paid and unpaid labor respectively that the tenant said she had working on the plot for throughout the season. The "Index" combines the three indicators by first standardizing each outcome into a z-score (by subtracting the control group mean at the corresponding survey round and dividing by the control group standard deviation), then takes the average of the z-scores, and again standardizes to the control group.

TABLE V: EFFECTS ON CROP CHOICE

	Maize (1)	Beans (2)	Peanuts (3)	Tomatoes (4)	Potatoes (5)
Panel A:			*Extensive Margin*		
High s (T1)	0.112**	0.049	0.055	0.021***	0.012
	(0.047)	(0.042)	(0.040)	(0.010)	(0.008)
	[0.025]	[0.253]	[0.212]	[0.008]	[0.201]
High w (T2)	0.090*	0.032	0.049	-0.001	0.002
	(0.048)	(0.041)	(0.038)	(0.004)	(0.003)
	[0.084]	[0.447]	[0.239]	[0.805]	[0.686]
H_0: T1 = T2	0.652	0.720	0.899	0.013	0.217
Mean Outcome (C)	0.620	0.300	0.327	0.000	0.000
Observations	479	479	479	479	479
Panel B:			*Intensive Margin: Number of Plants*		
High s (T1)	159.82	4.53	330.43	41.02**	3.40
	(145.70)	(391.33)	(179.11)	(19.14)	(2.85)
	[0.295]	[0.994]	[0.128]	[0.020]	[0.318]
High w (T2)	-66.01	-85.58	-39.70	1.48	0.67
	(131.88)	(362.02)	(154.24)	(10.48)	(1.31)
	[0.635]	[0.841]	[0.818]	[0.912]	[0.841]
H_0: T1 = T2	0.147	0.760	0.094	0.013	0.205
Mean Outcome (C)	861.96	867.83	577.09	0.00	0.00
Observations	479	479	479	479	479
Panel C:			*Intensive Margin: Value of Output*		
High s (T1)	4.51	5.40	32.77***	7.67*	0.27
	(4.85)	(6.17)	(11.04)	(4.23)	(0.24)
	[0.384]	[0.389]	[0.003]	[0.051]	[0.447]
High w (T2)	-2.43	1.78	4.72	-0.25	0.05
	(4.40)	(6.84)	(9.38)	(1.89)	(0.11)
	[0.591]	[0.820]	[0.655]	[0.917]	[0.814]
H_0: T1 = T2	0.152	0.613	0.065	0.074	0.318
Mean Outcome (C)	28.43	15.78	22.44	0.00	0.00
Observations	479	479	479	479	479

Notes: The table reports ordinary least square estimates based on specification (4). T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. In square brackets randomization inference p-values of the null hypothesis of no effect are provided; *** (**) (*) indicates significance of that test at the 1% (5%) (10%) level. Additionally the randomization inference p-value of a test of the null hypothesis that the effect of T1 and T2 are equal is provided for all estimating equations. Dependent variables in Panel A are dummy variables equal to 1 if at the time of the pre-harvest crop assessment survey, any harvestable plants of the specified crop were observed on the plot: maize in column (1), beans in column (2), peanuts in column (3), tomatoes in column (4), and potatoes in column (5). In Panel B, the dependent variable is the number of plants of the relevant crop; and in Panel C, the dependent variable is the output value from the relevant crop on the plot – calculated by multiplying the quantity of output of each crop with the price of the relevant crop measured on local markets. All monetary values are in PPP USD.

TABLE VI: SOCIO-ECONOMIC STATUS

	Labor income (1)	Consumpt. (2)	Cash savings (3)	Household income (4)	Household assets (5)
High s (T1)	4.07	4.43	56.83	33.04*	656.54*
	(7.33)	(9.60)	(35.39)	(18.34)	(332.13)
	[0.626]	[0.678]	[0.127]	[0.076]	[0.060]
High w (T2)	14.98*	-3.98	66.12	0.49	183.46
	(8.35)	(7.84)	(39.27)	(18.04)	(209.29)
	[0.086]	[0.652]	[0.102]	[0.982]	[0.396]
H_0: T1 = T2	0.214	0.372	0.852	0.064	0.164
Mean Outcome (C)	36.65	115.34	143.63	181.80	1242.61
Observations	424	421	427	398	427

Notes: The table reports ordinary least square estimates based on specification (4). T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. In square brackets randomization inference *p*-values of the null hypothesis of no effect are provided; *** (**) (*) indicates significance of that test at the 1% (5%) (10%) level. Additionally the randomization inference *p*-value of a test of the null hypothesis that the effect of T1 and T2 are equal is provided for all estimating equations. All monetary values are in PPP USD terms. "Labor income" is the average monthly labor income of the respondent during the 12 months preceding the survey. "Consumption" is the monthly consumption expenditure of the respondent; it is calculated by adding her monthly personal consumption on non-food items and services with her household's per-capita food consumption where monthly food consumption is imputed from previous 2 days' recall. "Cash savings" is the value of savings that the respondent had at the time of the survey. "Household income" is the response to the question "What is the total income of your household in a typical month?". "Household assets" is the value of durable assets owned by the respondent's household.

TABLE VII: SOIL QUALITY AT THE END OF THE EXPERIMENT

	N	K	P	Org. M.	Ph	Index
	(1)	(2)	(3)	(4)	(5)	(6)
High s (T1)	-0.11	-0.00	0.06	-0.06	0.05	-0.04
	(0.08)	(0.05)	(0.11)	(0.09)	(0.12)	(0.13)
	[0.216]	[0.975]	[0.598]	[0.515]	[0.685]	[0.793]
High w (T2)	-0.00	-0.02	0.10	0.01	-0.01	0.07
	(0.08)	(0.05)	(0.11)	(0.09)	(0.12)	(0.12)
	[0.993]	[0.711]	[0.369]	[0.912]	[0.917]	[0.574]
Within-Equation Test						
H_0: T1 = T2	0.185	0.760	0.779	0.476	0.592	0.441
Cross-Equations Test						
H_0: T1 = 0			0.711			-
H_0: T2 = 0			0.959			-
H_0: T1 = T2			0.797			-
Mean Outcome (C)	2.29	0.77	2.33	2.11	5.21	-0.00
Observations	324	322	323	321	324	318

Notes: The table reports ordinary least square estimates based on specification (4). T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. In square brackets randomization inference p-values of the null hypothesis of no effect are provided; *** (**) (*) indicates significance of that test at the 1% (5%) (10%) level. Cross-Equations Tests report the randomization inference p-value for a test of the hypothesis that the specified restriction holds in all estimating equations across columns. Within-Equation Tests report the randomization inference p-value for a test of the specified compound hypothesis. The dependent variables are the results of soil tests conducted on sampled of soil taken from the plots that were part of the experiment. For Nitrogen (N) the index is: 1=lack, 2=inadequate, 3=adequate; for Potassium (K): 0=deficient, 1=sufficient; for Organic Matter: 1=low, 2=high, 3=very high; for Phosphorous (P): 1=very low, 2=moderate, 3=adequate, 4=high. The Ph-level variable reports the ph level of the soil sample. The "Index" combines the five indicators by first standardizing each outcome into a z-score (by subtracting the control group mean at the corresponding survey round and dividing by the control group standard deviation), then takes the average of the z-scores, and again standardizes to the control group.

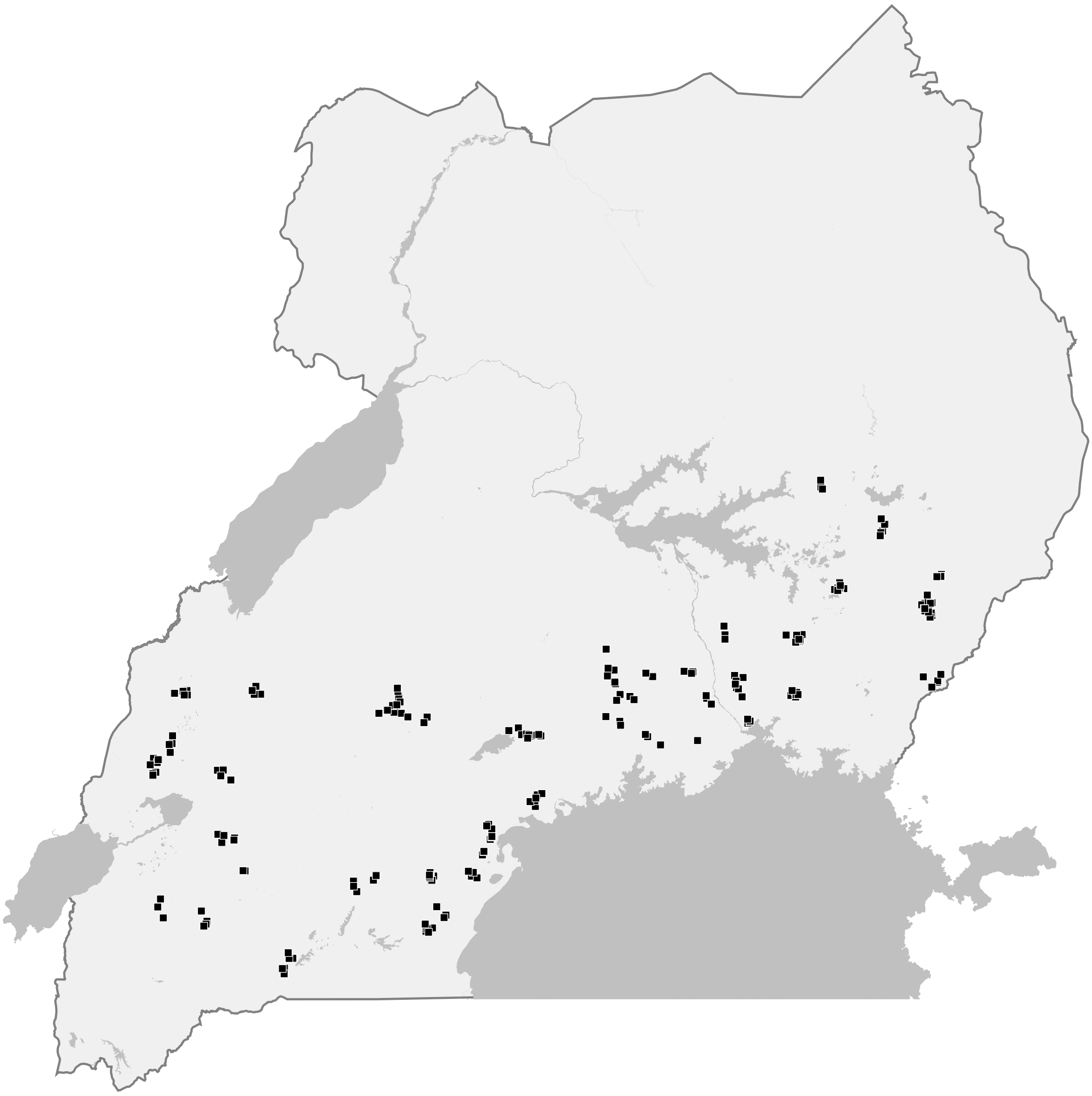
TABLE VIII: EFFECTS ON INPUT AND OUTPUT VALUES

	Capital (1)	Labor hours (2)	Land size (3)	Output (4)
Panel A:		*In Levels*		
High s (T1)	12.42**	72.94*	71.37	56.28***
	(5.08)	(38.34)	(59.95)	(18.52)
	[0.027]	[0.086]	[0.277]	[0.004]
High w (T2)	2.18	14.91	31.17	5.36
	(4.29)	(34.32)	(57.09)	(17.17)
	[0.646]	[0.686]	[0.639]	[0.765]
H_0: T1 = T2	0.048	0.167	0.481	0.023
Mean Outcome (C)	39.90	338.68	607.13	95.13
Observations	432	417	473	473
Panel B:		*In Logs*		
High s (T1)	0.20	0.10	0.29**	0.38**
	(0.12)	(0.10)	(0.13)	(0.17)
	[0.123]	[0.320]	[0.040]	[0.039]
High w (T2)	0.04	0.01	0.13	0.11
	(0.12)	(0.09)	(0.14)	(0.16)
	[0.751]	[0.944]	[0.395]	[0.536]
H_0: T1 = T2	0.204	0.365	0.199	0.122
Mean Outcome (C)	3.40	5.53	5.81	3.53
Observations	432	417	473	473

Notes: T1 is a dummy variable equal to 1 if the tenant/plot was randomized to receive high (75%) output share, T2 is a dummy variable equal to 1 if the tenant/plot was randomized to receive same output share as control (50%) and an additional cash transfer. All specifications control for strata fixed effects. Standard errors are clustered at the village level and given in round brackets. "Capital" is the monetary value (in PPP USD terms) of capital inputs used on the plot, obtained by summing up the values of fertilizer, insecticide and households tools. "Labor hours" is the total hours of labor used on the plot during each season, obtained by summing respondent's labor hours (hours worked in typical week during the season multiplied by 12 weeks/season) and hours of hired labor (numbers of days of hired labor used during the season multiplied by 8 hours/day). "Land size" is the size (in m^2) of the plot area cultivated by the tenant. "Output" is the monetary value of total output (in PPP USD terms) of the plot measured through the pre-harvest crop assessment survey (see notes to Table II for further details on this variable). In Panel B, all dependent variables are the natural logarithm of the value of the relevant variable.

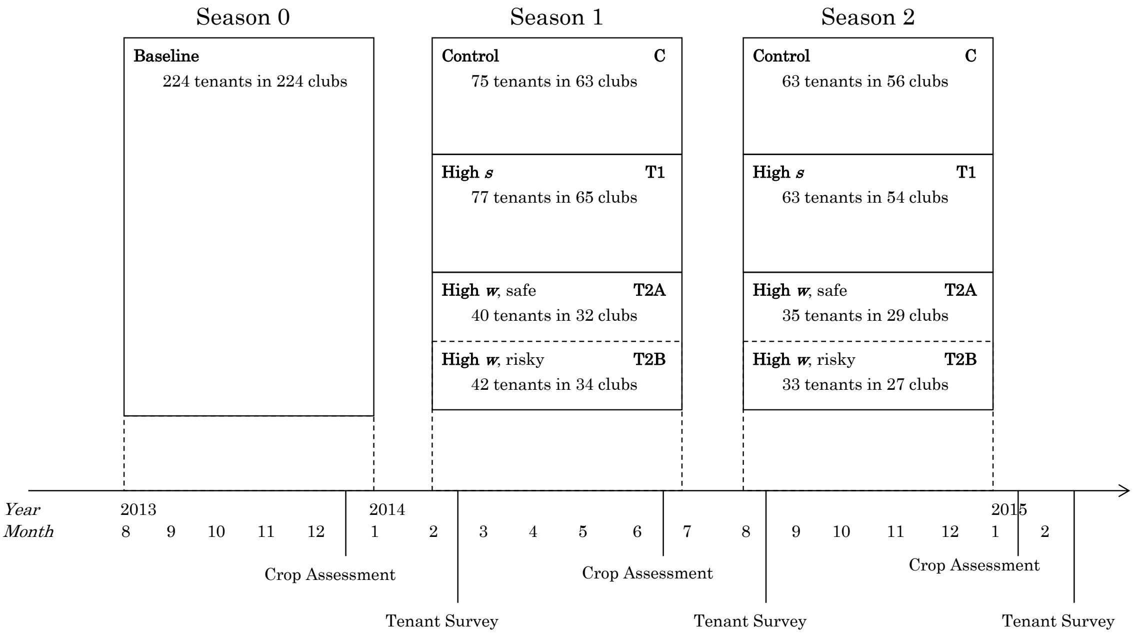
Figures

FIGURE I: LOCATION OF THE PLOTS



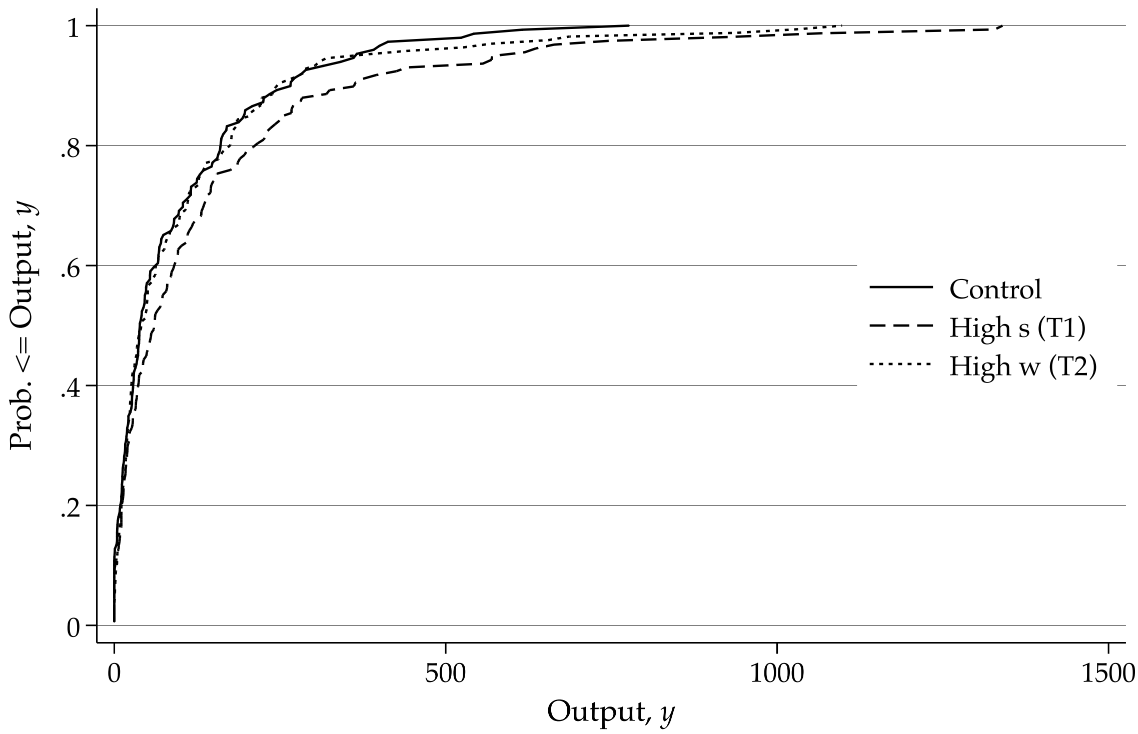
Notes: This map of Uganda shows as black squares the location of BRAC clubs whose farmers participated in the experiment and are covered in the plot-level analysis.

FIGURE II: EXPERIMENT SETUP



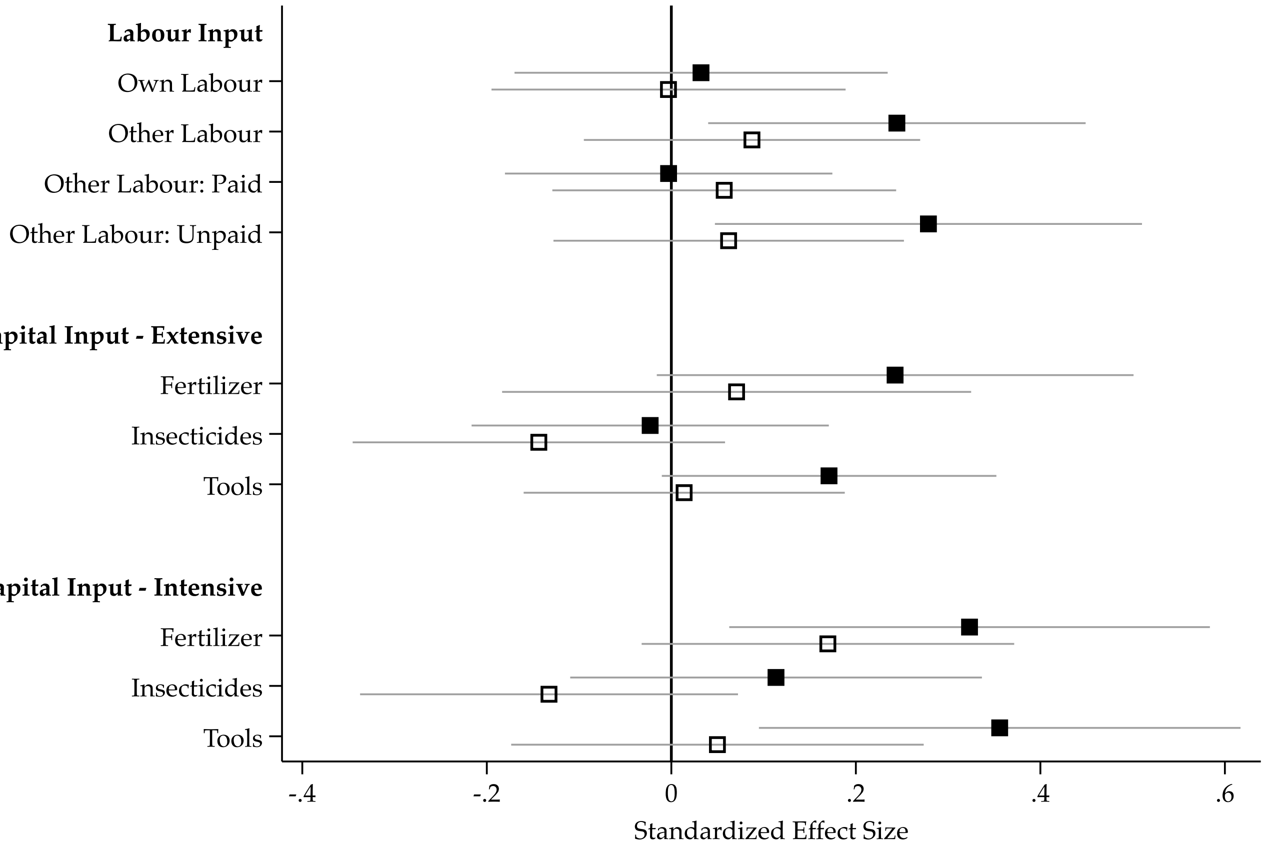
Notes: The figure shows the number of farmers and clubs included in the analysis in seasons 0, 1 and 2 by treatment assignment.

FIGURE III: CONTRACTS AND DISTRIBUTION OF *Output*, y



Notes: The figure plots the cumulative distribution function of expected output from the plots, by treatment status. Tenants in T1 are those who were randomized to receive high (75%) output share, tenants in T2 received the same output share as control C (50%) and an additional cash transfer. *Output*, y is the expected output of the plot measured through the pre-harvest crop assessment survey. It is calculated by multiplying the expected quantity of output of each crop with the price of the relevant crop measured on local markets, and summing over crops. Values are in PPP USD.

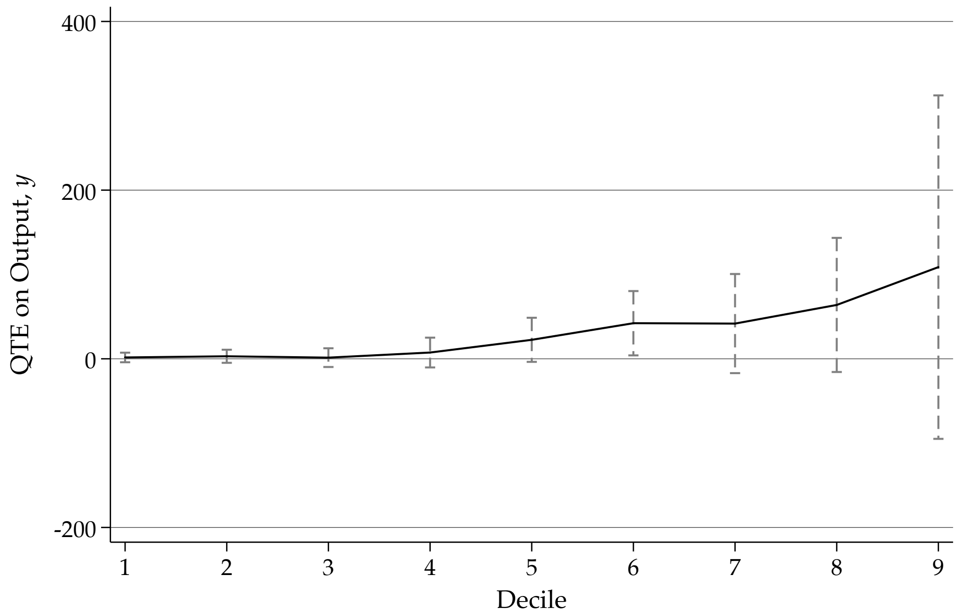
FIGURE IV: CONTRACTS AND INPUT CHOICE



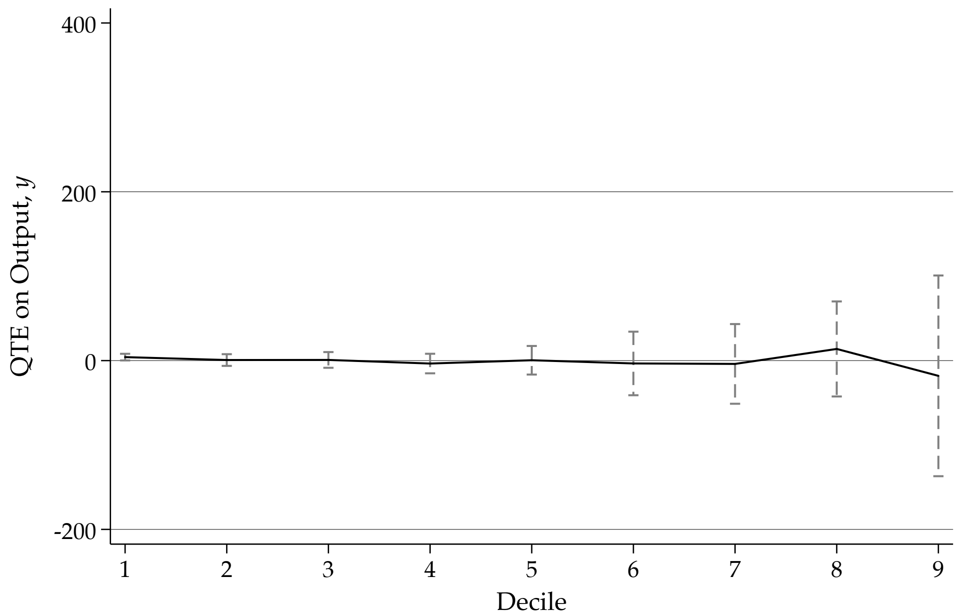
Notes: The figure plots the standardized effect sizes and 95% confidence intervals for labor and capital inputs used for cultivation. The solid squares show the effects of being selected to receive high (75%) output share (T1) relative to the control group, while the hollow squares show the effect of receiving the same output share as the control group (50%) plus an additional cash transfer (T2). The effects are estimated using ordinary least square estimates based on specification (4). All specifications control for strata fixed effects and standard errors are clustered at the village level. For capital inputs, the extensive margins correspond to dummy variables equal to 1 if the tenant used any fertilizer; any insecticide; if she bought any agricultural tools to cultivate her plot. The intensive margins are the monetary value (in PPP USD) of the inputs used on the plot. For tools, the intensive margin gives the value of agricultural tools that the tenant had at the time of the survey.

FIGURE V: HETEROGENEITY OF OUTPUT EFFECTS

(A) QUANTILE TREATMENT EFFECTS OF T1 VS. C

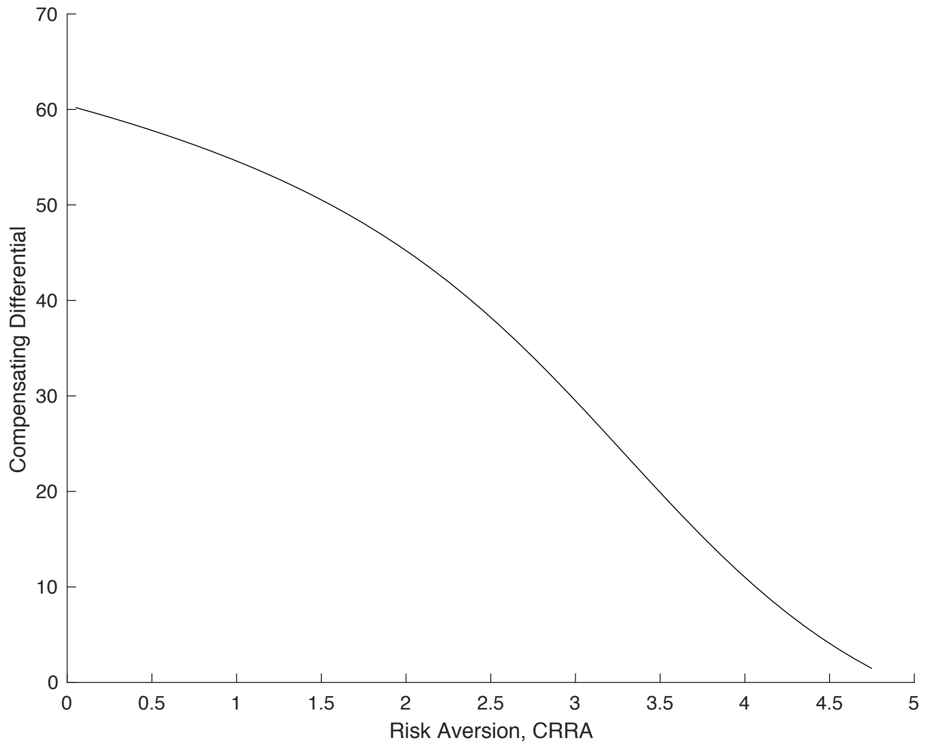


(B) QUANTILE TREATMENT EFFECTS OF T2 VS. C



Notes: The figure plots quantile treatment effect (QTE) estimates for *Output*, y and 90% confidence intervals based on bootstrapped (with 500 replications) standard errors clustered at the village level (unit of randomization). Each specification controls for the randomization strata. *Output*, y is the expected output of the plot measured through the pre-harvest crop assessment survey. It is calculated by multiplying the expected quantity of output of each crop with the price of the relevant crop measured on local markets, and summing over crops. Values are in PPP USD.

FIGURE VI: WELFARE – COMPENSATING DIFFERENTIAL C VS. T1



Notes: The graph shows the fixed amount (in PPP USD) that a tenants in control (C) would need to receive per season to be as well-off as tenants in the high share s treatment (T1), for a range of risk aversion levels. We calculate a distribution of potential income levels for each experimental plot in C and T1 (see Section V.C for details). We assume preferences are characterized by the iso-elastic utility function $u(c) = (c^{1-\eta} - 1)/(1 - \eta)$, where η is the constant coefficient of relative risk aversion, CRRA, shown on the x-axis above. Given any η, we find the certainty equivalent as the amount e of income such that tenants are, on average, indifferent between the income stream of tenants in C plus e, and the income stream in T1.