Risk-Sharing and Land Misallocation*

Davide Pietrobon

Universitat de Barcelona

Alessandro Ruggieri

Institut d'Anàlisi Econòmica (IAE-CSIC) and Barcelona School of Economics

May 26, 2026

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' land cultivation 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 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

*The authors thank Arpad Abraham, Giacomo De Giorgi, Andre Groeger, Maia Güell, Nezih Guner, Jonna Olsson, Nicola Pavoni, Josep Pijoan-Mas, Sevi Rodríguez-Mora, Raul Santaeulalia-Llopis, and seminar participants at CEMFI, CUNEF Universidad, Universitat Autonoma de Barcelona, the Workshop of the Spanish Macro Network, the Spanish Workshop on Development Economics, and the BSE Summer Forum, for their useful comments. Alessandro Ruggieri acknowledges financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2024-001476-S). 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), temporary migration (Morten, 2019), and engagement in non-farm entrepreneurship (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. While existing literature emphasizes frictions in land markets—such as inheritance rules, tax policy, and tenancy regulation—as key sources of misallocation (Adamopoulos and Restuccia, 2014; Adamopoulos et al., 2022), we show that limited consumption risk-sharing generates substantial land misallocation, output reductions, and welfare losses, *beyond* those attributable to factor market frictions. Our quantitative findings for rural India imply that completing insurance markets can greatly improve the allocation of land across farmers, leading to output and welfare gains of 19% and 29%, respectively.

In village economies, shocks from harvest failures, illness, and pests leave households vulnerable to severe hardship. Because formal borrowing and saving are very limited in these environments, insurance against idiosyncratic income fluctuations largely relies on informal arrangements such as gift exchanges and personal loans.[1] 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). Lack of insurance not only affects consumption but it can also have a distortionary effect 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*

[1] See Dercon (2002) for a discussion of the sources of idiosyncratic income risks and coping strategies in rural economies.

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 to cultivate to maximize expected profits. Thus, each farmer is driven to the familiar condition of equating the expected marginal product of land to its price, 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.

Building on the foundational works of Singh et al. (1985) and Singh et al. (1986), 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 household 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 landholdings. Under full insurance, each household-farm's equilibrium land choices maximize its expected income, making it impossible to redistribute land from one household to another without lowering aggregate expected income. Incomplete insurance distorts these choices away from expected income maximization by increasing the weight households place on states of the world in which income is low. As a result, the equilibrium expected marginal products of land are not equalized across households—that is, land is misallocated.

We provide suggestive evidence consistent with our model using the latest ICRISAT monthly panel data (2010–2014) from the Indian semi-arid tropics, offering evidence linking risk-sharing to land misallocation. First, we measure risk-sharing 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: 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 former is a well-known measure of allocative efficiency in the land market, where a higher value indicates that more productive farms cultivate more land (Chen et al., 2023). The latter 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. While we do not claim to estimate any causal effect, we provide suggestive evidence that villages with a more homogeneous caste structure are characterized by higher risk sharing and, in turn, better allocative efficiency of land, and we show that this correlation is robust to the inclusion of measures of asset accumulation, access to formal credit markets, and risk diversification.

Finally, we 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 across villages as an untargeted moment. Armed with the structural estimates, we conduct a counterfactual analysis to explore the impact of enhancing the functioning of 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. Output per unit of land increases by 41.5% 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 in rural China increases agricultural TFP by 53%. We check the robustness of our findings to incorporating farm-specific distortions, which capture the impact of land and output market frictions that disproportionately affect high-productivity households. We find that the aggregate output and efficiency gains from completing insurance remain substantial, even in this alternative framework, and account for between 29% and 45% of the overall efficiency gains that can be achieved from moving to a fully undistorted economy. Thus, 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.

Our analysis yields an important policy insight: when frictions operate in both factor and insurance markets, reforms that prioritize improving the functioning of factor markets need not deliver the largest welfare gains. The misallocation literature typically focuses on frictions that distort the prices faced by individual producers, such as transaction costs or weak property rights, leading to straightforward policy recommendations—such as land titling or rental market liberalization—aimed at reallocating land from low-productivity to high-productivity farmers. However, under incomplete risk-sharing, improving the functioning of insurance markets delivers dual gains: it promotes reallocation from less to more productive farmers, improving allocative efficiency, while also providing insurance to risk-averse households. Our quantification indicates that, while removing farm-specific distortions generates larger output gains (40.7% vs. 17.8%), completing insurance markets yields substantially larger welfare gains (13.7% vs. 6.5%). This result underscores the importance of explicitly accounting for frictions in multiple markets when assessing the welfare consequences of alternative policies aimed at improving the functioning of one particular market.

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, suggesting 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.[2] 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). 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 on a related mechanism 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 agricultural inputs, we highlight how these imperfections result in land misallocation.

Finally, our paper contributes to the understanding of how land gets allocated to farm-

[2]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.

ers 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.

Our paper relates to an extensive literature on the effects of risk on agricultural input use and technology adoption in developing countries (Feder et al., 1985; Foster and Rosenzweig, 2010; Magruder, 2018). Incomplete insurance distorts farmers' production choices by constraining input use below the level at which expected marginal benefits equal expected marginal costs, reducing investment in activities with high expected profits. Empirical evidence shows that credit provision or index insurance can relax the constraints imposed by distortions in insurance markets (Karlan et al., 2014; Emerick et al., 2016). Much of this literature centers on the same question that we also tackle in this paper—i.e., how uninsurable risk shapes production choices—but our focus is distinct both theoretically and empirically. From a theoretical standpoint, we focus on how incomplete insurance distorts the allocation of a *fixed* resource endowment among producers with heterogeneous productivities, rather than on how it depresses demand for risky inputs. In our framework, even when the aggregate supply of factors of production is fixed, incomplete insurance lead to a less efficient allocation of these factors across producers and, consequently, lower aggregate output. This misallocation channel is particularly transparent for land, which is available in fixed supply at the village level, in contrast to other inputs—such as machinery, fertilizer, and high-yielding crop varieties—which are typically available in elastic supply. For these more "mobile" inputs, changes in risk-sharing affect not only how efficiently they are allocated across farmers but also their aggregate use. Our empirical contribution is twofold. First, we use the well-known ICRISAT data to demonstrate that a specific type of insurance market friction—namely, incomplete risk-sharing—relates to misallocation of land among farmers. This evidence is novel: it shows that insurance markets are systematically related to the misallocation of factors of production across farmers, departing from the traditional focus of imperfect-insurance on technology adoption and instead focusing on the use of land—an input available in fixed supply at the village level. Second, we provide a new theoretical framework that links risk-sharing failures to misallocation in general equilibrium and we quantify the aggregate consequences of such frictions.

2 Model

We analyze a static economy in which households with heterogeneous productivities face idiosyncratic output shocks and can insure against these shocks by relying on a risk-sharing arrangement. Our theoretical framework is grounded in the seminal works of Singh et al. (1985) and Singh et al. (1986), which examine the behavior of households acting simultaneously as consumers and producers. To simplify the exposition, we consider an environment where land is the sole factor of production.[3] 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.[4,5] We use this model to illustrate how the degree of risk-sharing affects misallocation in the land market. All proofs are at the end of the paper. Appendix A.1 discusses some of our modeling choices. 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

[3] In Appendix A.2, we show that our results extend to economies that include additional inputs, such as materials and labor. Moreover, our model can be extended to include aggregate output shocks. Since these shocks are common across households, they can engage in risk-sharing to fully insure against the output variation coming from idiosyncratic shocks.

[4] We abstract from modeling borrowing or savings decisions, which are alternative means for households to self-insure. In village economies, there is often a restricted menu of savings instruments available (Alderman and Paxson, 1992). Moreover, in the context of rural India, households' ability to save in either financial or physical assets (such as livestock) appears to be heavily constrained (Rosenzweig and Wolpin, 1993; Morten, 2019).

[5] In Appendix A.3 we show that the effect of incomplete consumption insurance on factor misallocation carries over to environments where agents can borrow and save through a risk-free asset. See also Armangué-Jubert et al. (2025). This is also the case in the context we analyze, where the majority of households do not save (see Subsection 3.4).

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, $\widetilde{\ell}_i$. Let the total quantity of land available in the economy be $L = \int \widetilde{\ell}_i \mathrm{d}i$.[6,7] Households have identical preferences over consumption, represented by a 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 $\left[\underline{\rho}, \overline{\rho}\right] \subset \mathbb{R}_+$.

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

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

where ℓ_i is land cultivated by a household of type i, and $\alpha \in (0,1)$ denotes the land share; i.e., the elasticity of agricultural yields with respect to land cultivated.[8] Let r be the price of land, and let $\pi_{i\rho} = \theta_i\rho\ell_i^\alpha - r\left(\ell_i - \widetilde{\ell}_i\right)$ denote the income of a household of type i under output shock realization ρ. In the following, we assume that $\underline{\rho}$ is high enough so that π_{it} is bounded away from zero for all possible land cultivation choices.[9] 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 narrative in the misallocation literature, which typically attributes land misallocation to land or goods market distortions.

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

[7]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. In the empirical analysis, we account for heterogeneous land quality based on observable characteristics.

[8]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.

[9]This assumption guarantees that household consumption is always positive, regardless of the insurance regime, ensuring that utility remains well-defined. Alternatively, positive cash-on-hand can be ensured by introducing an asset and allowing households to borrow and save (see Appendix A.3).

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 (i.e., with full insurance) to one in which households are hand-to-mouth (i.e., with no risk-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 $Q_{\boldsymbol{\rho}}(\boldsymbol{\rho})$.[10] 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 $(\nu_i)_i$:

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

subject to the land availability constraint

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

and the feasibility constraint

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

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.[11] An optimal consump-

[10] 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.

[11] Pareto weights are type-specific, meaning that full risk-sharing does not necessarily result in an egalitarian allocation of consumption. In particular, households with larger land endowments or higher productivity may systematically consume more than those with smaller endowments or lower productivity.

tion 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: each household is allocated an amount of land such its 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 *ex-ante* 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:

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

subject to the budget constraint

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

where r^{IM} denotes the equilibrium price of land under no sharing. A competitive equilibrium for this economy is a plan for consumption, $(c_i(\boldsymbol{\rho}))_i$, an allocation of land holdings $(\ell_i)_i$ and a price of land r^{IM} such that:

- *land holding is chosen optimally; i.e.,*

$$\ell_i \in \arg\max_{\ell_i} \int \frac{\left(y_{i\rho} - r^{IM}\left(\ell_i - \widetilde{\ell}_i\right)\right)^{1-\sigma}}{1-\sigma} \mathrm{d}Q(\rho),\ \forall i;$$

- *land market clears; i.e.,*

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

Without risk-sharing, household's consumption and its marginal utility depend 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 internalize how land cultivation choices impact the distribution of consumption across different states of the world, the expected marginal product of land they cultivate will not equate its market price. When this is the case, there exists land misallocation—redistributing land across households can increase aggregate expected output.

Claim 3. *Under no sharing, there is ex-ante 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 *ex-ante* misallocation in the *land* market.

2.2 Partial insurance

More broadly, we can investigate the impact of risk-sharing on land misallocation for any degree of risk-sharing. 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 its price 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 the previous section. To model partial insurance, we define the following consumption function for a household i:

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

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. Beyond providing a flexible way to model the relationship between household consumption and income, the function in Equation (1) maps into the standard regression equation commonly estimated in the literature to test for efficient risk-sharing in village economies (Townsend, 1994; Chiappori et al., 2014). The following theorem shows that misallocation in the land market decreases with the degree of insurance.

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

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 households' marginal utility of consumption increasingly reflects the impact of its realized output shocks. This 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.

Specifically, two forces operate in the model. First, when risk-sharing is not perfect, land choices aim to increase expected utility subject to the downside risk of renting before uncertainty is realized. This makes high-productivity farmers—those whose expected utility is more sensitive to idiosyncratic shocks—to utilize relatively less land in order to avoid losses from unfavorable realizations of the shocks. A second force operates in general equilibrium: with land being in fixed supply, the reduced demand for land from high-productivity farmers in a low-risk-sharing scenario lowers the rental price of land relative to a scenario with higher risk-sharing. This, in turn, incentivizes low-productivity farmers—those whose expected utility is less sensitive to idiosyncratic shocks, to utilize relatively more land.

2.3 Robustness of theoretical results

The model in the previous section focuses on a setting where land is the only factor of production. This assumption is made purely for conceptual clarity: in Appendix A.2, we show that all of our results extend to a model that includes additional factors of production, such as materials and labor. Here, the key point is that as long as *some* factors of production are chosen before the realization of output shocks, incomplete consumption insurance distorts their allocation away from one that maximizes aggregate expected output. This result holds even if other inputs, such as labor, can be adjusted ex post. Moreover, insurance market incompleteness alone is sufficient to generate factor misallocation, even if *all* factor markets operate without frictions. Finally, as shown in Appendix A.3, the effect of incomplete consumption insurance on factor misallocation carries over to environments where households can borrow and save through a single risk-free asset.

3 Data

3.1 Background

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 and others annually. Importantly, the data allow us to construct *monthly* measures of consumption and income for households in different villages.[12] This feature makes it possible to estimate the level of risk-sharing specific to

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

each village and year. We exploit this possibility later in the section, where we relate the level of risk-sharing to the degree of land misallocation across villages and years. The data include information from 40 randomly selected households in each village, stratified by landholding size, and including 10 landless laborers, 10 small farmers, 10 medium farmers, and 10 large farmers. This classification is based on operational landholdings, defined as land owned and leased in, minus land leased out.[13] The data also provide detailed information on the quantity and value of all inputs and outputs in farm activities, as well as expenditures and incomes. We refer to Townsend (1994), Mazzocco and Saini (2012), and Morten (2019) for more detailed discussions of the data.[14] Appendix B describes the variables used in the analysis.

3.2 Land distribution

Central to our analysis is how land is distributed among farmers. Table 1 presents the distribution of cultivated land (in hectares) across farms in our data and compares it to the distributions in Malawi, Belgium, and the United States, as reported in the 1990 World Census of Agriculture (Adamopoulos and Restuccia, 2014).[15]

Compared to Belgium and the United States, the distribution of cultivated land in village India 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 aver-

[13] Because landless laborers, by definition, report zero land usage, we exclude them from our analysis.

[14] 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 years, 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.

[15] 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 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.

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

	2010-2014 ICRISAT	1990 World Census of Agriculture		
Farm size (hectares)	India	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	3.82	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 refer to the 2009-2014 ICRISAT panel data. The sample refers to Indian farms with positive amounts of cultivated land. Data for Malawi, Belgium, and the United States are from the 1990 World Census of Agriculture, as documented in Adamopoulos and Restuccia (2014).

age 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 Farmers' physical productivity

For our analysis, it is essential to accurately estimate each household-farm's physical productivity, also known as TFP-Q. We later use these estimates to infer land misallocation. Measuring productivity is a fundamental step in virtually all assessments of factor misallocation, whether direct or indirect (Restuccia and Rogerson, 2017). In turn, nearly all productivity measurements require imposing some structure—typically by specifying a production function. To assess household-farm productivity, we estimate the agricultural production function outlined in Appendix A.2, which incorporates materials and labor alongside land as factors of production. Specifically, consider the following agricultural production function:

$$y_{i\tau} = A_{v\tau}\theta_i\rho_{i\tau}k_{i\tau}^{\alpha^k}h_{i\tau}^{\alpha^h}\ell_{i\tau}^{\alpha^\ell}, \tag{2}$$

where $y_{i\tau}$ denotes the physical quantity of output (measured in kilograms) produced by household i in year τ; $A_{v\tau}$ captures factors common to village v and year τ; $k_{i\tau}$ is total value of materials; $h_{i\tau}$ denotes total hours of family labor dedicated to farming activities; $\ell_{i\tau}$ represents total land cultivated, adjusted for differences in quality.[16] We decompose the village-level aggregate productivity term as $A_{v\tau} = e^{\mu_\tau} e^{\vartheta \text{rain}_{v\tau}}$, where μ_τ capture common year-specific factors and $\text{rain}_{v\tau}$ denotes the amount of rain (in millimeters) in village v and year τ.[17] We assume that

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

where μ_i represents permanent unobserved heterogeneity specific to household i. As for the output shock, $\rho_{i\tau}$, we assume that

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

where $\epsilon_{i\tau}$ is an unobserved error term. Our approach involves treating the 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 re-arranging terms, Equation (2) becomes

$$\log(y_{i\tau}) = \alpha^k \log(k_{i\tau}) + \alpha^h \log(h_{i\tau}) + \alpha^\ell \log(\ell_{i\tau}) + \mu_i + \mu_\tau + \vartheta \text{rain}_{v\tau} + \epsilon_{i\tau}. \tag{5}$$

Finally, we assume $\epsilon_{i\tau}$ to be randomly distributed and estimate Equation (5) using OLS.

In Table 2, we report the dispersion of estimated farm productivity, $\log \widehat{\theta}_i$, among In-

[16] Specifically, we assume that $\ell_{i\tau} = q_{i\tau} a_{i\tau}$, where $a_{i\tau}$ is the area cultivated by household i in year τ, and

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

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 B for additional details on the construction of these variables). In Appendix C.3, we show that the findings in this section are robust to an alternative definition of land quality. Although land quality is notoriously difficult to measure, and any attempt to relate it to observable or unobservable characteristics is prone to specification error, it remains an unavoidable component in estimating the agricultural production functions.

[17] 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.

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 for the productivity of manufacturing plants 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 percentiles; 90-10 is the log difference between the 90 and 10 percentiles.

dian household farms, and compare it to the same dispersion among farms in Malawi (Chen et al., 2023) and the US (Adamopoulos and Restuccia, 2014), as well as to the dispersion of TFP-Q among manufacturing plants in both countries, as reported by Hsieh and Klenow (2009). There is substantial dispersion in TFP-Q among Indian farms: the standard deviation of log productivity (1.08) is comparable to that of farms in Malawi (1.18) and much larger than that of farms in the US (0.80). Moreover, variation in TFP-Q explains differences in agricultural output across Indian farms (see Table C1 in Appendix C.1).

3.4 Households' consumption and income

Finally, for our analysis, we need information on consumption and income across households and years. Table 3 reports the distributions of monthly real consumption and income across households and years in our sample. Both variables are deflated using the national CPI and expressed in 1975 Indian rupees.[18]

Households in the ICRISAT villages are poor, with an average real per capita income of 176 rupees per month, equivalent to 101 US dollars in 2016 terms, or about 3.4 US dollars

[18]In 1975, 8 Indian rupees were worth about 1 US dollar, which is about 4.60 dollars in 2016 (Bold and Broer, 2021).

Table 3: Income, consumption, and savings across households and years

	Observations (1)	Average (2)	Median (3)	St.dev. (4)	Min (5)	Max (6)
Household size	46,369	5.00	5	2.31	1	24
Age of household head	46,369	41.19	5 39	9.58	22	82
Income	46,369	702.61	395.22	2815.4	0	375208.9
Income, p.c.	46,369	176.00	97.434	716.89	0	65826.13
Consumption	46,369	495.98	373.08	976.06	24.9	59947.02
Consumption, p.c.	46,369	123.35	92.980	279.71	7.68	25477.59
Savings	45,781	1825.5	393.61	3706.5	0	31857.0
Savings, p.c.	45,781	482.97	102.34	1070.1	0	16196.6
No savings (dummy)	45,781	0.5057	1	0.500	0	1
Loans p.c., share of yearly income	43,462	0.346	0.095	0.783	0	8.895
Loans from banks (dummy)	43,462	0.675	1	0.468	0	1

Notes: Nominal variables are expressed in 1975 rupees. Per-capita income and consumption expressed in adult-equivalent terms, using the weights proposed in Townsend (1994). "No savings" is a dummy variable that equals 1 when the stock of savings is lower than monthly income. "Loans from banks" is a dummy variable taking value 1 when a farmer has an active loan from a bank within the year.

per day. Per capita real consumption averages 70% of real monthly income, amounting to 123 rupees (71 US dollars in 2016 terms). Per capita consumption is also much less dispersed than per capita income, whose standard deviation is twice as large. Finally, per capita savings are small: the median value is 102 rupees (59 US dollars in 2016 terms), and more than 50% of households in the sample have an overall stock of savings that is less than one month's worth of income. On the other hand, a significant share of farmers borrows from banks (about 67%), and the average amount borrowed amounts to 35% of the yearly income per capita.

4 Risk-sharing and misallocation in Indian villages

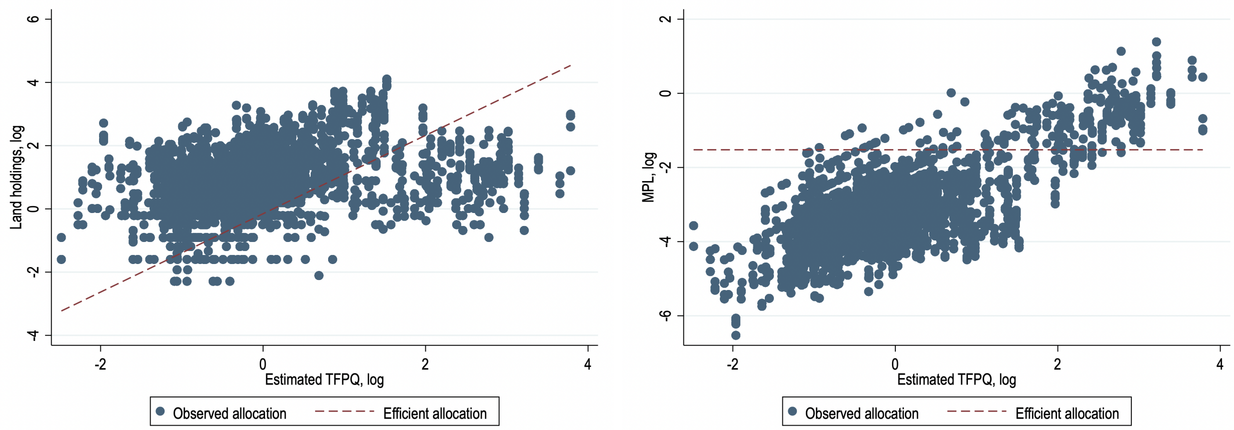
4.1 Land misallocation

We now present direct evidence of land misallocation among households in rural India. To do so, we employ data on operational (cultivated) landholdings, $\ell_{i\tau}$, together with our estimates of physical productivity (TFP-Q), $\widehat{\theta}_i$.

Figure 1 displays the allocation of operational lands across farmers. Specifically, panel A scatters the land operated by each farm against the estimated farm-level TFP-Q. Both

Figure 1: Land misallocation among Indian farmers

(A) Land size vs TPF-Q **(B)** (log) MPL



Notes: Panel A presents a scatterplot of the amount of land operated by each farm against estimated farm-level TFP-Q. Panel B presents a scatterplot of the marginal product of land for each farm against estimated farm-level TFP-Q.

variables are expressed in logs. Contrary to the efficient allocation, which predicts a tight relation between land size and TFP-Q, the correlation between the operational land size and farm TFP-Q is very low and equal to 0.29; i.e., low- (high-) productivity farmers operate relatively more (less) land relative to the efficient benchmark. As a result, the estimated marginal product of land is not equalized across farmers (Panel B), and the standard deviation of the log MPL amounts to 1.12. These findings are strong evidence of land misallocation across farmers in India.

We replicate the same analysis at a more disaggregated level. That is, we construct the unconditional correlation between the log of operational land size and the log of farm productivity; i.e.,

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

and the dispersion in the marginal product of land; i.e.,

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

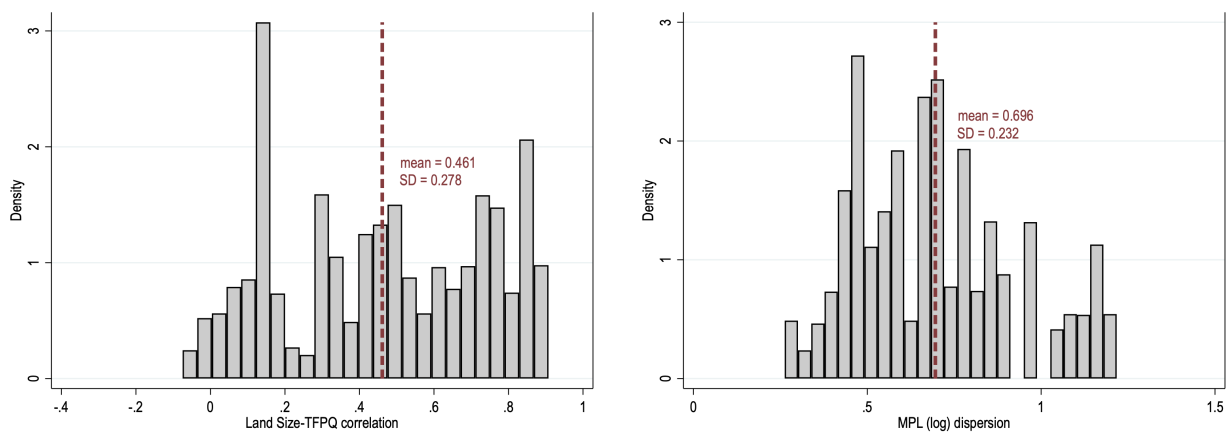
separately for every village v and year τ in the sample. As highlighted in Figure 1, low correlations and high dispersion of MPL are both suggestive of land misallocation.

There is a large variation in allocative efficiency of land across villages and years. Panel A in Figure 2 displays the distribution of estimated correlations. They range 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 2 reports the distribution of estimated marginal product dispersions across villages and years. Similarly to Panel A, the dispersion is small in certain village-year pairs, indicating little misallocation, while much greater in others.

Figure 2: Land misallocation in each village

(A) Land size-TPF-Q correlation

(B) Dispersion in the (log) MPL



Notes: This figure reports the distribution of misallocation measures, $\text{corr.}_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$ (panel A) and $\text{st.dev.}_{v\tau}\left[\log \text{MPL}_{i\tau}\right]$ (panel B), estimated for each village v and year τ. The red dashed line refers to the average of the distribution.

What accounts for the variation in land misallocation across villages and years? The theory developed in Section 2 suggests that misallocation in the land market is negatively correlated with the ability of households to share idiosyncratic risks. In what follows, we present evidence of imperfect risk-sharing among households, and show that land misallocation negatively correlate with the ability to share risks within villages.

4.2 Risk-sharing

It is commonly acknowledged that risk-sharing within villages in developing countries tends to be incomplete. This holds true for the ICRISAT villages as well. To see this, we consider how the per-capita consumption for household i in month t, c_{it}, comoves with its per-capita income, π_{it}. We start by outlining the construction of these two variables, which are critical to our measure of risk-sharing: the elasticity of consumption with respect to idiosyncratic income shocks. Rather than providing a direct measure of household consumption, the ICRISAT data record detailed individual expenditures on food and non-food items, as well as the monetary value of home-produced commodities that are consumed. This granular level of detail for each consumption item is one of the key advantages of this dataset, as it allows us to compile a comprehensive measure of household consumption by aggregating these recorded values at the household-month level. Notice that, as a result, our measures of household consumption are never imputed from household income, not even for the poorest households. This feature stands in contrast to many household surveys, which often rely on income-based measures of living standards (Deaton and Zaidi, 2002). To derive a measure of monthly household income, we adopt the budget-constraint approach of Mazzocco and Saini (2012). In particular, full income is calculated as total expenditure minus resources borrowed from various sources, plus resources saved in different accounts or lent to others, plus transfers given out, minus transfers received, plus taxes, and minus subsidies. To convert monthly household consumption and income into per-capita terms, we adjust them using the age-sex index proposed by Townsend (1994).[19] Then, we consider the following model:

$$\log c_{it} = \beta \log \pi_{it} + \chi_i + \chi_{vt} + \varepsilon_{it}. \tag{6}$$

In Equation (6), $\log c_{it}$ and $\log \pi_{it}$ denote the log per-capita consumption and log per-capita income, respectively, for household i in month t; χ_i are household fixed effects; and χ_{vt} represents village-month fixed effects that capture the average resources available to

[19]See Appendix B for additional details on the construction of the household consumption and income variables.

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 with respect 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.

Full risk-sharing is rejected. On average, 22.5% of idiosyncratic income fluctuations are passed through to consumption (Column 1, Table C2 in Appendix C.2). 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; Bold and Broer, 2021). The elasticities of consumption with respect to idiosyncratic income shocks are comparable across households with and without savings and across households with and without bank loans (Columns 2 and 3 of Table C2), suggesting that even households that accumulate assets or manage to access credit markets face difficulties insulating consumption from income shocks.

The average elasticity of consumption with respect to idiosyncratic income shocks in equation (6) 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 an elasticity of consumption with respect to income, $\beta_{v\tau}$, independently for each village v and year τ, using the following specification:

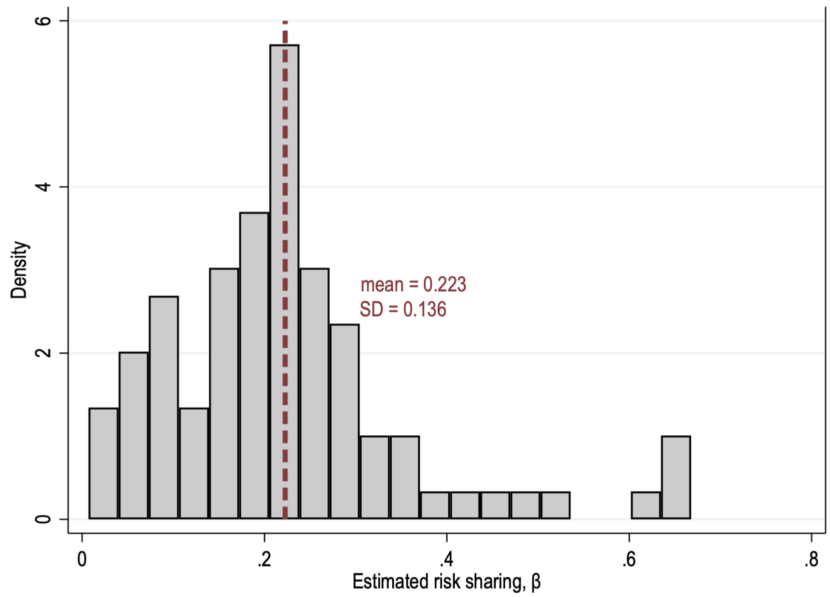
$$\log c_{it} = \beta_{v\tau(t)} \log \pi_{it} + \chi_i + \chi_t + \varepsilon_{it},$$

where index i denotes households within village v, while index t corresponds to the months within year $\tau(t)$, where $\tau(t)$ specifies the year associated with month t.[20] A high estimate of $\beta_{v\tau}$ indicates that *within village v during year τ*, the response of household consumption to household idiosyncratic income shocks was high—i.e., risk-sharing was low.

Figure 3 plots the distribution of the estimated elasticities of consumption with respect

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

Figure 3: Estimated degrees of risk-sharing in each village



Notes: This figure reports the distribution of risk-sharing parameters, $\widehat{\beta}_{v\tau}$, estimated for each village v and year τ. The red dashed line refers to the average of the distribution.

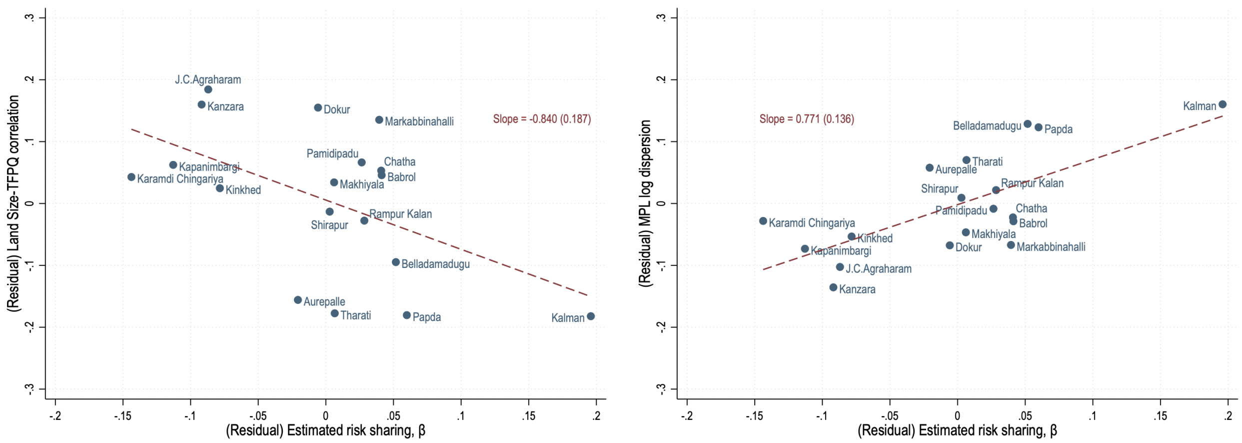
to idiosyncratic income shocks, $\widehat{\beta}_{v\tau}$, for each village-year pair. The average estimated $\widehat{\beta}_{v\tau}$ across villages and years is 0.223, indicating that, on average, a 1% idiosyncratic increase in income results in a 0.223% increase in consumption. As shown in the figure, 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 ($\widehat{\beta}_{v\tau} > 0.6$).

4.3 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 4.1 and denoted by $\omega_{v\tau}$, to our estimates of the elasticities of consumption with respect to idiosyncratic income shocks at the village-year level, $\widehat{\beta}_{v\tau}$, obtained in Section 4.2. In particular,

Figure 4: Risk-sharing and land misallocation across villages

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



Notes: Panel A scatter the (residualized) correlation between log farm size and log TFP-Q, $\text{corr.}_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$, against the (residualized) degree of risk-sharing, $\beta_{v\tau}$, across villages in the sample. Panel B scatters the (residualized) standard deviation of the log marginal product of land, $\text{st.dev.}_{v\tau}\left[\log \text{MPL}_{i\tau}\right]$, against the (residualized) degree of risk-sharing, $\beta_{v\tau}$, across villages in the sample. Residualized misallocation and risk-sharing are obtained controlling for village-level averages and dispersions of farm size, TFP-Q, and annual farmer income; total rainfall in the village; village-level shares of farmers without savings and of farmers with a banking loan. The point estimate of the slope is obtained with OLS. The standard error of the estimated slope (in parenthesis) is robust.

we estimate the following equation:

$$\omega_{v\tau} = \gamma \widehat{\beta}_{v\tau} + \varphi_v + \varphi_\tau + \varphi_v \times \tau + \Omega X_{v\tau} + \varsigma_{v\tau}, \tag{7}$$

where φ_v and φ_τ are village and year fixed effects, $\varphi_v \times \tau$ are village-specific linear year trends, $X_{v\tau}$ are various time-varying village-specific observables, and $\varsigma_{v\tau}$ are random disturbances.

Figure 4 scatters this relation along the cross-section of villages, after controlling for village-level observables. As we move from the village with the lower risk-sharing, (e.g., "Kalman") to the village with the higher one (e.g., "Karamdi Chingariya") land misallocation decreases: the correlation between TFP-Q and landholdings rises by 0.3 (panel A), while the dispersion of MPL drops by 0.2 (panel B).

Table 4 reports the estimation outcomes for different specifications of Equation (7), which exploits also time variation. There is a negative and significant correlation between land misallocation and the degree of risk-sharing. For example, the estimate in

Table 4: 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)	(9)	(10)
$\widehat{\beta}_{v\tau}$	-0.648***	-0.675***	-0.163**	-0.229***	-0.315**	0.732***	0.749***	0.161***	0.192***	0.269***
	(0.226)	(0.228)	(0.080)	(0.109)	(0.144)	(0.178)	(0.245)	(0.063)	(0.076)	(0.094)
Observations	90	90	90	90	90	90	90	90	90	90
R^2	0.106	0.118	0.897	0.942	0.958	0.218	0.223	0.902	0.948	0.964
Year FE		✓	✓	✓	✓		✓	✓	✓	✓
Village FE			✓	✓	✓			✓	✓	✓
Village time trends				✓	✓				✓	✓
Controls					✓					✓

Notes: The unit of analysis across all columns is a village-year pair. The first five columns present the results of regressing corr.$_{v\tau}$ $\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$ on the estimated village-and-year-specific consumption elasticities to idiosyncratic income shocks, $\widehat{\beta}_{v\tau}$. The following five columns show the results of regressing st.dev.$_{v\tau}$ $[\log \text{MPL}_{i\tau}]$ on $\widehat{\beta}_{v\tau}$. Controls include the village–year averages and dispersions of farm size, TFP-Q, and annual farmer income; total rainfall; and the shares of farmers without savings and with a banking loan in village v and year τ. Standard errors in parentheses are computed using village-level clustered bootstrap (5,000 replications), following the procedure in Cameron et al. (2008).

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 2, Panel A). The dispersion of the marginal product of land is negatively correlated with the degree of risk-sharing. The estimate in Column (9) suggests 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 $\text{MPL}_{i\tau}$. 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 2, Panel B).

These results are robust to controlling for village–year averages and dispersions of landholdings, productivity (TFP-Q), and annual income, total rainfall, and measures of financial depth—proxied by the shares of farmers without savings and with bank loans (see Columns (5) and (10) of Table 4). In particular, notice that our controls include the village-year mean and variance of household income, which allows us to capture households' adoption of income-smoothing strategies, such as crop diversification or the use of spatially separated land plots (Townsend, 1988; Morduch, 1995).

4.4 Risk-sharing and caste homogeneity

Our results in Section 4.3 show that land misallocation declines as risk-sharing improves. This naturally raises the question: where do differences in risk-sharing originate? In this subsection, we provide suggestive evidence that homogeneity in caste composition is an important determinant of the level of risk-sharing across villages. A large body of work has underscored the central role of caste networks in facilitating access to credit and providing informal insurance to their members (Ambrus et al., 2014; Munshi and Rosenzweig, 2016; Mazzocco and Saini, 2012; Munshi, 2019). The intuition underlying our approach is that more homogeneous communities may be better able to mitigate frictions that constrain risk-sharing—such as limited enforcement and private information. A similar perspective is adopted, for example, in Breza et al. (2019), where caste homogeneity proxies for social connectedness and workers' ability to sanction deviations from collusive arrangements.

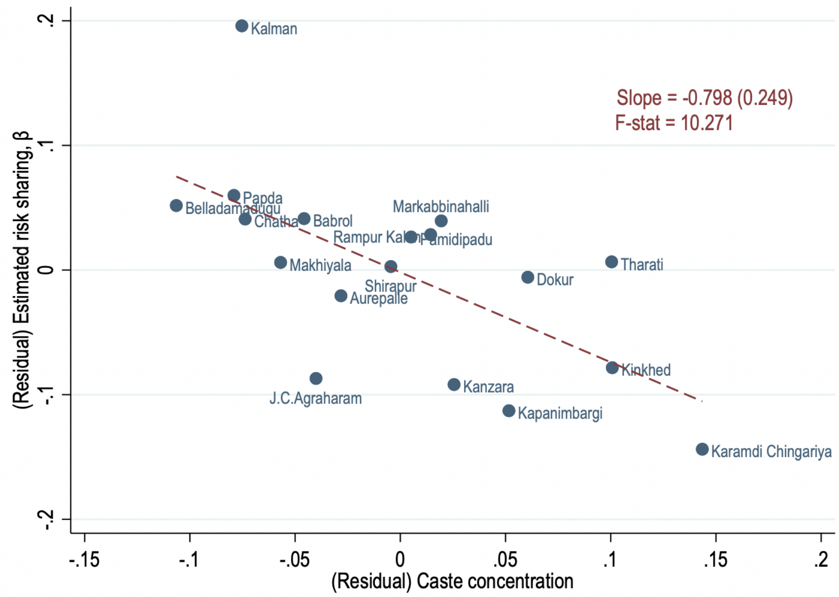
The ICRISAT data record each household head's caste (*jati*), sub-caste, and "broad" caste category (backward caste, forward caste, nomadic tribe, other backward caste, scheduled tribe, and special backward caste). We use the latter to construct a measure of caste concentration at the village level, z_v. Let $\mathcal{J}$ be an index set for the set of caste categories, where j denotes a typical element in this set. Let n_{jv} be the number of households belonging to caste category j in village v, with $\sum_{jv} n_{jv} = n_v$. We define caste concentration in village v using the normalized Herfindahl-Hirschman (HH) index;[21] i.e.,

$$z_v = \frac{\left(\sum_{j \in \mathcal{J}} \left(\frac{n_{jv}}{n_v} \right)^2 - \frac{1}{n_v} \right)}{1 - \frac{1}{n_v}}$$

By construction z_v varies between 0 and 1 and higher values suggest a higher caste concentration (or lower caste heterogeneity) in village v. Nothing prevents our measure of caste homogeneity from varying over time. However, because caste composition within villages is highly persistent in our data, we fix our index of caste concentration over time

[21] See Esteban and Ray (1994) and Esteban et al. (2012) for measures of concentration indices that distinguish between polarization and fractionalization.

Figure 5: Risk-sharing and caste diversity across villages



Notes: This figure scatters the estimated (residualized) risk-sharing against the (residualized) degree of caste concentration across villages in our sample. Residualized risk-sharing and caste diversity are obtained controlling for village-level averages and dispersions of farm size, TFP-Q, and annual farmer income; total rainfall in the village; village-level shares of farmers without savings and of farmers with a banking loan. The point estimate of the slope is obtained with OLS. The standard error of the estimated slope (in parenthesis) is robust.

and allow it to vary only across villages.

Figure 5 plots village-level estimates of risk-sharing against caste concentration, conditional on village-level observables. The figure confirms our hypothesis: villages with more diverse caste composition, i.e., low z_v, (such as "Kalman"), exhibit lower degrees of risk-sharing, i.e., high estimated β_v, as opposed to villages with a more homogeneous caste structure (such as "Karamdi Chingariya")

By altering the degree risk sharing, caste structure also affects the allocative efficiency of land. In Table 5, we regress our measures of land misallocation on the projection of village-specific risk-sharing, $\widehat{\beta}_v$, onto caste concentration, z_v. The estimates also suggest that villages with weaker risk-sharing (higher $\widehat{\beta}_v$) tend to exhibit substantially higher land misallocation. Specifically, moving from villages with full risk-sharing to those with no risk-sharing is associated with a 0.742 decrease in the within-village correlation between land and productivity and a 0.762 increase in the standard deviation of marginal products. These correlations are consistent with the intuition that limited risk-sharing distorts the allocation of land across farmers.

Table 5: Land misallocation and risk-sharing: The role of caste diversity

	$\text{corr.}_v\left[\log \ell_{it}, \log \widehat{\theta}_i\right]$	$\text{st.dev.}_v\left[\log \text{MPL}_{it}\right]$
	(1)	(2)
$\widehat{\widehat{\beta}_v}$	-0.742*	0.762***
	(0.401)	(0.266)
Observations	18	18
R^2	0.313	0.555
Controls	✓	✓

Notes: This table reports the OLS estimates from regressing $\text{corr.}_v\left[\log \ell_{it}, \log \widehat{\theta}_i\right]$ (column 1) and $\text{st.dev.}_v\left[\log \text{MPL}_{it}\right]$ (column 2) on the projection of village-specific risk-sharing, $\widehat{\beta}_v$, onto caste concentration, z_v. Controls include village-level averages and dispersions of farm size, TFP-Q, and annual farmer income; total rainfall; shares of farmers without savings and shares of farmers with a banking loan. Standard errors (in parentheses) are robust.

5 The gains from full insurance

How much would land efficiency and aggregate output increase with a better risk-sharing arrangement? In this section, we employ our theoretical model to quantify the aggregate gains from completing village consumption insurance markets. To proceed, we need to specify values for the model parameters, such as 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, we also need to define the distributions of household productivity, θ_i, and the output shocks, ρ. In the following, we describe what we do in detail.

Table 6: Parameters calibrated externally

Parameters	Description	Value	Source
α	Land share	0.193	Equation (5)
L	Aggregate land supply (hectares)	3.819	Table 1
β	Elasticity of consumption with respect to idiosyncratic income shocks	0.223	Equation (6)

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

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 6. We set the output elasticity of land, α, to 0.193, 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 with respect to idiosyncratic income shocks, β, is set to 0.223, which is the estimate obtained from Equation (6). The estimates from Equation (5) are used construct household-farms' physical productivity, θ_i, and output shocks, ρ_i. The distributions of these parameters are calibrated to match the empirical frequencies observed across households and years.[22]

Table 7: 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.

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

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

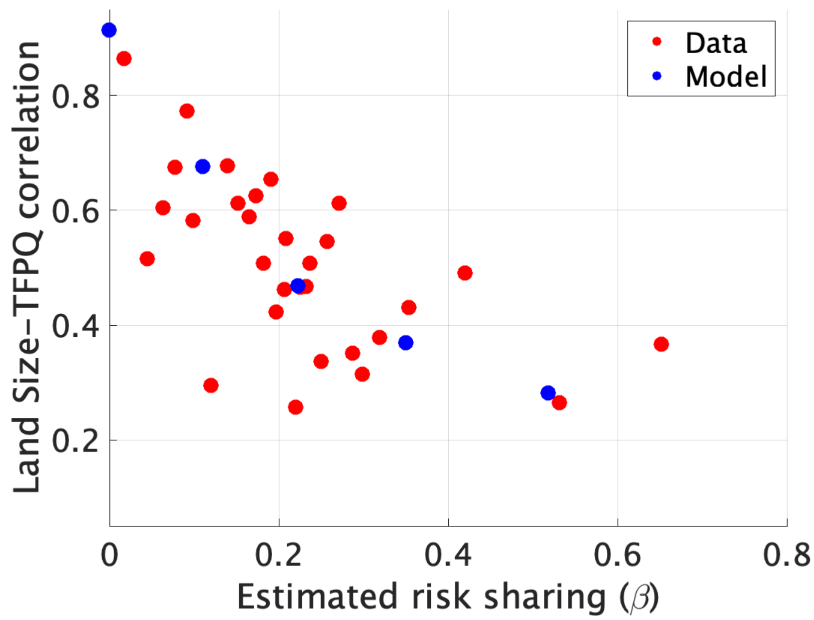
where $\mathcal{V}$ and $\mathcal{T}$ denote the numbers of villages and years in the data. Table 7 reports the point estimate and the model fit. We obtain a value of $\widehat{\sigma} = 1.60$, indicating moderate risk aversion. This estimate aligns with the findings in Holden and Quiggin (2017), who estimate a coefficient of relative risk aversion of 1.73 (Table A.4) for a sample of farmers in Malawi.[23]

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 6 plots the equilibrium

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

[23] 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 6: Risk-sharing and misallocation: Model vs. data



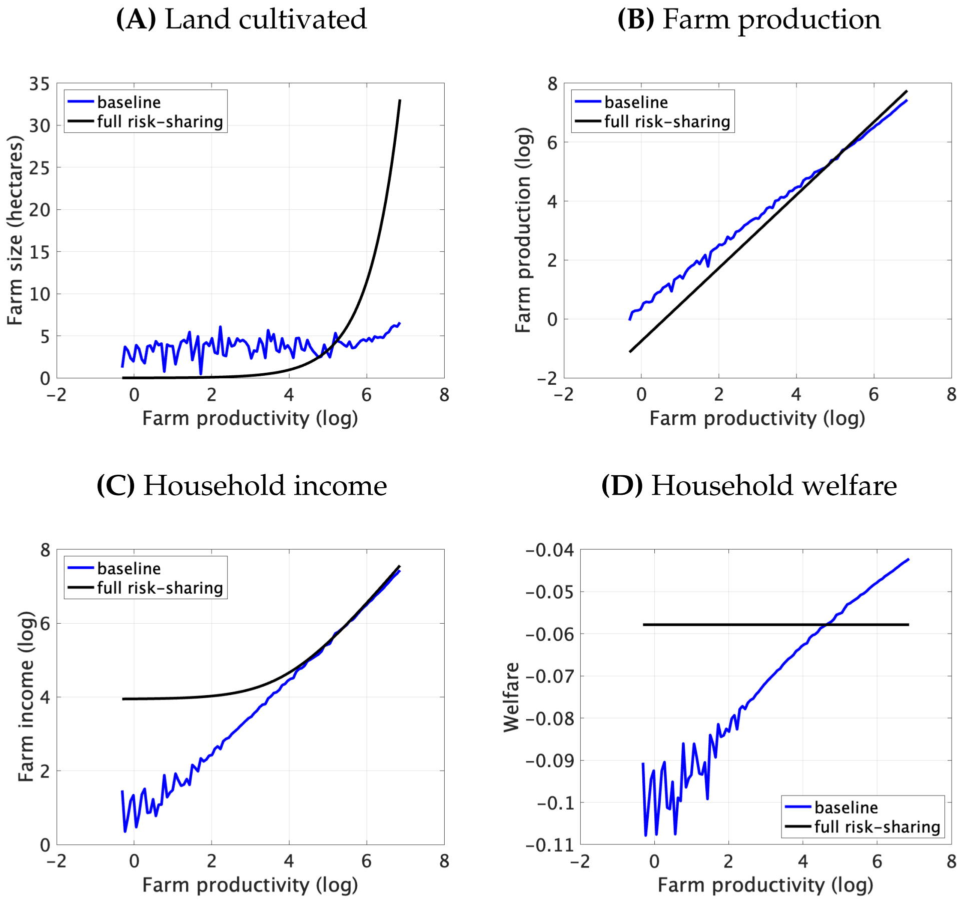
Notes: This figure plots the model-predicted equilibrium correlation between log farm size and log productivity for different levels of risk-sharing (blue dots) against the corresponding empirical estimates from Figure 4A (red dots).

correlation between log farm size and log productivity for different levels of risk-sharing (blue dots) against the corresponding empirical estimates from Figure 4A (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 estimation. This suggests that the model effectively captures the relationship between land misallocation and the degree of consumption insurance across villages and years.

5.1 Counterfactual exercise

How would households' land cultivation choices differ if village insurance markets were complete? To what extent does full insurance enhance allocative efficiency 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 output shocks constitutes a natural benchmark, being the conceptual counterpart to removing farm-specific distortions in traditional analyses of

Figure 7: Partial (baseline) vs. full risk-sharing



Notes: This figure 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).

land misallocation (Chen et al., 2023). Specifically, we contrast our baseline model to a counterfactual scenario in which $\beta = 0$, corresponding to perfect risk-sharing. We keep all the other parameters at their baseline values, including the overall land supply, and assume identical Pareto weights across households, ensuring that, in the full insurance equilibrium, each household consumes the economy's average income.[24]

Figure 7 plots the amount of land cultivated (Panel A), farm output (Panel B), house-

[24]If we conceptualize the full insurance equilibrium as the solution to a planner's problem, as outlined in Section 2, we can imagine the planner having complete information and the ability to maximally punish households that opt out of the risk-sharing arrangement. As a result, she can implement this solution without facing incentive compatibility or participation constraints. The egalitarian allocation, in which each household consumes a constant fraction of aggregate income, is generally not implementable as a decentralized solution in a complete markets economy. Instead, in such a solution, households of the same type (i.e., those who are ex-ante identical) consume the same amount in expectation.

hold 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 as under partial insurance, increasing from just over 5 hectares to nearly 35. Conversely, low-productivity farms 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 8: Counterfactual exercise

	Baseline (partial insurance) (1)	Counterfactual (full risk-sharing) (2)
β	0.223	0
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

Notes: This table compares the baseline (partial insurance) and counterfactual (full-insurance) economies across several key dimensions, such as aggregate output and welfare.

In Table 8, we further compare the baseline and counterfactual economies across several key dimensions. With full insurance, land misallocation is sizably reduced: the correlation between log farm size and log productivity is nearly twice as large the one in the baseline. Under full insurance, the distribution of cultivated land becomes significantly more unequal: the share of total available land allocated to firms at the top 1% of the productivity distribution increases by approximately five times, while the share going to the top 10% of farms increases by approximately four times. The standard deviation of (log) cultivated land increases approximately sixfolds. Improved insurance leads to an increase in aggregate efficiency of 41.5%, to output gains of 18.6%, while the overall welfare gains, measured in consumption-equivalent terms, are equal to 28.6%.

Efficiency and output gains from full risk-sharing are also robust to incorporating farm-specific distortions in the model. Specifically, in Appendix D.1, we extend our model by introducing farm-specific distortions taking the form of output wedges that are correlated to farm productivity. We then conduct the same comparative analysis as for the baseline model. We report the results in Table D2. We find that the efficiency and output gains from achieving perfect risk-sharing remain large and are equal to 39.3% and 18.3%, respectively, accounting for between 30% and 45% of the overall gains that can be achieved from moving to a fully undistorted economy.[25] These figures are comparable to those quantifying the 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.

In Table D2, we compare the welfare gains from completing insurance markets with those from eliminating farm-specific distortions. Although removing these distortions raises output relatively more (40.7% vs. 17.8%), completing insurance markets delivers much larger welfare gains (13.7% vs. 6.5%). This occurs because improved risk-sharing generates dual benefits: better allocative efficiency across farmers and direct insurance against idiosyncratic shocks. Our analysis suggests that in economies characterized by frictions in consumption insurance and factor markets, policy interventions may be most effective when prioritizing improvements in risk-sharing institutions.

6 Conclusions

This paper bridges the gap between the literatures on risk-sharing and resource misallocation. We begin with two key observations. First, insurance markets in rural villages are

[25] Results are also robust to an alternative calibration of the land share, α. See Appendix D.2.

incomplete, leading household income shocks to significantly affect consumption. Second, there is substantial misallocation of factors of production among farmers, resulting in reduced agricultural productivity in village economies. We argue that these two phenomena are deeply interconnected. Specifically, we see the limited functioning of consumption insurance markets as a key factor contributing to land misallocation in rural villages.

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.

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. (2020). Land Reform and Productivity: A Quantitative Analysis with Micro Data. *American Economic Journal: Macroeconomics*, 12(3):1–39.

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.

Alderman, H. and Paxson, C. H. (1992). Do the poor insure?: a synthesis of the literature on risk and consumption in developing countries.

Ambrus, A., Chandrasekhar, A. G., and Elliott, M. (2014). Social investments, informal risk sharing, and inequality. Technical report.

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

Armangué-Jubert, T., Pietrobon, D., and Ruggieri, A. (2025). Incomplete Markets as Correlated Distortions. Working paper.

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.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921.

Bold, T. and Broer, T. (2021). Risk-Sharing in Village Economies Revisited. *Journal of the European Economic Association*, 19(6):3207–3248.

Breza, E., Kaur, S., and Krishnaswamy, N. (2019). Coordination without organization: Collective labor supply in decentralized spot markets. Technical report.

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.

Chari, A., Liu, E. M., Wang, S.-Y., and Wang, Y. (2021). Property Rights, Land Misallocation, and Agricultural Efficiency in China. *The Review of Economic Studies*, 88(4):1831–1862.

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.

Chiappori, P.-A., Samphantharak, K., Schulhofer-Wohl, S., and Townsend, R. M. (2014). Heterogeneity and Risk Sharing in Village Economies. *Quantitative Economics*, 5(1):1–27.

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

Deaton, A. and Zaidi, S. (2002). *Guidelines for constructing consumption aggregates for welfare analysis*, volume 135. World Bank Publications.

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.

Emerick, K., de Janvry, A., Sadoulet, E., and Dar, M. H. (2016). Technological Innovations, Downside Risk, and the Modernization of Agriculture. *American Economic Review*, 106(6):1537–1561.

Esteban, J., Mayoral, L., and Ray, D. (2012). Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–1342.

Esteban, J. and Ray, D. (1994). On the measurement of polarization. *Econometrica: Journal of the Econometric Society*, pages 819–851.

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

Feder, G., Just, R. E., and Zilberman, D. (1985). Adoption of Agricultural Innovations in Developing Countries: A Survey. *Economic Development and Cultural Change*, 33(2):255–298.

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.

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural Decisions After Relaxing Credit and Risk Constraints. *Quarterly Journal of Economics*, 129(2):597–652.

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.

Magruder, J. R. (2018). An assessment of experimental evidence on agricultural technology adoption in developing countries. *Annual Review of Resource Economics*, 10(1):299–316.

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

Midrigan, V. and Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American economic review*, 104(2):422–458.

Morduch, J. (1995). Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives*, 9(3):103–114.

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.

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.

Singh, I., Squire, L., and Strauss, J. (1985). Agricultural household models: A survey of recent findings and their policy implications. Working paper, Yale University, Economic Growth Center.

Singh, I., Squire, L., and Strauss, J. (1986). *Agricultural household models: Extensions, applications, and policy*. Johns Hopkins University Press, Baltimore.

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

Townsend, R. M. (1988). Models as economies. *The Economic Journal*, 98(390):1–24.

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

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.

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,

$$c_i\left(\boldsymbol{\rho}\right) = \lambda^{-\frac{1}{\sigma}} \nu_i^{\frac{1}{\sigma}}, \tag{8}$$

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 (8), 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:

$$\iota + \lambda \int \frac{\partial y_{i\rho}}{\partial \ell_i} \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}\right) = 0, \tag{9}$$

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:

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

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

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

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

Proof of Claim 4. The proof that each household consumes a constant fraction of aggregate output, proportional to its Pareto weight, follows the same steps as the proof of Claim 2. Therefore, we omit the details to avoid repetition.

The first-order conditions for ℓ_i and k_i read as follows:

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

and

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

where ι^{ℓ} and ι^{k} are the Lagrange multipliers attached to the land and material availability constraints, respectively. Thus, the expected marginal products of land are equalized across households, as are the expected marginal products of materials. The first-order conditions for $h_i\left(\boldsymbol{\rho}\right)$ read as follows:

$$-\phi+\lambda\frac{\partial y_{i\rho}}{\partial h_i\left(\boldsymbol{\rho}\right)}=0.$$

To maximize aggregate expected output, we can solve the following programming problem:

$$\max_{\boldsymbol{\ell},\boldsymbol{k},\left(\boldsymbol{h}\left(\boldsymbol{\rho}\right)\right)_{\boldsymbol{\rho}}}\int\int y_{i\rho}\mathrm{d}Q_{\rho}\left(\boldsymbol{\rho}\right)\mathrm{d}i$$
$$\text{s.t.}\quad \int \ell_i\mathrm{d}i=L,\quad \int k_i\mathrm{d}i=K,\quad \int h_i\left(\boldsymbol{\rho}\right)\mathrm{d}i=H\left(\boldsymbol{\rho}\right),\quad \forall\boldsymbol{\rho},$$

where $H\left(\boldsymbol{\rho}\right)$ is the total amount of labor the households are willing to supply under state

$\boldsymbol{\rho}$. The first-order conditions for ℓ_i and ℓ_j imply that

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

i.e., to maximize aggregate expected output, the expected marginal products of land must be equalized across households. Similarly, it follows directly that the equalization of expected marginal products of materials across households is a necessary condition for maximizing aggregate expected output. The first-order conditions for $h_i\left(\boldsymbol{\rho}\right)$ and $h_j\left(\boldsymbol{\rho}\right)$ imply that

$$\frac{\partial y_{i\rho}}{\partial h_i\left(\boldsymbol{\rho}\right)} = \frac{\partial y_{j\rho}}{\partial h_j\left(\boldsymbol{\rho}\right)},$$

for each $\boldsymbol{\rho}$. □

Proof of Claim 5. Under no sharing, the first-order conditions for ℓ_i and k_i read as follows:

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

and

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

On the other hand, the first-order conditions for $h_i\left(\boldsymbol{\rho}\right)$ read as follows:

$$c_i\left(\boldsymbol{\rho}\right)^{-\sigma} \frac{\partial y_{i\rho}}{\partial \ell_i} - \phi = 0.$$

Thus, unless the households are risk neutral ($\sigma = 0$), the expected marginal products of land and materials are not necessarily equalized across households. The misallocation of land and materials spills over to labor. That is, even though the marginal products of labor are equalized across households—implying that standard marginal product equalization tests would fail to detect labor misallocation—labor allocation across farms is indirectly distorted by the misallocation of land and materials. □

Proof of Theorem 2. For each $\beta \in (0,1]$, the first-order conditions for ℓ_i and k_i are

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

and

$$\int (c_i(\boldsymbol{\rho}))^{-\sigma} \frac{\partial c_i(\boldsymbol{\rho})}{\partial \pi^P_{i\boldsymbol{\rho}}} \left(\frac{\partial y_{i\boldsymbol{\rho}}}{\partial k_i} - q^P \right) \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}) = 0,$$

where

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

Letting

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

we can rewrite these first-order conditions as

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

and

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

These equation show that the effect of partial insurance on optimal land cultivation and choice of materials can be interpreted as the introduction of type- and state-specific distortions. These distortions become more pronounced the further $T_{i\boldsymbol{\rho}}$ deviates from being constant across states of the world. Notice that an increase in β amplifies this variance. □

Proof of Theorem 3. For each $\beta \in (0,1]$, the first-order conditions for b_{mit} are

$$-\int \left(c_{mit\boldsymbol{\rho}_t}\right)^{-\sigma} \frac{\partial c_{mit\boldsymbol{\rho}_t}}{\partial \pi^P_{mit\boldsymbol{\rho}_t}} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}_t) + \Xi\left(1+R^P\right) \int \left(c_{mit+1\boldsymbol{\rho}_{t+1}}\right)^{-\sigma} \frac{\partial c_{mit+1\boldsymbol{\rho}_{t+1}}}{\partial \pi^P_{mit+1\boldsymbol{\rho}_{t+1}}} \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho}_{t+1}) = 0,$$

which is a standard Euler equation. The first-order conditions for $\ell_{mit}(b_{mit-1})$ and $k_{mit}(b_{mit-1})$

are

$$\int \left(c_{mit\boldsymbol{\rho}_t}\right)^{-\sigma} \frac{\partial c_{mit\boldsymbol{\rho}_t}}{\partial \pi^P_{mit\boldsymbol{\rho}_t}} \left(\frac{\partial y_{mit\boldsymbol{\rho}_t}}{\partial \ell_{mit}\left(b_{mit-1}\right)} - r^P_t\right) \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}_t\right) = 0$$

and

$$\int \left(c_{mit\boldsymbol{\rho}_t}\right)^{-\sigma} \frac{\partial c_{mit\boldsymbol{\rho}_t}}{\partial \pi^P_{mit\boldsymbol{\rho}_t}} \left(\frac{\partial y_{mit\boldsymbol{\rho}_t}}{\partial k_{mit}\left(b_{mit-1}\right)} - r^P_t\right) \mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}_t\right) = 0.$$

The rest of the proof follows the same steps as the proof of Theorem 2. □

Supplementary Appendix

A Theory appendix

A.1 Discussion of modeling assumptions

Household types. Our model features households that are ex-ante heterogeneous in their "permanent" agricultural productivity. The importance of heterogeneity in farm productivity has been emphasized in various fundamental papers, including Adamopoulos and Restuccia (2014), Chen et al. (2023), and Adamopoulos et al. (2022). By focusing our model on ex-ante heterogeneity in productivity, we can examine the relationship between risk-sharing and the misallocation of inputs chosen ex ante. We adopt this approach because the production decisions we are interested in are naturally viewed as choices made before any shocks to agricultural yields, such as rainfall or pests, occur.

Land markets. The model in Section 2 assumes that land markets are undistorted. 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 D.1. The incomplete markets economy 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. The presence of these frictions implies that the cash obtained from selling land cannot be perfectly smoothed over time or across states of the world, and that a farmer who sells land might subsequently be 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 aligns with our theoretical framework, focusing on misallocation in operational landholdings that encompass both owner-cultivators and renters.

Consumption functions. We model the relationship between consumption and income using consumption functions, as specified in Equation (1). We adopt this approach for two main reasons. First, it directly corresponds to our empirical specification for estimating the elasticity of consumption with respect to idiosyncratic income shocks, as introduced in Equation (6), which is a standard specification in the literature (see, e.g., Morten (2019)). This correspondence is crucial because, in our quantitative exercise (Subsection 5), we calibrate the consumption functions so that β matches the estimated elasticity of consumption with respect to idiosyncratic income shocks. Second, this approach falls within the tradition of the literature on exogenously incomplete markets[26] and partial insurance (Blundell et al., 2008). In particular, these consumption 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. 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

[26]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.

specific forms of the consumption functions, which determine the level of insurance in the economy, focusing instead on how different degrees of insurance influence land misallocation. Finally, beyond providing a flexible way to model the relationship between household consumption and income, the consumption function in Equation (1) has the advantage of directly corresponding to our strategy to estimate risk-sharing in Subsection 4.2. Recall that, in our model, $\underline{\rho}$ is set sufficiently high to ensure that household income remains strictly positive for all possible land cultivation choices. Alternatively, positive cash-on-hand can be ensured by introducing an asset that allows households to borrow and save (see Appendix A.3).

A.2 Model with other factors of production

In this appendix, we show that all the results from Section 2 remain valid in a model that incorporates materials and labor as factors of production, where we also allow labor to be chosen after the realization of output shocks.

Consider a version of the model in Section 2 where household types are defined by a productivity level θ_i and an endowment

$$\left(\widetilde{\ell}_i, \widetilde{k}_i\right),$$

where $\widetilde{k}_i$ represents initial materials for a household of type i.

Households decide how much land to cultivate and how much material to buy before output shocks are realized. Since households of the same type are ex-ante identical and make these choices prior to the realization of shocks, their land and material allocations are identical within each type. Thus, we can refer to ℓ_i and k_i as the land cultivated and materials utilized by a household of type i. Besides choosing land and materials, households decide how many hours to work. Labor supply decisions are made after the output shocks are realized, in contrast to materials and land. We assume households have complete information, so they can choose the amount of labor to employ conditional on the realization of all output shocks in the economy, $\boldsymbol{\rho}$. Since households of the same type are ex-ante identical, we can unambiguously define the function $h_i(\boldsymbol{\rho})$, which represents the farm labor supply of a household of type i when state $\boldsymbol{\rho}$ is realized. The output produced by a household of type i in state $\boldsymbol{\rho}$ is given by

$$y_{i\boldsymbol{\rho}} = A\theta_i \rho \ell_i^{\alpha^\ell} k_i^{\alpha^k} \left(h_i(\boldsymbol{\rho})\right)^{\alpha^h},$$

where A is an aggregate productivity term, $\alpha^\ell, \alpha^k, \alpha^h \in (0,1)$, and $\alpha^\ell + \alpha^k + \alpha^h < 1$. Let r be the price of land and q the price of materials. Given these prices, let

$$\pi_{i\boldsymbol{\rho}} = y_{i\boldsymbol{\rho}} - r\left(\ell_i - \widetilde{\ell}_i\right) - q\left(k_i - \widetilde{k}_i\right)$$

denote the income of a household of type i when state $\boldsymbol{\rho}$ is realized.

Let $\boldsymbol{k} = (k_i)_i$ and $\boldsymbol{h}(\boldsymbol{\rho}) = (h_i(\boldsymbol{\rho}))_i$. 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 $(\nu_i)_i$:

$$\max_{(\boldsymbol{c}(\boldsymbol{\rho}))_{\boldsymbol{\rho}}, \boldsymbol{\ell}, \boldsymbol{k}, (\boldsymbol{h}(\boldsymbol{\rho}))_{\boldsymbol{\rho}}} \int \nu_i \int \left[\frac{(c_i(\boldsymbol{\rho}))^{1-\sigma}}{1-\sigma} - \phi h_i(\boldsymbol{\rho})\right] \mathrm{d}Q_{\boldsymbol{\rho}}(\boldsymbol{\rho})\, \mathrm{d}i,$$

subject to the land and materials availability constraints

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

and

$$\int k_i \mathrm{d}i = \int \widetilde{k}_i \mathrm{d}i = K,$$

and the feasibility constraint

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

The following claim describes an allocation of resources under complete markets.

Claim 4. *Under full insurance, each household consumes a constant fraction of aggregate output, with the fraction being proportional to its Pareto weight. Moreover,*

1. *the expected marginal products of land are equalized across households,*

2. *the expected marginal products of materials are equalized across households,*

3. *the marginal products of labor are equalized across households, in each state of the world.*

Finally, aggregate expected output is maximized.

Next, we analyze the allocation of land that arises in a competitive equilibrium under no risk-sharing. In this case, the income of a household of type i in state $\boldsymbol{\rho}$ is given by

$$\pi_{i\boldsymbol{\rho}}^{IM} = y_{i\boldsymbol{\rho}} - r^{IM}\left(\ell_i - \widetilde{\ell}_i\right) - q^{IM}\left(k_i - \widetilde{k}_i\right),$$

where r^{IM} and q^{IM} denote the equilibrium prices of land, materials, and labor under incomplete markets. The household's optimization problem is then given by:

$$\max_{\ell_i, k_i, (h_i(\boldsymbol{\rho}))_{\boldsymbol{\rho}}} \int \left[\frac{\left(\pi_{i\boldsymbol{\rho}}^{IM}\right)^{1-\sigma}}{1-\sigma} - \phi h_i(\boldsymbol{\rho})\right] \mathrm{d}Q_{\rho}(\rho).$$

The following result demonstrates that incomplete insurance markets lead to factor misallocation, even when *all* factor markets operate without frictions.

Claim 5. *Under no sharing, there is land, materials, and labor misallocation.*

One implication of this result is that the presence of factor markets where inputs are chosen after the realization of output shocks does not undo the misallocation caused by missing insurance markets. The reason is that if some factors of production are chosen before shocks are realized (e.g., land and materials), but insurance markets are missing, households distort their ex-ante input choices to self-insure against risk. After output shocks are realized, households adjust their ex-post input choices (e.g., labor) by equating the realized marginal product of labor to its price, thereby maximizing their income. However, as long as inputs chosen ex post are not perfect substitutes for those chosen ex ante, the distortions in ex-ante input choices caused by missing insurance markets "spill over" to those chosen ex post.

Finally, consider an environment with partial insurance. Let the income of a household of type i in state $\boldsymbol{\rho}$ is given by

$$\pi_{i\boldsymbol{\rho}}^{P} = y_{i\boldsymbol{\rho}} - r^{P}\left(\ell_i - \widetilde{\ell}_i\right) - q^{P}\left(k_i - \widetilde{k}_i\right),$$

where r^P and q^P denote the equilibrium prices of land, materials, and labor under partial insurance. We define the following consumption function for a household of type i:

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

The following theorem extends Theorem 1 to settings with multiple factors of production.

Theorem 2. *Land market misallocation increases in the elasticity of consumption with respect to own income, β.*

A.3 Model with borrowing and saving

In this section, we show that Theorem 2 extends to environments where households can borrow and save through a risk-free asset. To do so, we build on the static model in Appendix A.2 by introducing time in a discrete setting. In this framework, households have additively separable preferences over time and discount the future at a constant rate Ξ. Let $\boldsymbol{\rho}_t$ be the state of the world in period t, assumed to be identically and independently distributed over time. Households can borrow and save through a risk-free asset. Let b_{mit} denote the amount of bonds held by household m of type i in period t (where a negative value indicates borrowing). This amount must be repaid in the following period at the interest rate R^P. In each period, households decide how much land to cultivate and how much material to employ before the realization of $\boldsymbol{\rho}_t$. After $\boldsymbol{\rho}_t$ is realized, they choose how much labor to supply to their own farm. Let these choices be denoted by

$$\left(\ell_{mit}\left(b_{mit-1}\right), k_{mit}\left(b_{mit-1}\right), h_{mit}\left(\left(b_{mit-1}\right), \boldsymbol{\rho}_t\right)\right).$$

Denote household m's farm output when state $\boldsymbol{\rho}_t$ realizes by

$$y_{mit\boldsymbol{\rho}_t} = A\theta_i\rho_t \left(\ell_{mit}\left(b_{mit-1}\right)\right)^{\alpha^{\ell}} \left(k_{mit}\left(b_{mit-1}\right)\right)^{\alpha^{k}} \left(h_{mit}\left(\left(b_{mit-1}\right), \boldsymbol{\rho}_t\right)\right)^{\alpha^{h}}.$$

The income of a household m of type i in period t when state $\boldsymbol{\rho}_t$ realizes is given by

$$\pi^P_{mit\boldsymbol{\rho}_t} = y_{mit\boldsymbol{\rho}_t} - r^P_t\left(\ell_{mit}\left(b_{mit-1}\right) - \widetilde{\ell}_i\right) - q^P_t\left(k_{mit}\left(b_{mit-1}\right) - \widetilde{k}_i\right) + \left(1 + R^P\right) b_{mit-1} - b_{mit}.$$

Let consumption for household m of type i in period t when state $\boldsymbol{\rho}_t$ realizes be

$$c_{mit\boldsymbol{\rho}_t} = \exp\left\{\beta \log\left(\pi^P_{mit\boldsymbol{\rho}_t}\right) + (1-\beta)\log\left(\frac{\nu_i^{\frac{1}{\sigma}}}{\int \nu_j^{\frac{1}{\sigma}}\mathrm{d}j}\int\int\int \pi^P_{njt\boldsymbol{\rho}_t}\mathrm{d}Q_{\boldsymbol{\rho}}\left(\boldsymbol{\rho}_t\right)\mathrm{d}n\mathrm{d}j\right)\right\}.$$

The following theorem generalizes Theorem 2 to environments where households can borrow and save through a risk-free asset.

Theorem 3. *Land market misallocation increases in the elasticity of consumption with respect to own income, β.*

B Data appendix

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 4.2, 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 productivity (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's value of intermediate agricultural inputs. 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 farm-specific measure of yearly value of materials for each household. We use this variable in the estimation of household-farms' physical productivity, 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 productivity, 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 productivity, 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 documents every monetary inflow and outflow for each household on a monthly basis, along with the monetary value of all home-produced commodities. Rather than providing a direct measure of household consumption, the expenditure segment records detailed individual expenditures on food and non-food items, as well as the monetary value of home-produced commodities that are consumed. This level of granularity in consumption items is one of the key advantages of the ICRISAT dataset, as it allows us to compile a comprehensive measure of monthly household consumption by aggregating these recorded values at the household-month level. As a result, measures of household consumption are never imputed from household income, not even for the poorest households. To derive a measure of monthly household income, we adopt the budget-constraint approach of Mazzocco and Saini (2012). 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 4.2 to estimate the level of risk-sharing in Indian villages.

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).

C Further empirical results

C.1 Agricultural output decomposition

In Table C1, 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 C1: 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	0.161
Land quality, Var $[\log \widehat{q}_{i\tau}]$	0.103	3.469
Household-farms' physical productivity: Var $\left[\log \widehat{\theta}_i\right]$	0.614	41.57
Family labor: Var $[\log h_{i\tau}]$	0.382	19.03
Materials: Var $[\log k_{i\tau}]$	0.292	10.08
Landholdings: Var $[\log \ell_{i\tau}]$	0.277	10.02

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.

C.2 Imperfect risk-sharing in the ICRISAT villages

Column (1) of Table C2 reports the OLS estimates for equation (6). Column (2) reports the OLS estimates of equation (6) extended with an interaction between log of per-capita income, $\ln \pi_{it}$, and a dummy variable $[1 - \text{No-saving}_{it}]$ taking value 0 if the household i does not have savings at time t (i.e., its stock of savings is lower than average monthly income) and 1 otherwise; that is,

$$\log c_{it} = \beta \log \pi_{it} + \delta \log \pi_{it} \times [1 - \text{No-saving}_{it}] + \mu[1 - \text{No-saving}_{it}] + \chi_i + \chi_{vt} + \varepsilon_{it}.$$

Column (3) reports the OLS estimates of equation (6) extended with an interaction between log of per-capita income, $\ln \pi_{it}$, and a dummy variable $[\text{Loan}_{it}]$ taking value 0 if the household i has borrowed from a bank at time t and 0 otherwise; that is,

$$\log c_{it} = \beta \log \pi_{it} + \delta \log \pi_{it} \times [\text{Loan}_{it}] + \mu[\text{Loan}_{it}] + \chi_i + \chi_{vt} + \varepsilon_{it}.$$

Finally, Column (4) reports the OLS estimate an alternative regression for household i within village v in month t:

$$\Delta \log c_{it} = \beta \Delta \log \pi_{it} + \chi_{vt} + \varepsilon_{it}$$

where, $\Delta \log c_{it}$ and $\Delta \log \pi_{it}$ denote log changes in per-capita consumption and per-capita income, respectively, for household i between two consecutive months.

C.3 Alternative measure of land quality

Here, we perform a robustness check on the results in Section 3 using an alternative measure of land quality. Following De Giorgi et al. (2024), we construct this measure based on plot-level indices of agricultural suitability, which are estimated from a comprehensive regression of plot rental values on land characteristics, while controlling for village-level permanent unobserved heterogeneity and year-specific fixed effects. The idea behind this approach is that the rental value of a plot reflects valuable information about its suit-

Table C2: Risk-sharing in the ICRISAT villages

	$\ln c_{it}$			$\Delta \ln c_{it}$
	(1)	(2)	(3)	(4)
$\ln \pi_{it}$	0.223***	0.265***	0.250***	
	(0.018)	(0.021)	(0.0103)	
$\ln \pi_{it} \times [1 - \text{No-saving}_{it}]$		-0.042		
		(0.020)		
$\ln \pi_{it} \times [\text{Loan}_{it}]$			-0.016	
			(0.019)	
$\Delta \ln \pi_{it}$				0.206***
				(0.019)
Household FE	✓	✓	✓	
Village-time FE	✓	✓	✓	✓
Observations	46,369	45,781	43,462	41,263
R-squared	0.681	0.690	0.689	0.319

Notes: The unit of analysis across all columns is the household-year. Column (1) presents the results of regressing monthly log consumption per capita on monthly log income per capita while controlling for household and village-month fixed effects. Column (2) extends the regression in Column (1) by interacting monthly log income per capita with a dummy variable $[1 - \text{No-saving}_{it}]$, taking value 1 if the household i is does not save in year t. Column (3) extends the regression in Column (1) by interacting monthly log income per capita with a dummy variable $[\text{Loan}_{it}]$, taking value 1 if the household i has borrowed from a bank in year t. Column (4) refers to the specification in log changes. Standard errors (in parentheses) are clustered at the village-year level across all columns.

ability for agricultural production; i.e., its "quality." Specifically, we run the following regression:

$$\begin{aligned}\log\left(\text{rental}_{pi\tau}\right) &= f\left(\text{type}_{pi\tau}, \text{fertility}_{pi\tau}, \text{slope}_{pi\tau}, \text{degradation}_{pi\tau}, \text{depth}_{pi\tau}; \boldsymbol{\psi}\right) \\ &\quad + \varphi_v + \varphi_\tau + \epsilon_{pi\tau},\end{aligned}$$

where $\text{rental}_{pi\tau}$ represents the self-reported rental value of plot p cultivated by household i in year τ; "type" denotes a categorical variable reflecting the plot's soil type; "fertility" is an ordered categorical variable measuring the plot's fertility on a scale from 0 to 4; "slope" is an ordered categorical variable indicating the degree of the plot's slope; "soil degradation" is a categorical variable indicating whether the plot is subject to soil degradation and specifies its type; "depth" refers to the average depth of the plot measured in centimeters; f is a function that represents linear and higher-order terms for the included variables, along with multiple interaction terms between those variables; $\boldsymbol{\psi}$ is a parameter vector; φ_v are village fixed effects; and φ_τ are year fixed effects. We use the predicted val-

Table C3: Input contributions to agricultural output—alternative land quality measure

Variable	R-squared (1)	Shapley value (%) (2)
Rainfall shocks: Var $\left[\widehat{\vartheta}\text{rain}_{v\tau}\right]$	0.004	0.191
Land quality, Var $[\log \widehat{q}_{i\tau}]$	0.186	7.500
Household-farms' physical productivity: Var $\left[\log \widehat{\theta}_i\right]$	0.584	36.91
Family labor: Var $[\log h_{i\tau}]$	0.382	19.43
Materials: Var $[\log k_{i\tau}]$	0.292	9.513
Landholdings: Var $[\log \ell_{i\tau}]$	0.277	10.72

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.

ues of this regression, $\log\left(\widehat{\text{rental}}_{pit}\right)$, as plot-level indices of agricultural suitability. The intuition is that these predicted values contain information about the extent to which the interactions of multiple soil characteristics, village-level permanent unobserved factors, and year-level aggregate shocks influence the suitability of a plot for agricultural production. To generate a household-year level land quality index, we average these suitability indices as follows:

$$q'_{i\tau} = \frac{\sum_{p \in P_{i\tau}} \left(\log\left(\widehat{\text{rental}}_{pi\tau}\right) \times a_{pi\tau}\right)}{\sum_{p \in P_{i\tau}} a_{pi\tau}},$$

where $a_{pi\tau}$ denotes the area of plot p cultivated by household i in year τ, and $P_{i\tau}$ is an index set for the set of plots cultivated by household i in year τ.

We use this alternative measure of land quality to reproduce two main findings from the main text. First, in Table C3 we report the contribution of each input to the agricultural output. While our alternative measure of land quality explains twice as much of the overall output variation across farms as before (7.5 versus 3.5%, see Table C1), household-farms' physical productivity still represents the biggest source of output variation, with a Shapley value of about 37%.

Second, we construct the two measures of land misallocation using our alternative proxy for land quality and relate them to the degree of risk-sharing across villages and years. Specifically, we estimate different specifications of Equation (7) in the main text, and report the outcomes in Table C4. All the new estimates are virtually unchanged.

Table C4: Risk-sharing and land misallocation—alternative land quality measure

	corr.$_{v\tau}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$				st.dev.$_{v\tau}$ $[\log \mathrm{MPL}_{i\tau}]$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\widehat{\beta}_{v\tau}$	-0.639***	-0.675***	-0.132*	-0.187*	0.691***	0.715***	0.150***	0.178***
	(0.234)	(0.230)	(0.078)	(0.096)	(0.175)	(0.230)	(0.056)	(0.069)
Observations	90	90	90	90	90	90	90	90
R^2	0.100	0.116	0.892	0.942	0.220	0.229	0.901	0.944
Year FE		✓	✓	✓		✓	✓	✓
Village 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).

D Further counterfactual results

D.1 Incomplete insurance vs. farm-specific distortions

In this appendix, we discuss the robustness of our results to the inclusion of generic farm-specific distortions (wedges). These wedges capture the effects of frictions in output or land markets that may disproportionately affect more productive farmers, such as land ceilings (Adamopoulos and Restuccia, 2020), restrictions on land rental activities (Chari et al., 2021; Acampora et al., 2022), borrowing constraints (Midrigan and Xu, 2014), or size-dependent policies that directly tax/subsidize output (Guner and Ruggieri, 2022). Formally, we assume households of type i are subject to an "output tax," τ_i. This tax is given by

$$\tau_i = 1 - \theta_i^{-\zeta}, \tag{10}$$

where ζ governs the correlation between output distortions and household 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 productivity, meaning high-productivity households face relatively higher distortions. Conversely, when $\zeta < 0$, high-productivity households face relatively lower distortions. When $\zeta = 0$, there are no output distortions in the economy, and output production reverts to the scenario described in the main text.[27]

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 5. As in Section 5, we estimate σ by matching the average correlation between log

[27] For some combination of ζ and θ_i, τ_i can be negative. In this case, distortions take the form of a subsidy towards household i. For further applications of the functional form in Equation (10) to describe firm-level distortions in developing countries, see Guner and Ruggieri (2022), among others.

farm size and log productivity, while ζ is estimated by targeting the share of households operating land smaller than 5 hectares.

Table D1: 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.

Table D1 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 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. Specifically, 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 D2 reports the outcomes of the same counterfactual exercise described in Section 5 implemented within 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 ($\zeta = 0.052$). Column (2) refers to a counterfactual scenario where risk-sharing is perfect ($\beta = 0$), keeping everything else equal. Column (3) refers to a counterfactual scenario where distortions are absent ($\zeta = 0$), keeping everything else equal. Column (4) refers to a counterfactual scenario where risk-sharing is perfect ($\beta = 0$) and distortions are absent ($\zeta = 0$).

The aggregate output gains from improving risk-sharing in an economy with farm-specific distortions amount to 17.8%, closely aligned with the gains observed in an economy without wedges (Table 8 in Section 5). When we move to a fully undistorted scenario, the output gains are equal to 61.4%. The efficiency gains of full insurance, when accounting for output distortions, are 39.3%, compared to 89.6% in the distortion-free scenario.

This implies that incomplete risk-sharing accounts for between 29.8% and 43.9% of the overall gains that can be achieved from moving to a fully undistorted economy. These

Table D2: Counterfactual exercise: risk-sharing vs. correlated distortions

	Baseline	Counterfactuals			Explained by β, %
	(1)	(2)	(3)	(4)	(5)
β	0.223	0	0.223	0	-
ζ	0.052	0.052	0	0	-
Share of land, top 1% productive farms	0.017	0.082	0.022	0.086	-
Share of land, top 10% productive farms	0.151	0.608	0.194	0.627	-
Land dispersion (st.dev.[$\log \ell_{i\tau}$])	0.314	2.458	0.369	2.589	-
Aggregate efficiency (output per hectare)	1	1.393	1.178	1.896	43.86%
Aggregate output	1	1.183	1.407	1.614	29.80%
Aggregate welfare	1	1.137	1.065	1.158	86.71%

Notes: This table reports the outcomes of the same counterfactual exercise described in Section 5 implemented within a model with output distortions. The table compares the baseline economy (Column (1)) with three counterfactual economies: one in which insurance markets are completed (Column (2)) one in which output distortions are removed (Column (3)), and one in which both reforms are implemented (Column (4)). Column (5) reports the percentage of changes in outcomes explained by changes in β, and it is computed as 100 times the ratio of the difference between Column (2) and Column (1) and the difference between Column (4) and Column (1).

results show that assuming the existence of farm-specific distortions only marginally reduces the estimated output, efficiency, and welfare gains from improving risk-sharing.

D.2 Robustness to alternative land share

In this section, we test the robustness of our main counterfactual results to an alternative value of the land share, α. In particular, we fix $\alpha = 0.50$. We keep the values of all the other parameters unchanged, relative to the main calibration, except for the coefficient of relative risk aversion, σ, which is re-estimated to match the average correlation between log farm size and log productivity. We find a value for σ of 1.675, not far from our baseline estimate of 1.60 (Table 7 in the main text).

Table D3: Estimated risk aversion—alternative land share

Parameters	Description	Value	Target	Data	Model
σ	Relative risk aversion	1.675	$\overline{\text{corr.}}\left[\log \ell_{i\tau}, \log \widehat{\theta}_i\right]$	0.461	0.470

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.

Table D4 reports selected outcomes for the baseline and counterfactual economies. Again, Column (1) corresponds to the baseline scenario, where consumption insurance is partial ($\beta = 0.223$) while Column (2) corresponds to a counterfactual scenario in which risk-sharing is perfect ($\beta = 0$).

Table D4: Counterfactual exercise—alternative land share

	Baseline (partial insurance) (1)	Counterfactual (full risk-sharing) (2)
β	0.223	0
Share of land, top 1% productive farms	0.017	0.135
Share of land, top 10% productive farms	0.162	0.797
Land dispersion (st.dev.[$\log \ell_{i\tau}$])	0.320	4.180
Aggregate efficiency (output per hectare)	1	2.131
Aggregate output	1	1.671
Aggregate welfare	1	1.519

Notes: This table compares the baseline (partial insurance) and counterfactual (full-insurance) economies across several key dimensions, such as aggregate output and welfare, in a model in which $\alpha = 0.5$.

All the results are qualitatively the same as those in the main text. On the other hand, a larger calibrated value for α implies larger output and welfare gains from completing insurance markets, equal to 67% and 52%, respectively.