Risk-Sharing and Land Misallocation

Pietrobon, Davide; Ruggieri, Alessandro

2024

Document Version:
Other version

Link to publication

Citation for published version (APA):
Pietrobon, D., & Ruggieri, A. (2024). *Risk-Sharing and Land Misallocation*. (Working Papers; No. 2024:7).

Total number of authors:
2

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2024:7

Department of Economics
School of Economics and Management

Risk-Sharing and Land Misallocation

Davide Pietrobon
Alessandro Ruggieri

Sep 2024

Risk-Sharing and Land Misallocation[*]

Davide Pietrobon
Lund University

Alessandro Ruggieri
CUNEF Universidad

September 18, 2024

Abstract

We study the impact of incomplete consumption risk-sharing on land misallocation in rural economies. We develop a general equilibrium model of land cultivation choices, where heterogeneous agricultural households face idiosyncratic output shocks and insure themselves by participating in a risk-sharing arrangement. Incomplete insurance distorts households' choices, leading them away from maximizing expected incomes and resulting in land misallocation. Using the latest ICRISAT panel data from rural India, we quantify the losses attributable to limited risk-sharing. Completing insurance markets leads to output and welfare gains of 19% and 29%, respectively. Improving the functioning of consumption insurance markets within an otherwise undistorted economy can yield gains comparable to those achieved by removing distortions in factor markets.

Keywords: Misallocation, risk-sharing, agriculture, productivity, welfare
JEL Classification: O11, D61, Q12, D52

[*]Pietrobon: Department of Economics, Lund University, Tycho Brahes väg 1, Lund, Sweden, d.pietrobon91@gmail.com. Ruggieri: Department of Economics, CUNEF Universidad, Calle Leonardo Prieto Castro 2, Madrid, Spain, alessandro.ruggieri@cunef.edu. We thank Giacomo De Giorgi, Andre Groeger, Maia Güell, Nezih Guner, Jonna Olsson, Nicola Pavoni, Sevi Rodríguez Mora, and Raul Santaeulalia-Llopis for their useful comments. The usual disclaimers apply.

1 Introduction

Markets in the developing world are plagued by several frictions, and insurance markets are no exception to this general rule. Barriers to risk-sharing affect the allocation of consumption (Townsend, 1994; Udry, 1994; Fafchamps, 2011), production choices (Benjamin, 1992), migration decisions (Morten, 2019), and engagement in the non-farm sector (De Giorgi et al., 2024). Incomplete insurance also diminishes the incentives for acquiring risky inputs—an argument originally advanced by Arrow (1971), lowering productivity and increasing consumption inequality among farmers in developing countries (Donovan, 2021).

In this paper, we show that imperfections in insurance markets, taking the form of incomplete consumption risk-sharing, have implications for the allocative efficiency of land among farmers, resulting in large output and welfare losses. A growing literature highlights how farm-specific distortions contribute to factor misallocation in agriculture (Adamopoulos and Restuccia, 2014; Adamopoulos et al., 2022).[1] In the presence of limited consumption risk-sharing, land misallocation can exist even in an economy with otherwise *undistorted* markets. Our quantitative findings for rural India imply that completing insurance markets can substantially reduce land misallocation, leading to output and welfare gains of 19% and 29%, respectively.

In village economies, shocks from harvest failures, price shifts, illness, and pests leave households vulnerable to severe hardship. Insurance against idiosyncratic income fluctuations often relies on informal arrangements like gift exchanges and personal loans.[2] The literature indicates that imperfections in these arrangements are pervasive: households are unable to fully insure against idiosyncratic risks (Townsend, 1994; Udry, 1994; Fafchamps, 2011). This lack of insurance not only affects consumption but can also have distortionary effects on the allocation of factors of production (Foster and Rosenzweig, 2010; Donovan, 2021). Building on this evidence, we study the impact that limited insurance has on *land* markets, focusing on allocative efficiency—the potential to redistribute cultivated plots among farms thereby increasing overall agricultural yields.

To illustrate our line of reasoning, consider a village economy with a fixed supply of land that is bought and sold (or rented in and out) in an undistorted, competitive market. Under full insurance, household-farms' production decisions are separable from their consumption, ensuring that, in equilibrium, each household chooses how much land

[1] The literature highlights the impact of various factors, including inheritance laws, tax and subsidies, and land and tenancy regulations, as potential sources of distortions to the allocation of land across farms in developing countries. See Subsection 1.1 for a review.

[2] See Dercon (2002) for a review of the source of idiosyncratic income risks and coping strategies in rural economies.

to cultivate to maximize its expected profits. Expected profit maximization drives each farmer to the familiar condition of equating the expected marginal product of land to its price. Thus, the expected marginal products of land are equalized across farms, resulting in an efficient allocation of land and maximal aggregate expected output. Under incomplete insurance, the "separation property" breaks apart: households' equilibrium land choices are not generally characterized by an expected profit maximization condition, which prevents the equalization of the expected marginal products of land across farms. Thus, imperfections in insurance markets lead to land misallocation and lower aggregate output.

We outline a general equilibrium model of risk-sharing in which household-farms with heterogeneous productivities insure against idiosyncratic output shocks by sharing the incomes they generate from operating their farms. Each farmer chooses how much land to buy before the shocks are realized. We characterize the equilibrium land allocation across a range of risk-sharing levels, ranging from full to no insurance. Besides decreasing the expected utility of buying land, lower insurance weakens the link between farm productivity and land holdings. Under full insurance, each household-farm's equilibrium land choices maximize its expected income, making it impossible to redistribute land from one farmer to another without lowering aggregate expected income. Incomplete insurance distorts these choices away from expected income maximization by increasing the weights that households attach to states of the world in which income is low. These distortions imply that, in equilibrium, the expected marginal products of land are not equalized across households—i.e., land is misallocated.

We test our model using the latest ICRISAT monthly panel data (2009–2014) from the Indian semi-arid tropics, offering evidence linking risk-sharing to land misallocation. First, we measure risk-sharing across villages and time by estimating the elasticity of household consumption with respect to idiosyncratic income shocks for each village and year. We find evidence of limited consumption insurance in rural India, consistent with the literature: on average, 22.5% of idiosyncratic income fluctuations are passed through to consumption. Second, we quantify land market misallocation in each village and year using two metrics of factor misallocation: the correlation between total land cultivated and household-farm physical productivity, and the variance of the marginal product of land for each village-year pair. The correlation between productivity and land cultivated is a well-known measure of allocative efficiency in the land market, where a higher correlation indicates that more productive farms cultivate more land, on average (Chen et al., 2023). The variance of the marginal product of land quantifies the deviation from a benchmark scenario in which land's marginal products are equalized across farms, indicative of an efficient allocation of land to production units (Restuccia and Rogerson, 2017). Consistent with our theory, we find a significant positive correlation between risk-sharing and the correlation between productivity and land cultivated, and a significant negative correlation between risk-sharing and the variance of the marginal product of land.

We then leverage the structure of our model to quantify output and welfare gains resulting from improving the functioning of consumption insurance markets in village economies. While parsimonious, our model successfully replicates the negative correlation between risk-sharing and land misallocation as an untargeted moment. Armed with the structural estimates, we conduct a counterfactual analysis to explore the impact of improving consumption insurance markets in Indian villages. We examine how these improvements affect the functioning of land markets and contribute to gains in aggregate output and welfare. Completing insurance markets leads to output and welfare gains of 19%, and 29%, respectively. Agricultural productivity (yields per unit of land) increases by 45% under full insurance. This figure is comparable to other estimates in the literature: e.g., Adamopoulos et al. (2022) finds that eliminating farm-specific distortions coming from land market frictions in rural China average farm productivity by 53%. These results are robust to incorporating both imperfect risk-sharing and household-specific distortions in the output market: we find that the aggregate output and efficiency gains from completing insurance remain virtually unchanged even when distortions are modeled. Our counterfactual exercise allows us to conclude that imperfections in *consumption insurance* markets can be as important as farm-specific distortions in explaining the gains from reallocating inputs across farms in developing countries.

1.1 Related literature

Our paper belongs to the growing literature on misallocation of inputs in agriculture. Gollin et al. (2002) and Restuccia et al. (2008) emphasize the role of the agricultural sector in economic development and its importance in explaining cross-country productivity and income differences. The broad theme of factor misallocation and its influence on cross-country productivity differences is explored in Restuccia and Rogerson (2008), Restuccia and Rogerson (2013), and Restuccia and Rogerson (2017). Chen et al. (2023) find that capital and operational land size are essentially unrelated to farm productivity in Malawi, implying the existence of misallocation in the land market. Chen et al. (2022) show how land rental market imperfections in Ethiopia lead to land misallocation, highlighting the output and welfare gains from land certification reforms. Acampora et al. (2022) provide experimental evidence that cultivation rights decrease land misallocation in Kenya. Adamopoulos et al. (2022) argue that within-village frictions in the capital and land markets, linked to land institutions, disproportionately constrain productive farmers in rural China.[3] This body of literature generally explains factor misallocation as a consequence of generic distortions in input or output markets (i.e., "wedges"), or institutions that constrain the choices of productive firms. Our research enriches this narrative by introducing a potential micro-foundation for these wedges, highlighting the role of incomplete consumption insurance markets.

Deviations from perfect risk-sharing within village economies are well documented (see Townsend (1994), Udry (1994), and Fafchamps (2011), among others). A body of work has provided several explanations for imperfect risk-sharing, rationalizing them as consequences of primitive frictions such as action unverifiability (Ligon, 1998), limited commitment (Ligon et al., 2002), hidden income (Kinnan, 2021), and localized information constraints (Ambrus et al., 2022). The degree of risk-sharing within a village can affect several aspects of its economy. Morten (2019) studies the interaction between informal insurance and temporary migration as a self-insurance strategy. Mazur (2023) explores the relationship between risk-sharing and irrigation investments. Pietrobon (2024) examines how informal insurance impacts fertilizer use when fertilizer is risky and effort-complementary, and insurance may crowd out effort supply. Donovan (2021) examines the impact of insurance on the use of agricultural intermediates, and suggests that completing financial markets allows farmers to invest in risky inputs, leading to significant increases in labor productivity and input share. We build upon a mechanism similar to the one in Donovan (2021) to argue that the lack of insurance might distort farmers' land cultivation choices. However, our emphasis is distinct: rather than focusing on how imperfect insurance can decrease investments in land, we highlight how these imperfections result in land misallocation.

Finally, our paper contributes to the understanding of how land gets allocated to farmers in developing countries. In the semi-arid tropics of India, land markets exhibit a rich diversity, with many farm households engaging in buying or selling of land, or participating in the land rental market to some extent (see Ray (1998), Chapter 12). The salience of rental markets is emphasized in the literature on sharecropping practices (e.g., Lamb (2003)). Our research intersects with this topic by exploring the interaction between imperfect consumption insurance and input allocation in the land market.

[3] Misallocation of inputs in agriculture extends beyond the markets for capital and land: for example, Adamopoulos and Restuccia (2022) estimate substantial aggregate productivity gains from the spatial reallocation of crop production.

2 Model

We analyze a static economy in which households with heterogeneous productivities face idiosyncratic output shocks and can insure against these shocks by only relying on a risk-sharing arrangement.[4] Each household operates a farm and decides how much land to purchase for cultivation before its output shock is realized. This choice affects the distribution of the income generated by the farm, which is calculated as the value of agricultural output net of the cost of acquiring land. The risk-sharing pool allows households to share their incomes to hedge against the idiosyncratic output shocks. We use this model to illustrate how the degree of risk-sharing affects misallocation in the land market. In Appendix A, we briefly discuss some modeling choices. Appendix B contains all the proofs. In the discussion that follows, when we refer to households, we specifically mean agricultural households that also operate a farm.

Consider a static economy populated by a unit measure of household *types* indexed by i. For each type i, there is a unit mass of ex-ante identical households. Each household type i is characterized by a productivity level, θ_i, and initial land holdings, $\tilde{\ell}_i$. Let the total quantity of land available in the economy be $L = \int \tilde{\ell}_i \mathrm{d}i$.[5] Households have identical preferences over consumption, represented by a constant-relative risk averse (CRRA) utility function with a coefficient of relative risk aversion σ. Households of the same type, which are ex-ante identical, differ ex-post only with respect to the realization of an idiosyncratic (household-specific) output shock, ρ, drawn from a cumulative distribution function $Q_\rho(\rho)$ and support on some interval $\left[\underline{\rho}, \overline{\rho}\right] \subset \mathbb{R}_{++}$. We assume that $\underline{\rho}$ is high enough so that household consumption is bounded away from zero.

A household of type i produces output according to the following decreasing-returns-to-scale production function:

$$y_{i\rho} = \theta_i \rho \ell_i^\alpha,$$

where ℓ_i is land cultivated by a household of type i,[6] and $\alpha \in (0,1)$ denotes the land share—the elasticity of agricultural yields with respect to land cultivated.[7] Let r be the

[4] We abstract from modeling borrowing or savings decisions, which are alternative means for households to self-insure. In the context of rural India, households' ability to borrow and save appears to be heavily constrained (Rosenzweig and Wolpin, 1993).

[5] With a fixed land supply, risk-sharing does not affect aggregate equilibrium land cultivation, as the land price adjusts to balance supply and demand. In this way, we can theoretically isolate the effect of risk-sharing on land misallocation.

[6] Since households of the same type are ex-ante identical and make land cultivation choices before the output shocks are realized, these choices are identical for all households of the same type. Hence, referring to ℓ_i as the land cultivated by a household of type i is unambiguous.

[7] Our model can be extended to include an output shock with both aggregate and idiosyncratic components: since all households receive the same aggregate shock, they can engage in risk-sharing to fully

price of land, and let $\pi_{i\rho} = \theta_i \rho \ell_i^{\alpha} - r\left(\ell_i - \tilde{\ell}_i\right)$ denote the income of a household of type i under output shock realization ρ. All the land cultivated by each household is bought and sold in a land market before the output shocks are realized. Unlike most of the misallocation literature, we assume this market to be competitive and frictionless. Additionally, there are no distortions or "taxes" in the output market. These assumptions enable us to distinguish our findings from the more conventional narratives in the misallocation literature, which typically attribute land misallocation to land or output market distortions.

2.1 Full insurance vs. no sharing

To isolate the impact of insurance on land misallocation, we begin by comparing an economy with complete markets (full insurance) to one in which households are hand-to-mouth (no sharing). Starting from the former, let $c_i(\boldsymbol{\rho})$ denote the consumption of a household of type i when the state of the world is $\boldsymbol{\rho}$, where $\boldsymbol{\rho}$ represents the collection of realizations of the output shock for each household in the economy, drawn from the joint cumulative distribution function $Q_{\boldsymbol{\rho}}(\boldsymbol{\rho})$.[8] Moreover, let $\boldsymbol{c}(\boldsymbol{\rho}) = (c_i(\boldsymbol{\rho}))_i$ and $\boldsymbol{\ell} = (\ell_i)_i$ represent the collections of consumptions (under state of the world $\boldsymbol{\rho}$) and land cultivation choices of all household types. To characterize an allocation of resources under complete markets, we solve the following planner's problem for a given collection of type-specific Pareto weights $(\nu_i)_i$:

$$\max_{(\boldsymbol{c}(\boldsymbol{\rho}))_{\boldsymbol{\rho}}, \boldsymbol{\ell}} \int \nu_i \int \frac{(c_i(\boldsymbol{\rho}))^{1-\sigma}}{1-\sigma} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho})\, \mathrm{d}i,$$

subject to the land availability constraint

$$\int \ell_i \mathrm{d}i = \int \tilde{\ell}_i \mathrm{d}i = L$$

and the feasibility constraint

$$\int \int c_i(\boldsymbol{\rho})\, \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho})\, \mathrm{d}i = \int \int y_{i\rho} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho})\, \mathrm{d}i.$$

insure against the output variation that arises from the idiosyncratic components of the shocks. For simplicity, we abstract from modeling farm capital and labor. In the empirical section, we incorporate data on these other factors of production, land quality, and rainfall shocks to estimate household-farm productivity. See Subsection 3.3 for further details.

[8] The careful reader will observe that our notation implies identical household consumption for households of the same type, conditional on the realization of the output shock. This assumption holds if the equilibrium allocation of resources under full insurance can be computed as the solution to a planner's problem with a weighted utilitarian social welfare function, using type-specific Pareto weights. We invoke this assumption below.

Under full insurance, households can completely eliminate the effects of idiosyncratic output shocks. Each household consumes a fixed fraction of the constant aggregate output, where this fraction is proportional to its Pareto weight. An optimal consumption allocation satisfies the well-known Borch rule, which states that the ratio of any two households' marginal utilities of consumption is constant across all states of the world.

Claim 1. *Under full insurance, each household consumes a constant fraction of aggregate output, with the fraction being proportional to its Pareto weight.*

An optimal consumption allocation under full insurance ensures that each household's consumption remains constant across all states of the world. Consequently, the planner can disregard how land cultivation decisions affect the distribution of consumption, implying that an optimal allocation of land across households simply requires each farmer to cultivate an amount of land where the expected marginal product equals its shadow price. In a decentralized complete-market economy, this outcome would result from the separation theorem, which states that each household-farm makes production decisions to maximize its expected income. (Bardhan and Udry, 1999). Given that the expected marginal products of land are equalized across households, an allocation of land under full insurance features no misallocation and maximizes aggregate expected output.

Claim 2. *Under full insurance, the expected marginal products of land are equalized across households and aggregate expected output is maximized.*

Next, we consider the allocation of land that obtains in a competitive equilibrium under no sharing. When risk-sharing is absent, the problem of a household of type i reads as follows:

$$\max_{c_i(\rho), \ell_i} \int \frac{\left(c_i\left(\rho\right)\right)^{1-\sigma}}{1-\sigma} \mathrm{d}Q_\rho\left(\rho\right)$$

subject to the budget constraint

$$c_i\left(\rho\right) = y_{i\rho} - r^{IM}\left(\ell_i - \widetilde{\ell}_i\right),$$

where r^{IM} denotes the equilibrium price of land under no sharing (incomplete markets). Without risk-sharing, each household's consumption (and marginal utility of consumption) depends on the realization of its output shock. This dependency distorts households' land cultivation decisions, implying that an equilibrium land allocation does *not* maximize expected income for each household. In particular, because households consider how land cultivation choices impact the distribution of consumption across different states of the world, it generally will not hold that each household equates the expected marginal product of land with its market price. When the expected marginal products of land do not uniformly match the price of land across different households, there exists land misallocation—redistributing land across households can increase aggregate expected output.[9]

Claim 3. *Under no sharing, there is land misallocation.*

The key takeaway from Claim 3 is that distortions in *consumption insurance* markets (i.e., lack of insurance) alone are sufficient to cause misallocation in the *land* market, even when the land market itself operates without any other distortions.

2.2 Partial insurance

More broadly, we can investigate the impact of risk-sharing on land misallocation for any degree of risk-sharing. More specifically, we explore the relationship between the elasticity of consumption with respect to own income—used as a measure of lack of insurance—and misallocation in the land market, as measured by the extent to which the marginal returns of land are distorted away from zero across household types and states of the world.

We consider an environment with partial insurance, an intermediate situation between the full insurance and no sharing scenarios discussed in Subsection 2.1. To model partial insurance, we define the following consumption function for a household of type i:

$$c_i\left(\boldsymbol{\rho}\right) = \exp\left\{\beta\log\left(\pi_{i\rho}\right) + (1-\beta)\log\left(\frac{\nu_i^{\frac{1}{\sigma}}}{\int \nu_j^{\frac{1}{\sigma}}\mathrm{d}j}\int\int \pi_{j\rho}\mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right)\mathrm{d}j\right)\right\}. \tag{1}$$

In this formulation, β represents the elasticity of consumption with respect to individual income, while $1-\beta$ is the elasticity of consumption with respect to aggregate income. Under full insurance, $\beta = 0$; under no sharing, $\beta = 1$. Any β value between these extremes represents varying degrees of partial insurance, with higher β values indicating worse insurance. The following theorem shows that misallocation in the land market decreases with the degree of insurance.

Theorem 1. *Land market misallocation increases in the elasticity of consumption with respect to own income, β.*

[9] Our model assumes homogeneous land quality for the sake of simplification, but the main results of the paper would remain valid even if land parcels had heterogeneous productivities. However, when estimating households' agricultural production functions in Subsection 3.3, we account for heterogeneous land quality based on observable characteristics.

In the full-insurance benchmark, where each household's consumption is constant across all states of the world, the expected marginal products of land are equalized across households. As we deviate from this benchmark, the distribution of each household's marginal utility of consumption increasingly reflects the impact of its realized output shocks. Amplifying the variability in households' marginal utilities of consumption across different states of the world is isomorphic to imposing distortions that affect the marginal return on land in each state. Thus, as households' marginal utilities become increasingly tied to their realized output shocks, land allocation decisions deviate further from the full insurance benchmark.[10]

3 Risk-sharing and misallocation in Indian villages

3.1 Background and data

We use household panel data collected under the Village Dynamics in South Asia (VDSA) project by the International Crop Research Institute for the Indian Semi-Arid Tropics (ICRISAT). The data are derived from detailed survey interviews conducted between 2009 and 2014 in 18 villages in the Indian semi-arid tropics. Some components of the survey were administered monthly, while others were administered annually. As discussed in Subsection 3.5, these data allow us to construct *monthly* measures of consumption and income for households in different villages.[11] This feature offers the advantage of estimating the level of risk-sharing specific to each village and year. We exploit this possibility in Subsections 3.5 and 3.6, where we relate the level of risk-sharing in each village and year to land misallocation at the village-year level. The data include information from 40 randomly selected households in each village, stratified by landholding size. Specifically, the 40 households include 10 landless laborers, 10 small farmers, 10 medium farmers, and 10 large farmers.[12] For our empirical exercise, we require data on the households' agricultural inputs and corresponding outputs. We are particularly interested in farm output, farm capital, and land. These data fit our needs because they provide detailed information on households' farming activities: for each plot and each operation performed

[10]Our model shows that enhancing risk-sharing reduces land misallocation in equilibrium, but it is silent about the distributional implications of such improvements. While a reduction in risk-sharing could theoretically cause land to shift from less to more productive households, this situation would still lead to greater misallocation. This result follows from the fact that under full insurance, the expected marginal products of land are equalized. Therefore, any distortion in households' choices due to decreased risk-sharing would lead to deviations from this benchmark, exacerbating misallocation.

[11]As in Section 2, the term 'household' refers to an agricultural household operating a farm.

[12]This classification is based on operational landholdings, which equals the size of own land plus that of land leased/shared in and minus that of land leased/shared out.

in a plot, the data reports the quantity and value of all inputs used by the household cultivating the plot. The data also contains comprehensive information on expenditures and incomes, which we use to construct variables for consumption and income at the household-month level. We refer to Townsend (1994), Mazzocco and Saini (2012), and Morten (2019) for more detailed descriptions of the data.[13] Section C in the Appendix reports a detailed description of the variables used in the analyses.

Table 1: Farm size distributions (% of farms by size)

Farm size (hectares)	2009-2014 ICRISAT India	1990 World Census of Agriculture Malawi	Belgium	United States
≤ 1	17.93	77.7	14.6	0.0
1–2	23.88	17.3	8.5	0.0
2–5	38.26	5.0	15.5	10.6
5–10	13.45	0.0	14.8	7.5
≥ 10	6.49	0.0	46.6	81.9
Average	3.819	0.7	16.1	187.0

Notes: This table presents the percentage distribution of farm sizes in hectares for India, Malawi, Belgium, and the United States. Data for India are derived from our computations based on the 2009-2014 ICRISAT panel data from the Indian semi-arid tropics. Data for Malawi, Belgium, and the United States are from the 1990 World Census of Agriculture, as documented in Adamopoulos and Restuccia (2014).

3.2 Land distribution

Table 1 presents the distribution of cultivated land, measured in hectares, for each farmer with positive amounts of cultivated land. This table compares these figures with those from Malawi, Belgium, and the United States, as documented in the 1990 World Census of Agriculture (Adamopoulos and Restuccia, 2014).[14]

[13] As pointed out by Mazzocco and Saini (2012), it can be difficult to compare some of the information contained in the data (e.g., expenditures) across households and over time, since (1) the frequency of the interviews varies, and (2) the interview dates differ across respondents. Some recall periods can be longer than a month (e.g., a household in Aurepalle reported the amount spent on rice from July 1 to November 8, 2009). Hence, it is impossible to determine how the information provided is distributed over the months that make up recall periods longer than a month. Fortunately, from 2010 onward, the survey gives information on the month to which every piece of information refers. Therefore, we drop the observations that pertain to the year 2009.

[14] In the World Census of Agriculture, "[a]n agricultural holding is an economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form or size" (`https://www.fao.org/4/x0187e/x0187e01.htm`). Given this definition, when comparing the ICRISAT data to the World Census of

Compared to Belgium and the United States, the distribution of land cultivated is more left-skewed. In our sample, around 80% of farms cultivate less than 5 hectares of land, and nearly 95% of the farmlands are under 10 hectares. The average land size in our sample is 3.8 hectares, against a value of 0.7 hectares for Malawian farms, and much larger values of 16.1 and 187.0 hectares for Belgian and American farms, respectively.

3.3 Physical productivities across households and years

For our analysis, it is essential to accurately estimate each farmer's physical productivity, also known as TFP-Q. To achieve this, we propose and estimate an agricultural production function that incorporates additional sources of variability from observable characteristics compared to the one presented in Section 2. Specifically, following Chen et al. (2023), we posit the following agricultural production function:

$$y_{i\tau} = e^{\kappa_t} e^{\vartheta \text{rain}_{v\tau}} \theta_i \rho_{i\tau} (k_{i\tau}^{\gamma} h_{i\tau}^{1-\gamma})^{1-\alpha} (q_{i\tau} \ell_{i\tau})^{1-\alpha}, \quad \alpha, \gamma \in (0,1), \tag{2}$$

where $y_{i\tau}$ denotes the agricultural output of household-farm i in year τ;[15] κ_τ capture common year-specific factors; $\text{rain}_{v\tau}$ denotes the amount of rain (in millimeters) in village v and year τ;[16] θ_i denotes household i's time-invariant productivity; $\rho_{i\tau}$ is a multiplicative, idiosyncratic output shock, which we define below (see Equation (4)); $k_{i\tau}$ is capital, measured by the total value of farm equipment owned by the household; $h_{i\tau}$ denotes total hours of family labor dedicated to farming activities; $\ell_{i\tau}$ represents total land cultivated, measured in hectares; and $q_{i\tau}$ is land quality, which we define below (see Equation (6)). We assume that

$$\theta_i = e^{\mu_i}, \tag{3}$$

Agriculture, it is more appropriate to consider farm size in terms of cultivated land rather than owned area. We also maintain that cultivated area is inherently a more accurate measure of farm size because it avoids measurement errors arising from not including plots that are cultivated but not owned including plots that are owned but not cultivated.

[15]In the ICRISAT data, information on agricultural yields is available only at an annual frequency. On the other hand, information on some inputs used in agricultural production, such as labor, is available at a monthly frequency. To estimate households' physical productivities, we aggregate these higher-frequency variables to an annual frequency.

[16]Rainfall shocks are major sources of transitory variation in agricultural output in semi-arid tropical India, where the vast majority of land plots are rain-fed. To measure them, we use daily recordings of rainfall levels at the nearest weather station to each village and derive the total annual rainfall for each village by summing these daily measurements over the year.

where μ_i represents permanent unobserved heterogeneity specific to household i. Similarly, we let the shock to output, $\rho_{i\tau}$, to be equal to

$$\rho_{i\tau} = e^{\varepsilon_{i\tau}}, \tag{4}$$

where $\varepsilon_{i\tau}$ is an unobserved error term. Our approach involves treating the idiosyncratic output shock, $\rho_{i\tau}$, as a residual after accounting for farm-specific fixed effects and other variation in output originating from observable sources. After taking logs and rearranging terms, Equation (2) becomes

$$\log(\pi_{i\tau}) = (1-\alpha)\log(k_{i\tau}^{\gamma} h_{i\tau}^{1-\gamma}) + \alpha \log(q_{i\tau}\ell_{i\tau}) + \kappa_\tau + \vartheta \text{rain}_{v\tau} + \mu_i + \varepsilon_{i\tau}. \tag{5}$$

As for land quality, we posit that

$$\log q_{i\tau} = \delta_1 \text{depth}_{i\tau} + \delta_2 \text{slope}_{i\tau} + \delta_3 \text{fertility}_{i\tau} + \delta_4 \text{degradation}_{i\tau}, \tag{6}$$

where $\text{depth}_{i\tau}$, $\text{slope}_{i\tau}$, $\text{fertility}_{i\tau}$, and $\text{degradation}_{i\tau}$ represent measures of the average soil depth, slope, fertility, and degree of degradation, respectively, for the plots cultivated by household i in year τ (see Appendix C for additional details on the construction of these variables). Finally, we assume that μ_i and $\varepsilon_{i\tau}$ are independent of each other.

Table 2: TFP-Q dispersion across farms and manufacturing firms

	Farms			Manufacturing firms		
	India 2010-2014	Malawi 2010-2011	US 1990	India 1987	China 1998	US 1977
St.dev., log	1.08	1.18	0.80	1.16	1.06	0.85
75-25 log ratio	1.02	1.39	1.97	1.55	1.41	1.22
90-10 log ratio	2.50	2.89	2.50	2.77	2.72	2.22

Notes: The first column reports statistics of the estimated farm productivity using the 2009-2014 ICRISAT panel data for the Indian semi-arid tropics The second column reports statistics of farm productivity in Malawi from Chen et al. (2023). The third column reports statistics of farm productivity in the United States from the calibrated distribution in Adamopoulos and Restuccia (2014) to U.S. farm-size data. The third and fourth columns report statistics of manufacturing plants' productivities in Hsieh and Klenow (2009). St.dev. refers to the standard deviation of log productivity; 75-25 is the log difference between the 75 and 25 percentile; 90-10 is the log difference between the 90 and 10 percentile.

In Table 2, we report the dispersion of the estimated farm (log) productivity, $\widehat{\mu}_i$. We compare the dispersion of TFP-Q among Indian household-farms with the dispersion of TFP-Q across farms in Malawi (Chen et al., 2023) and in the US (Adamopoulos and Restuccia, 2014), as well as with the dispersion of TFP-Q across manufacturing plants in the US and India, as reported by Hsieh and Klenow (2009). The dispersion of physical productivity across Indian farms is large: the standard deviation of log productivity (1.08) is comparable to that of farms in Malawi (1.18) and significantly larger than that of farms in the US (0.80).[17]

3.4 Land misallocation

We present direct evidence of the degree of misallocation of land among households in rural India. To do so, we employ data on operation (cultivated) landholdings, $\ell_{i\tau}$, together with our estimates of physical productivity, $\widehat{\theta}_i$, and construct two measures of allocative efficiency in the land market. The first is the village-year unconditional correlation between the log of operational land size, $\ell_{i\tau}$, and the log of farm productivity:

$$\text{corr.}_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right].$$

Larger misallocation implies a lower correlation between farm size and TPF-Q. The second measure is the dispersion in the marginal product of land in each village-year pair:

$$\text{st.dev.}_{v\tau}\left[\log \text{MPL}_{i\tau}\right],$$

where $\text{MPL}_{i\tau}$ is constructed using the production function specified in Equation (2). Large misallocation implies a higher dispersion of the marginal product of land across farms (Restuccia and Rogerson, 2017).

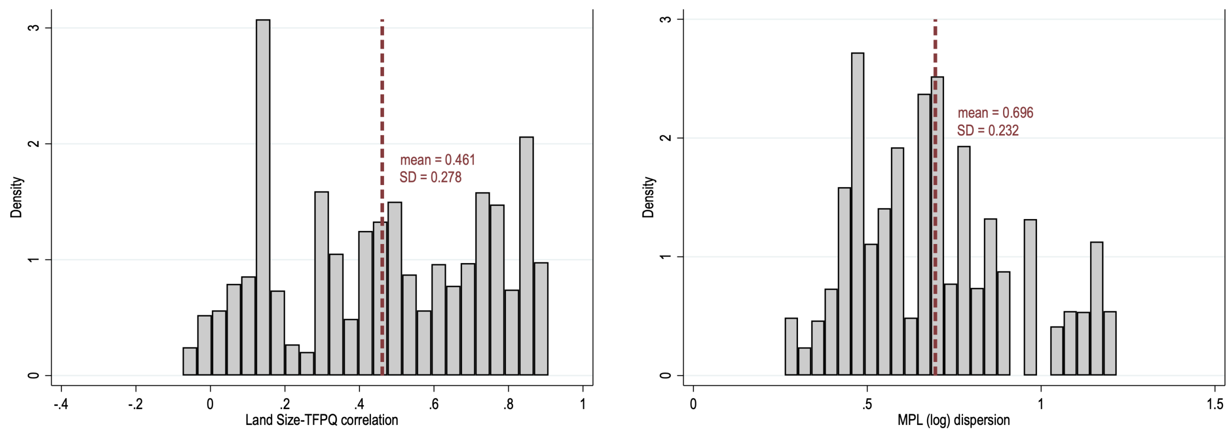
There is a large variation in land misallocation across villages and years. Panel A in Figure 1 displays the distribution of $\text{corr.}_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$ across villages and years. The correlation ranges from negative values in some village-year pairs to nearly 1 in others. That is, while in certain village-year pairs more productive household-farms cultivate, on average, *less* land than less productive households, in other village-year pairs, there is an almost one-to-one relationship between operational landholdings and physical productivity, indicating minimal land misallocation. Panel B in Figure 1 reports the distribution $\text{st.dev.}_{v\tau}\left[\log \text{MPL}_{i\tau}\right]$ across villages and years. Similarly to Panel A, this dispersion is small in certain village-year pairs, indicating little misallocation, while much greater in others.

What accounts for the variation in land misallocation across villages and years? The

[17] In Appendix D, Table 7, we show that differences in TFP-Q largely account for differences in agricultural output across Indian farms.

Figure 1: Land misallocation in each village

(A) Land size-TPF-Q correlation **(B)** Dispersion in the (log) MPL



Source: VDSA survey (ICRISAT) and own calculations.

theory developed in Section 2 suggests that misallocation in the land market is negatively correlated with the degree to which households are able to share idiosyncratic risks. In the following subsection, we present evidence on imperfect risk-sharing and how households' ability to share risks varies across different villages and years. In Subsection 3.6, we show that our estimates of risk-sharing across villages and years are negatively correlated with misallocation in the land market within those villages and years, as implied by our model.

3.5 Risk-sharing

A commonly acknowledged fact is that risk-sharing within villages in developing countries tends to be incomplete. This holds true for the ICRISAT villages as well. To see this, consider the following model:

$$\log c_{it} = \psi + \beta \log \pi_{it} + \mu_i + \zeta_{vt} + \epsilon_{it}. \tag{7}$$

In Equation (7), $\log c_{it}$ and $\log \pi_{it}$ denote the log per-capita consumption and log per-capita income, respectively, for household i in month t; μ_i are household fixed effects; and ζ_{vt} represents village-month fixed effects that capture the average resources available to each village in each month. We can interpret $1 - \beta$ as the level of risk-sharing in village economies, where a higher β means a higher elasticity of consumption to idiosyncratic income shocks (indicating a lower degree of risk-sharing). Under perfect risk-sharing, household income should not affect household consumption, conditional on total resources at the village-month level. In Appendix E, we present the results of estimating equation (7). Full risk-sharing is rejected. On average, 22.5% of idiosyncratic income fluctuations are passed through to consumption. These values are aligned with what the literature has already documented for Indian villages using alternative empirical specifications (Townsend, 1994; Ravallion and Chaudhuri, 1997; Morduch, 2005).

The average elasticity of consumption with respect to idiosyncratic income shocks in equation (12), hides considerable variation in risk-sharing across villages and years. To uncover this variation, we estimate the degree of risk-sharing across households within each village *separately*. More specifically, we estimate the following equation independently for each village v and year τ:

$$\log c_{it} = \beta_{v\tau(t)} \log \pi_{it} + \mu_i + \kappa_t + \epsilon_{it}. \tag{8}$$

Here, with two slight abuses of notation, index i denotes households within village v, and index t corresponds to the months within year $\tau(t)$, where $\tau(t)$ specifies the year associated with month t.[18] As usual, c_{it} and π_{it} indicate the consumption and income, respectively, of household i during month t, and μ_i and κ_t are household and month fixed effects (where, again, note that 'months' here refers to the months within the specific year $\tau(t)$). A high estimate of $\beta_{v(t)}$ indicates that *within village v during year $\tau(t)$*, the elasticity of household consumption in response to idiosyncratic shocks to household income is high—i.e., risk-sharing is low in village v and year $\tau(t)$.

Figure 2 plots the distribution of the estimated elasticities of consumption with respect to idiosyncratic income shocks, $\widehat{\beta}_{v\tau}$, for each village-year pair. The average estimated $\widehat{\beta}_{v\tau}$ across villages and years is 0.223, indicating that, on average, a 1% idiosyncratic increase in income results in a 0.223% increase in consumption. As shown in the figure, there is considerable variation in risk-sharing across villages and years, with our estimates ranging from full insurance ($\widehat{\beta}_{v\tau} \approx 0$) to several others showing very high elasticities of consumption with respect to idiosyncratic income shocks ($\widehat{\beta}_{v\tau} > 0.6$).

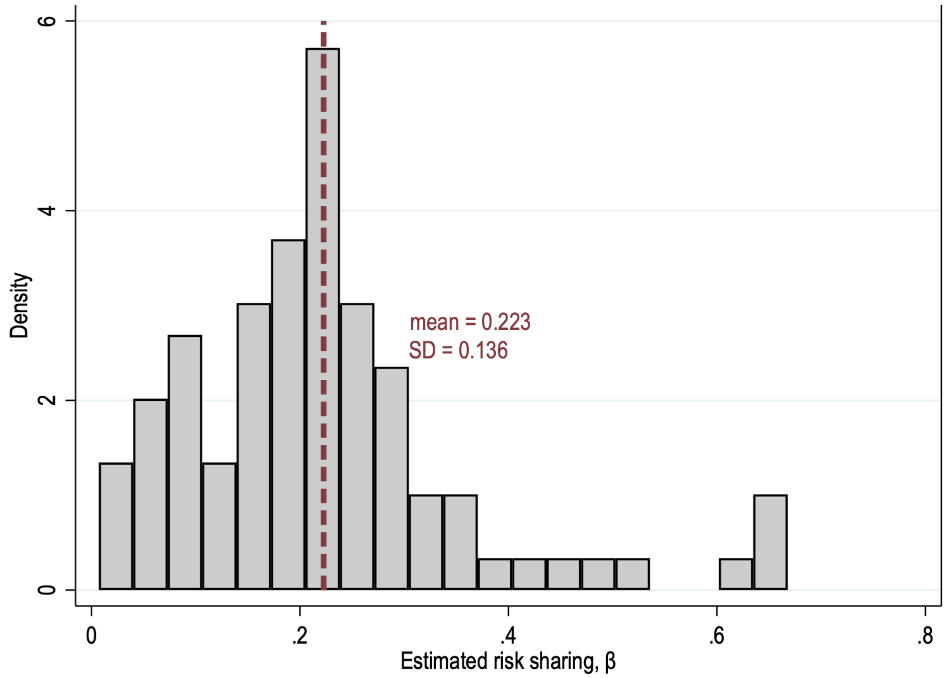
According to the model in Section 2, villages in years with a low degree of risk sharing face larger land misallocation. In the following subsection, we test this hypothesis.

3.6 Linking land misallocation to risk-sharing

We are now ready to test whether a higher degree of risk-sharing within villages and years is associated with better allocative efficiency in the land market. To do so, we relate our two measures of misallocation in each village and year (obtained in Section 3.4), which we denote by $\omega_{v\tau}$, to our estimates of the elasticities of consumption with respect

[18]For example, if t corresponds to October of the year 2010, then $\tau(t) = 2010$.

Figure 2: Estimated degrees of risk-sharing in each village



Notes: This figure reports the distribution of risk-sharing parameters estimated for each village and year using equation (8), $\widehat{\beta}_{v\tau}$, for each v and τ. Source: VDSA survey (ICRISAT) and own calculations.

to idiosyncratic income shocks at the village-year level (obtained in Section 3.5), $\widehat{\beta}_{v\tau}$. In particular, we estimate the following equation:

$$\omega_{v\tau} = \varrho + \gamma\widehat{\beta}_{v\tau} + \mu_v + \mu_\tau + \varphi_v \times \tau + \varepsilon_{v\tau}, \tag{9}$$

where μ_v and μ_τ are village and year fixed effects, and $\varphi_v \times \tau$ are village-specific linear year trends.

Figure 3 binscatters the unconditional relation between the estimated degrees of risk-sharing, $\widehat{\beta}_{v\tau}$, and

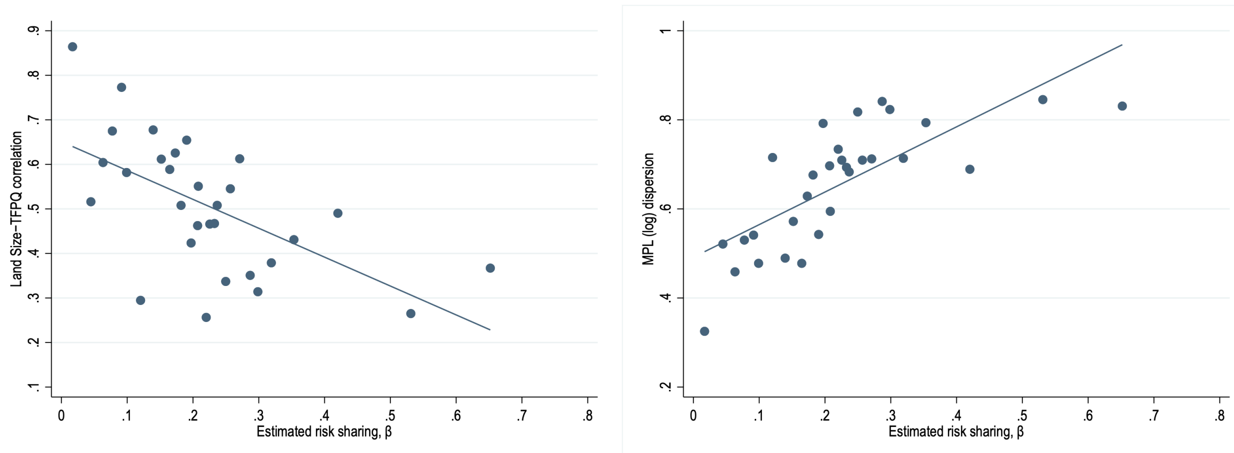
1. the correlation between log farm size and log TFP-Q, $\text{corr.}_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$, in Panel A;

2. the standard deviation of the log marginal product of lands, $\text{st.dev.}_{v\tau}\left[\log \text{MPL}_{i\tau}\right]$, in Panel B.

For each dot, the x-axis shows the average estimated risk-sharing for village-year pairs within a given percentile bin, while the y-axis gives the average measure of land misallocation. As we move from village-year pairs with no risk-sharing ($\widehat{\beta}_{v\tau} = 1$) to pairs with full insurance ($\widehat{\beta}_{v\tau} = 0$), land misallocation decreases: the correlation between farmer

Figure 3: Risk-sharing and land misallocation

(A) Land size-TPF-Q correlation vs. risk-sharing **(B)** Dispersion in the MPL vs. risk-sharing



Notes: Panel A scatters the unconditional relation between the degree of risk-sharing, β, and the correlation between log farm size and log TFP-Q. Panel B scatters the unconditional relation between the degree of risk-sharing, β, and the standard deviation of the log marginal product of land. In both panels, each dot refers to the average village-year pair in a given 4 percent bin of the estimated risk-sharing. Source: VDSA survey (ICRISAT) and own calculations.

TFP-Q and land holdings rises from 0.3 to around 0.7, while the dispersion of MPL drops from 0.8 to approximately 0.4.

Table 3 reports the estimation outcomes for different specifications of Equation (9). Columns (1) to (4) refer to corr.$_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$, and (5) to (8) to st.dev.$_{v\tau}\left[\log \mathrm{MPL}_{i\tau}\right]$. Standard errors (in parentheses) are computed using village-level clustered bootstrap with 5,000 replications.

There is a negative correlation between land misallocation and the degree of risk-sharing. For example, column (4) indicates that moving from village-year pairs with full insurance ($\widehat{\beta}_{v\tau} = 0$) to no risk-sharing ($\widehat{\beta}_{v\tau} = 1$) reduces the correlation between farm size and productivity by approximately 0.229 points, after controlling for village fixed effects, year fixed effects, and a village-specific linear time trend. In terms of magnitudes, the effect is equal to 0.82 times the standard deviation of the correlation between land size and farm productivity across villages and years (Figure 1, Panel A). The dispersion of the marginal product of land is negatively correlated with the degree of risk-sharing. The estimates in column (8) suggest that moving from full insurance to no risk-sharing is associated with an increase of approximately 0.192 points in the standard deviation of the log of $\mathrm{MPL}_{i\tau}$. Similarly to before, this increase represents 0.83 times the standard deviation in the dispersion of the marginal product of land across villages and years (Figure 1, Panel B).[19]

[19]In Appendix F, we report the estimates from an IV strategy that exploits variation in caste diversity

Table 3: Risk-sharing and land misallocation

	corr.$_{v\tau}$ $\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$				st.dev.$_{v\tau}$ $[\log \text{MPL}_{i\tau}]$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\widehat{\beta}_{v\tau}$	-0.648***	-0.675***	-0.163**	-0.229***	0.732***	0.749***	0.161***	0.192***
	(0.226)	(0.228)	(0.080)	(0.109)	(0.178)	(0.245)	(0.063)	(0.076)
Observations	90	90	90	90	90	90	90	90
R^2	0.106	0.118	0.897	0.942	0.218	0.223	0.902	0.948
Village FE		✓	✓	✓		✓	✓	✓
Year FE			✓	✓			✓	✓
Village time trends				✓				✓

Notes: The unit of analysis across all columns is a village-year pair. The first four columns present the results of regressing our first measure of land misallocation on the estimated village-and-year-specific consumption elasticities to idiosyncratic income shocks. The following four columns show the results using our second measure of land misallocation. Standard errors in parentheses are computed using village-level clustered bootstrap (5,000 replications) following the procedure in Cameron et al. (2008).

Our results show a clear negative correlation between the degree of risk-sharing in village economies and the misallocation in the land markets of those villages. What does this negative correlation imply in terms of output and welfare gains from improving risk-sharing in Indian villages? In the following section, we leverage the structure of our model to address these questions in detail.

4 The gains from full insurance

In this section, we employ the model described in Section 2 to quantify the aggregate gains from completing village consumption insurance markets. In what follows, we assume that households exhibit constant relative risk aversion; i.e.,

$$u\left(c_i\right) = \frac{c_i^{1-\sigma}}{1-\sigma}.$$

where σ denotes the (common) coefficient of relative risk aversion. To proceed, we must specify values for the model parameters: the land share α, the aggregate (fixed) supply of land L, the level of risk-sharing $1-\beta$, and the coefficient of relative risk aversion σ. In addition to this, we must also define the distributions of farmer physical productivity, θ_i, and the output shock, ρ_i. Below we describe in detail what we do.

We fit the model to the average village in our data. Some parameters are externally calibrated without solving the model. These parameters are listed in Table 4. We set

across villages to address potential endogeneity in the relation between risk-sharing and misallocation.

the output elasticity of land, α, to 0.282, based on estimates obtained from Equation (5). The aggregate land supply, L, is set to the average farm size of 3.819 hectares (Table 1). The elasticity of consumption to idiosyncratic income shocks, β, is set to 0.223, which is the estimate obtained from Equation (7). The estimates from Equation (5) are used to derive values for household-farms' physical productivities, θ_i, and output shocks, ρ_i. The distributions of these parameters are calibrated to match the empirical frequencies observed across households and years.[20]

Table 4: Parameters calibrated externally

Parameters	Description	Value	Source
α	Land share	0.282	Equation (5)
L	Aggregate land supply (hectares)	3.819	Table 1
β	Elasticity of consumption to idiosyncratic income shocks	0.223	Equation (7)

Notes: This table reports the parameters that are externally calibrated without solving the model and their sources.

We are left with only one parameter, the coefficient of relative risk aversion, σ, which we estimate using the simulated method of moments (SMM). In particular, we look for values of σ to match the average correlation between log farm size and log productivity, which we denote by

$$\overline{\text{corr.}}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right] = \frac{\sum_{v\tau} \text{corr.}_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]}{\mathcal{V}\mathcal{T}},$$

where $\mathcal{V}$ and $\mathcal{T}$ denote the numbers of villages and years in the data. We obtain an estimate of $\widehat{\sigma} = 1.6$, indicating moderate risk aversion. This estimate aligns with the findings in Holden and Quiggin (2017), who estimate a coefficient of relative risk aversion of 1.73 (Table A.4) for a sample of farmers in Malawi.[21]

Table 5: Estimated risk aversion

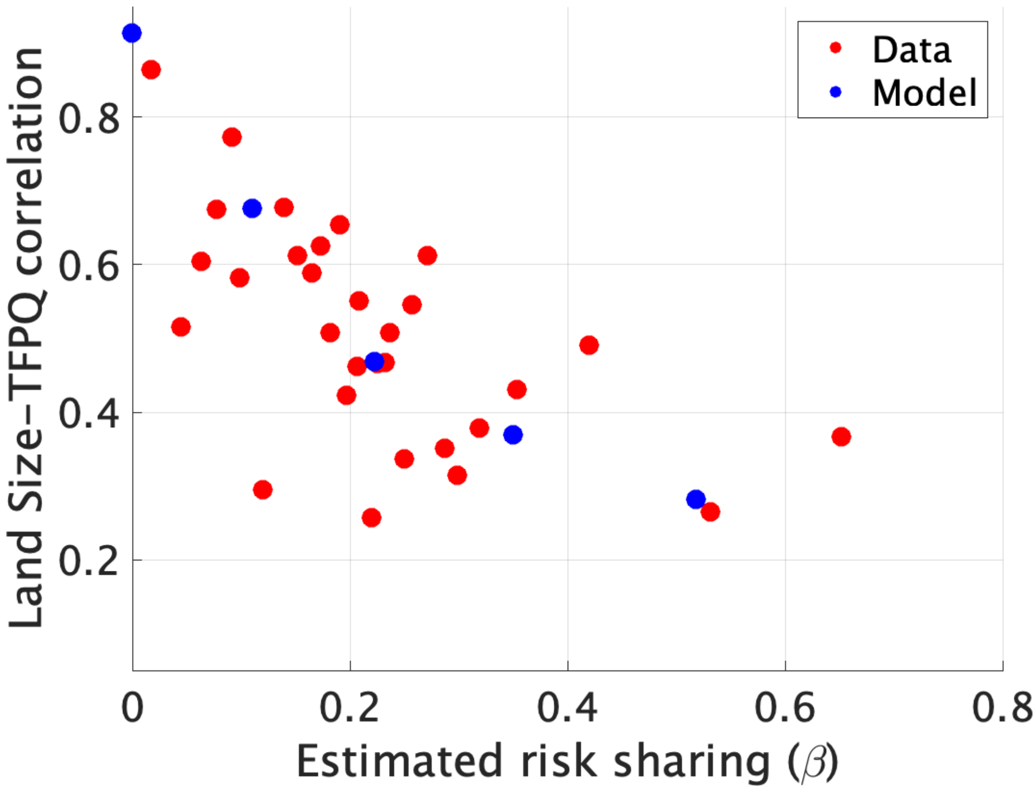
Parameters	Description	Value	Target	Data	Model
σ	Relative risk aversion	1.600	$\overline{\text{corr.}}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$	0.461	0.469

Notes: This table reports the value of the coefficient of relative risk aversion that is estimated to match the average correlation between log farm size and log productivity.

[20]To solve the model, we discretize the possible values of physical productivities and output shocks into 100 and 50 bins, respectively, each corresponding to different percentiles within their distributions.

[21]As shown in Equation (1), solving the model requires us to take a stance on the households' Pareto weights, $(\nu_i)_i$. In all the exercises performed in this section, we assume that $\nu_i = 1$, for each i.

Figure 4: Risk-sharing and misallocation: Model vs. data



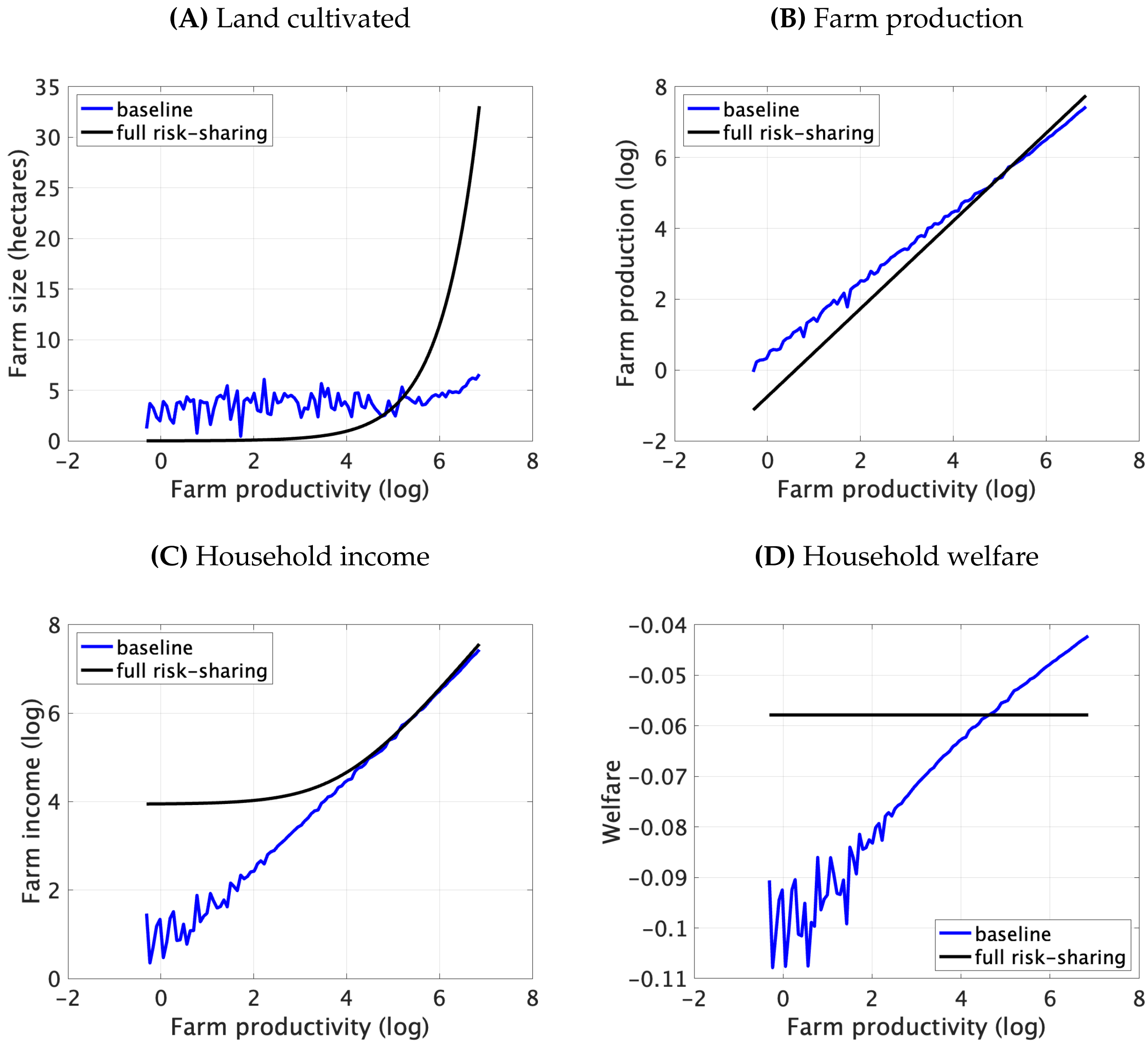
Source: VDSA survey (ICRISAT) and own calculations.

To validate the model, we assess its ability to replicate the observed correlation between the estimated levels of risk-sharing, $\widehat{\beta}_{v\tau}$, and the correlation of farm size and productivity across villages and years, $\text{corr.}_{v\tau}\left(\log \ell_{i\tau}, \log \widehat{\theta}_i\right)$. To accomplish this, we solve replicas of our model that differ only in the values of β. Figure 4 plots the equilibrium correlation between log farm size and log productivity for different levels of risk-sharing (blue dots) against the corresponding empirical estimates from Figure 3A (red dots). Our model can replicate the negative correlation between risk-sharing and land misallocation observed in the data, even though this correlation was not explicitly targeted in the model's estimation. This suggests that the model effectively captures the relationship between land misallocation and the degree of consumption insurance across villages and years.

4.1 Counterfactual exercise

How would households' choices and the allocation of resources differ if village insurance markets were complete? To what extent does full insurance enhance allocative efficiency

Figure 5: Partial (baseline) vs. full risk-sharing



Source: VDSA survey (ICRISAT) and own calculations.

in the land market? We answer these questions with a counterfactual exercise where we improve the functioning of consumption insurance markets in village economies. Completing the market for insurance against shocks to agricultural output is a natural benchmark of "financial deepening" (see Townsend and Ueda (2006) on this concept). In particular, we examine a counterfactual scenario within the model where $\beta = 0$, which represents perfect risk-sharing. We contrast this scenario with the baseline model setting, where β is set to 0.223, the estimated elasticity of consumption to idiosyncratic income shocks obtained from Equation (7). We keep all the other parameters at their baseline values, including the overall land supply.

Figure 5 plots the amount of land cultivated (Panel A), farm output (Panel B), household income (Panel C), and household expected utility (Panel D) on the y-axes against (log) farm productivity (on the x-axis) for both the baseline (blue line) and full-insurance economies (black line). As we move from partial to full risk-sharing, land reallocates from low- to high-productivity farms (Panel A). Under full insurance, the most productive household cultivates more than four times as much land compared to what it does under partial insurance, increasing its cultivated area from just over 5 hectares to nearly 35 hectares. Conversely, those with low productivity cultivate less land under full insurance than under partial insurance. Improved risk-sharing decreases land misallocation, leads to greater output dispersion across farms (Panel B), and simultaneously reduces the dispersion in household income (Panel C). Panel D shows that most households, particularly those with low productivity, experience substantial welfare gains under full sharing compared to partial insurance. Conversely, the most productive households face welfare losses when participating in the full sharing arrangement rather than the partial insurance scheme.

Table 6: Counterfactual exercise

	Baseline (partial insurance) (1)	Counterfactual (full risk-sharing) (2)
β	0.223	0
$\overline{\text{corr.}}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$	0.469	0.914
Share of land, top 1% productive farms	0.017	0.086
Share of land, top 10% productive farms	0.155	0.627
Land dispersion (st.dev.[$\log \ell_{i\tau}$])	0.399	2.589
Aggregate efficiency (output per hectare)	1	1.415
Aggregate output	1	1.186
Aggregate welfare	1	1.286

Source: VDSA survey (ICRISAT) and own calculations.

Table 6 offers a comparative analysis of two different economies: one featuring partial insurance (the baseline) and the other characterized by full risk-sharing, evaluated across various dimensions. In the counterfactual economy with full insurance, land misallocation is sizably reduced, as evidenced by the correlation between log farm size productivity and log productivity nearly doubling compared to the baseline. Under full insurance, the distribution of cultivated land becomes significantly more unequal: the share of total available land allocated to the top 1% of farms, based on productivity, increases by approximately six times, while the share going to the top 10% of farms increases by approximately four times. The variance in the distribution of (log) cultivated land increases approximately ninefold. Improved insurance leads to an increase in output of 18.57%, while the overall welfare gains, measured in consumption-equivalent terms, are equal to 28.75%.[22] These figures are comparable to those quantifying the welfare gains from eliminating distortions in the land markets (e.g., Adamopoulos et al. (2022)). Thus, our counterfactual analysis suggests that inefficiencies in consumption insurance markets may be as significant as land market distortions in explaining the potential gains from improving land allocation across households. This result is confirmed by the 41.51% increase in output per unit of land associated with improved risk-sharing.

5 Conclusions

This paper bridges the gap between the literature on risk-sharing and resource misallocation. We begin with two key observations. First, insurance markets in rural villages are often incomplete, leading to significant impacts of household income shocks on consumption. Second, there is substantial misallocation of factors of production among farmers, which reduces overall productivity in the agricultural sector of developing countries. We argue that these two phenomena are deeply interconnected. Specifically, we see the limited functioning of consumption insurance markets in village economies as a key factor contributing to land misallocation in these communities.

We explore how imperfections in insurance markets affect land misallocation. Our theoretical results show that incomplete consumption insurance can increase land misallocation, even when land markets operate without distortions. Empirically, we quantify the losses attributable to limited risk-sharing using the latest ICRISAT data from rural India. Our findings suggest that fully developed insurance markets could significantly enhance the allocation of land, resulting in output and welfare gains of 19% and 29%, respectively. Thus, improving risk-sharing within an otherwise undistorted economy can yield gains comparable in magnitude to those achieved by removing distortions in factor or output markets (Adamopoulos et al., 2022).

From the perspective of the misallocation literature, we offer an alternative explanation for land misallocation beyond the well-known land market frictions or farm-specific distortions that disproportionately affect more productive farmers. Our explanation is grounded in the observation that village-level risk-sharing is incomplete, and our analy-

[22] In Appendix G, we study whether efficiency and welfare gains from ensuring full risk sharing are robust to incorporating farm-specific distortions in the model, as seen in more traditional misallocation literature (Chen et al., 2023). Specifically, we extend our model by introducing farm-specific distortions in the form of output wedges that are correlated to farm productivity. We then conduct a comparative analysis of i) completing insurance markets versus ii) removing distortions. We find that the efficiency and welfare gains from achieving perfect risk-sharing remain large and are equal to 39% and 14%, respectively.

sis quantifies how imperfect risk-sharing contributes to land misallocation in rural Indian communities. From a policy perspective, our emphasis on imperfect risk-sharing as a cause of misallocation highlights the potential of financial deepening as a way to enhance the allocation of factors of production.

References

Acampora, M., Casaburi, L., and Willis, J. (2022). Land rental markets: Experimental evidence from kenya. Technical report, National Bureau of Economic Research.

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2022). Misallocation, Selection, and Productivity: A Quantitative Analysis with Panel Data from China. *Econometrica*, 90(3):1261–1282.

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–1697.

Adamopoulos, T. and Restuccia, D. (2022). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. *The Review of Economic Studies*, 89(4):1629–1653.

Ambrus, A., Gao, W., and Milán, P. (2022). Informal Risk Sharing with Local Information. *The Review of Economic Studies*, 89(5):2329–2380.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375):598–606.

Arrow, K. J. (1971). The Theory of Risk Aversion. In *Essays in the Theory of Risk-Bearing*, pages 90–120.

Bardhan, P. and Udry, C. (1999). *Development Microeconomics*. Oxford University Press.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: testing for separation in agricultural household models. *Econometrica: Journal of the Econometric Society*, pages 287–322.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics*, 90(3):414–427.

Chen, C., Restuccia, D., and Santaeulàlia-Llopis, R. (2022). The Effects of Land Markets on Resource Allocation and Agricultural Productivity. *Review of Economic Dynamics*, 45:41–54.

Chen, C., Restuccia, D., and Santaeulàlia-Llopis, R. (2023). Land Misallocation and Productivity. *American Economic Journal: Macroeconomics*, 15(2):441–465.

De Giorgi, G., Di Falco, S., and Pietrobon, D. (2024). Farming, Non-Farm Enterprises, and Migration: Incomplete Markets and Structural Transformation. Working paper.

Dercon, S. (2002). Income risk, coping strategies, and safety nets. *The World Bank Research Observer*, 17(2):141–166.

Donovan, K. (2021). The Equilibrium Impact of Agricultural Risk on Intermediate Inputs and Aggregate Productivity. *The Review of Economic Studies*, 88(5):2275–2307.

Fafchamps, M. (2011). Risk Sharing Between Households. In *Handbook of Social Economics*, pages 1255–1279. Elsevier.

Foster, A. D. and Rosenzweig, M. R. (2010). Microeconomics of Technology Adoption. *Annual Review of Economics*, 2(1):395–424.

Gollin, D., Parente, S., and Rogerson, R. (2002). The role of agriculture in development. *American economic review*, 92(2):160–164.

Guner, N. and Ruggieri, A. (2022). Misallocation and inequality. *CEPR Discussion Paper No. DP17113*.

Holden, S. T. and Quiggin, J. (2017). Climate risk and state-contingent technology adoption: shocks, drought tolerance and preferences. *European Review of Agricultural Economics*, 44(2):285–308.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Kinnan, C. (2021). Distinguishing Barriers to Insurance in Thai Villages. *Journal of Human Resources*. Forthcoming.

Lamb, R. L. (2003). Inverse Productivity: Land Quality, Labor Markets, and Measurement Error. *Journal of Development Economics*, 71(1):71–95.

Ligon, E. (1998). Risk Sharing and Information in Village Economies. *Review of Economic Studies*, 67(4):847–864.

Ligon, E., Thomas, J. P., and Worrall, T. (2002). Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. *Review of Economic Studies*, 69(1):209–244.

Mazur, K. (2023). Sharing Risk to Avoid Tragedy: Theory and Application to Farming. *Journal of Development Economics*, 161:103030.

Mazzocco, M. and Saini, S. (2012). Testing Efficient Risk Sharing with Heterogeneous Risk Preferences. *American Economic Review*, 102(1):428–468.

Morduch, J. (2005). *Consumption smoothing across space: Testing theories of risk-sharing in the ICRISAT study region of South India*. Oxford: Oxford University Press.

Morten, M. (2019). Temporary Migration and Endogenous Risk Sharing in Village India. *Journal of Political Economy*, 127(1):1–46.

Munshi, K. (2019). Caste and the Indian Economy. *Journal of Economic Literature*, 57(4):781–834.

Munshi, K. and Rosenzweig, M. (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review*, 106(01):46–98.

Pietrobon, D. (2024). The Dual Role of Insurance in Input Use: Mitigating Risk versus Curtailing Incentives. *Journal of Development Economics*, 166:103203.

Ravallion, M. and Chaudhuri, S. (1997). Risk and insurance in village india: Comment. *Econometrica: Journal of the Econometric Society*, pages 171–184.

Ray, D. (1998). *Development Economics*. Princeton University Press.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic dynamics*, 16(1):1–10.

Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–174.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics*, 55(2):234–250.

Rosenzweig, M. R. and Wolpin, K. I. (1993). Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in india. *Journal of political economy*, 101(2):223–244.

Sleet, C. (2006). Endogenously Incomplete Markets: Macroeconomic Implications. *Mimeo. Tepper Business School*.

Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica*, 62(3):539–591.

Townsend, R. M. and Ueda, K. (2006). Financial deepening, inequality, and growth: a model-based quantitative evaluation. *The Review of Economic Studies*, 73(1):251–293.

Udry, C. (1994). Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria. *Review of Economic Studies*, 61(3):495–526.

A Brief discussion of modeling assumptions

Household types. Our model features household types that encode their "permanent" productivity, making households ex-ante heterogeneous. We adopt this approach because the production decisions we model are naturally considered ex-ante choices: the decision of how much land to cultivate is typically made before any shocks to agricultural yields, such as rainfall or pests, occur. Thus, our analysis focuses on the relationship between risk-sharing and misallocation in the ex-ante chosen factors of production. Alternatively, we could model an economy where households are ex-ante homogeneous, receive productivity shocks, and make land cultivation decisions ex-post (after their productivity is revealed). This framework would still allow us to analyze the impact of insurance on misallocation; however, the misallocation would be contingent on the realization of those shocks.

Land markets. Our model maintains the assumption of undistorted land markets throughout the analysis. This modeling choice allows us to distinguish our findings from most of the results in the misallocation literature, where land misallocation typically stems from land market frictions. We relax this assumption in Appendix G.

The economy with no risk sharing introduced in Subsection 2.1 features an environment where households initially possess land endowments and engage in trading these endowments within a competitive land market before production occurs. Alternatively, we may imagine that competitive moneylenders initially own all the plots and sell them to households before farming takes place. Our findings apply in both contexts.

Land ownership and tenancy. In our model, purchasing (respectively, selling) land is essentially equivalent to renting in (respectively, renting out) land; i.e., there is no difference between ownership and tenancy. A dynamic model may feature channels through which imperfect risk-sharing influences the decision to sell versus rent land. In a frictionless environment, a standard arbitrage condition dictates that the selling price of land should equal the net present value of its expected future rental earnings. Missing insurance markets, borrowing constraints, and imperfections in the saving technology might deter farmers from selling land, which serves as a buffer stock. Indeed, the presence of these frictions implies that the cash obtained from selling land cannot be perfectly smoothed over time or states of the world and that a farmer who sells land may be subsequently forced to engage in renting. In this context, the insurance value of owning land may contribute to land misallocation by affecting the relationship between a landowner's productivity and the amount of land owned. Our empirical analysis closely mirrors our theoretical framework by focusing on misallocation in operational landholdings, encompassing both owner-cultivators and renters.

Consumption functions. Our approach to modeling the dependence of consumption on income, based on the existence of the consumption functions detailed in Equation (1), is well in line with the literature on exogenously incomplete markets.[23] These functions formalize the idea that households participate in a risk-sharing arrangement, allowing them to pool their agricultural incomes to insure against idiosyncratic shocks and are flexible enough to capture a whole range of possible risk-sharing arrangements, from no sharing to full insurance.[24] Note that risk-sharing does not have to be egalitarian: the Pareto weights $(\nu_i)_i$ allow different types of households to receive different fractions of the constant aggregate output. We deliberately sidestep detailed explanations of the underlying reasons for the specific forms of the consumption functions, which determine the level of insurance in the economy, focusing instead on how different degrees of insurance influence the equilibrium in the land market.

[23]Contrast the approach where $(c_i)_i$ are primitives of the model with the perspective taken in the literature on optimal risk-sharing (Townsend, 1994) and endogenously incomplete markets (Sleet, 2006), where agents' consumption functions are derived from optimal consumption allocation problems featuring deeper primitive constraints on monitoring or enforcement technologies.

[24]While our empirical exercise assumes that the village is the relevant risk-sharing unit, we may think of it as a caste or kinship network at this point.

B Proofs

Proof of Claim 1. The first-order conditions for $c_i(\boldsymbol{\rho})$ read as follows:

$$\nu_i \left(c_i\left(\boldsymbol{\rho}\right)\right)^{-\sigma} - \lambda = 0,$$

or, equivalently,

$$c_i\left(\boldsymbol{\rho}\right) = \lambda^{-\frac{1}{\sigma}} \nu_i^{\frac{1}{\sigma}}, \tag{10}$$

where λ is the Lagrange multiplier attached to the feasibility constraint. Thus, each farmer's consumption is constant across states of the world $\boldsymbol{\rho}$. Integrating the last equation over all farmer types j and states of the world ρ, we get that

$$\int\int c_j\left(\boldsymbol{\rho}\right) \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) \mathrm{d}j = \int\int \left(\frac{\nu_j}{\lambda}\right)^{\frac{1}{\sigma}} \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) \mathrm{d}j = \lambda^{-\frac{1}{\sigma}} \int \nu_j^{\frac{1}{\sigma}} \mathrm{d}j.$$

Combine this equation with the feasibility constraint to obtain

$$\lambda^{-\frac{1}{\sigma}} = \frac{\int\int y_{j\rho} \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) \mathrm{d}j}{\int \nu_j^{\frac{1}{\sigma}} \mathrm{d}j}.$$

Substituting this expression back into Equation (10), we get

$$c_i\left(\boldsymbol{\rho}\right) = \frac{\nu_i^{\frac{1}{\sigma}}}{\int \nu_j^{\frac{1}{\sigma}} \mathrm{d}j} \int\int y_{j\rho} \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) \mathrm{d}j.$$

□

Proof of Claim 2. The first-order conditions for ℓ_i read as follows:

$$\iota + \lambda \int \frac{\partial y_{i\rho}}{\partial \ell_i} \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) = 0, \tag{11}$$

where ι is the Lagrange multiplier attached to the land availability constraint. Thus, the expected marginal products of land are equalized across households. To maximize aggregate expected output, we can solve the following programming problem:

$$\max_{\boldsymbol{\ell}} \int\int y_{i\rho} \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) \mathrm{d}i$$
$$\text{s.t.} \quad \int \ell_i \mathrm{d}i = L.$$

The first-order conditions for ℓ_i and ℓ_j imply that:

$$\int \frac{y_{i\rho}}{\partial \ell_i} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}) = \int \frac{y_{j\rho}}{\partial \ell_j} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho});$$

i.e., an allocation of land that maximizes aggregate expected output is such that the expected marginal products of land are equalized across households. □

Proof of Claim 3. Under no sharing, the first-order conditions for ℓ_i read as follows:

$$\int (c_i(\rho))^{-\sigma} \left(\frac{\partial y_{i\rho}}{\partial \ell_i} - r^{IM} \right) \mathrm{d}Q_\rho(\rho) = 0.$$

Thus, unless the households are risk neutral ($\sigma = 0$), the expected marginal products of land are not necessarily equalized across households. □

Proof of Theorem 1. For each $\beta \in (0, 1]$, the first-order conditions for ℓ_i are

$$\int (c_i(\boldsymbol{\rho}))^{-\sigma} \frac{\partial c_i(\boldsymbol{\rho})}{\partial \pi_{i\rho}} \left(\frac{\partial y_{i\rho}}{\partial \ell_i} - r^P \right) \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}) = 0,$$

where

$$\frac{\partial c_i(\boldsymbol{\rho})}{\partial \pi_{i\rho}} = \exp\left\{ \beta \log(\pi_{i\rho}) + (1-\beta) \log\left(\frac{\nu_i^{\frac{1}{\sigma}}}{\int \nu_j^{\frac{1}{\sigma}} \mathrm{d}j} \int \int \pi_{j\rho} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}) \, \mathrm{d}j \right) \right\} \frac{\beta}{\pi_{i\rho}}.$$

Letting

$$T_{i\rho} = (c_i(\boldsymbol{\rho}))^{-\sigma} \frac{\partial c_i(\boldsymbol{\rho})}{\partial \pi_{i\rho}}$$

we can rewrite these first-order conditions as

$$\int T_{i\rho} \left(\frac{\partial y_{i\rho}}{\partial \ell_i} - r^P \right) \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}) = 0.$$

This equation shows that the effect of partial insurance on optimal land cultivation choices can be interpreted as the introduction of type- and state-specific distortions, affecting the marginal return of land in each state of the world. These distortions imply that the expected marginal returns of land are not equalized to zero across farms. Instead, they vary in proportion to the type- and state-specific distortions. Notice that these distortions become more pronounced the further $T_{i\rho}$ deviates from being constant across states of the world. Since

$$\log\left(\frac{\nu_i^{\frac{1}{\sigma}}}{\int \nu_j^{\frac{1}{\sigma}}\mathrm{d}j}\int\int \pi_{j\rho}\mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right)\mathrm{d}j\right)$$

is a constant, an increase in β amplifies the variance of $T_{i\rho}$ across states of the world. $\square$

C Data

We use information from the "Village Dynamics Studies in South Asia" (VDSA) project by ICRISAT, a widely used panel data set (Townsend (1994), Mazzocco and Saini (2012), and Morten (2019), among many others). The data is collected through different modules (general endowment, cultivation schedule, rainfall schedule, among others) containing questions on different topics generally asked to the household head. Questions asked only to the household head generally refer to information about the whole household. Some modules ask questions to a greater subset of the household members (e.g., the questions in the employment schedule are asked to all members who completed 6 years of age).

Most modules are collected at a monthly frequency (e.g., the employment schedule) while others only come at a yearly frequency (for instance, the general endowment schedule, and the questions in the cultivation schedule that refer to agricultural output). We use data from July 2010 to June 2015. We aggregate the individual-level data to the household level. We end up with monthly household-level panel data, containing information on farming, expenditure, and income for families in 18 villages in the Indian semi-arid tropics.

General endowment schedule. This schedule provides annual, individual-level data on various characteristics of household members, including age, sex, education, and primary and secondary occupations. Additionally, it offers household-level details on landholdings, such as ownership status, total and irrigable areas, and various soil characteristics. We leverage the data on these characteristics to build the measures of average soil depth, slope, fertility, and degree of degradation introduced in Subsection 3.3. The schedule also contains yearly household-level data on livestock, farm equipment, buildings, durable consumption goods. stocked items (like crops, cooking fuel, and agricultural inputs), assets, liabilities, gender roles, and coping strategies employed in response to self-reported negative income shocks.

We employ the individual-level demographic data in this schedule to construct an age-sex index at the household-year level, following the methodology described in Townsend (1994). Specifically, we assign individual weights based on age and sex as follows: 1 for males over 18 years, 0.9 for females over 18 years, 0.94 for males aged 13 to 18, 0.83 for females aged 13 to 18, 0.67 for children aged 7 to 12, 0.52 for children aged 4 to 6, 0.32 for toddlers aged 1 to 3, and 0.05 for infants under one year. We then calculate the household-year age-sex index by aggregating these weights for each household annually. In Subsection 3.5, we utilize this index to adjust household-level consumption and income variables to per capita terms.

The general endowment schedule provides detailed information on each household's landholdings annually, with the plot as the unit of observation. This includes data on ownership status—whether owned, leased, shared, or mortgaged—and identifies the household members associated with each plot. It also details both total and irrigable areas, proximity to the house, irrigation sources and their distances. Additionally, the schedule provides information on plot attributes such as soil type (e.g., red, shallow black, medium black, deep black), fertility (an ordered scale from 0 to 4), slope (an ordered categorical variable indicating the degree of the plot's slope), soil degradation (a categorical variable indicating whether the plot is subject to soil degradation and specifying its type), and average soil depth in centimeters. The schedule also notes the presence of bunds, number of trees, if the plot is owned or leased, potential sale revenue, actual rent paid or received, and an imputed rental value for owned plots. We utilize the detailed information on plot attributes to construct measures of average soil depth, slope, fertility, and degree of degradation, which we use to construct a measure of land quality used in the estimation of household-farms' physical productivities (see SubSection 3.3). Specifically, define fertility_{pit}, slope_{pit}, degradation_{pit}, and depth_{pit} as the fertility, slope, degradation, and soil depth for plot p cultivated by household i in year τ. For each $x \in \{\text{fertility}, \text{slope}, \text{degradation}, \text{depth}\}$, we define

$$x_{it} = \sum_{p \in P(it)} \frac{\ell_{pit}}{\sum_{p' \in P(it)} \ell_{p'it}} x_{pit},$$

where $P(it)$ is an index set for the set of plots cultivated by household i in year τ.

Finally, we leverage the farm equipment section of the general endowment schedule to construct a measure of farm capital. Each year, the household head is asked to report the names and values of all farm equipment items owned by the household, including plows, sprayers, dusters, electric motors, diesel pumps, bullock carts, tractors, trucks, threshers, pipelines, rice mills, and flour mills, among others. We aggregate this data at the household-year level to create a measure of farm capital for each household annually. We use this variable in the estimation of household-farms' physical productivities, as detailed in Subsection 3.3.

Cultivation schedule. The cultivation schedule is divided into two main sections: inputs and outputs. We start with an overview of the input section. This part of the schedule gathers detailed monthly data on the inputs utilized by each household for every operation conducted on each plot they farm. Specifically, interviewers asks the household head to detail all operations carried out on each cultivated plot in each month. For every operation, they collect data on the quantities and costs of the inputs used. Thus, the unit of analysis for this section is the operation conducted on each plot by each household each month.

One fundamental piece of information we can obtain from the input section of the cultivation schedule is the household's labor supply to their farm, which we use in the estimation of household-farms' physical productivities, as detailed in Subsection 3.3. This section meticulously details the total labor hours devoted to each farming operation, categorizing them by family, hired, and exchange labor—specifically distinguishing contributions from females, males, and children—as well as labor provided by bullocks, motors, and other sources. To calculate the total labor supplied by a household to their farm, we aggregate the labor hours contributed by family members to each operation across all plots at the household-month level.

The output section gathers data differently, focusing not on individual operations each month but on crop production each season for each plot. Specifically, interviewers collect information from household heads regarding the quantity (in kilos) and value of each crop harvested during the defined agricultural seasons: Rabi, Kharif, annual, perennial, and summer. A critical measure derived from this section is the total annual output quantity per household, which serves as the dependent variable in Equation (2). To obtain this variable, we compute the total output produced by each household across all cultivated plots and each season throughout the year, aggregating this data at the household-year level. Another crucial variable compiled from the output section of the cultivation schedule is the total size of the plots cultivated by each household annually. We use this variable as a measure of land size in the estimation of household-farms' physical productivities, as explained in Subsection 3.3. An advantage of this variable is that it reflects the total land cultivated by the households, independent of ownership title, legal status, or other formal distinctions (see Subsection 3.2).

Transaction schedule. This schedule meticulously chronicles every monetary in-flow and out-flow for each household on a monthly basis. The expenditure segment collects information on both food and non-food purchases, enabling us to compile a measure of total monthly household consumption by aggregating the values of these purchases. To derive a measure of monthly household income, we adopt Mazzocco and Saini (2012)'s budget-constraint approach. Specifically,

- We use the section on financial transactions to track monthly household cash flows from lending and borrowing activities.
- The section on loans allows us to keep track of monthly inflows from loans and repayments by the household.

- From the section on government benefits, we determine monthly receipts of state-provided aid to households.
- The section on product and livestock sales allows us to measure the monthly revenue households earn from agricultural and livestock sales.
- The section on asset sales and purchases allows us to monitor the households' monthly financial activities related to the trading of capital goods.

We construct our measure of total monthly household income, we compute:

$$\begin{aligned}\text{Income}_{it} = &\ \text{Consumption}_{it} - \text{Cash received}_{it} + \text{Cash lent}_{it} \\ &- \text{Loans received}_{it} + \text{Loans repaid}_{it} - \text{Government benefits received}_{it}.\end{aligned}$$

We use the age-sex index defined above to convert the household-level monthly consumption and income variables to per capita terms. These variables are employed in Subsection 3.5 to test for the presence of full insurance at the village level.

Rainfall schedule. The rainfall schedule provides detailed information on rainfall levels (measured in millimeters) for each village daily, derived from readings at the nearest weather station. We aggregate these daily measurements over a year, to generate total village-specific annual rainfall. This aggregation yields the total annual rainfall for each village, denoted as rain_{vt}. We utilize this variable to parameterize the impact of observable environmental shocks on output, as specified in Equation (4).

D Agricultural output decomposition

In Table 7, we decompose the variance of agricultural output into its different sources. We employ two measures. In column (1) we report the R-squared from regressing household-farms' annual physical output on various production inputs separately, one by one. The R-squared suits our purpose as it indicates the proportion of variance in the dependent variables explained by the regressors.

In column (2) we report the Shapley value (expressed in %) of each production input. The Shapley value quantifies the average marginal contribution of each variable to the explained variance in agricultural output, considering all possible combinations of the explanatory variables. Each Shapley value is computed by averaging the incremental changes in R-squared when an explanatory variable is added to a subset of other variables across all possible subsets.

Table 7: Input contributions to agricultural output

Variable	R-squared (1)	Shapley value (%) (2)
Rainfall shocks: Var $\left[\widehat{\vartheta}\text{rain}_{v\tau}\right]$	0.004	0.160
Land quality, Var $[\log \widehat{q}_{i\tau}]$	0.103	3.480
Household-farms' physical productivities: Var $\left[\log \widehat{\theta}_i\right]$	0.614	41.68
Family labor: Var $[\log h_{i\tau}]$	0.382	19.07
Capital: Var $[\log k_{i\tau}]$	0.292	10.10
Landholdings: Var $[\log \ell_{i\tau}]$	0.277	10.05

Notes: Column (1) reports the R-squared from regressing household-farms' annual physical output on various production inputs separately, one by one. Column (2) reports the Shapley value (expressed in %) of each production input.

Differences in estimated physical productivity stand as the major sources of variation in production yield across household-farms. Using the R-squared, differences in estimated physical productivity can explain around 60% of the variation in annual yields across households. Using the Shapley value, more than 40% of total output variation across farms can be attributed to differences in estimated physical productivity.

E Imperfect risk-sharing in the ICRISAT villages

Table 8 reports our estimates for equation (7). To complement these results, we also estimate the following regression for household i within village v in month t (following Townsend (1994)):

$$\Delta \log c_{it} = \psi + \beta \Delta \log \pi_{it} + \zeta_{vt} + \epsilon_{it} \tag{12}$$

Here, $\Delta \log c_{it}$ and $\Delta \log \pi_{it}$ denote changes in log per-capita consumption and log per-capita income, respectively, for household i between two consecutive months. We estimate equation (12) through ordinary least squares and using the first three lags of the change of log income as an instrument to correct for measurement error.[25]

Table 8: Risk-sharing in the ICRISAT villages

	$\ln c_{it}$	$\Delta \ln c_{it}$	
	OLS	OLS	IV
	(1)	(2)	(3)
$\ln \pi_{it}$	0.223***		
	(0.0181)		
$\Delta \ln \pi_{it}$		0.206***	0.230***
		(0.0194)	(0.00483)
Household FE	✓		
Village-month FE	✓	✓	✓
Observations	46369	41263	29601
R-squared	0.681	0.319	0.193
First-stage F-statistic			574.54

Notes: The unit of analysis across all columns is the household year. The first column presents the results of regressing monthly consumption on monthly log income while controlling for village-month fixed effects. The second column presents a regression analysis where the monthly change in log consumption is regressed against the monthly change in log income, controlling for village-month fixed effects. The third column shows the second-stage results of an instrumental variable regression, where the first three lags of the monthly change in log income serve as instruments for the contemporaneous change in log income. Standard errors (in parentheses) are clustered at the village-year level across all columns.

[25]This strategy was inspired by (Ravallion and Chaudhuri, 1997), who instrument changes in household income using year dummies and the changes in the components of household income from sources other than agricultural cultivation. Our approach is similar to Anderson and Hsiao (1981)'s instrument in panel data econometrics.

F Instrumental variable strategy

We claim that risk-sharing reduces misallocation in the land market. In Section 3, we provide evidence that risk-sharing is negatively correlated with land misallocation at the village-year level. In this section, we propose an instrumental variable strategy that may get us closer to identifying the causal effect of risk-sharing on land misallocation. Specifically, we exploit information on the caste composition of each village as an instrument for risk-sharing. We propose that villages with a more homogeneous caste structure are better able to provide insurance to their villagers. Considering relevance first, several studies (Munshi and Rosenzweig (2016), Mazzocco and Saini (2012), Munshi (2019)) argue that caste networks play an important role in providing credit and insurance to its members. Secondly, exclusion restriction requires caste composition to be uncorrelated with land misallocation once we control for risk-sharing. To satisfy this criterion, we must ask whether caste networks tend to be more homogeneous in villages that have less land misallocation, irrespective of the capacity of more homogeneous networks to enhance the functioning of credit and insurance markets. A possible concern is that households permanently migrate between villages to capitalize on the benefits of operating in more efficient rural land markets. Even though permanent migration from one rural area to another is not completely uncommon in India, caste composition in our villages is very persistent, with an average coefficient of variation below 0.10. Thus, it seems unlikely that differences in the ability of village factor markets to efficiently allocate production factors among households would affect caste composition through permanent migration. On the other hand, although the caste structure of a village may certainly influence its ability to access resources (e.g., members of upper castes may have better access to financial products, such as loans), we believe there is no clear reason to suggest that it *directly* impacts the extent to which the community manages to efficiently allocate factors of production. Certainly, one could argue that taste-based discrimination could influence how factors of production are allocated among competing uses, thus suggesting a direct role for caste diversity in the extent of land misallocation in village economies. However, the available evidence, particularly in urban labor markets, indicates that caste discrimination is statistical (Munshi (2019)), and there is no compelling evidence to suggest that discrimination patterns would differ significantly in other factor markets, including those in rural economies.

Below we highlight our instrumental variable strategy in detail. The ICRISAT data records each household head's caste (*jati*), sub-caste, and caste group (backward caste, forward caste, nomadic tribe, other backward caste, scheduled tribe, and special backward caste). Although we share Munshi (2019)'s concerns regarding the analysis of caste networks based on broad caste groups rather than *jatis*, our limited sample size necessitates the use of these broader categories. Let $\mathcal{J}$ be an index set for the set of caste groups, where j denotes a typical element in this set. Let n_{jv} be the number of households belonging to a caste group j in village v, with $\sum_{jv} n_{jv} = n_v$. We construct a measure of caste diversity in village v as 1 minus the Simpson's Diversity Index:

$$z_v = 1 - \frac{\sum_{j \in \mathcal{J}} n_{jv} \left(1 - n_{jv}\right)}{n_v \left(1 - n_v\right)}.$$

Because of the very limited variation over time in caste composition, we let our instrumental variable vary across villages only.

Finally, we estimate the following equation independently for each village v and year τ

$$\log c_{it} = \beta_{v\tau} \log \pi_{it} + \mu_i + \kappa_t + \epsilon_{it}.$$

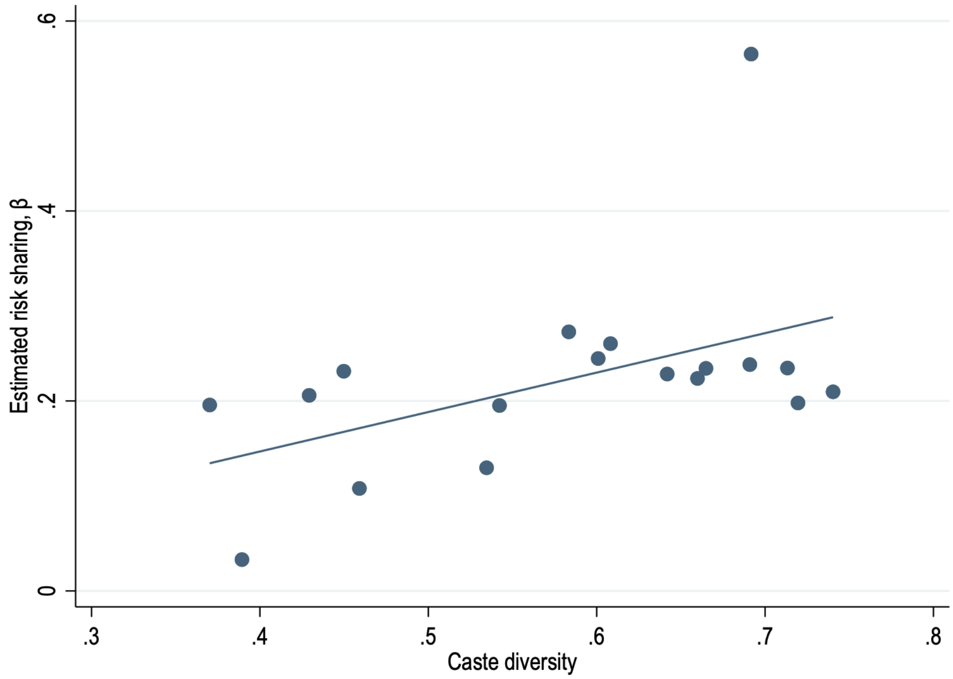
Then, we relate our two measures of misallocation in each village and year, denoted by $\omega_{v\tau}$, to the elasticities of consumption with respect to idiosyncratic income shocks at the village level, $\widehat{\beta}_{v\tau}$:

$$\omega_{v\tau} = \varrho + \gamma \widehat{\beta}_{v\tau} + \mu_\tau + \epsilon_\tau,$$

where we use z_v as an instrumental variable for $\widehat{\beta}_{v\tau}$.

Figure 6 scatters the first-stage regression: caste diversity is significantly and positively correlated with the estimated elasticity of consumption with respect to income at the village and year level. Table 9 reports first and second-stage IV estimates together with F-stat. A one-unit increase in our measure of caste diversity results in a 0.440-point increase in the village-specific elasticity of consumption with respect to idiosyncratic income shocks. Our instrumental-variable strategy suggests that a one-unit increase in this elasticity results in a 1.331-point decrease in the correlation between farm size and productivity and a 2.095-point increase in the dispersion of the marginal product of lands.

Figure 6: Risk-sharing and caste diversity



Source: VDSA survey (ICRISAT) and own calculations.

Table 9: Land misallocation and risk-sharing

	corr.$_v$ $\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$ (1)	st.dev.$_v$ $[\log \text{MPL}_{i\tau}]$ (2)
$\widehat{\beta}_v$	-1.331**	2.095***
	(0.574)	(0.579)
Observations	90	90
R^2	0.783	0.8584

First-stage regression	$\widehat{\beta}_v$ (1)
z_v	0.415***
	(0.126)
Observations	90
F-statistics	10.87

Notes: Standard errors are robust. Source: VDSA survey (ICRISAT) and own calculations.

G Risk-sharing vs. output distortions

In this appendix, we discuss the robustness of our results to the inclusion of generic household-specific distortions (wedges) in the output market. To do so, we assume household of type i are subject to an "output tax" τ_i that is correlated to its productivity, θ_i, and equal to

$$\tau_i = 1 - \theta_i^{-\zeta}, \tag{13}$$

where ζ governs the correlation between output distortions and farmer productivity. Then, the output of a farmer of type i is

$$y_{i\rho} = (1 - \tau_i)\,\theta_i \rho \ell_i^{\alpha} = \theta_i^{1-\zeta} \rho \ell_i^{\alpha}.$$

Notice that when $\zeta > 0$, distortions are positively correlated with household productivities, meaning high-productivity households face relatively higher output taxes. Conversely, when $\zeta < 0$, high-productivity households face relatively lower output taxes. When $\zeta = 0$, there are no output distortions in the economy, and output production reverts to the scenario described in the main text.[26]

To quantify the gains from full insurance using the model described in this appendix, we estimate two parameters: the coefficient of relative risk aversion, σ, and the correlation between productivity and output wedges, ζ. All other parameter values are as in Section 4. As in Section 4, we estimate σ by matching the average correlation between log farm size and log productivity, while ζ is estimated by targeting the share of households operating land smaller than 5 hectares.

Table 10: Estimated parameters

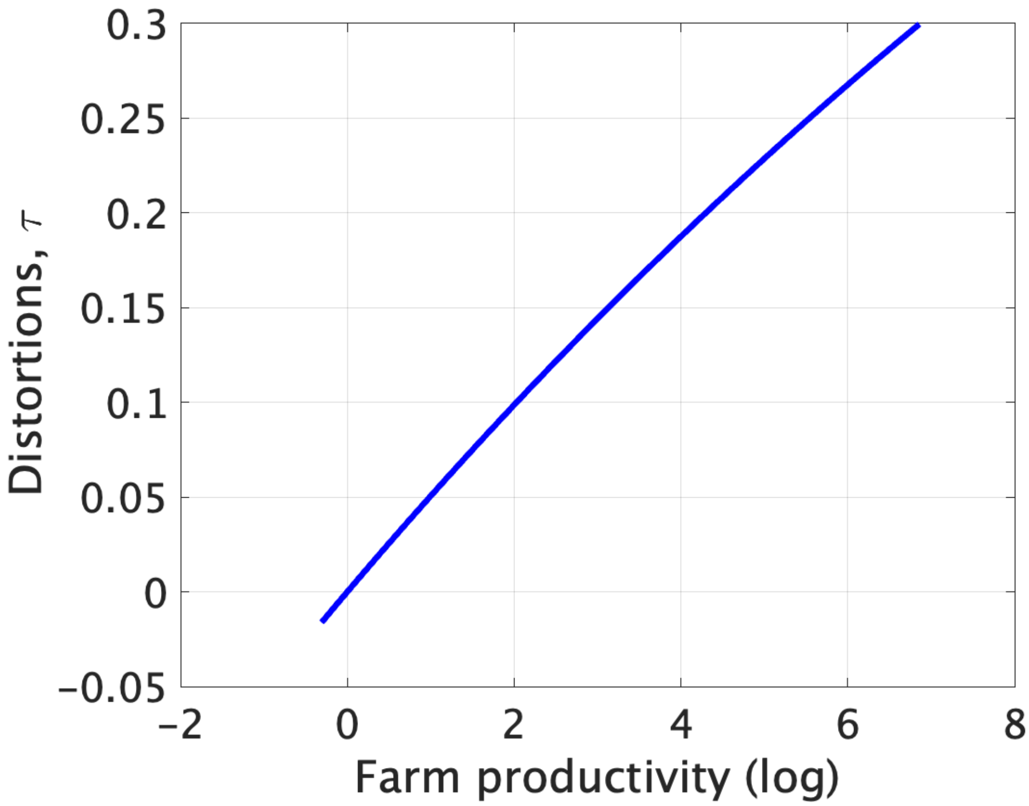
Parameters	Description	Value	Target	Data	Model
σ	Relative risk aversion	1.647	$\overline{\text{corr.}}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$	0.461	0.452
ζ	Distortion correlation	0.052	Land $\leq$ 5 hectares, share of households	0.801	0.810

Notes: This table reports the estimates for the coefficient of relative risk aversion and the correlation between output distortions and farm productivity, and the targets used in estimation; i.e., the average correlation between log farm size and log productivity and the share of households operating with land smaller than 5 hectares. Source: VDSA survey (ICRISAT) and own calculations.

Table 10 presents the estimates of σ and ζ, together with the empirical and simulated values of their respective targeted moments. The estimated coefficient for σ is slightly

[26] For some combination of ζ and θ_i, τ_i can be negative. In this case, distortions take the form of an output subsidy towards household i. For further applications of function (13) to describe firm-level output distortions in developing countries, see Guner and Ruggieri (2022), among others.

Figure 7: Estimated distortions



Source: VDSA survey (ICRISAT) and own calculations.

higher than the value obtained in the model without distortions, at 1.65 compared to 1.60. We estimate ζ at 0.052, which implies a positive correlation between distortions τ_i and (log) productivity θ_i across households of about 0.9. Figure 7 shows this pattern: distortions take the form of an output tax as big as 30% for households with the highest productivity and of a subsidy of 1.5% for households with the lowest productivity.

Table 11 reports the outcomes of the same counterfactual exercise described in Section 4 but in a model with output distortions. Column 1 refers to a baseline scenario where consumption insurance is partial, ($\beta = 0.223$), and households' land decisions are distorted by wedges that are correlated to their productivity ($\zeta = 0.052$). Columns 2 refer to a counterfactual scenario where risks sharing is perfect ($\beta = 0$), keeping everything else equal. Columns 3 refer to a counterfactual scenario where risks sharing is perfect ($\beta = 0$) and distortions are absent ($\zeta = 0$).

As shown, the aggregate output gains from improving risk-sharing in an economy with output distortions are 18%, nearly identical to the 19% gains in an economy without these wedges, as presented in Section 4. The welfare and efficiency gains of full insurance,

Table 11: Counterfactual exercise

	Baseline	Counterfactual	
	(1)	(2)	(3)
β	0.223	0	0
ζ	0.052	0.052	0
Aggregate efficiency (output per hectare)	1	1.393	1.896
Aggregate output	1	1.186	1.614
Aggregate welfare	1	1.137	1.158

Source: VDSA survey (ICRISAT) and own calculations.

when accounting for output distortions, are 13% and 39%, respectively, compared to 29% and 45% in the results presented in Section 4. Thus, while neglecting output distortions may slightly overstate the welfare gains from improving risk-sharing, the aggregate output and efficiency gains remain virtually unchanged, whether or not these wedges are considered.