Title
Essays on Productivity and Consumption Smoothing Under Imperfect Markets

Author
Silver, Jedediah

Publication Date
2024

Essays on Productivity and Consumption Smoothing Under Imperfect Markets

by

Jedediah Silver

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ethan Ligon, Chair
Professor Jeremy Magruder
Professor Edward Miguel

Spring 2024

Essays on Productivity and Consumption Smoothing Under Imperfect Markets

Abstract

Essays on Productivity and Consumption Smoothing Under Imperfect Markets

by

Jedediah Silver

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Ethan Ligon, Chair

Perhaps the most central insight of development economics is that, absent a complete set of perfect markets, households' economic activities cannot be neatly "separated" into those of a profit-maximizing firm and a utility-maximizing consumer (Singh et al., 1986). In particular, risk-averse farm households face a tradeoff between maximizing farm profits and smoothing consumption across states of the world. Balancing these motives is important not only for these households, who constitute a massive share of the world's poor, but for aggregate productivity as well. However, little is known about how to diagnose the market failures that create these tradeoffs, quantify their costs, and prescribe robust policies to address them. This dissertation seeks to provide methodological and empirical progress from the micro to the macro levels.

Chapter 1 focuses on identifying how distinct market failures affect aggregate productivity in Thai agriculture. Agricultural markets often fail to allocate resources efficiently across farm households in developing countries. However, policymakers require knowledge of which markets fail and how the distortions they generate are correlated. In this chapter, I use data from rural Thailand to characterize how distortions in land, labor, credit, and insurance markets each contribute to factor misallocation. I use moments in household consumption and production data to separately identify these distortions and then quantify their impacts on aggregate productivity through an equilibrium model of misallocation. I find that the efficient allocation would increase aggregate productivity by 31% relative to the status quo, while only 15% (7%) gains could be achieved by eliminating financial (input) distortions in isolation. Positive interaction effects from addressing multiple distortions simultaneously account for the remaining 9% TFP gains. Meanwhile, other common methods would produce larger estimates of misallocation and suggest that a financial market intervention would decrease aggregate productivity. Accounting for multiple correlated distortions is therefore crucial for measuring misallocation and designing policies to address it.

In Chapter 2, coauthored with Ethan Ligon, we move from Thailand to Northeastern Nigeria

and move from the growing season to the lean season spanning harvests to study another important tradeoff between consumption smoothing and investment. In particular, we conduct a randomized control trial offering postharvest loans (PHLs) to farm households in Gombe State.[1] The purpose of these loans is to enable households to shift from exhausting grain stocks and buying them back at high prices to becoming net arbitrageurs. While such programs have increased household incomes in Kenya (Burke et al., 2019) and Tanzania (Channa et al., 2022), their theory of change relies on grain prices rising, which is a highly uncertain proposition across sub-Saharan Africa. During our study period, prices of maize and other major crops stayed flat. While we find that the loans induced households to store more crops later into the season, we do not find significant effects on sales or overall welfare. While this is an example of the downside risk of PHLs being realized, we also use a simple model of intertemporal arbitrage to show how *ex ante* risk can have ambiguous effects on the demand for PHLs, depending on whether households are more vulnerable in states with high vs. low prices.

Chapter 3, based in part on work coauthored with Ethan Ligon, focuses on production function estimation when input choices are distorted. These estimators, which are used to estimate the production function in Chapter 1, extend the canonical approach in industrial organization (Ackerberg et al., 2015; Gandhi et al., 2020) to risk-averse producers facing imperfect markets. In particular, they proxy for unobserved productivity by inverting the demand function for a flexible input from the setting with profit-maximizing firms in competitive markets to risk-averse households, possibly facing distorted input markets. The method involves combining consumption and production data to model input demands as a function of unobserved productivity and a stochastic discount factor, which includes the covariance between production shocks and consumption at harvest. Essentially, the consumption side of the household's problem provides information to help us identify the production side. Three main specifications are considered: the canonical Cobb-Douglas with Hicks-neutral shocks, a heteroskedastic generalization of Cobb-Douglas that allows for differentially risky inputs, and a dynamic multi-stage Cobb-Douglas featuring sequential shocks. The differences across specifications show the importance of accounting for risk, both overall and input- and stage-specific, to consistently estimate production functions and draw inferences about efficiency and misallocation.

Together, these three chapters show how better understanding households' tradeoffs between productivity and consumption smoothing can improve policies to address both micro-level food insecurity and macro-level productivity.

[1]This study was registered in the American Economics Association RCT registry as AEARCTR #8022.

To all of those who tolerated six years of awful puns, and especially to those courageous enough to tell me to they weren't funny.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my dissertation committee members: Ethan Ligon, Jeremy Magruder and Ted Miguel. Their invaluable guidance, insightful feedback, and unwavering support have been invaluable throughout my time at Berkeley. I would especially like to thank Ethan Ligon for his contribution to Chapters 2 and 3.

I am also grateful to my cohort-mates — Daniel Agness, Pierre Biscaye, Michelle Chen, Joel Ferguson, Wei Guo, Qingyang Huang and Leila Safavi — whose camaraderie, comments, commiseration and cooking have provided intellectual, emotional and culinary sustenance. Honorary cohort members Matthew Suandi and Elif Tasar also merit recognition.

Furthermore, I extend my heartfelt thanks to Umar Abubakar, Parth Ahya and Justin Graham from the Taimaka Project, for their collaboration, generosity and hospitality during RCT for Chapter 2. I am also grateful for funding from the Structural Transformation and Economic Growth (STEG) initiative for generous financial support that helped produce Chapters 2 and 3. I am also grateful to the Agricultural Technology Adoption Initiative, The Weiss Fund for International Development, and the International Growth Center for financial support on other projects undertaken during my PhD.

Lastly, I would like to thank my parents and step-parents for their unwavering love, patience, and encouragement throughout this academic endeavor.

Chapter 1

Farm Household Misallocation

1.1 Introduction

Farm households in developing countries face many different market failures, but how does each matter for aggregate productivity? Decades of research in development economics has provided robust empirical evidence of incomplete credit, insurance, land, labor, fertilizer, equipment, seed, and other markets, often occurring simultaneously.[1] However, these market failures rarely operate in a vacuum; in equilibrium, they combine to misallocate resources across farms. While the resulting misallocation may be extremely costly (Restuccia and Rogerson, 2008; Adamopoulos and Restuccia, 2014), how can policymakers distinguish between its many possible sources?

Doing so is especially important, yet challenging, because distortions generated by different market failures may compound or offset each other in equilibrium. The theory of the second best implies that the effects of reducing distortions in *any* market are ambiguous and depend on the underlying distribution of distortions in *all* markets (Lipsey and Lancaster, 1956). What determines a policy's effectiveness is not how much it reduces a particular distortion, but whether it moves producers closer to or further from the efficient allocation. For example, correcting distortions in land markets may have limited or negative effects if the households that expand their landholdings are already inefficiently large due to preferential access to credit. Since considering a single market failure in isolation can lead to inefficient and even harmful policy recommendations, it is important to distinguish them empirically.

This paper separately identifies a wide range of distortions in Thai agriculture and characterizing how they combine to generate misallocation in equilibrium. Such a task requires a structural model:[2] Specifically, I estimate distortions in input (e.g. land, labor, and equip-

[1] See Magruder (2018) and Suri and Udry (2022) for recent overviews.Goldstein and Udry (2008); Breza et al. (2021); Karlan et al. (2014); Mobarak and Rosenzweig (2013); Diop (2023); Caunedo and Kala (2021); and Bold et al. (2017) provide excellent examples of each of these market failures, respectively. Emerick et al. (2016) and Jones et al. (2022) are examples providing experimental evidence on how these market failures can compound each other.

[2] If there are K potential (binary) market failures, the ideal experiment would require 2^K treatment

ment) and financial (credit and insurance) markets.[3] Under general production and utility functions, distortions in these markets each affect households' input demands through distinct wedges. However, the full set of input and financial wedges cannot be separately identified using solely production data (Hsieh and Klenow, 2009) — there generally is no way to tell whether a household uses less of an input because it cannot obtain it at the market price or because it is financially constrained. In particular, analyses that treat farm households as profit-maximizing firms cannot separately identify the distortions induced by uninsured risk.

However, unlike typical firms, farm households are also consumers. Under imperfect markets, household consumption enters their investment decisions and thus contains information about how production is distorted (Benjamin, 1992). I leverage this information to identify how credit constraints and uninsured risk distort households' productive choices distinctly from frictions in input markets. In particular, credit constraints enter as a wedge between the marginal utilities of consumption at planting and at harvest (reflecting the inability to smooth consumption across time by borrowing against future harvests). Meanwhile, uninsured risk enters through the covariance between production shocks and the marginal utility of consumption at harvest (reflecting the dependence of consumption on realizations of output when households cannot use insurance to smooth consumption across *states of the world).* On the other hand, input frictions function like a tax or subsidy and can be identified from dispersion in input *composition* across households.

I then estimate the production function using a novel structural method, which I introduce in Chapter 3. I use these estimates of the production function and distortions to calculate aggregate TFP under the observed allocation, the efficient allocation, and counterfactual distributions of distortions. Crucially, my estimation strategy is only possible when both input and financial distortions are well-specified. Otherwise, the common alternative is to calibrate the production function using revenue shares from a setting in which perfect markets are assumed to hold, such as the US or Canada (e.g. Adamopoulos and Restuccia, 2020; Chen et al., 2023), or use lagged inputs as instruments (e.g. Shenoy, 2017, 2021; Manysheva, 2021).

I implement my approach with the Townsend Thai Data, which is a 196-month panel of rural households in 16 Thai villages (with annual surveys in another 48 villages over the same period) from 1998 to 2014. Many studies have used the Townsend Thai Data to provide evidence of credit constraints (Kaboski and Townsend, 2011, 2012) and imperfect risk-sharing (Kinnan and Townsend, 2012; Karaivanov and Townsend, 2014; Samphantharak and Townsend, 2018; Kinnan et al., 2024). Shenoy (2017) also estimates a lower bound on input misallocation of about 11% of TFP. I interpret these findings as evidence of both imperfect financial and input markets in Thailand and view this paper as the first full decomposition of their costs.

---

arms, at the village (or higher) level of aggregation.

[3] These are the distortions I find to be most relevant in the Thai context. In general, the model I develop in Section 1.2 can accommodate distortions in financial markets and $K-1$ input markets if there are K inputs.

However, many of the institutional features common in other studies of misallocation, such as restrictive land policy and absence of credit markets, do not apply.[4] This makes Thailand a useful benchmark for less developed countries; finding nontrivial amounts of misallocation suggests that favorable institutions alone do not guarantee efficiency. The level of misallocation in Thailand may therefore be a more realistic counterfactual for institutional reforms in these settings than full efficiency.

I present four main empirical findings: First, I find that going from the observed to efficient allocation increases aggregate TFP by 31%. This is similar to estimates of total misallocation of 19% in Shenoy (2017) from Thailand (albeit using different methodologies and data), but substantially lower than estimates of 53% from China (Adamopoulos et al., 2022b), 97% from Ethiopia (Chen et al., 2022), 259% from Malawi (Chen et al., 2023), and 286% from Uganda (Aragon et al., 2022). These gains increase to about 35% when allowing the aggregate supply of tradable inputs to respond to increased aggregate TFP, as in Donovan (2021).

Second, I decompose these gains into the effects of eliminating either friction in isolation and the interaction effect from eliminating them simultaneously. I find that removing financial distortions while holding observed input wedges fixed would achieve 15-18% TFP gains relative to the observed allocation while removing input distortions alone would achieve 7-11% gains. Thus, TFP can be increased by a further 5–9% (relative to baseline) by addressing both sets of distortions together. While the sign of these interaction effects is theoretically ambiguous, in the data it is positive because more financially constrained households are relatively subsidized in input markets.[5]

Third, I model the effects of incrementally reducing distortions in one or more markets. This may represent a more realistic policy scenario when budgetary, political, or feasibility constraints make it impossible to eliminate some distortions entirely. While reducing both input and financial distortions simultaneously yields large complementarities relative to reducing either in isolation, most of these complementarities are only unlocked after large improvements to both sets of markets. In contrast, small reductions to one or both sets of input and financial frictions would have modest effects on aggregate productivity. This suggests that there are diminishing returns to addressing a single distortion in isolation and locally increasing returns from "big-push" policies that achieve major improvements from multiple markets simultaneously.

Finally, I analyze some distributional implications of reducing distortions. In the data, wealthier households tend to have much larger farm sizes. Each counterfactual leads to a more concentrated farm size distribution in which most households contract, but which households expand depends on which distortions are reduced. Reducing financial frictions weakens the correlation between farm size and baseline income by reallocating from the

[4] Thai agriculture features important distortions at the sectoral level, including heavy price supports for rice and fertilizer. However, this would only affect conclusions from the model in Section 1.2 to the extent it creates variation in prices across households in the same location, which is unlikely to be the case.

[5] This is consistent with evidence that poorer households over-supply labor to their own farms because the shadow value of their time is lower (Dillon et al., 2019; Jones et al., 2022).

wealthiest households to those at the middle of the income distribution. However, removing input frictions alone further concentrates resources towards wealthier households, exacerbating inequality. Under the efficient allocation, the progressivity of financial reform outweighs the regressivity of input reform, reducing the correlation between farm size and baseline wealth.

This paper's main contribution is developing a framework and estimation strategy to attribute misallocation to failures in distinct markets. Doing so is important not only for understanding where misallocation comes from but for developing policies to address it. This is because unmodeled distortions can bias estimates of misallocation and even suggest harmful policies, depending on how the measured distortions are correlated with unmeasured ones. Recent advances in the misallocation literature (e.g. Carrillo et al., 2023; Sraer and Thesmar, 2023; Hughes and Majerovitz, 2023) show how misallocation can nonparametrically be estimated from (quasi-)experimental variation but are generally unable to trace misallocation to its different sources. There is also a growing literature applying quantitative misallocation models to microdata in agriculture (Adamopoulos and Restuccia, 2020; Adamopoulos et al., 2022a,b; Aragon et al., 2022; Chari et al., 2021; Chen et al., 2017, 2022, 2023; Donovan, 2021; Gottlieb and Grobovšek, 2019; Manysheva, 2021; Shenoy, 2017). However, these papers typically model a single distortion in isolation or combine all distortions into a composite wedge. Notable exceptions are Manysheva (2021), who models the explicit dependence of credit constraints and land distortions through the collateral channel, and Shenoy (2017) who derives bounds for input and financial misallocation under assumptions on the joint distribution of distortions in Thailand. In contrast, I estimate a more complete range of distortions and model how the effects of counterfactual policies depend on their underlying distribution. Importantly, I show how my results differ substantially from the conclusions one would draw using other methods.

An important advantage of this framework is that it allows me to remain agnostic towards the specific institutions that generate distortions. These distortions have many potential, possibly simultaneous, causes and conclusions may depend on which ones a model specifies. For example, recent empirical work has identified expropriation risk (Goldstein and Udry, 2008), incomplete contracting (Burchardi et al., 2019), an explicit cap on landholdings (Adamopoulos and Restuccia, 2020), lack of titling (Chen et al., 2022), land fragmentation (Bryan et al., 2022), and others, as contributing to imperfect land markets. It would be impossible to capture all of these explicitly in a single model. Instead, my method allows me to diagnose how distortions in each market affect aggregate productivity without strong assumptions about their root causes.

I also contribute to the recent literature on how measurement error can inflate estimates of misallocation by using a model to separate between financial frictions and input mismeasurment. Rotemberg and White (2021) and Bils et al. (2021) find large upward biases due to measurement error in U.S. and Indian manufacturing. Meanwhile, Gollin and Udry (2021) argue that up to 70% of observed productivity dispersion in Ugandan and Tanzanian agriculture is due to measurement error and unobserved heterogeneity. This is supported by evidence of large and systematic measurement error in survey measures of agricultural land, labor, and output (e.g. Arthi et al., 2018; Desiere and Jolliffe, 2018; Abay et al., 2019, 2021).

Estimating a wider range of distortions helps me overcome these concerns and avoid having to infer them from a noisy residual. In particular, observed productivity dispersion is a (nonlinear) function of true misallocation and measurement error. When estimating a model with only a subset of distortions, e.g. only input distortions, the residual contains both financial distortions and measurement error. In other words, measurement error looks like a distortion in the data — and will tend to inflate estimates of misallocation.[6] However, directly estimating financial distortions allows me to distinguish between measurement error and true misallocation in this residual.[7] Without estimating both input and financial distortions, one would not be able to make this distinction.[8] I find that this would produce slightly larger estimates of misallocation than my model does and would suggest that eliminating financial distortions would *lower* aggregate productivity. This occurs due to the correlation between financial distortions and measurement error.

The rest of this chapter is organized as follows: In Section 1.2, I present the theoretical framework and derive expressions for financial and input wedges at the household level, showing how they map to aggregate misallocation. Section 1.3 provides more information about the Thai data and context. Section 1.4 presents the estimation framework I develop and the results. Section 1.5 shows the counterfactuals that I evaluate and Section 1.6 concludes.

1.2 Model

I propose a dynamic farm household model to characterize how frictions in financial and input markets generate distinct wedges in households' input demands. In equilibrium, these create dispersion in marginal revenue products (TFPR in the language of Hsieh and Klenow (2009)) across households, lowering aggregate TFP relative to the case of perfect markets. The model is dynamic and features many possible sources of distortions, but allows their effects on each market to be separately identified from three sets of first-order conditions.

While this allows me to estimate distortions while remaining agnostic towards the specific institutions that generate them I cannot prescribe specific policies without further assumptions on the root causes of distortions in each market. Doing so would require distinguishing between, for example, limited commitment or asymmetric information in risk-sharing networks and expropriation risk and lack of titling in land markets. While further research is

[6] The effect of measurement error on misallocation is theoretically ambiguous, but measurement error would need to be sufficiently negatively correlated with true distortions to create a downward bias.

[7] Of course, estimated quantities (TFP and wedges) contain error as well. However, TFP estimates (by design) remove much of the error in raw input measurements and are therefore less noisy. Moreover, having estimates of financial frictions allows me to compute both TFP-based and input-based estimates of aggregate productivity under any allocation.

[8] In the expression I derive for misallocation in Section 1.2, mismeasurement in inputs appears like a distortion in the sense that moves inputs either away from or closer to the efficient allocation. If it is correlated with other distortions and household productivity, the effects on measured misallocation are ambiguous, much like with two correlated "true" distortions.

required to further distinguish between these sources of distortions, quantifying the misallocation within each market may nonetheless be useful for policymakers.

1.2.1 Environment

There are V villages[9] and time, indexed by t, is discrete. For simplicity, each village v has a fixed number of households N_v, indexed by j. Agriculture is the only sector in the villages and uses $K \geq 3$ inputs to produce a single numéraire good[10] I assume for simplicity that the supply of land $\bar{Q}_{1vt}$ and labor $\bar{Q}_{2vt}$ is fixed within villages. There is an urban sector with stand-in firms that produce a vector of other consumption goods, indexed by i, and the remaining $K-2$ inputs used in agriculture.[11] Each of these can be imported to the village at exogenous prices p_{ivt} for goods i and $\bar{w}_{kvt}$ for inputs k. However, households may face different (effective) prices for each input, as I describe below.

1.2.2 Production

Production is given by

$$Y_{jt+1} = F(q_{jt}, \varphi_{jt+1}) \tag{1.1}$$

where q_{jt} is a vector of K inputs applied by household$j$ at time t, and φ_{jt+1} is a shock realized at $t+1$, prior to harvesting output Y_{jt+1}. As is standard, I assume that $F_k > 0, F_\varphi > 0$,and $F_{kk} < 0$ for each k. I assume that F is common across households and fixed over time, but households may have heterogeneous time-varying productivity. Note that I treat all inputs as static – in a benchmark economy with complete rental markets, households' input use at time t would not depend on their endowments or previous seasons' input choices.

I assume that $\bar{w}_{vkt}$ is the (endogenously determined) market price of each input k in village v at time t. However, households may face idiosyncratic taxes or subsidies such that they face prices $s_{jkt}\bar{w}_{vkt}$. Households may also be subject to upward or downward rations on inputs such that $\underline{q}_{jkt} \leq q_{jkt} \leq \bar{q}_{jkt}$.

While I only directly model the agricultural sector, allowing households to earn income from other sources is important to match the income diversification observed in the data. Households can invest in a portfolio of assets b_{jmt} with uncertain returns r_{jmt+1}. They may also be subject to borrowing constraints such that $\sum_m b_{jmt} \geq \bar{B}_{jt}$. b_{jmt} should also be thought of as capturing formal and informal insurance with state-contingent payouts. As with inputs, frictions in the asset market can be modeled by writing returns as $r_{jmt+1} \equiv \chi_{jmt}\bar{r}_{vmt+1}$, where $\bar{r}_{vmt+1}$ is the (endogenously determined and possibly stochastic) average

[9] I use the word villages for exposition but the unit of analysis I use in the empirical section is the tambon (township) (see Samphantharak and Townsend, 2018).

[10] This implicitly assumes that all farmers face the same output price, which I show in Section 1.4 is a reasonable approximation in the Thai setting.

[11] The urban sector plays no substantive role in the model but captures that many goods are not produced in the village.

return in village v.[12] Let B_{jt} denote a household's portfolio of assets and R_{jt+1} be the return to that portfolio. I denote the set of primitive taxes and rations that generate the distortions I derive below as $\mathcal{D} \equiv \{\chi, s, \underline{q}, \bar{q}, \bar{B}\}$.[13] Note that the estimation strategy I develop in section 1.4 does not depend on which frictions in $\mathcal{D}$ generate λ and τ. In section 1.5, I discuss how whether input frictions act as taxes or rations affects counterfactuals and compute results both ways.

1.2.3 Dynamic Program

I assume households j have time-separable, von Neumann-Morgenstern preferences with discount factor δ and per-period utility function $u(c, l)$, which I assume is continuously differentiable, strictly increasing, and concave in consumption c and leisure l. At time t, they maximize

$$\mathbb{E}_t \left[\sum_{s=t}^{\infty} \delta^{s-t} u(c_{js}, l_{js}) \right]$$

subject to the following budget constraint, in which M is total assets.

$$M_{jt+1} = M_{jt} + Y_{jt+1} - w'_{jt} q_{jt} - p'_t c_{jt} + R_{jt+1} B_{jt+1} - B_{jt} \tag{1.2}$$

which holds in each state of the world.

The household's value function satisfies the Bellman equation

$$V(Y, M, w, p, \varphi, R, \mathcal{D}) = \max_{c,q,B} u(c) + \delta \mathbb{E}_t V(Y', M', w', p', \varphi', R', \mathcal{D}') \tag{1.3}$$

subject to the budget constraint (1.2), borrowing constraint $\bar{B}$, and possible rations on hiring inputs in or out, $\underline{q}, \bar{q}$. Taking first-order conditions with respect to the choice variables c, q, and B:

$$(c) \qquad u_i(c) = \lambda p_i \tag{1.4}$$

$$(q) \qquad \delta \mathbb{E} \left[\frac{\partial V}{\partial Y}(Y', k', w', p', \varphi', R', \mathcal{D}) F_k(q, \varphi') \right] = \lambda w_k + \underline{\mu}_k - \bar{\mu}_k \tag{1.5}$$

$$(B) \qquad \delta R \mathbb{E} \left[\frac{\partial V}{\partial B}(Y', k', w', p', \varphi', R', \mathcal{D}') \right] + \mu^B = \lambda \tag{1.6}$$

where λ, μ^B, $\underline{\mu}_k$, and $\bar{\mu}_k$ are the Lagrange multipliers on the budget constraint, borrowing constraint $\bar{B}$, and rations on hiring inputs in and out, $\underline{q}, \bar{q}$, respectively. The first FOC simply states that households equate the marginal utility of expenditure on each good consumed within a period to a common Lagrange multiplier λ. The second implies that households equate the marginal utility of expenditure on each input to the expected marginal utility of its marginal product, unless an input ration binds. The third is simply the Euler equation with the possibility of binding borrowing constraints.

[12] $\chi_{jmt} = -\infty$ implies a household never purchases asset m.

[13] While the elements of $\mathcal{D}$ cannot be separately identified without many additional assumptions, they microfound the distortions the markets in credit, insurance and the k input markets I derive below.

1.2.4 Input Demands and Wedges

Applying the envelope theorem to the first-order condition (FOC) for q with simple substitutions yields the following expression for input demands:

$$\bar{w}_{vkt}\tau_{jkt} = \delta\mathbb{E}_t\left[F_k(q_{jt}, \varphi_{jt+1})\right]\Lambda_{jkt} \tag{1.7}$$

in which

$$\tau_{jkt} \equiv s_{jkt} + \frac{\underline{\mu}_{jkt} - \bar{\mu}_{jkt}}{\lambda_{jt}\bar{w}_{vkt}} \tag{1.8}$$

$$\Lambda_{jkt} \equiv \frac{\mathbb{E}_t[\lambda_{jt+1}]}{\lambda_{jt}} + \frac{cov_t(\lambda_{jt+1}, F_k(q_{jk}, \varphi_{jt+1}))}{\lambda_{jt}} \tag{1.9}$$

(1.7) simply states that households equate the marginal utility of expenditure on input k to the discounted expected marginal utility of its marginal product. Under input frictions, the (shadow) cost of each input k differs from the common market price by τ_{jkt} as defined by (1.8). Meanwhile, Λ captures how credit constraints and uninsured risk affect input demands through the two terms in (1.9), respectively. When credit constraints bind, (1.5) implies that $\lambda_{jt} > \mathbb{E}_t[\lambda_{jt+1}]$ since households cannot borrow against expected future earnings. Likewise, absent full insurance, consumption at $t+1$ will depend on the realization of production shocks, creating a non-zero covariance between λ_{jt+1} and (stochastic) marginal products, $F_k(q_{jt}, \varphi_{jt+1})$. This covariance may differ across inputs for a general production function. However, it will be negative if households are prudent ($u'''(c) > 0$), input k does not reduce risk ($F_{k\varphi} \geq 0$), and agriculture is not a hedge against overall portfolio risk. In this case, both mechanisms would reduce input demands relative to the case of perfect financial markets.

Λ_{jkt} and τ_{jkt} fully characterize the distortions generated by $\mathcal{D}$ in the markets for each input k. To see this, compare (1.7) to the benchmark of perfect markets, in which it reduces to expected profit maximization.

$$\bar{w}_{vkt} = \delta\mathbb{E}_t[F_k(q_{jt}, \varphi_{jt+1})] \tag{1.10}$$

This is identical to (1.7) when $\Lambda_{jkt} = \tau_{jkt} = 1$ for all j, k, t. In this case, ratios of marginal utilities λ are constant across households and cancel out and all households equalize expected marginal products to the common price of each input ($\tau = 1$). The equalization of marginal products across households implies the allocation is efficient. Note how deviations from efficiency are completely characterized by Λ_{jkt} and τ_{jkt}, which together define the distortions in the market for each input k.

I have thus far kept the model as general as possible to illustrate how financial and input frictions create distinct wedges under very general conditions. However, estimating the model requires functional form assumptions for F and u. While I discuss functional forms for utility in Section 1.4, I assume output is determined by the following Cobb-Douglas production function:

$$F(q, \phi) = A_{jt}\varphi_{jt+1}\prod_k q_{jkt}^{\alpha_k} \tag{1.11}$$

where A_{jt} is (possibly time-varying) household-specific TFP that is known ex-ante and φ_{jt+1} is an unanticipated shock with mean 1 realized after input decisions are made.[14] I assume decreasing returns to scale with $\gamma \equiv \sum_k \alpha_k < 1$.[15]

Under the Cobb-Douglas assumption. I can rewrite (1.7) to obtain the demand function for each input k.

$$q_{jkt} = \frac{\delta\alpha_k}{\bar{w}_{vkt}\tau_{jkt}} \frac{\mathbb{E}_t[\lambda_{jt+1}Y_{jt+1}]}{\lambda_{jt}} \tag{1.12}$$

(1.12) can also be expressed as

$$q_{jkt} = \frac{\delta\alpha_k}{\bar{w}_{vkt}\tau_{jkt}} \mathbb{E}_t[Y_{jt+1}]\Lambda_{jt} \tag{1.13}$$

where $\Lambda_{jt} = \dfrac{\mathbb{E}_t[\lambda_{jt+1}\varphi_{jt+1}]}{\lambda_{jt}}$ is now constant across inputs k.[16]

Meanwhile, distortions in the market for each input k enter through τ_{jkt}. In contrast, financial frictions Λ_{jt} distort the *scale* of production while the *composition* of inputs is only distorted by τ. To see this, take the ratio of demands for any two inputs, k and l:

$$\frac{q_{jkt}}{q_{jlt}} = \frac{\alpha_k}{\alpha_l} \frac{\bar{w}_{vlt}}{\bar{w}_{vkt}} \frac{\tau_{jlt}}{\tau_{jkt}} \tag{1.14}$$

Input ratios are solely a function of technology and relative market prices, which under perfect markets are constant across households in the same village-year. Thus any dispersion in input ratios can be attributed to τ.[17] This is a feature of any homothetic production function.[18]

[14] This is equivalent to writing

$$Y_{jt+1} = \tilde{A}_{jt} e^{\phi}_{jt+1} \prod_{k=1}^{K} q_{jkt}^{\alpha_k}$$

where $A_{jt} = \tilde{A}_{jt}\mathbb{E}_t[e^{\phi_{jt+1}}]$ and and $\varphi_{jt+1} \equiv \dfrac{e^{\phi_{jt+1}}}{\mathbb{E}_t[e^{\phi_{jt+1}}]}$. The normalization I use more clearly delineates the expected and unexpected components of TFP and guarantees that φ is strictly positive with mean 1.

[15] If $\gamma \geq 1$, then the efficient allocation is degenerate with only the most productive producer producing.

[16] To see this, it is useful to write the expectation in the numerator as $\mathbb{E}_t[\lambda_{jt+1}] + cov_t(\lambda_{jt+1}, \varphi_{jt+1})$ (since φ is mean 1 by construction). Also note that (1.13) can be written in closed-form by substituting (1.11) for Y_{jt+1} and solving the system of equations implied by (1.12)

$$q_{jkt} = \frac{\alpha_k}{\bar{w}_{vkt}\tau_{jkt}} \left(A_{jt}\Lambda_{jt} \prod_l \left(\frac{\alpha_l}{\bar{w}_{vlt}\tau_{jlt}} \right)^{\alpha_l} \right)^{\eta}$$

where $\eta \equiv \frac{1}{1-\gamma}$

[17] Note that $s, \underline{q}, \bar{q}, \bar{B}$, and χ are the primitives that determine the distortions τ and Λ.

[18] Note that under CES production, the ratio of τs on the right-hand side of (1.14) is raised to the elasticity of substitution σ.

1.2.4.1 Nonhomothetic Production

While the misallocation literature typically assumes a homothetic production function with Hicks-neutral shocks, this implies that all inputs contribute proportionally to the variance as outputs. Maintaining this assumption not only increases tractability but allows me to directly compare my results to others in the literature, showing how modeling the consumption side produces drastically different conclusion, holding the model fixed. However, a stark implication of homotheticity is that households facing the same input prices would use the same input mix and financial distortions would only affect the scale of production. To relax this assumption, I assume production takes the following generalized Cobb-Douglas form following Just and Pope (1978, 1979).

$$Y_{jt+1} = A_{jt} \prod_k^K q_{jkt}^{\alpha_k} + \varphi_{jt+1} B_t \prod_k^K q_{jkt}^{\beta_k} \tag{1.15}$$

where Y_{t+1} is output realized the period following production, q_{kt} is the quantity of input k at time t, A is TFP, and φ_{t+1} is a mean 0 shock realized before harvest and consumption at $t+1$. I assume that expected returns to scale $\gamma \equiv \sum_k \alpha_k < 1$ to ensure the socially optimal allocation is nondegenerate. The main difference between this and the workhorse Cobb-Douglas specification is that the variance of output now depends on input composition. Inputs are differentially risky if $\alpha \not\propto \beta$. In particular, α_k can be thought of as the elasticity of the expectation of output with respect to input k, while β_k is the elasticity of the *standard deviation* of output with respect to input k.[19]

$$q_{jkt} = \frac{\alpha_k \mathbb{E}_t[Y_{jt+1}]\mathbb{E}_t[\lambda_{jt+1}] + \beta_k cov_t(\lambda_{jt+1}, Y_{jt+1})}{\lambda_{jt} \bar{w}_{kvt} \tau_{jkt}} \tag{1.16}$$

Note how when $\alpha = \beta$ this reduces to (1.12). The only difference is that (1.16) assigns different coefficients to the expected and stochastic components of $\mathbb{E}_t[\lambda_{jt+1} Y_{jt+1}]$. Inputs with higher β contribute more to the variability of output, causing their demand to be disproportionately affected by imperfect insurance. In contrast, the separability of the shocks in the standard Cobb-Douglas means that the same Λ_{jt} applies to demand for each input.[20] The first term can be thought of as the wedge created by the inability to intertemporally smooth consumption and is constant across inputs. For example, if a household faces a binding borrowing constraint, then $\mathbb{E}_t[\lambda_{t+1}]$ would generally be lower than λ_{t+1}. The second term captures how uninsured risk affects demand. Again, one would expect the covariance term to be negative,[21] but this is amplified by how risky a given input is.

[19] I prefer this specification to that recently introduced by Bohr et al. (2023), since this functional form allows for a first order effect of uninsured risk on input demand as shown below. Note that this functional form nests the workhorse Cobb-Douglas specification $Y_{t+1} = A_t e^{\phi_{t+1}} \prod_{k=1}^K q_{kt}^{\alpha_k}$ if $\alpha = \beta$ and $B = A/\mathbb{E}[e^{\phi}]$.

[20] This is true for any homothetic production function.

[21] unless $u'''(c) \leq 0$ or returns from agriculture are sufficiently negatively correlated with those from other investments

However, this no longer allows the straightforawrd identification of τ from (1.14), requiring an alternative set of identification assumptions, which I discuss in Appendix A.2. I also show results from this more general specification and the results are broadly similar to those under the standard Cobb-Douglas.

1.2.5 Equilibrium

I now show how this model of farm-household distortions maps to aggregate misallocation. Let $\eta \equiv \frac{1}{1-\gamma}$, which is a nonlinear transformation of returns to scale that approaches ∞ as production approaches CRS. In what follows, I drop time subscripts to ease notation. A decentralized allocation yields the following expression for the share of factor k in a given location allocated to household j.[22]

$$\omega_{jk} \equiv \frac{\frac{1}{\tau_{jk}} \left(A_j \Lambda_j \prod_l \tau_{jl}^{-\alpha_l} \right)^\eta}{\sum_{h=1}^{N_v} \frac{1}{\tau_{hk}} \left(A_h \Lambda_h \prod_l \tau_{hl}^{-\alpha_l} \right)^\eta} \tag{1.17}$$

(1.17) is obtained by aggregating household first-order conditions (1.13) and implies that any allocation can be defined as a function of technology α, household TFP A, and distortions Λ and τ.[23]

An important distinction is whether factor stocks are fixed within locations or determined through general equilibrium.[24] In the base case, I assume that stocks of all inputs are fixed at the township level. I then continue to assume that land and labor are fixed but allow fertilizer, equipment, and seeds to be supplied from outside the village at an exogenous price while maintaining fixed stocks of land and labor at the township level.[25] In this case, which essentially treats villages as small open economies, demand for each input is pinned down by exogenous import prices $\bar{w}$ rather than endowments $\bar{Q}$. Definition 1 formalizes an equilibrium in either case.

Definition 1. *A decentralized equilibrium is defined by a set of prices* $\{\bar{w}_{vkt}, p_{it}, R_v t\}$, *an input allocation* $\{q_{jkt}\}$, *and a consumption allocation* $\{c_{jt}\}$ *such that*

1. *Households choose inputs, consumption and borrowing following (1.4)-(1.6) given initial asset holdings, prices, initial productivity and beliefs over future shocks.*

2. *Input demands* q_{jkt} *equal* $\omega_{jkt}\bar{Q}_{vkt}$, *where* ω_{jkt} *is given by (1.17) and* $\sum_{j=1}^{N_v} \omega_{jkt} = 1$ *for each* v

[22]Note that both the constant market price of each input $\bar{w}_{vkt}$ and aggregate supply $\bar{Q}_{kvt}$ are constants that cancel out of (1.17).

[23]Again, note that τ and Λ capture how primitive distortions $\mathcal{D}$ affect the equilibrium input allocation.

[24]The latter is the mechanism through which uninsured risk generates dispersion in fertilizer intensity even with perfect input markets in Donovan (2021).

[25]In a full spatial model, trade costs would determine the response of market-level demand to changes in within-market aggregate TFP, while migration costs would also be needed to determine counterfactual reallocation of labor across villages.

3. Λ_{jt} and τ_{jkt} are defined as in (1.8) and (1.13)

given a set of initial asset holdings M_{jt} and primitive distortions $\mathcal{D}$.

This also implies that when there are no distortions (i.e. $\Lambda_j = \tau_{jk} = 1$ for all inputs and households), the optimal allocation is

$$\omega_j^* \equiv \frac{A_j^\eta}{\sum_{i=1}^{N_v} A_j^\eta}\ \forall k \in \{1, \dots, K\}. \tag{1.18}$$

In this case, each input is allocated proportionally to household TFP, transformed by returns to scale.[26]

However, deviations of Λ and τ away from 1 in either direction lead to misallocation.

In equilibrium, expected aggregate productivity in a given village is:

$$\mathbb{E}[TFP_v] = \sum_{j=1}^{N_v} A_j \prod_k \omega_{jk}^{\alpha_k} = \frac{\sum_j \left(A_j \Lambda_j^\gamma \prod_l \tau_{jl}^{-\alpha_l}\right)^\eta}{\prod_k \left(\sum_{j\in v} \frac{\Lambda_{jk}}{\tau_{jk}} \left(A_j \Lambda_j^\gamma \prod_l {\tau_{jk}}^{-\alpha_l}\right)^\eta\right)^{\alpha_k}} \tag{1.19}$$

as opposed to the case of perfect markets in which this reduces to

$$\mathbb{E}[TFP_v^*] = \left(\sum_{j=1}^{N_v} A_j^\eta\right)^{\frac{1}{\eta}} \tag{1.20}$$

My base definition of misallocation is the percentage by which aggregate TFP would need to be increased to attain the efficient allocation, summed across locations and time periods.[27] Formally:

$$\mathcal{M} \equiv \frac{\sum_{v=1}^V \sum_{t=1}^T \mathbb{E}[TFP_{vt}^*]}{\sum_{v=1}^V \sum_{t=1}^T \mathbb{E}[TFP_{vt}]} - 1 \tag{1.21}$$

1.3 Empirical Setting and Data

I use monthly survey data from the Townsend Thai Monthly Survey, which covers 196 months of production and consumption in 16 villages from four tambons (townships), each in a different changwat (province). Two changwats (Chachoengsao and Lobpuri) are located in relatively developed Central Thailand and the other two (Buriram and Sisaket) are in the more rural North. The data span 1998 to 2014, during which substantial growth and structural change occurred after the Asian financial crisis. Table A.1.1 and Table A.1.2

[26] This is a standard result in the misallocation literature.

[27] Note that in the case where all inputs are in fixed supply within each location, aggregate TFP is proportional to aggregate output. Otherwise, aggregate demand for intermediate inputs is increasing in allocative efficiency, which further augments aggregate TFP.

provide some summary statistics of household demographics and agricultural production. There are a total of 791 households in the data, of which 568 engage in agriculture during the sample period. Over 68% of plots are grown with rice. In addition to crop production, households also earn income from wages, livestock and aquaculture, and other businesses. The average agricultural household sample in the household earns slightly less than half its income from crop cultivation. Importantly, the estimation procedure I develop in the following section can account for this feature of the data. In particular, it is robust to households endogenously selecting into production in a given year and does not impose a 1-to-1 mapping between farm income and consumption.

The data in Table A.1.2 show that markets for land, labor, equipment (mainly tractors, power tillers, and pumps), fertilizer, and seed exist. However, land and labor markets are much more active in the Central region and appear quite thin in the North. The average farm (defined as all of a household's plots in a given year) hires about 28% of its labor input, although more than two-thirds of farms hire some labor in a given season. Fertilizer, commercial seed, and mechanization use is widespread and is frequently acquired from outside the tambon. Land market participation is fairly low, with about 16% of farms renting any plots in a given season. However, this masks substantial regional heterogeneity: nearly 40% of farms rent land in Chachoengsao while only 2.5% rent land in Sisaket. About 89% of farms use fertilizer and over 90% of farms use equipment, which can be owned or hired.

There is quite active participation in both formal and informal finance, with people obtaining loans from government banks and credit schemes as well as neighbors and informal lenders. However, only 5.7% of loans are collateralized. The data include input quantities and expenditures (for transacted inputs), which allows me to calculate prices even though I do not observe them directly.[28] With this in mind, the data show a large degree of price dispersion in land and labor transacted on the market in all tambons, while the law of one price appears to hold for other inputs and output. In Table A.1.3, I plot the coefficients of variation for the price of each input and output for the average year in each tambon. There is very little variation in the prices of fertilizer, seed, and rice, but large variation in wages, land rents and tractor rental rates.[29] This lends support to my assumption that output, fertilizer, and seed are perfectly tradable within townships while other factors are not.[30]

For the main analysis, I treat the township as the level of aggregation, since villages within townships are often quite integrated (Kaboski and Townsend, 2011; Samphantharak and Townsend, 2018). I focus on the sample of households cultivating annual crops during the main season, which I define as crops taking fewer than 8 months from planting to harvesting

[28]I discuss how I value households' own inputs in the following section. While it is unclear to what extent input market frictions are pecuniary distortions that show up in these expenditures, I do not need to take a stand on this to estimate the homothetic production function. I discuss

[29]Much of this variation may also be coming from imputing prices as expenditures divided by quantities and averaging across months.

[30]Thailand did not have a targeted fertilizer subsidy during the sample period. While price controls were enacted in 2008 and 2011 (with the latter not binding), these would not violate my assumption since price controls would apply equally to all farmers in a township.

I drop all plots that do not report using land or labor.

In the main analysis, I also differentiate between labor at different stages of the production process, essentially treating planting, weeding, and harvest labor as separate inputs.[31] While stopping short of a fully sequential production function, this allows me to capture some of the seasonality in rural labor markets, where there may be tightness in planting and harvesting seasons but slack at other times. This gives me a total of 7 inputs: land, fertilizer, equipment, seed, and planting, weeding, and harvesting labor. I then aggregate inputs up to the farm-season level, since the model implies that the shadow prices of inputs and consumption apply to all plots cultivated by a household at a given time.[32] This gives me a panel of 6,223 farm-level observations across 16 years. Marginal utilities of consumption, λ are estimated using the procedure I describe in Section 1.4.1 from monthly expenditures on 47 food and non-food goods. I merge these estimates into the production panel to match the months of input use and harvests.

1.3.1 Evidence of Imperfect Markets in Thailand

Other authors have used the Townsend Data to study imperfect risk-sharing, borrowing constraints, and factor market imperfections. Kaboski and Townsend (2011, 2012) find that a microcredit expansion that occurred during the sample period partially relaxed binding credit constraints. Meanwhile, several papers suggest that kinship networks manage to share idiosyncratic risk fairly well (Kinnan and Townsend, 2012; Karaivanov and Townsend, 2014; Samphantharak and Townsend, 2018) but far from perfectly, as idiosyncratic shocks propagate through labor supply and financial networks (Kinnan et al., 2024). Meanwhile, Shenoy (2017) argues that input frictions reduce aggregate productivity by at least 6%.

Additionally, I implement two canonical tests of complete markets before imposing the structure of my model. First, Townsend (1994) provides a test of full insurance, under which a regression of log consumption on log income with household and village-year fixed effects should yield a coefficient of 0. Second, Benjamin (1992) tests the null hypotheses of a full set of complete markets, under which households' production decisions should be fully separable from their consumption decisions. In this case, household composition (and other variables associated with households' preferences) should be independent of labor use. While rejection of this null hypothesis does not identify which market fails, the common interpretation in Benjamin (1992) and related papers (Dillon et al., 2019, e.g.) is frictions in labor markets causing households with larger labor endowments to use more farm labor. Column (1) of Table A.1.4 presents the results of the Townsend (1994) test while columns (2) and (3) present the results of the Benjamin (1992). The former rejects at all levels of significance while the latter rejects at the 10% level when using household size as the single right-hand

[31] I use "weeding" as a shorthand for all midseason labor tasks, including fertilizing, irrigating, and pest control.

[32] See Gollin and Udry (2021) and Aragón et al. (2022) for further discussion of aggregation at different levels and its advantages/disadvantages with regard to measurement error. For robustness, I also compute all results using plots as the unit of aggregation.

side variable and at the 5% level when using the counts of household members in different age-sex bins.

While the regression coefficients in these tests do not have structural interpretations, it is useful to examine whether consumption is more or less sensitive to income shocks in villages where labor intensity depends more on household endowments. To test this, I run both tests cutting the sample into 64 village×4-year blocks and plot each of the coefficients against each other in Figure A.1.1. The coefficients appear negatively correlated with each other, suggesting that the joint distribution of distortions merits further structural analysis.

1.4 Estimation Framework

I now describe how each of the key components of the model λ, τ, α, A, and Λ are estimated in four steps. First, I estimate realized marginal utilities λs from the full sample of expenditure data in Section 1.4.1. I do so under the assumption of CRRA preferences as well as under the more flexible Constant Frisch Elasticity system of Ligon (2020). Second, I estimate input wedges τ from dispersion in input ratios within a township-year, as in (1.14), in Section 1.4.2. While inferring input distortions from factor ratios is standard in the misallocation literature, I discuss additional steps I take to avoid misattributing measurement error and unobserved heterogeneity to τ. Having estimated λ and τ, the production coefficients α are now identified from the moment conditions for input demands (1.12). In Section 1.4.3, I use the linear GMM specification derived in Chapter 3 to estimate α from these moment conditions and show the robustness of results to several alternative specifications. This allows me to back out TFP A and production shocks φ. The last step, which I discuss in Section 1.4.4 is to estimate the composite financial wedge Λ_{jt}, which depends on the covariance between the realizations of φ_{jt+1} and the marginal utility of consumption at harvest λ_{jt+1}.

1.4.1 Estimating marginal utilities (λ)

While the model in Section 1.2 doesn't require any particular structure on preferences over goods, estimation requires mapping disaggregated expenditure data into a measure of welfare, λ_{jt}.[33] This requires choosing a functional form for utility. To place as minimal structure as possible on preferences, I use the constant Frisch elasticity (CFE) demand system proposed by Ligon (2020). I discuss the theoretical properties and estimation of this demand system in section A.2. An advantage of the CFE demand system is that it flexibly accounts for non-homotheticity and can be estimated from incomplete data on expenditures and prices. However, I obtain very similar results when estimating λ assuming CRRA preferences, which, like many other commonly used demand systems, are a special case of CFE.

[33]Since all households are assumed to face constant prices for output and other goods, what matters for misallocation in the model are intertemporal and risk preferences. How different consumption goods are aggregated matters for accurately mapping disaggregated expenditures into MUEs, but does not otherwise influence misallocation.

I estimate λ using the full 196-month panel featuring 47 food and non-durable consumption goods.[34] The estimation also allows demands to vary with household composition, as measured by the counts of members in different age-sex bins. Figure A.2.1, which plots the time series of the average $\log \lambda$ in each township, shows that the estimates capture substantial variation in the MUE across tambons, over time, and across seasons. I also compute results using CRRA for robustness. Figure A.2.2 plots estimated $\log \lambda$ against log consumption expenditure, controlling for month fixed-effects. The elasticity of λ to total consumption value is (minus) the coefficient of relative risk aversion under von Neumann-Morgenstern preferences. Imposing CRRA preferences leads to an estimate of $\theta = 1.5$. To ensure that my results are not being driven by the choice of demand system, I compute all results using both CFE and CRRA λs. Reassuringly, the estimates of both the production function and counterfactuals are extremely similar.

1.4.2 Identifying factor frictions

I now describe how I use the dispersion in input ratios to separately identify τ.[35] Recall that Λ_{jt} is common across all inputs and plots used by a household in a given period. Therefore, it affects the overall scale of production but not input composition and cancels out of *relative* input demands (1.14). However, input ratios may be measured with error ν, such that we observe

$$\frac{\tilde{q}_{jkt}}{\tilde{q}_{jlt}} = \frac{\alpha_k}{\alpha_l} \frac{\bar{w}_{vlt}}{\bar{w}_{vkt}} \frac{\tau_{jlt}}{\tau_{jkt}} e^{\nu_{jkt} - \nu_{jlt}} \tag{1.22}$$

where $\tilde{q}$ denotes measured inputs and ν may include misreported quantities of inputs or heterogeneous input quality.[36] Since α_k and $\bar{w}_{kvt}$ are not household-specific, (1.22) shows that any dispersion in input ratios across households is either due to differences in the ratio of τs, unobserved quality or measurement error. However, (1.22) also highlights two challenges for identifying τ.

First, τs for K inputs cannot be identified with $K-1$ ratios. Because of this, most papers in the misallocation literature are only able to identify the *relative* distortion of land to labor (Hsieh and Klenow, 2009; Adamopoulos et al., 2022a). However, if at least one input, say K, were perfectly tradable within townships such that $\tau_{jKt} = 1$ for all households, the remaining $K-1$ τs are identified. This appears plausible for both seed and fertilizer in the Thai context. The survey asks households whether they have had trouble acquiring any inputs. Fewer than 1% of households answer yes for fertilizer or seed in a given year. Additionally, Table A.1.3 shows minimal price dispersion for both fertilizer and seed within a given township-year.[37] This allows me to compute results using either fertilizer or seed as

[34] While consumption of durable goods may be a concern in other cases, the CFE demand system can be consistently estimated from only a subset of goods.

[35] While this approach leverages the assumption of a homothetic production function, I discuss an alternative method that relaxes this assumption in Appendix A.2.

[36] It may be useful to think of q as a measure of effective input quantity.

[37] Much of this dispersion may also come from imputing prices by dividing expenditures by quantities.

the normalizing input. I use fertilizer in the main specifications, since it is less susceptible to unobservable quality but show that results are quite similar when using seed.[38]

I now describe my approach to distinguish true input distortions, unobserved heterogeneity, and noise. Results in both the micro and macro literatures recognize the potential for heterogeneous land quality to bias estimation (Benjamin, 1995; Gollin and Udry, 2021). I address this issue using a hedonic approach. Specifically, I train a model to predict rental values from observed plot features on a random sample of rented plots. These features include area, soil type, and quality, histories of drought, flood, erosion, and fertilizer application, proximity to water sources, roads, and the household, and (self-reported) sale values.[39] I use cross-validated boosted trees and test the model's fit on a holdout sample, achieving an R^2 of 0.54. I then use the model to assign rental values to plots that were cultivated by the owner, for which no rental price is observed. I then use observed and predicted rental prices as a measure of quality-adjusted land quantities.

There are some caveats to this procedure. First, distorted land markets may not accurately reflect true land quality in prices. While this approach allows for land distortions to take the form of either an implicit tax or a ration, it essentially assumes that there is no distortion to the *relative* prices of observable plot attributes, such as soil and proximity to water sources. Nevertheless, there is no a priori reason to assume that relative prices of different attributes should be distorted in a particular direction. Another concern is that transacted plots may be selected on unobservable physical attributes. However, the model would capture the value of these attributes to the extent they are correlated with observable attributes.

I then turn to input measurement. There is evidence of considerable misreporting of inputs in household surveys (e.g. Beegle et al., 2012; Carletto et al., 2013, 2015; Arthi et al., 2018; Abay et al., 2019, 2021). However, other papers in the misallocation literature either attribute all variation in observed input ratios to τ or only attribute the time average of distortions for each household in a panel to τ.[40] I therefore take a more intermediate approach and attempt to capture only the systematic variation in τs.[41] Although τs are unlikely to be fixed over time, they are likely to be highly serially correlated and also depend on household composition.[42] I therefore model τ as following an AR(1) process, conditional on household

[38] Although farmers use different varieties of fertilizer, for simplicity I use the market value of the total fertilizer used by households to compute τs. Note that since τs are computed relative to the village-year average, this does not affect the results under the model's assumptions as long as farmers' mix of fertilizer varieties is not distorted.

[39] A similar approach is applied by Gordeev and Singh (2023).

[40] While more conservative with respect to measurement error, the latter approach discards the time-varying components of true distortions. If τ represents a binding input ration, then the *shadow* price implied by the ration will depend on other time-varying state variables even if the ration itself stays fixed. Moreover, household fixed effects may pick up permanent differences in land quality in addition to average input distortions.

[41] This exercise is in a similar spirit to Bils et al. (2021), who leverage time-series variation to isolate the predictable part of distortions.

[42] LaFave and Thomas (2016) show that even mechanical changes to household composition in Indonesia

characteristics X_{jt}, with the following equation of motion.

$$\tau_{jkt} = \rho\tau_{jkt-1} + \kappa_k X_{jt} + \xi_{jkt} \tag{1.23}$$

The AR(1) model can be thought of as a coarse way of capturing how τ depends on unobserved market institutions and household state variables that may evolve over time. Substituting into (1.14) implies that $\log \tau_{jkt}$ can be written:

$$\begin{aligned}\log \tau_{jkt} &= \log\left(\frac{\bar{w}_{Kvt}q_{jKt}}{\bar{w}_{kvt}q_{jkt}}\right) + \log(\alpha_k/\alpha_K) + \nu_{jkt} \\ &= \rho_k\left(\log\left(\frac{\bar{w}_{Kvt-1}q_{jKt-1}}{\bar{w}_{kvt-1}q_{jkt-1}}\right) + \log(\alpha_k/\alpha_K) + \kappa_k\Delta X_{jt} + \nu_{jkt}\right) + \xi_{jkt}\end{aligned} \tag{1.24}$$

This simply states that τ, net of measurement error, is proportional to the ratio of the *market* value of input K to k used by household j at time t,[43] which can be expressed as a lagged dependent variable model after moving measurement and constants ν_{jkt} to the right-hand side.

$$\log(q_{jKt}/q_{jkt}) = \rho_k \log\left(q_{jKt-1}/q_{jkt-1}\right) + \kappa_k\Delta X_{jt} + \iota_{kvt} + \upsilon_{kvt} \tag{1.25}$$

where ι_{kvt} is a location-input-time fixed effect that combines constants and υ_{kvt} is the composite error term corresponding to $\rho\nu_{jkt-1} - \nu_{jkt} + \xi_{jkt}$.

I estimate this using both 2SLS and standard dynamic panel GMM approaches (Blundell and Bond, 1998). I use the predicted values of $\frac{q_{jKt}}{q_{jkt}}$ — normalizing by their location year averages — as my estimate of τ_{jkt}.[44]

1.4.2.1 τ Estimation Results

In Figure A.2.3 and Figure A.2.4, I plot kernel densities of the estimated τs for land and labor alongside those derived from raw input measurements, using the time-series average input ratio for each household as time-invariant measure of τ, and for the estimated τ for land not accounting for heterogeneous land quality. Each specification reduces the variation in measured input ratios relative to the raw data. The standard deviations of the estimated τs for land and labor are about one-third of those calculated from raw input ratios. Much of this difference is likely due to error in raw input measurements. Overall, my preferred estimates may offer a more robust approach to dealing with measurement error in inputs

due to the aging of members significantly predict land/labor ratios.

[43]Note that since $\bar{w}_{kvt}$ is constant across households in the same location-year by construction, they can also be subsumed into location-time fixed effects.

[44]This normalization implies that τ is the deviation from village-average factor ratios. While this is consistent with a one-sector model, it rules out common cases in which the shadow wage for farm-households is below the market wage, such as labor rationing (Breza et al., 2021; Agness et al., 2022). In this case, the τs I estimate would be too high and this would bias the production function coefficients upward in the procedure I describe in Chapter 3. However, the coefficients I estimate for labor are already quite low, suggesting that this may not be a major issue in my sample.

without discarding time variation in input wedges. Nevertheless, it is possible that they do not capture all of the idiosyncratic variation in the true underlying τ. However, the estimation and counterfactual results are quite robust across various specifications.

1.4.3 Production function estimation

A reasonable estimate of the production function is crucial for any analysis of misallocation. As in similar models, the elasticity of aggregate output to wedges is $\eta \equiv \frac{1}{1-\gamma}$, which goes to infinity as returns to scale approach 1. This means that even small biases in production can greatly affect estimates of misallocation. The challenge is that the usual identification concerns that plague production function estimation can be even more severe when input choices are subject to multiple distortions. However, the literature has typically calibrated the production function using input shares from settings where markets are assumed to function well, (Chen et al., 2023; Adamopoulos and Restuccia, 2020; Adamopoulos et al., 2022b), or used lagged instruments to estimate the production function in-sample (Shenoy, 2017; Manysheva, 2021). The issues with the former approach are that the underlying production function may be different in the U.S. and Canada than in Sub-Saharan Africa and Southeast Asia. The latter approach is valid in theory (Shenoy, 2021) but relies on strong assumptions about the nature of unobserved shocks (i.e. autoregressivity).

Meanwhile, structural methods in the spirit of Olley and Pakes (1996) overcome endogeneity concerns by using the firm's optimal choice of a flexible input to proxy for anticipated productivity shocks, both observable and unobservable. However, this requires the firm's optimization problem to be well-defined. Commonly, this amounts to assuming that firms maximize profits in competitive input markets (Gandhi et al., 2020) or under certain types of markups (Asker et al., 2019). Section 1.3 already shows evidence that these conditions do not hold in Thai agriculture.

To estimate the production function for farm households, I adapt the structural approach to directly account for the ways in which input and financial frictions distort households' input choices, through the λs and τs that I've estimated. In Chapter 3, I show how the first-order conditions for input demands provide moment conditions that can be exploited to recover the production function parameters under rational expectations using linear GMM in the spirit of Hansen and Singleton (1982). In Chapter 3, I further show how this estimator can be applied to the generalized non-homothetic Cobb-Douglas specification in subsubsection 1.2.4.1 and a dynamic multi-stage Cobb-Douglas with sequential shocks.

1.4.3.1 Production Function Estimates

With estimates of λ and τ, I am able to estimate the production function following the procedure in Chapter 3. In the main specification, I use continuously updated GMM (Hansen et al., 1996) with planting, weeding, and harvesting labor, land, fertilizer, equipment, and seed as inputs, with lags of λ from the previous 5 months and tambon dummies as instruments.[45] Given that the estimator relies on generated variables, I compute standard errors by block bootstrapping the entire estimation procedure, including estimates of λ and τ, at the household level.

I compute the main results assuming the annual time-preference discount factor $\delta = .95$. I also show robustness to Kaboski and Townsend (2011)'s estimate of $\delta = .926$ using the same data and 1. Since the median season covers 5 months, I convert the annual δ to its 5-month equivalent. Note that δ doesn't affect the results qualitatively, since it is constant across households and cancels out of (1.17). However, lower values of δ would lead to higher estimates of returns to scale and larger estimates of misallocation across specifications.[46]

The results are presented in Table 1.1. Column 1 presents the main results, using the CFE demand system to estimate λs and fertilizer as the normalizing input, restricting the sample to rice plots and aggregating to the farm level. The coefficients all take reasonable values for agricultural production functions and together imply returns to scale $\gamma \approx 0.83$, which is larger than other papers in the literature.[47] I test the overidentifying restrictions of the full model against one with a single lag of λ and tambon dummies as instruments. While I reject the null hypothesis that all instruments are exogenous, this appears to arise from the Cobb-Douglas specification struggling to capture heterogeneity across regions. I fail to reject the validity of the lagged λs as instruments when applying a difference-in-J test (what Hayashi (2011) calls a C test). In Table A.2.1, I also show robustness to using seed rather than fertilizer as the normalizing input for τ, using CRRA to estimate λs instead of the more general CFE specification, restricting to rice plots, treating all labor as a single input, and aggregating to the plot rather than farm level. All specifications produce extremely similar results.

In Columns 2 and 3, I show the estimates of α and β from the generalized Cobb-Douglas specification in Appendix A.2. The αs are quite similar across specifications, suggesting that standard Cobb-Douglas would fit the data well if households were fully insured or risk-neutral. This suggests that the bias from failing to account for differentially risky inputs is relatively small. Nevertheless, there are important differences between the two specifications. Recall that the generalized production function reduces to Hicks-neutral Cobb Douglas when $\alpha = \beta$, meaning that the elasticity of expected output with respect to input k is the same that of the standard deviation of expected output(Just and Pope, 1978, 1979). Inputs with larger β_k relative to α_k can be considered relatively "risk-augmenting." The results in Table 1.1 suggest that inputs chosen at planting (land, seed, fertilizer and planting labor) appear to be risk augmenting (although I cannot reject equality of α and β for land). The difference between β and α is most striking for planting labor, suggesting that its returns are highly variable. Meanwhile, other inputs appear neither risk-enhancing or risk-reducing, based on

[45] Given that t corresponds to a season in the model in Section 1.2, the lagged λs should be thought of as occurring within different subperiods prior to planting.

[46] I show in Section 1.5 that while a lower δ increases my estimates of misallocation by a few percentage points, it doesn't alter any of the qualitative conclusions.

[47] Note that a lower value of γ would lower estimated misallocation because inputs are optimally allocated proportionally to $1/(1-\gamma)$.

the similarities between α and β.[48]

1.4.4 Recovering TFP and financial wedges

With the production coefficients in hand, the next step is to recover household TFP A and financial wedges Λ. This is substantially more challenging than estimating the production function because it requires taking a more explicit stance on what households do and do not anticipate in each period, as opposed to relying on sample averages. Notably, these issues affect any quantitative analysis of misallocation. I first take the average of realized TFP, computed using the estimated αs as $\bar{A}_j \equiv \frac{1}{T}\sum_{t=1}^{T} Y_{jt+1}/\prod_k q_{jkt}^{\alpha_k}$. I then try and predict deviations of realized household TFP in each period from $\bar{A}$ using variables in households' information sets $\mathcal{I}_{jt}$. Both ridge regressions and boosted trees using a rich set of features achieve an R^2 of close to zero, suggesting that $\bar{A}_j$ is a good approximation to anticipated TFP. Using this approximation means that production shocks $\varphi_{jt+1} = Y_{jt+1}/\prod_k \bar{A}_j q_{jkt}^{\alpha_k}$.

Recall from Section 1.2 that

$$\Lambda_{jt} = \frac{\mathbb{E}_t[\lambda_{jt+1}\varphi_{jt+1}]}{\lambda_{jt}}$$

While the denominator of Λ_{jt} has already been estimated, the numerator is an (unobserved) subjective expectation conditional on time t information. λ_{jt+1} is a function of φ_{jt+1} as well as households' other sources of income (including returns from other investments and payouts from insurance networks) which may be correlated with realizations of φ_{jt+1}. Therefore $\mathbb{E}_t[\lambda_{jt+1}\varphi_{jt+1}]$ can also be thought of as a function of households' state variables at time t integrated over the distribution of φ_{jt+1}.[49] I use supervised machine learning to approximate this function as flexibly as possible using the rich set of time t information. This is a valid approximation under rational expectations under similar conditions as in Section 1.4.3 — essentially realized shocks must be uncorrelated on average with the state variables used as predictors. Dividing these predictions by the observed λ_{jt} identifies Λ_{jt}.[50]

[48] One might expect harvest labor to be fairly insensitive to risk. However, there is still substantial uncertainty over the value of output due to price fluctuations and postharvest losses in developing country agriculture (Aggarwal et al., 2018; Omotilewa et al., 2018; Burke et al., 2019; Channa et al., 2022). Also refer to 2. While this paper uses a static production function that does not permit attributing risk to different stages of production, see (Felkner et al., 2012) and 3 for estimates of a sequential production function that permits this.

[49] For example, under CRRA utility

$$\mathbb{E}_t[\lambda_{jt+1}Y_{jt+1}] = \int_\varphi \frac{\varphi}{(R_{jt+1}(\varphi)B_{jt} + A_{jt}\varphi\prod_k q_{jkt}^{\alpha_k} - B_{jt+1}(\varphi) - \sum_k w_{jkt+1}(\varphi)q_{jkt+1}(\varphi))^\theta} d\varphi$$

where the possible dependence of $t+1$ variables on realizations of φ is made explicit.

[50] An alternative would be to model Λ as a function of returns to agriculture, other assets, and state-contingent transfers integrated over the distribution of the shocks. However, this would require further assumptions on preferences and the distribution of shocks, which is beyond the scope of this paper.

I predict Λ_{jt} with gradient boosted trees (Friedman, 2001), using estimates of A_j, the lagged λs used as instruments in Section 1.4.3, and a rich set of information from household's balance sheets as features. This includes agricultural and non-agricultural assets, cumulative income from agricultural and non-agricultural investments. The R^2 of this prediction is 0.35, while the R^2 when predicting λ_{jt+1} alone is 0.63. Of course, a perfect model of households' subjective expectations of future consumption *shouldn't* have an R^2 close to 1 under incomplete insurance. Nevertheless, the results suggest that consumption is fairly predictable despite substantial uncertainty in production (the R^2 when predicting φ is negligible). I also obtain similar results when using a ridge regression instead of boosted trees.

In Tables A.2.2 and A.2.3, I show that these estimates of Λ are correlated with untargeted observables in the data on borrowing, saving and mutual gift-giving (insurance) networks. In particular, it appears that those with higher Λ (less constrained) have larger loans and make larger informal transfers (referred to as "gifts" in the survey) in typical years. This holds across specifications of Λ and also when splitting it into credit and risk wedges. I also show that positive (negative) production shocks are associated with gift outflows (inflows).[51]

Figure 1.1 shows the distribution of Λ. The mean of Λ in the main specification is 0.86, with a median of 0.77. While these estimates are close to 1, as would be the case under perfect financial markets, raising them to the elasticity $\eta \approx 6$ implies that the average (median) household only produces at 42% (23%) of its desired scale. This is consistent with evidence of functional but incomplete credit markets and risk-sharing in this setting (Kaboski and Townsend, 2011; Karaivanov and Townsend, 2014; Samphantharak and Townsend, 2018; Kinnan et al., 2020). It also suggests that for the 27% of households with $\Lambda_{jt} > 1$, agriculture is a hedge against other sources of income, which is also consistent with evidence from other countries that households use off-farm labor to smooth consumption (Kochar, 1999) or substitute on-farm for off-farm labor when seasonal consumption constraints bind (Fink et al., 2020). Moreover, households in my sample have fairly diversified income streams that may be negatively correlated with returns to crop production.[52]

1.5 Results and Counterfactuals

Estimates of financial distortions Λ, input wedges τ, production coefficients α, and TFP A allow misallocation to be computed using the expression for aggregate TFP (1.19) relative to the efficient allocation (1.20). The model in Section 1.2 implies that overall misallocation depends on the joint distribution of Λ, τ and A.[53] Before delving into counterfactuals, I provide some descriptive graphical evidence to characterize this distribution.

[51]By remaining agnostic to the primitives that cause distortions, it is unclear which moments in the data the wedges I estimate should map to. While taking such a stand may help discipline the model, it may rule out other important channels.

[52]Imposing that $\Lambda \leq 1$ does not change the qualitative conclusions in the counterfactuals in Section 1.5, although it lowers estimates of misallocation.

[53]This is an extension of results in Hsieh and Klenow (2009) and Adamopoulos et al. (2022b). regarding the covariance between wedges as a sufficient statistic for misallocation.

Descriptive Results

Figure 1.2 plots 2D histograms of TFP-weighted input and financial distortions and reports their correlation coefficients.[54] The top left panel plots the Cobb-Douglas price index of τs, $\prod_l \tau_{jlt}^{\alpha_l}$ against the estimates of financial distortions Λ, each weighted by TFP A. The top right panel plots the τ for land against Λ while the bottom left plots the index of τ for the three types of labor (planting, weeding, and harvesting) considered. The bottom right panel plots the unweighted histogram of the τ price index and Λ. The positive correlation between τ and Λ suggests that, on average, more financially constrained households are relatively *subsidized* on inputs. More productive households also appear to be less financially constrained and more taxed on inputs. This corresponds to the conventional wisdom that poorer households oversupply labor to their own farms under imperfect labor markets (LaFave and Thomas, 2016; Breza et al., 2021; Jones et al., 2022).

This implies that the observed distortions partially offset each other — relaxing credit constraints would disproportionately direct capital toward farms that are effectively subsidized on inputs. The direct gains from relaxing credit constraints are large enough to swamp this effect but are smaller than they would be if credit constraints were uncorrelated with input distortions.[55] The results also show that distortions for land and labor are positively correlated. Most of the misallocation literature rules this out by assumption, modeling τ as a distortion in the *relative* price of land and labor. However, I am able to relax this assumption by using fertilizer and seed as normalizing inputs when estimating τs.

Main Counterfactuals

I now proceed to compute counterfactual *expected* aggregate productivity following (1.21) under the following four scenarios: (1) the first best allocation; (2) the baseline allocation, with all of the distortions I measure; (3) an allocation with perfect financial markets and the observed input wedges; (4) an allocation with perfect input markets and the observed financial wedge. I consider counterfactual allocations within township-years and then sum up these gains across townships in each of the 16 years of the sample.

I provide four main sets of results. First I characterize overall misallocation in Thailand. Second, I decompose misallocation into input distortions, financial distortions, and interactions between them. I then show other methods that are more susceptible to measurement error in inputs yield starkly different results. Finally, I use the model to approximate the marginal returns to incremental reductions in one or both sets of distortions. Note that the results below all refer to expected TFP since the realizations of ex-post shocks cannot be considered misallocation.

[54] In equilibrium, the influence of each of these distortions is weighted by household TFP.

[55] TFP governs the incidence of these distortions; since it is the sole determinant of scale under the efficient allocation, multiplicative wedges such as Λ or τ exert a large influence on the aggregate economy when it affects firms that command more inputs. In **??**, I show that results are similar without weighting distortions by TFP.

The gains from reallocation depend on whether one assumes that the stock of tradable inputs is held fixed or can respond to changes in counterfactual demand. The results also depend on whether one assumes input frictions take the form of implicit taxes or rations. I show how results depend on each of these cases below.

Baseline Misallocation

Figure 1.3 plots the gains from reallocation under each counterfactual as a percentage of (expected) aggregate TFP in the observed allocation. The three counterfactuals I consider are (1) eliminating financial distortions (i.e. setting $\Lambda = 1$) holding input frictions τ fixed; (2) eliminating input distortions (setting $\tau = 1$) while holding Λ fixed; and (3) eliminating all distortions. The blue (left) bars show results holding the aggregate supply of all inputs fixed, as if villages are in autarky. In this case, aggregate TFP is directly proportional to aggregate output. This is a relatively conservative assumption because it excludes gains from the increased aggregate demand for tradable inputs. The green (right) bars allow intermediate inputs (fertilizer, seed, and equipment) to be imported from outside the village at a constant price (as if the village were a small open economy). Confidence intervals from 200 bootstrap replications are shown for each specification.

The gains from full reallocation are 31% in the baseline case and 35% when aggregate supply of tradable inputs is allowed to adjust. The baseline estimates are similar to Shenoy (2017)'s estimates from Thailand, which I discuss below. On the other hand, my results are an order of magnitude lower than some estimates from Africa of up to 286% gains from reallocation (Chen et al., 2023; Aragon et al., 2022). The additional gains from allowing the aggregate supply of tradable inputs to adjust are much smaller than those in Carrillo et al. (2023), where they account for almost all the estimated misallocation.[56]

Decomposing Misallocation

It is clear from the first two groups of bars in Figure 1.3 that both sets of markets contribute significantly to misallocation in isolation. Perfecting financial markets while holding observed input distortions intact achieves about 48% of the possible efficiency gains, or 15% of observed TFP. Similarly, removing input distortions holding observed financial frictions intact achieves about 23% of these gains (7% of TFP).

Notably, these two gains sum to less than 100%, meaning the gains from full reallocation are more than the sum of its parts. This is because Λ and τ are positively correlated (when weighted by TFP). In other words, the most financially constrained households are relatively subsidized in input markets, especially labor, as shown in Figure 1.2 and **??**.[57] The effect of relaxing financial constraints is thus attenuated — but not offset — by reallocating resources to farms made inefficiently large by other distortions. Overall, these patterns suggest that

[56]See Donovan (2021) for a more detailed discussion of this channel where the price of intermediates is endogenous in general equilibrium.

[57]This reflects the common finding that poorer households tend to oversupply labor to their own plots.

the effects of policies targeting a single market failure would be attenuated, rather than amplified, by failures in other markets.

I also compute counterfactuals relaxing the distortions for some inputs but not others. Table 1.2 shows the results of removing wedges from each of these markets, with and without relaxing financial constraints. Reducing frictions in labor markets is slightly more effective than for land markets, despite them accounting for roughly equal expenditure shares. The sum of gains from reducing individual frictions is also more than the gains from reducing all of them simultaneously. While input frictions are negatively correlated with financial distortions, they are positively correlated with each other. In other words, reducing frictions in land markets also indirectly addresses labor market distortions by reallocating resources toward households that are relatively taxed.

Intermediate Policies

The results above consider the gains from completely eliminating one set of distortions while holding others fixed at observed values. However, while policymakers have a menu of policy instruments to choose from, they may not be able to fully eliminate distortions. The model allows me to estimate aggregate TFP under any values of Λ and τ. I therefore conduct a simple illustrative exercise in Figure 1.4, in which I plot the TFP gains from uniform partial reductions in τs and λs. This approximates the marginal returns to reductions in distortions. However, modeling the effects of a specific policy would require assumptions on the specific institutions underlying the distortions I measure, which also govern the second-order effects of how a change in τ affects Λ (and vice versa).

Figure 1.4 illustrates the complementarities between policies that reduce both sets of distortions. In particular, it shows that the marginal returns to reducing either distortion alone limited, moving along either horizontal axis. However, the marginal returns are much higher after both sets of distortions have been reduced substantially, suggesting that small reductions to one or both sets of distortions may have limited effects and that significant improvements to both sets of markets may be required to unlock large large gains. If one knew the relative costs of reducing each distortion, the gradient of Figure 1.4 would define an expansion path for the social planner in terms of which distortions to target as its budget shifts out. Additionally, Figure 1.4 shows that these marginal returns are not monotonic: at baseline levels of input (financial) distortions, going from 10% of observed financial (input) distortions to perfect financial (input) markets actually worsens efficiency.

1.5.1 Methodological Differences and Measurement Error

I now describe how estimating both Λ and τ helps alleviate concerns about measurement error. With both Λ and τ, counterfactual aggregate productivity can be computed in two ways: taking the observed allocation and then "removing" a distortion or taking the first-best allocation and "adding a distortion". To see this, note that the efficient allocation (1.18), which is just a function of A_{jt}, can also be written as a function of observed input demands and wedges by inverting (1.12) as a function of A and dividing out constants

$$\omega^*_{jt} = \frac{q_{jkt}\tau_{jkt}\left(\frac{\prod_l \tau_{jlt}^{\alpha_l}}{\Lambda_{jt}}\right)^{1-\gamma}}{\sum_{h=1}^{N_{vt}} q_{jkt}\tau_{jkt}\left(\frac{\prod_l \tau_{jlt}^{\alpha_l}}{\Lambda_{jt}}\right)^{1-\gamma}}. \tag{1.26}$$

Likewise, under the status quo, rewriting (1.17) should simply yield

$$\omega_{jkt} = \frac{q_{jkt}}{\sum_{h=1}^{N_{vt}} q_{jkt}} \tag{1.27}$$

This allows me to compute TFP using either (1.17) or (1.26) and then aggregating using 1.19) for any counterfactual values of Λ and τ. However, this requires estimates of both Λ and τ.

If inputs were measured perfectly and τ and Λ were estimated without error, then these two approaches should produce identical estimates. The difference is that the former approach (1.17 and 1.18) uses estimated TFP while the latter (1.26 and 1.27) uses raw input measurements. Which estimate is preferable depends on how severe measurement error in inputs is relative to the errors in estimated objects. Given that estimates of TFP are less noisy than the raw inputs used to estimate them, one would therefore expect estimates using the TFP-based measures in (1.17) and (1.18) to be more reliable than the input-based measures in (1.26) and (1.27). I confirm this using Monte Carlo simulations in Figure A.3.1, which shows that the TFP-based measure is approximately unbiased and less noisy than the input-based measure, which is biased upwards.

How different are the conclusions these measures produce in the data? To make this comparison, it will be useful to denominate misallocation by the *attainable* output (equivalent to TFP when aggregate input supply is fixed) forgone due to distortions in each scenario. Figure 1.5 compares results from the TFP-based results in the solid bars and the input-based results in the shaded bars. The solid bars simply recast the estimates from Figure 1.3. The blue bars show the percent of attainable output foregone in the observed allocation, while the orange (green) bars show allocations with only the observed input (financial) frictions. By definition, the optimum allocation achieves all the attainable output so there is no solid purple bar.

Now contrast these TFP-based results with the shaded bars, which are computed using the input-based measure. As discussed in Section 1.2, these two panels would yield identical results if there were no measurement error and the model was perfectly specified. However, the differences between the two panels are quite striking when comparing bars of the same color in Figure 1.5. First, measured misallocation in the status quo is 59% larger using the input-based rather than the TFP-based measure. Second, it appears that perfecting financial markets would *worsen* misallocation. Most strikingly though, the implied “optimum” allocation is not only suboptimal but actually performs worse than the observed allocation.

How is this possible? Recall that counterfactuals using the input-based measure are computed by adding distortions to the observed allocation, which includes mismeasured inputs. The shaded green bar is calculated by equalizing factor ratios in a way that preserves scale across farmers: this is the model of an exchange economy that serves as a lower bound on factor misallocation in Shenoy (2017). The purple bar is then calculated by reweighting those demands by $1/\Lambda$, removing estimated financial frictions.[58] The input-based estimates are higher across the board than those using only estimated quantities. The conflicting result that removing financial frictions would worsen misallocation can be explained by their negative correlation with input measurement error. In other words, measurement error looks like a distortion that is partially offset by financial frictions — removing financial wedges thus makes this spurious distortion appear worse.[59]

Second, if there were no measurement error, then estimates of misallocation should be similar at the plot and farm level. Aragón et al. (2022) argue that plot-level data amplifies the potential for measurement error. Meanwhile Gollin and Udry (2021) argue that since optimization implies that households should be indifferent between allocating marginal expenditures towards one plot or another, differences in input intensity across plots of the same crop grown by the same farmers are likely to be either measurement error or unobserved heterogeneity. This suggests, that if households, or at least individuals, are truly optimizing and measurement error is not a concern, then plot-level data should not increase estimates of misallocation.

Figure A.3.2 shows the main results using the plot rather than the household as the unit of analysis. This assumes that the same input and financial wedges apply equally to all plots a household cultivates simultaneously as in Gollin and Udry (2021). Table A.2.1 shows that this produces nearly identical estimates of the production function as the farm-level specifications. Naturally, the solid bars in Figure A.3.2 show slightly lower estimates of misallocation than the farm-level analysis in Figure 1.5. This is because the joint distribution of wedges and TFP is the same as in the farm-level analysis, except that the estimate of η is higher using plot-level data and that households with more plots (which tend to be less distorted) are oversampled. However, in the shaded bars, the estimates of misallocation using raw inputs nearly double. The reason for this is switching from farm-level aggregates to raw plot-level measurements introduces additional measurement error. Notably, there is no longer a significant difference between estimates from the observed allocation and when removing financial distortions.

These differences between the TFP and input-based measures are quite robust across

[58]Note that the same wedges are used in each set of results but for different specifications. Input wedges are used to compute the orange and blue solid bars and the green and purple bars in the right panel. Meanwhile, financial wedges are used to compute the blue and green solid bars and the purple and orange shaded bars.

[59]Arthi et al. (2018) find that labor inputs are more upwardly biased for smaller farms. Since Figure 1.2 shows that these households are more financially constrained, financial constraints would then be negatively correlated with the measurement bias. Counterfactually relaxing these constraints would therefore allocate more resources to farms that appear artificially large in the raw data.

specifications. Together, these results underscore the importance of separately identifying both input and financial distortions. Without a credible estimate of financial distortions, one would need to rely on noisily measured inputs and arrive at qualitatively different conclusions about the effects of counterfactual policies.

1.5.2 Alternative specifications and robustness checks

In Figures A.3.3-A.3.5, I show results under the alternative assumptions about the normalizing input for τ, the demand system used to estimate λ and sample restrictions. While the magnitudes of misallocation differ slightly across specifications, the qualitative results are broadly consistent.

Taxes vs. Rations

While the estimation procedure doesn't require taking a stand on whether input wedges operate as taxes or rations, this affects how households adjust different inputs under counterfactuals. In particular, a household facing a downward labor ration, as in Breza et al. (2021), would not use additional credit to hire more labor. The results in Figure 1.3 treat all inputs as flexible, as if input frictions functioned as taxes. Figure A.3.6 shows the counterfactual gains from reallocation if land were a fixed factor or labor were rationed from below, relative to the case where both factors are mobile yet subject to distortions. The blue (left) bars in each group reproduce the results from the baseline case of Figure 1.3. The green (middle) bars show the results assuming land is a fully fixed factor in all specifications. However, the differences relative to the case of a tax are fairly small and statistically insignificant, as can be seen from the left-most group of bars in the figure. Even though households facing a downward labor ration would use additional credit to acquire other inputs until the ration no longer binds, the price of these other inputs also increases in equilibrium.

Levels of aggregation

So far I have assumed that reallocation occurs within townships, in which stocks of land and labor are fixed. I argue that this is a realistic level of aggregation since village boundaries within townships are fairly arbitrary (Kaboski and Townsend, 2011). However, I now consider how these results would change if reallocation could only occur within villages, or if reallocation could also take place across regions of Thailand. The latter should be viewed as an upper bound on the gains from reallocation since fundamental trade and migration costs cannot be considered misallocation. However, if these gains are large, it suggests that investments in roads and other infrastructure that promotes market integration may be effective at reducing misallocation.

Figure A.3.7 shows the potential gains from full reallocation if allocation only occurs within villages or occurs at the national level.[60] The gains from reallocation across regions are more than three times as large as those from reallocation within townships. However, there appears to be very little misallocation across villages within townships, consistent with other evidence that villages in the same area are fairly integrated.

1.5.3 Distributional Effects

The above counterfactuals only consider efficiency gains. What are the distributional implications of reallocation? Although a full treatment of welfare impacts is beyond the scope of this paper, Figure A.3.8 and Figure 1.6 show how the distribution of land changes under the main counterfactuals. First, wealthier households tend to have much larger landholdings. While eliminating financial frictions makes the land distribution more equal across levels of baseline welfare, reducing frictions in land markets alone strengthens the correlation between welfare and farm size. This is because input frictions disproportionately affect wealthy households, who may wish to explain their landholdings but be unable to do so. However, many of these households are already inefficiently large ex-ante because of their position in financial markets. Second, the concentration of farmland increases in all scenarios, meaning that the average household contracts its landholdings. This causes many farms to become infinitesimal, effectively exiting agriculture.[61] About 33% of households produce less than 1 rai (.125 ha) under perfect input markets and about 16% do under perfect financial markets. This is only 8% of farmers under the efficient allocation, in which the land distribution is more equal relative to reducing input frictions alone. This suggests that a single-market intervention may also induce inefficient levels of exit from agriculture. Nevertheless, I note that a richer model is required to fully capture the welfare effects of these channels.

1.6 Conclusion

In this paper, I estimate distinct distortions affecting farm households in Thailand and quantify how they each contribute to misallocation. This is necessary for policymakers to consider, as the welfare effects of interventions in a single market are ex-ante ambiguous. First, the model yields a novel, theory-consistent production function estimation approach that holds when input choices are distorted. My approach flexibly allows for TFP shocks unobserved to the econometrician. Empirically, I find relatively low levels of misallocation in Thai agriculture: In my preferred specification, the gains from optimal reallocation are 31%. Perfecting financial markets while leaving input distortions unchanged would achieve 48% of these gains while perfecting financial markets holding input distortions fixed would

[60]Note that since only 16 villages from 4 tambons are included in the sample, this should not be considered representative of a national-level reallocation.

[61]In the model, these households would continue to earn their non-agricultural income. However, I do not capture the potential entry by previously constrained households.

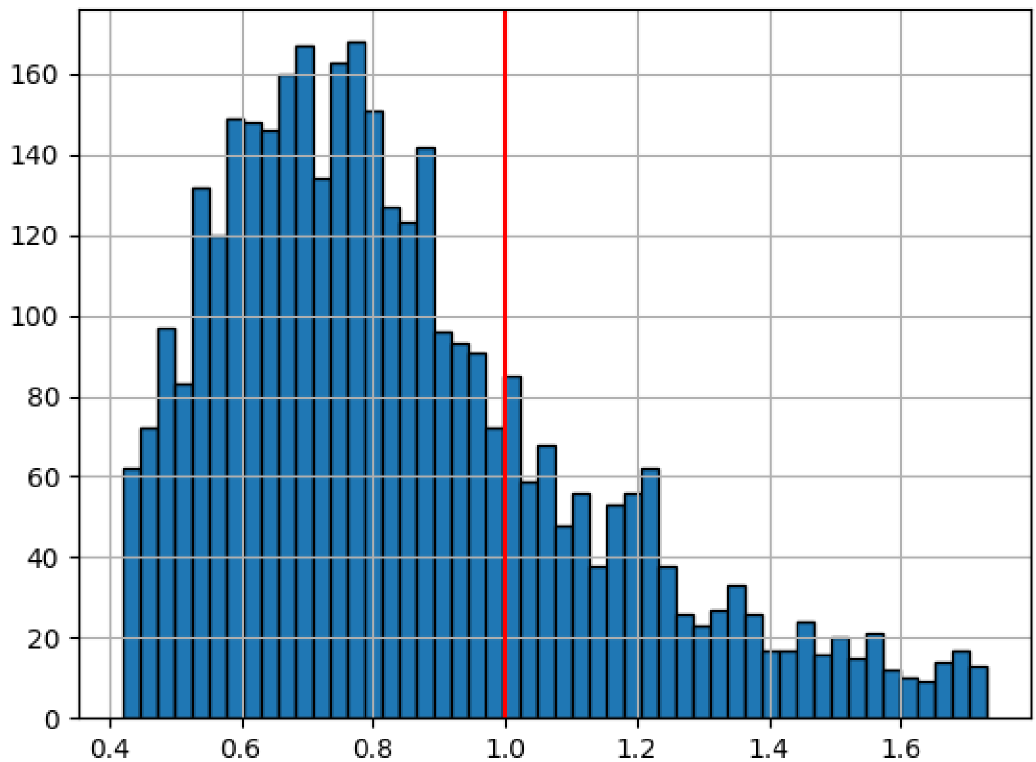
achieve 23% of them. These gains sum to less than one because more financially constrained farmers are relatively subsidized in input markets, particularly for labor. This suggests that policies that seek to alleviate both distortions may be more effective than those targeted towards a single one.

Directly estimating financial distortions rather than inferring them from a residual allows me to avoid attributing measurement error in inputs to misallocation. I find that not accounting for measurement error using the full model would lead to 59% larger estimates of misallocation and, in contrast to my preferred approach, suggest that removing financial frictions alone would worsen misallocation. While the model explicitly allows for such a possibility, my preferred results show that this is not the case.

This paper leaves many additional topics for future research. In particular, more work is required to understand the distributional implications of productivity-enhancing policies. Another open question is how misallocation in agriculture interacts with climate change, given that it increases production uncertainty but increasing agricultural production may create climate externalities. Finally, while the paper provides a broad framework for diagnosing the effects of a general set of distortions, more research is needed to understand specific policies to address the relevant institutions in different contexts.

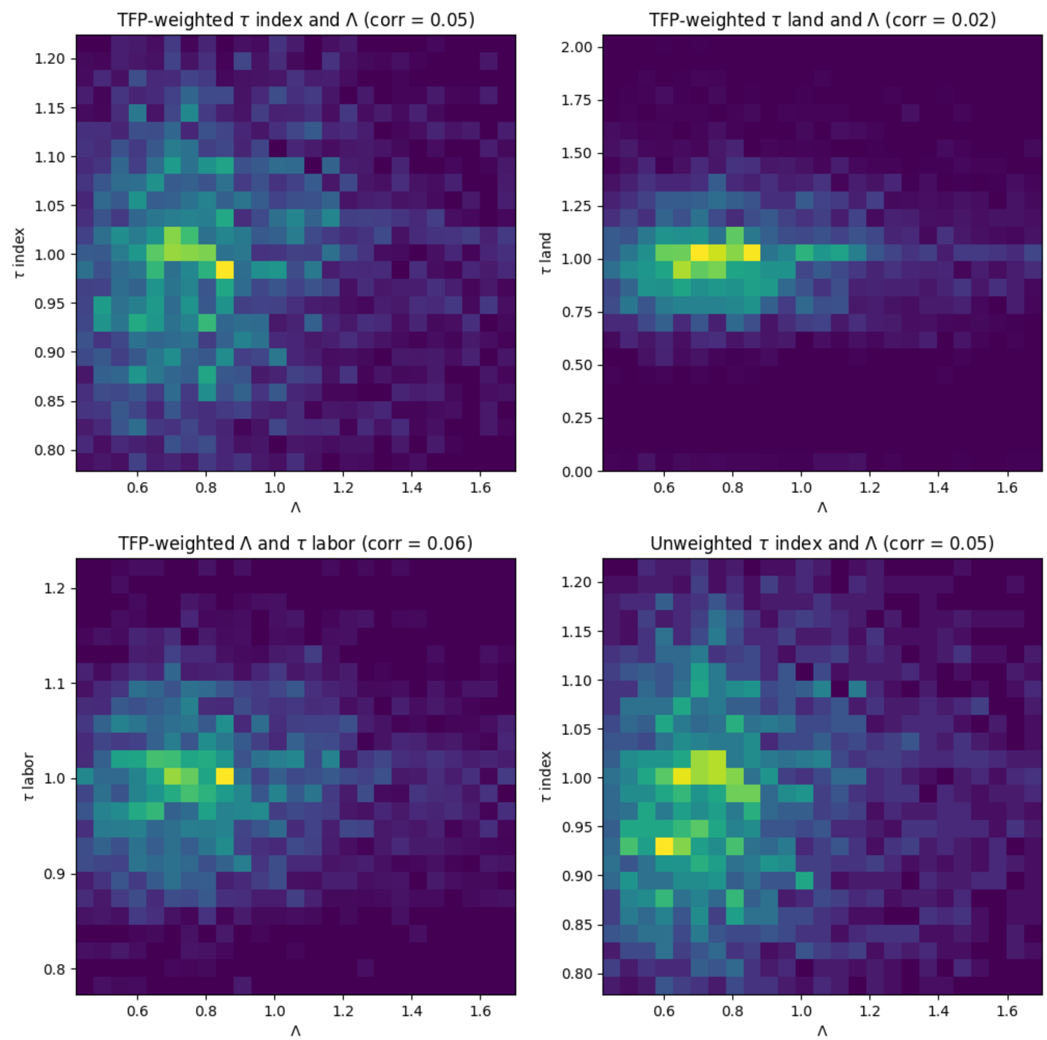
1.7 Figures and Tables

Figure 1.1: Histogram of Λ



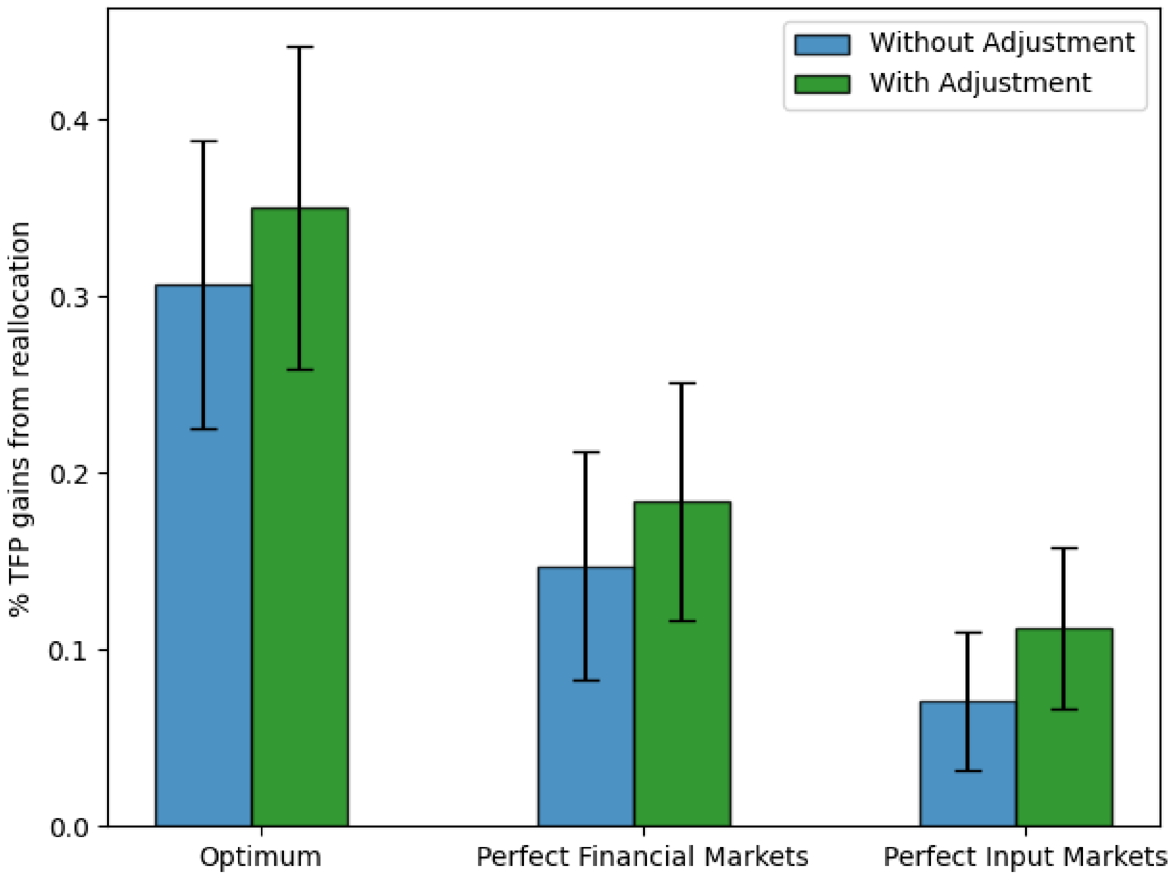
This figure plots the distribution of the estimated Λ_{jt} as described in Section 1.4.4. Perfect financial markets would imply a value of 1 for all households, while lower values reduce demand for risky inputs. Values above 1 suggest that agriculture is a hedge against some other income stream. Values are trimmed at the 5% upper and lower tails.

Figure 1.2: Joint distribution of TFP-weighted τ and Λ



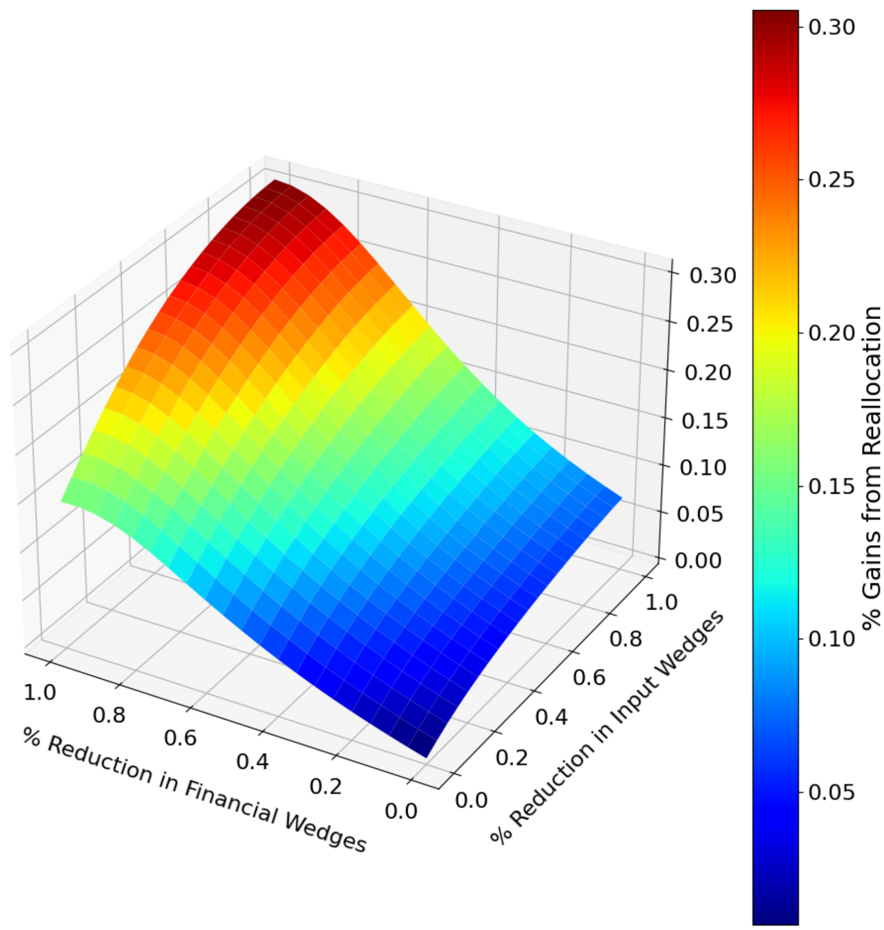
This figure plots TFP-weighted histograms of Λ and τ in 25×25 bins and reports their correlation coefficients. The top left panel plots the Cobb-Douglas price index of τs, $\prod_l \tau_{jlt}^{\alpha_l}$ against the estimates of financial distortions Λ, each weighted by TFP A. The top right panel plots the τ for land against Λ while the bottom left plots the index of τ for the three types of labor (planting, weeding, and harvesting) considered. The bottom right panel plots the price index of τs against Λ without weighting by TFP.

Figure 1.3: Counterfactual TFP gains from reallocation



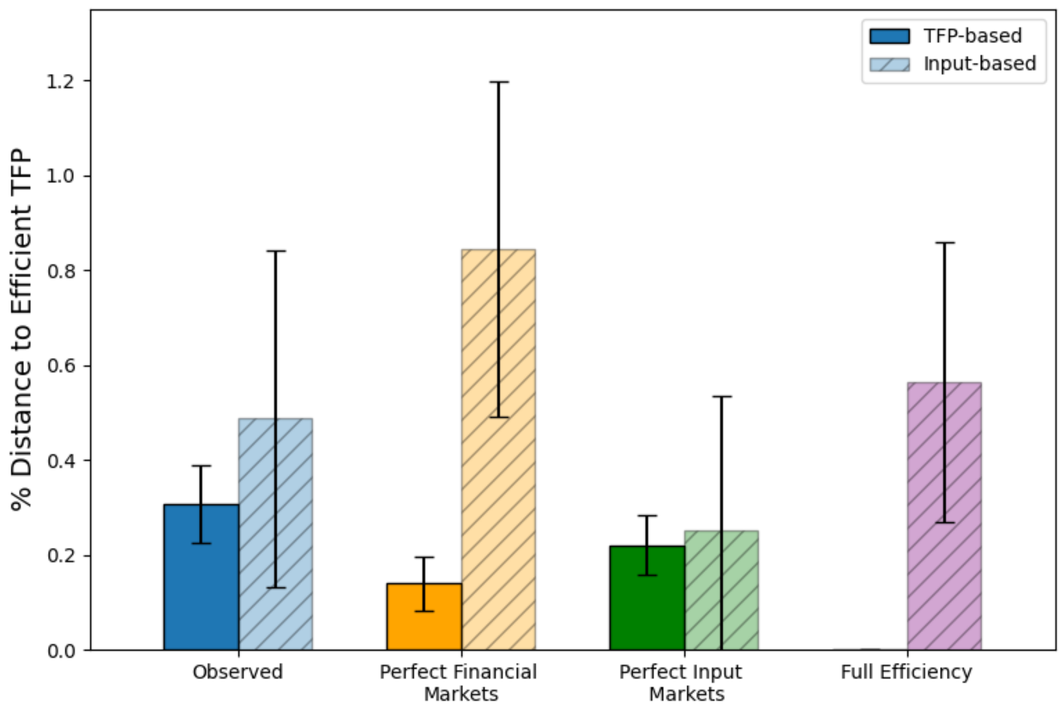
The figure shows the aggregate TFP gains from the main counterfactuals summed up across years, as a percentage of status quo aggregate TFP. The first group of columns shows results under perfect financial markets but with the observed input frictions. The second shows results under perfect input markets but with the observed financial distortions. The third shows the results under a full set of perfect markets. The blue (left) bars in each group show the gains holding aggregate supply fixed at the township level for all inputs while the green (right) bars show the gains allowing the aggregate supply of seed, fertilizer, and equipment to increase (holding their prices constant). The results are computed using fertilizer as the normalizing input for τ, CFE demands, and all crops, aggregated to the farm level.

Figure 1.4: Gains from partial reductions of τ and Λ



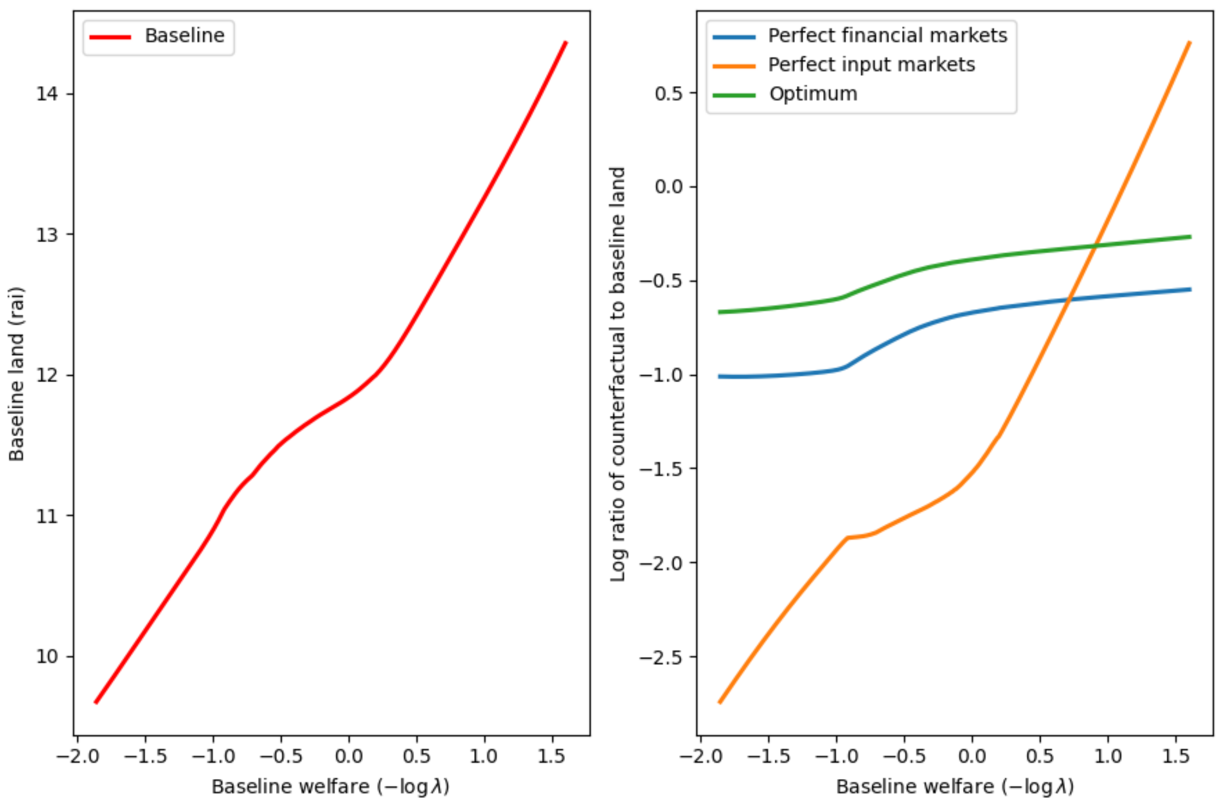
The figure shows counterfactual gains from reallocation using the TFP-based measure under different reductions of input and financial wedges. I compute aggregate TFP under each scenario shrinking Λ and τ towards unity by increments of .05. The origin corresponds to the status quo allocation and (1,1) corresponds to the efficient allocation. The vertical axis shows the percent increase in aggregate TFP relative to the status quo allocation. The figure uses fertilizer as the normalizing input for τs, CFE demands and includes all crops , aggregating to the farm level.

Figure 1.5: Aggregate TFP relative to optimum, with and without input mismeasurement



The figure shows the percentage of foregone attainable output from the four main counterfactuals (observed allocation, efficient allocation, perfect financial markets with input wedges intact, and perfect input markets with financial wedges intact). The solid bars compute these using the TFP-based measure of misallocation, using (1.17). The shaded bars are calculated by taking raw input observed in the data and augmenting them by the estimated τ and Λ, where relevant. 95% confidence intervals from 200 bootstrap replications are plotted. Results are computed using CFE demands, fertilizer as the normalizing input for τs, all crops, and aggregating to the farm level.

Figure 1.6: Changes in Land Distribution



The left panel shows the distribution of land under the baseline, denominated in rai (.125 ha), as a function of baseline welfare, which is the negative of the log MUE. The right panel shows the log ratio of land under the main counterfactuals to land at baseline. The plots show a lowess fit. This is shown for the closed economy case, using fertilizer as the normalizing input, CFE demands, andall crops at the farm level.

Table 1.1: GMM results

	α CD	α NH	β NH
Equip.	0.084	0.161	0.144
	(0.005)	(0.013)	(0.048)
Fert.	0.089	0.103	0.110
	(0.002)	(0.004)	(0.016)
Harv. Labor	0.225	0.175	0.181
	(0.006)	(0.028)	(0.077)
Land	0.208	0.219	0.362
	(0.004)	(0.069)	(0.208)
Plant. Labor	0.117	0.120	0.210
	(0.004)	(0.045)	(0.430)
Seed	0.092	0.087	0.130
	(0.002)	(0.005)	(0.028)
Weed. Labor	0.013	0.041	0.050
	(0.001)	(0.017)	(0.029)
J-stat	35.06	36.41	
p-val	0.465	0.132	
γ	0.828	0.906	
s.e.	(0.01)	(0.09)	

This table presents results from the main GMM specifications used to estimate the production function under both the Hicks-neutral Cobb-Douglas specification in the main text and the generalized Cobb-Douglas in Appendix A.2. Column 1 shows the estimates of the Cobb-Douglas coefficients α from (3.17) The second and third columns show estimates of α and β from (3.20), which are the elasticities of the mean and standard deviation of output with respect to each input. All specifications use tambon dummies and lags of λ_{jt} from the 5 months before input k is first applied as instruments. An annual discount factor of $\delta = .95$ is assumed. Results are computed using fertilizer and seed as the reference input for the estimation of τ from (1.25) (only relevant for Column 1), using rice plots only and CFE λs at the farm level. The J-statistic and p-values reported are from a test of the model with the full instrument set against one with only tambon dummies and a single lag of λ_{jt}. γ is the returns to scale parameter implied by the sum of the production coefficients. Standard errors are computed from 234 bootstraps of the full estimation procedure at the household level.

Table 1.2: Decomposition of Gains by Input Market

	Financial Constraints	Perfect Financial Markets
All	0.095	0.313
Land	0.047	0.234
Labor	0.068	0.273
Plant. Labor	0.020	0.191
Weed Labor	0.003	0.161
Harv. Labor	0.055	0.248
Equip	0.011	0.174
None	0.000	0.157

This table shows the gains from removing distortions τ_{jkt} in individual input markets, both with the observed financial constraints and under perfect financial markets. This is shown for the closed economy case, using fertilizer as the normalizing input, CFE demands, and all cropsat the farm level.

Chapter 2

The Welfare Effects of Postharvest Loans Under Price Risk

2.1 Introduction

Consumption is highly seasonal for poor farmers in developing countries, where hungry seasons are characterized by acute poverty and malnutrition. Seasonal hunger does not only create immediate deprivation, but its level and variability negatively affect long-run outcomes such as health and human capital formation (Christian and Dillon, 2018). While we know this is a severe problem, typical poverty measures often underestimate the seasonality of poverty and hunger (Merfeld and Morduch, 2023). Focusing on the seasonality of hunger is especially important in agricultural settings, where farmers typically receive lumpy income from harvests and have to smooth it over the lean season, choosing when to optimally consume or sell crops. This is especially challenging because both consumption needs and the value of stored crops can change unexpectedly over time. In particular, farmers frequently sell their harvests during peak seasons and buy back those crops at much higher prices during subsequent lean seasons after depleting their stocks – or in the words of Burke et al. (2019) "sell low and buy high."

A number of studies, most prominently Burke et al. (2019) but also Basu and Wong (2015); Aggarwal et al. (2018); Omotilewa et al. (2018); and Channa et al. (2022), have found generally positive effects of postharvest loan (PHL) and storage RCTs in sub-Saharan Africa. However, a simple model of intertemporal consumption smoothing—which is not spelled out in any of these studies—suggests that these results are far from obvious. This is because seasonal price increases vary heavily from year to year, making borrowing against them a risky proposition. Cardell and Michelson (2022) illustrate using price data from many sub-Saharan African grain markets and simulations to show that even a mild level of risk aversion would rationalize the observed lack of arbitrage. Empirically, there are also concerns that the treatment effects these studies find on profits do not neatly map to welfare improvements across the season. There are also supply-side concerns over the sustainability of such programs: if prices are spatially correlated, then lenders may be exposed to simultaneous default when prices fail to rise.

It is important for academics, policymakers and donors to view these programs through a broad and theory-consistent model of consumption smoothing and intertemporal arbitrage. We therefore partnered with the Taimaka Project, an NGO in Gombe State, Nigeria, that was implementing a postharvest loan program modeled after the One Acre Fund scheme studied in Burke et al. (2019). Taimaka randomly assigned offers of either a cash loan, a similarly-valued in-kind loan of maize, or no loan to groups of applicants. We collected 8 rounds of high-frequency consumption, investment, and storage data over a 12 month period (pre- and post-treatment) from a sample of 935 farmers in Gombe State. In addition, we experimentally elicited households' intertemporal marginal rates of substitution, by offering one-month "bonds" at different interest rates. Together with consumption data, this gives us two (independent) ways to measure households' welfare at each point throughout the season. We also have highly localized price data for the full range of crops grown and stored by sample households at the local market level within Gombe and the time series of prices over a longer period across several markets in Nigeria from USAID's FEWS project.

First, we estimate the treatment effects of Taimaka's loan program, which contained both a cash loan and a similarly-valued in-kind loan of maize grain. Unlike in other studies, we find large effects on grain storage, but insignificant effects on sales, consumption, and welfare. In particular, households that received the cash treatment appear to have increased their storage by up to 100,000 Naira, (twice the loan value) although the results are noisy. In contrast, we find a null effect of the in-kind loan on stocks, including for maize. However, we find no significant effects of either treatment on crop sales. Unlike in other contexts in which PHLs have been studied, prices stayed flat between when the loans were disbursed after the 2021 harvest and when repayment was due in mid-2022. It instead appears that these (cash) loans induced households to reschedule their consumption over the season as opposed to becoming arbitrageurs. As a result, many households defaulted on their loan and Taimaka decided not to continue the program the following year.

We also find minimal effects of the program on various measures of welfare. In particular, we detect no significant changes to average consumption, estimated welfare (marginal utility of expenditure or MUE), self-reported hunger, or experimentally elicited intertemporal marginal rates of substitution (IMRS). However, when breaking these results out by period, we find that estimated household MUEs are significantly higher for households in both treatments at endline, consistent with repaying loans or forfeiting assets after defaulting.

While the results from our RCT show that PHL programs do not always lead to major increases in profits and welfare, we would like to say something more general about their expected returns under price uncertainty. The data we've collected allow us to estimate a simple yet general model of seasonal consumption smoothing and arbitrage and test the simpler version of the model suggested by other papers in which price risk doesn't affect demand for postharvest loans.

First, we reject the null hypothesis that households could not increase *ex-ante* welfare by holding different portfolios of stocks. We find that the average household underinvests in arbitraging grain, relative to what a risk-neutral household would choose to invest. In particular, we find a positive (and statistically significant) correlation between crop price increases and the marginal utility of consumption, even for households holding positive stocks. In other words, these households are worse off when their assets appreciate, consistent with them expecting to deplete their stocks in future periods and become net consumers. This creates an important selection channel that limits the impacts of PHLs — exposure to price risk limits the demand for credit, possibly among the most vulnerable households that such programs seek to target.

However, we also do not find any significant effects of holding these additional stocks on households' (realized) intertemporal marginal rates of substitution (or *ex-post* welfare) due to the PHL treatment. In other words, even though households would have liked to increase their stocks *ex-ante*, this would not have increased their welfare given the observed (lack of) price increases.

These results primarily contribute to a large literature on the seasonality of income and consumption in agricultural settings, including several papers on PHL and storage interventions. Basu and Wong (2015) find that providing households with storage drums and in-kind staple food loans leads to increases in expenditure and income, but that only credit led to smoother lean season consumption. Aggarwal et al. (2018) and Omotilewa et al. (2018) find that pure storage interventions using hermetic PICS bags increase maize storage volumes, duration, and revenues in Kenya and Uganda, respectively. Channa et al. (2022) also include a treatment arm that only provided PICS bags and find slightly smaller and insignificant positive effects on maize storage than for households who received both credit and PICS bags (but cannot reject equality of the effects). Channa et al. (2022) and Burke et al. (2019) find that access to postharvest credit improves farmers' incomes, but do not detect any effects on consumption. We not only provide a cautionary tale about the effects of PHLs in years with minimal price increases, but use high-frequency consumption data to test for effects on welfare throughout the season.

We also contribute to a growing literature on the external validity of experimental estimates in a stochastic world (Rosenzweig and Udry, 2020). In particular, we verify that a PHL program similar to others studied in the literature is not effective in a season when prices do not rise. We further use the structure of the model to conclude that despite the risk of prices not increasing, even treated households would have liked to hold more stocks *ex ante*. This suggests that there is still a role for policies that relax seasonal credit constraints, even if price risk reduces demand for them somewhat.

These results have important implications for our understanding of seasonal poverty and policies to reduce them. First, they provide a cautionary tale of PHLs as a specific policy. More broadly, the result that even treated households nevertheless could have improved *ex ante welfare* by storing more, despite low *ex post* returns, suggests that risk may not have been the binding constraint against arbitrage. Instead, credit constraints appear more relevant, despite the treatment effects of the loans. Future analysis could consider alternative policies that help address both risk and credit constraints, In particular, a loan combined with a forward contract could perhaps provide liquidity while indemnifying households against the risks of prices not rising.

2.2 Theoretical Framework

We consider the context of a single agricultural season with T periods. Assume a household has per-period utility function $u(c_t, x_t)$ over an agricultural good c_t that can be bought and sold at price p_t (which is a random variable as of $t-1$) or stored and a non-agricultural good x_t that is purchased on spot markets with price 1. The household harvests an amount $H > 0$ of grain in period 0, of which it can choose to store amount $s \in [0, H]$ or sell for price p_0 to finance consumption or purchase a safe asset a with return R the next period. Assume the household cannot borrow so $a_t \geq 0\ \forall t$.

The household's budget constraint in each period (with H replacing s_{t-1} in period 0) is

$$p_t(s_{t-1} - s_t - c_t) + Ra_{t-1} = x_t + a_t \tag{2.1}$$

which simply states that (net) sales of stock plus the returns to the last period's safe asset equal current consumption and investment expenditure. Note that this budget constraint holds for any realization of p_t, with an associated Lagrange multiplier λ_t that is also a random variable.

The household solves

$$\max_{a,s,c,x} V(t, s, a) = u(c_t, x_t) + \beta E_t[V(t+1, s',' a')] \tag{2.2}$$

subject to the budget constraint, the credit constraint $a_t \geq 0$, and a non-negativity constraint on stocks $s_t \geq 0$.

This yields the first-order conditions

$$u_c(c_t, x_t)/p_t = u_x(c_t, x_t)\lambda_t\ \forall\ t, \tag{2.3}$$

where λ_t is the Lagrange multiplier on the household's budget constraint in each period, plus

$$\lambda_t = \beta\frac{R}{p_t}E_t[V_s(t+1, s', a')] + \mu_t^a \tag{2.4}$$

where μ_t^a is the Lagrange multiplier on the borrowing constraint, and

$$p_t\lambda_t = \beta E_t[p_{t+1}V_a(t+1, s', a')] + \mu_t^s \tag{2.5}$$

where $\mu_t{}^{\mathrm{s}}$ is the Lagrange multiplier on the non-negativity constraint for stocks.

Applying the envelope theorem implies that when the household is at an interior solution

$$RE_t[\frac{\lambda_{t+1}}{p_{t+1}}] = \frac{1}{p_t}E_t[p_{t+1}\lambda_{t+1}] \tag{2.6}$$

Simple comparative statics illustrate that the effects of p_{t+1} on welfare at harvest depends on s_t. In particular, if households enter period 1 with large stocks of s, then a positive shock to p_{t+1} simply earns them higher returns on their assets. In contrast, if a household has small stocks of s, this makes supplementing their food consumption with purchases more costly, which can swamp the benefits of higher returns to their small investments.

Now assume a lender is willing to offer the household a loan h at period 0 to be repaid with interest $\tau \in (R, E[p_1/p_0])$ at period 0. As such, the loan is unattractive when the household can borrow at R but is profitable in expectation if it cannot. We are interested in how uncertainty over p_1 affects demand for this loan. The simple model yields the following three facts.

Proposition 1.

1. *A risk-neutral credit-constrained household would always take the loan*
2. *Risk over p makes the loan less attractive to a (prudent) household when it expects to be a net seller*
3. *Risk over p makes the loan more attractive to a (prudent) household when it expects to be a net buyer*

(a) follows direction from the assumptions. To verify (b) and (c) note that the sign of the covariance between λ_{t+1} and p_{t+1} depends on s_t. To see this, first note that $\frac{\partial s_0}{\partial h} > 0$, otherwise the household would never be able to repay. Also note that staple consumption in period T $c_T = \frac{Ra_{T-1}}{p_T} + p_T s_{T-1} - x_T$. As $\frac{s_{T-1}}{s_{T-1}+a_{T-1}} \to 0$, $\frac{\partial c_T}{\partial p_T} < 0$ and as $\frac{s_{T-1}}{s_{T-1}+a_{T-1}} \to 1$, $\frac{\partial c_T}{\partial p_T} > 0$. Similar logic holds iterating forward to earlier periods. This means that if households are prudent, $cov(\lambda_{t+1}, p_{t+1})$ is increasing in s_t and positive (negative) for households with low (high) s_t relative to a_t. Since the loan increases s, $\frac{\partial cov(\lambda_{t+1}, p_{t+1})}{\partial h} < 0$. Meanwhile, a mean-preserving increase in the variance of p_{t+1} increases the *absolute value* of $cov(\lambda_{t+1}, p_{t+1})$. This therefore reduces risk for households with low s and increases risk for households with high s

1 states that in addition to being a profitable investment in expectation, the loan can be either risk-reducing or risk-augmenting for households depending on how much they expect to store. In other words, when the loan moves households into the net-seller regime (as in Burke et al., 2019), price risk reduces the welfare effects of the loan by causing the most vulnerable to select out. On the other hand, when households expect to remain in the net-buyer regime, poor households increasingly select in because of the insurance motive it provides. Furthermore, the importance of these channels also depends on the skewness of price shocks. If, as in the data we observe for Gombe, positive price shocks are extreme but negative price shocks are relatively mild (and the cost of default is low), this strengthens the insurance motive of the loan for prudent households. This is because prudent households place more weight on states with with extremely high marginal utilities of expenditure. This makes them more averse to having to purchase from the market when prices skyrocket relative to the potential windfall from being able to sell stocks in these cases (the opposite would hold for imprudent households, i.e. with $u''' < 0$). Therefore, while PHLs have been marketed as a way to increase commercial activity for the moderately poor, price risk reduces

the potential for this channel but creates an additional insurance motive that may benefit even poorer households.

We have direct data on demand for credit from two distinct sources. First, our experimental sample is comprised of households who applied for the Taimaka loan, so comparing characteristics of these households to the broader population (e.g., households in the Nigerian Household Survey) tells us something about loan demand. However, we can do better at characterizing demand for credit within our experimental sample, as we elicit households' intertemporal marginal rates of substitution in each survey wave, which captures their demand for smaller amounts of credit at a more frequent scale. We also have the consumption and investment data to test the null that households are behaving as profit-maximizing arbitrageurs, and whether their behavior is consistent with the net-buyer or the net-seller regime. With further structure, we can estimate the marginal propensity to invest (vs. consume), which we can then use to estimate the *ex ante* welfare effects of post-harvest loans and other policies to combat seasonal hunger. In future work, we also aim to evaluate counterfactuals such as forward contracts, which essentially indemnify borrowers against states of the world in which prices fail to rise.[1]

2.3 Experimental Design and Data

2.3.1 Sample Frame and Household Selection

Communities were selected from the 10 percent poorest locations in Gombe state as predicted by satellite data following Aiken et al. (2020) (about 50 sites). Of these, 20 were randomly selected for a rapid rural appraisal by Taimaka. Half (10) of these sites were chosen for household listing based on perceived need and accessibility. Then six of these sites were selected for program implementation based on the household listing.

In each of the six sample sites, households were selected as follows. Taimaka visited the sites to advertise the loan program and met with traditional leaders after obtaining their assent to move forward to the program. After two days, they returned to hold an informational session designed to emphasize the terms of the loan, the loan's theory of change, and group indemnity. They then advised interested farmers to start forming groups of 5 and that they would return in a few days to begin taking applications.

Each group completed an enrollment form, which included the 10 questions from the Poverty Probability Index (PPI) for Nigeria, some brief questions about farming, and demographic information and included pictures of farmers. After receiving applications, Taimaka developed a ranked list of groups in order of desirability, based on ability to repay and need for the loan.

[1] Karlan et al. (2011) piloted an experiment offering price-indemnified production loans in Ghana but found no differential takeup relative to ordinary loans. However, price risk may be a much more important concern for PHLs than production loans.

Taimaka then met with traditional leaders to verify that the members of groups were indeed residents of that location, were indeed farmers, and were known to be creditworthy. If a single member was deemed unqualified, groups were given an opportunity to choose a replacement individual. If more than two candidates were deemed unqualified, then the group was struck. The ranked list was then updated accordingly.

Loan officers were then given a target number of clients to enroll in the program in each site. They went down the ranked list, visiting three farmers' households in each group to verify the information given on their applications. If any farmer was found to have made material misrepresentations in their application, the group was dropped from the list. Otherwise, the group was enrolled, which made them eligible to be selected for the sample.

The ordered list was then partitioned into strata of 6 adjacently ranked groups. A randomly selected pair of groups in each stratum was assigned to receive the cash loan, another was assigned to receive the in-kind maize loan, while the remaining pair was assigned to a control group. This draws on the idea of a finely stratified assignment mechanism advocated by Athey and Imbens (2017). This led to a sample of 935 individuals from 187 groups.

2.3.2 Treatment

The treatments were an offer of a joint-liability loan of up to 50,000 Naira (approximately USD 100) in value to each of the 5 group members. The terms for the maize and cash loan were slightly different.

For the cash loan, each farmer was asked to commit up to 4 bags to store until July 15th, 2022, the due date of the loan. Farmers received 11,900 Naira in cash plus a hermetic PICS bag priced at 600 Naira for each bag they committed. The vast majority of farmers in this arm chose to commit 4 bags and received the maximum loan value of 50,000 Naira (approximately \$100). The loans were to be repaid with a 15% user fee in July 2022. Delivery of the loans took place between September 16th and October 10th, 2021. Farmers were also required to make monthly repayments of at least 3,000 Naira starting in December 2021.

For the in-kind maize loan, all farmers were offered a loan of 3 100kg bags of maize. Farmers in this arm were also required to repay the loan in July 2022 with a 15% user fee. The amount required to pay was equal to the value of the price at which Taimaka purchased the maize, plus the cost of transporting to the household plus the cost of 3 PICS bags. In total, this ranged from 62,000 to 63,000 Naira, depending on the group's distance from the market. Due to logistical issues on Taimaka's end, these loans were disbursed in late November 2021 and repayment was required to begin in January 2022. Otherwise, the conditions of the loan were identical.

Members of the control group were not offered any loan by Taimaka.

2.3.3 Survey Data

We collected 8 rounds of household surveys at approximately two-month intervals between August 2021 and November 2022. Each survey included modules on household composition, grain stocks, food acquisitions, other expenditures, and measures of seasonal hunger. During baseline and endline surveys, we also asked questions about agricultural inputs and output from the previous season, and assets, including livestock.

2.3.3.1 Food Acquisition and Stocks

In each wave, we elicited information on recent acquisitions of a list of 22 different goods that households in this setting consume. However, while many of these acquisitions are fairly clearly for consumption (e.g., milk, sugar) in the fairly near future, others may be held for some combination of present consumption and investment. In particular, households often hold positive stocks of maize, millet, beans, guinea corn (sorghum), and less frequently hold stocks of rice, Bambara nut, and groundnut. A few households also reported holding stocks of cassava. During each wave, we asked households for information on these stocks. At baseline, we asked for the amount of each crop that the household had stored. In waves 1-3, we also asked how much of each crop they had harvested, purchased in bulk, and sold (or given away) since the previous visit, in addition to asking them about their stocks. After wave 3, enumerators informed us that households appeared to consider questions about their stocks sensitive. Therefore, for waves 4-6 we asked about how much of households' stocks they had consumed since the previous visit instead of asking about their current position. This gives us an account of households' grain flows at each period, allowing us to impute stocks. At endine, we added back in the question about stocks.

Unsurprisingly, households' reported stocks and flows over time do not always balance over time in an accounting sense. As a result, we had to make substantial imputations, which we describe in Appendix B.1.

2.3.3.2 IMRS Elicitation

We also attempted to measure individuals' intertemporal marginal rates of substitution by measuring willingness to pay for bonds at different interest rates. After each survey wave, enumerators asked individuals whether they would be willing to invest 500 Naira during the following survey to receive $500(1 + x)$ Naira for each x between -0.1 and 1 in increments of 0.1. The enumerator then used an app to randomly select an interest rate. If the individual had agreed to the selected rate, then they were required to bring at least 500 but up to 2,500 Naira during the following survey wave. The money was then given to the enumerator and the money was repaid with interest the following survey wave. Otherwise, no deal was implemented. Households who had agreed to the interest rate that was drawn were told that they would be barred from future payouts if they failed to honor the deal. These real-stakes choices identify the range of households' intertemporal marginal rates of substitution during each survey round.

2.3.3.3 Other variables

We also asked whether households had made any large non-food purchases of at least 10,000 Naira ($20), whether they engaged in another business, and if so, what their expenses and revenues were, and whether and how much money they borrowed and repaid. We also pre-specified the Reduced Coping Strategies Index (RCSI) Maxwell et al. (2014), which asks households questions about the number of days in a month they restricted food consumption because of lack of resources, as a measure of seasonal hunger.

2.3.4 Other Data sources

We use two sources of price data for our analysis. Taimaka collected prices of staple commodities from local markets in Gombe State on a (roughly) monthly basis. We also use weekly price data from the Famine Early Warning System Network (FEWS-Net) for Nigeria, which covers major regional markets, including Gombe Town. We also use Taimaka's administrative records on loan applications (including the scoring of groups), disbursement, and repayment.

2.4 Descriptive Statistics

2.4.1 Prices

Using over 10 years of monthly price data from FEWS Network, as well as data collected from markets in Gombe State during 2021-22, Figure 2.1 shows that the magnitude of seasonal price increases is highly variable. While on average they increase by 61% from floor to peak, in the median year, the increase is 38%. This difference is driven by extreme price increases of 239% and 187% in 2016 and 2020, respectively. On the other hand, price increases were lower than Taimaka's interest rate of 15% in three of 11 years. Thus, intertemporal arbitrage typically yields moderate positive returns but is highly influenced by tail events at both extremes. As Cardell and Michelson (2022) argue, it appears that moderate degrees of risk aversion could rationalize the lack of intertemporal arbitrage.

Notably, maize prices in the 2021-2022 season increased by about 20% (from 16,500 to 22,000 Naira per 100kg bag) between November and January, then stayed between 18,000 and 20,000 Naira, where they stayed throughout the loan period. At least 2/3 of the maize in the sample was sold after the January peak, meaning that households would have obtained a return on harvested maize stored between 11 and 23%, depending on the week they sold, not factoring in depreciation. In practice, 33% of maize sold in the sample was sold below 18,000 Naira per bag and 83% was sold for below 20,000 Naira per bag. Overall, few people made significant profits from arbitrage and many made negative profits net of Taimaka's interest.

The price data also exhibit strong spatial correlation, both within Gombe and across markets in Nigeria, suggesting there is minimal scope for spatial arbitrage. Using FEWS data on prices at weekly frequencies we can reject the null hypothesis of no spatial correlation in maize prices (using a Skillings-Mack test; p=0.035); when prices go up in one location they are very likely to go up in another. This can be seen from Section 2.8. This can also be seen for and for smaller markets within Gombe throughout the study period. The wedges between prices in rural markets and Gombe town also stay roughly constant throughout the season, likely reflecting transport costs (Figure 2.3). This implies that despite high potential returns to intertemporal arbitrage, there is little scope for spatial arbitrage. These patterns are also not unique to maize as can be seen from Figure 2.3 and section 2.8.

One implication is that a lender operating within a single state, such as Taimaka, is exposed to correlated default risk across lending sites. Indeed, nearly 40% of Taimaka's borrowers did not complete their repayments by the deadline of August 1st, with 18% of total balances outstanding. As of May 2023, 23% of borrowers had not paid the full amount, accounting for 6.4% of the total balance. Avoiding the risk for such default would have required spreading their portfolio across a much wider geographic range.

2.4.2 Prices and Intertemporal Marginal Rates of Substitution

The conventional account is that in rural sub-Saharan Africa seasonality in crop production is reflected not just in prices for those seasonal crops, but also in consumption— with increasing hunger and scarcity until the next harvest. Such a story is consistent with the variation in maize prices described in Figure 2.1, but not implied by it. Variation in local grain prices will depend on a mix of both supply and demand factors, and even if prices are high because supply is low and demand is high, this need not imply great need, as people faced with a shortage of grain may simply substitute toward other foods.

Figure 2.4 provides some evidence on this point. Here we see (log) returns (i.e., $\log p_{t+1}/p_t$) to holding maize over roughly six-week intervals. The planting season in Northern Nigeria roughly coincides with the period of highest prices, in April through June, while the harvest period is in late summer, and is slightly anticipated by the period of sharpest decline in maize prices. At the same time, the average intertemporal marginal rate of substitution, $\bar{m}$, signals relative need in periods that largely coincide with higher maize prices. In this economy in which the staple crop of maize plays a central role, we do see evidence of economic need in periods when maize prices are signaling a shortage.

2.4.3 Sample

Table B.2.1 shows some summary statistics for the sample and tests for balance between each treatment arm and the control group. Overall, the results are largely balanced, and where there are baseline differences, they are largely washed out by strata fixed effects in the main analysis in Section 2.5 and by the controls selected by Double Post LASSO (Belloni et al., 2012) in robustness checks (Section 2.5.5).

Households have an average size of 6 members and heads are on average slightly below 40 years old, 93% male, and 60% have no education. Households report owning an average of

5.3 hectares of land and an average of 2 cows and 5 goats. Around 18% operated a business and 19% had outstanding loans at baseline. Nevertheless, households were quite poor at baseline, which was conducted right before the 2021 harvest, at least relative to subsequent periods. The estimated marginal utility of expenditure was much higher than other periods and households also reported high levels of food insecurity, although this may include some strategic (mis)reporting.

2.5 RCT results

Below, we estimate average treatment effects of Taimaka's cash and maize loans. Overall, the takeaway is that the loans enabled households to reschedule consumption to later periods but did not have large effects on overall income and welfare.

We define the treatment variables as follows:

1. Cash_i is a dummy variable equal to 1 if household i was offered the cash loan.
2. Maize_i is a dummy variable equal to 1 if household i was offered the maize loan.

The main analysis is based on the following intent to treat specification

$$Y_{ist} = \beta_1^t \text{Cash}_i + \beta_2 \text{Maize}_i + X_i \gamma + \delta_t + \delta_s + \varepsilon_{ist}$$

where Y_{ist} is outcome Y for household i in stratum s at time t, δ_t and δ_s are survey wave and stratum fixed effects, respectively, and $T = 7$.

To examine outcomes at higher frequency, we also run the following dynamic specification to estimate separate treatment effects for each period

$$Y_{ist} = \sum_{\tau=0}^{7} \beta_1^\tau \text{Cash}_i + \sum_{\tau=0}^{7} \beta_2^\tau \text{Maize}_i + X_i \gamma + \delta_t + \delta_s + \varepsilon_{ist}$$

2.5.1 Effects on Stocks

2.5.1.1 Cumulative Results

First, we find that both the cash and maize loan treatments had negligible effects on the total amount of grain sold, consumed, and harvested. In our main specification, the total increase in the value of grain sold was about 18,000 Naira (approximately the value of one bag of maize) for the cash loan group and less than 2,000 Naira ($4) for the maize loan group (Table 2.1). However, both figures are imprecisely estimated. From Table 2.2 there appears to be a large, increase in the consumption value of *own* grain for the cash arm of about 92,000 Naira ($180). While this estimate is noisy, the increases are mostly driven by statistically significant increases in millet and bean consumption.[2] However, these estimates are much smaller and insignificant for the maize arm. Finally, we also see large increases in the value of grain households in the cash arm reported bringing back home from the field of about 112,000 Naira and 47,000 Naira in the control group. However, these results are not statistically significant for any crop. Nevertheless, the large point estimates are perplexing, since is not likely that the treatment caused households to harvest more grain (although anecdotally the loans were often used to pay for harvest labor) given that they were not aware of the treatment assignment when they planted.[3] Rather, we believe that this reflects that households were less likely to sell off grain immediately after harvesting (e.g. to repay debts) or from maize cribs in the field.[4]

2.5.1.2 Within-season results

We also use the high-frequency data to shed light on households' asset positions throughout the season. First, we find that total stocks increase by large amounts for the cash loan group relative to the control group throughout the season. The treatment effect on the value of total stocks peaks at about 100,000 Naira (twice the value of the loan) in March 2022 but remains large throughout the season. Interestingly, this is not driven by maize, but by beans and to some extent millet. This is also not driven by the crowding in of additional purchases for the purposes of arbitrage (Figure 2.10), but rather by reducing the outflows of grain. For the maize loan group, the point estimates on maize stocks in later waves are close to 0, albeit with wide confidence intervals.[5] Together, these results suggest that households preferred investing their loans in a fairly diversified portfolio of crops and that offering in-kind loans of a single crop is less effective for encouraging arbitrage.

Further results shed light on what treated households did with their stocks. From Figure **??**, we see small insignificant increases in sales of maize and other crops in the in-kind loan group at the start of the planting season in March (although this does not translate into significantly higher input expenditures). For the cash loan group, we instead see increases in the consumption of own grain at this point (Figure 2.9), which dwarf the marginally significant increases in sales towards the end of the lean season. This is consistent with the idea that the loan helped farmers maintain positive stocks further into the lean season but did not lead them to become net arbitrageurs in a season in which prices did not appreciate.

[2]Consumption of own stocks was only directly elicited following wave 3, and was imputed as a residual between previous stocks and other flows prior to that.

[3]Indeed, the sample is balanced on area planted and production expenditures from the prior season.

[4]The survey question specifically asked about grain brought home from one's field, which may not have captured any transactions that did not first involve bringing grain home, which are common in Gombe. In this case, increases in reported harvests may also capture postponed sales.

[5]It is surprising that we do not observe significant increases in the quantity of maize stored for the in-kind loan group, which suggests that these households either disposed of it immediately after receiving it or that cash and control households contemporaneously acquired similar amounts of maize. However, we do see an increase in overall sales right after treatment (Figure B.3.2.

While the stock data relies heavily on imputations, these effects can also be seen from the consumption data. Households receiving the maize loan increase their consumption of own-produced crops shortly after treatment, but this effect fades by the lean season. Meanwhile, we see households in the cash arm consuming fewer of their own crops until the start of the lean season, and then consuming more during the lean season (Figure 2.11). Together, these results suggest that the loan helped farmers maintain positive stocks further into the lean season but did not lead them to become net arbitrageurs in a season in which prices did not appreciate.

2.5.2 Effects on Consumption

We now evaluate the program's effect on consumption and welfare. We use four main measures: the log of monthly expenditure, the IMUE as estimated from Ligon (2020), and our experimentally elicited IMRS.

We find no significant effects of either treatment on the average values of any of these measures for households across the sample period.[6] None of the point estimates suggest a change in welfare of greater than 5% and we can rule out large effects on the IMUE, seasonal hunger index, and elicited IMRS.

We also see few significant effects when breaking coefficients out by treatment wave, except for an increase in the IMUE at endline for both treatment arms. While the results on the IMUE are inconclusive for other periods, this is consistent with treated households reducing their consumption when repaying loans or forfeiting assets after defaulting.

Our data also allow us to look at whether households are consuming from their own stocks or the market. Consistent with this, households in the cash treatment arm consumed significantly more crops from their own stock compared to both the maize and control arms throughout the lean season, also mostly driven by beans, guinea corn, and millet. In contrast, households receiving the in-kind loan start consuming additional maize immediately, but the effects disappear by the start of the lean season. By the end of the lean season, households in both treatment arms are more likely to consume purchased beans, but households in the in-kind treatment arm are less likely to consume purchased millet or guinea corn. Given that consumption of own-produced millet and guinea corn are not higher for these households, this indicates that they were less likely to consume these crops. This further supports the conclusion that the cash treatment helps households reschedule consumption but not become arbitrageuers while the in-kind treatment had null or negative impacts on late-season welfare.

2.5.3 Effects on Other Outcomes

Other pre-specified outcomes of interest include farm investment and profits from the 2022 planting season, non-agricultural business expenditures, borrowing and lending, and semi-

[6]The p value on the coefficient for the effect of the cash treatment on the elicited IMRS is 0.108, but this is only the case when imputing the maximum interest rate in the choice set (100%) as the IMRS for households who didn't choose any rate.

durable/durable purchases. Treatment effects are presented

In Table 2.6, we see some significant increases in business revenues for both treatment arms earlier in the season, but these effects fade towards the end of the season. We also see that households in both arms spend more on durable goods towards the end of the season. We do not see any significant effects on whether or how much households borrowed (apart from Taimaka), but the point estimates are positive, especially towards the end of the season, consistent with some households borrowing to repay their Taimaka loans. In Table 2.5, we also do not see any significant effects on agricultural investment for the 2022 planting season but the point estimates are positive for the cash loan arm and negative for in-kind loan arm. Harvest values are also higher for the both arms but not significantly so.

2.5.4 Heterogeneity

We pre-specified two dimensions of heterogeneity: gender and baseline wealth. In Appendix B.3 we report results interacting each treatment with a dummy for female household head and a standardized index of baseline assets, respectively. Since only 7% of household heads in the sample are female, the results are quite noisy. Nevertheless, we find that the maize loan significantly increases female-headed households' expenditure and reduces their IMRS (Table B.3.15, even though the point estimates on stocks grain flows are negative (Table B.3.14). We also see negative interaction effects between the cash treatment and baseline assets on grain sales and significantly higher log λs (seeTables B.3.20 and B.3.21).

2.5.5 Robustness

We also pre-specified robustness checks using the Double Post LASSO method of Belloni et al. (2012) to select controls, as well as controlling directly for the score Taimaka assigned to groups rather than using the strata fixed effects. We also show robustness to controlling for baseline outcomes in an ANCOVA specification following McKenzie (2012). The results, none of which alter the main conclusions, are presented in Appendix B.3.

2.6 Structural Estimation

In the previous section, we established that while the cash treatment may have helped households smooth consumption, neither treatment led to an actual significant increase in average arbitrage profits. But *could* households have made themselves better off *ex ante* by holding more maize? Though price changes in maize during 2021–22 were not large by historical standards, a purchase of maize in April 2022 followed by a sale in August would have given a return of about 10% over four months, or an annualized yield of 33%. It would have been profitable to finance this using the loans offered by Taimaka (15% interest) or commercial banks. But of course, these are *ex post* returns. We would like to say something about the *expected* welfare effects of investing in stocks of grain given the uncertainty over seasonal price increases. We now turn to testing the predictions of a simple model of intertemporal arbitrage with our high-frequency consumption and storage data.

Let p_j^t denote the price of a storable or financial asset $j \in \{1, \dots, J\}$ at time t. For any household holding a positive stock of asset j at time t (i.e., not at a corner), changes in the price of the asset will affect the household's budget and so its marginal utility of expenditures λ_t. The optimal portfolio for household i at time t will satisfy the Euler asset pricing equation

$$\beta \mathbb{E}_t \left[\frac{\lambda_{it+1}}{\lambda_{it}} \frac{p_{t+1}^j}{p_t^j} \right] = 1 \qquad \text{for all } j. \tag{2.7}$$

We have observations of prices p_t^j, assumed to be common, and have measures of $\log \lambda_{it}$ estimated from each household's consumption portfolio.

Then a direct idea is to work directly with the orthogonality conditions implied by the Euler equation. Let $\delta_{it}^j \in \{0, 1\}$ indicate whether household i holds positive stocks of asset j at time t. Let m_{it} denote i's intertemporal marginal rate of substitution at time t, and let $\bar{m}_t^j$ denote the average IMRS of those hold positive stocks of the asset. In the aggregate we have $\mathbb{E}_{t-1}[\bar{m}_t^j R_t^j] = 1$. In this form, this is the standard asset pricing condition. And then in the cross-section we have $\mathbb{E}_[t-1; (m_{it} R_t^j - 1)\delta_{it-1}^j] = 0$, and (via substitution)

$$\mathbb{E}_{t-1}[(m_{it} - \bar{m}_t^j) R_t^j \delta_{it-1}^j] = 0.$$

Interpreting this, focus first on the factor $(m_{it} - \bar{m}_t^j)$. If all households were fully insured this term would always be equal to zero—though aggregate shocks to Gombe might change IMRSs, they would change in precisely the same way for every household. Where i's IMRS differs from the aggregate we might think this was because he held a different portfolio of assets from other households. If, for example, a household held proportionally more maize in its portfolio than did others, then $(m_{it} - \bar{m}_t^j)$ would be negatively correlated with returns to maize. Thus, we're looking for evidence that some households have IMRSs which predictably respond differently to returns than does the average household.

The central prediction is that realizations of $y_{it} = (m_{it} - \bar{m}_t^j) R_t^j \delta_{it-1}^j$ should be uncorrelated with any variable z_{it-1}^j in the time $t-1$ information set. One particular variable that should be uncorrelated under the null, but not under some important alternative hypothesis is lagged IMRS. If, for example, credit constraints were important for some households then the IMRSs would be lower than for other households, and the credit constraint would alter the portfolio of investments. Another is the assignment to treatment: does giving a bag of maize to a farmer make that farmer better off when prices rise? And a third is lagged stocks—it seems natural to suppose that there's considerable persistence in stocks held, so if a farmer held more maize last period then we might expect his IMRS differently to changes in maize prices than for the average farmer in Gombe.

2.6.1 Test of Euler Equation

Table 2.9 reports results of this test. We are interested in testing the null hypothesis that forecast errors y_{it} are orthogonal to a variety of different variables that are in the time $t-1$ information set. The first column asks whether or not treatment status (receiving a loan of cash or kind) is correlated with these errors; there is no evidence at all of such a correlation. The second asks whether or not twice-lagged m_{it-2} is orthogonal—one might think that having more or fewer resources in the past might help to predict the forecast error, but again, it does not. We next consider the experimental measures of IMRS we elicited, lagged one period. There is some weak evidence of such a correlation, but not enough for us to reject the null hypothesis ($p = 0.15$). Then finally we put all of these variables together; unsurprisingly, we again fail to reject the null.

2.7 Conclusion

PHLs have gained attention as a potential way to help farm households earn additional income and smooth consumption across the season. However, this theory of change rests on grain prices rising enough to cover the loans' interest rate, which is a highly risky proposition in sub-Saharan Africa. This has ambiguous implications for the expected returns to PHL programs. In addition to the possibility of low returns and high default when prices fail to rise *ex post*, loans also shift *ex ante* risk from states of the world with extremely high prices to those with low prices. As such, demand for the loans can be influenced by how the household expects its consumption to covary with grain prices.

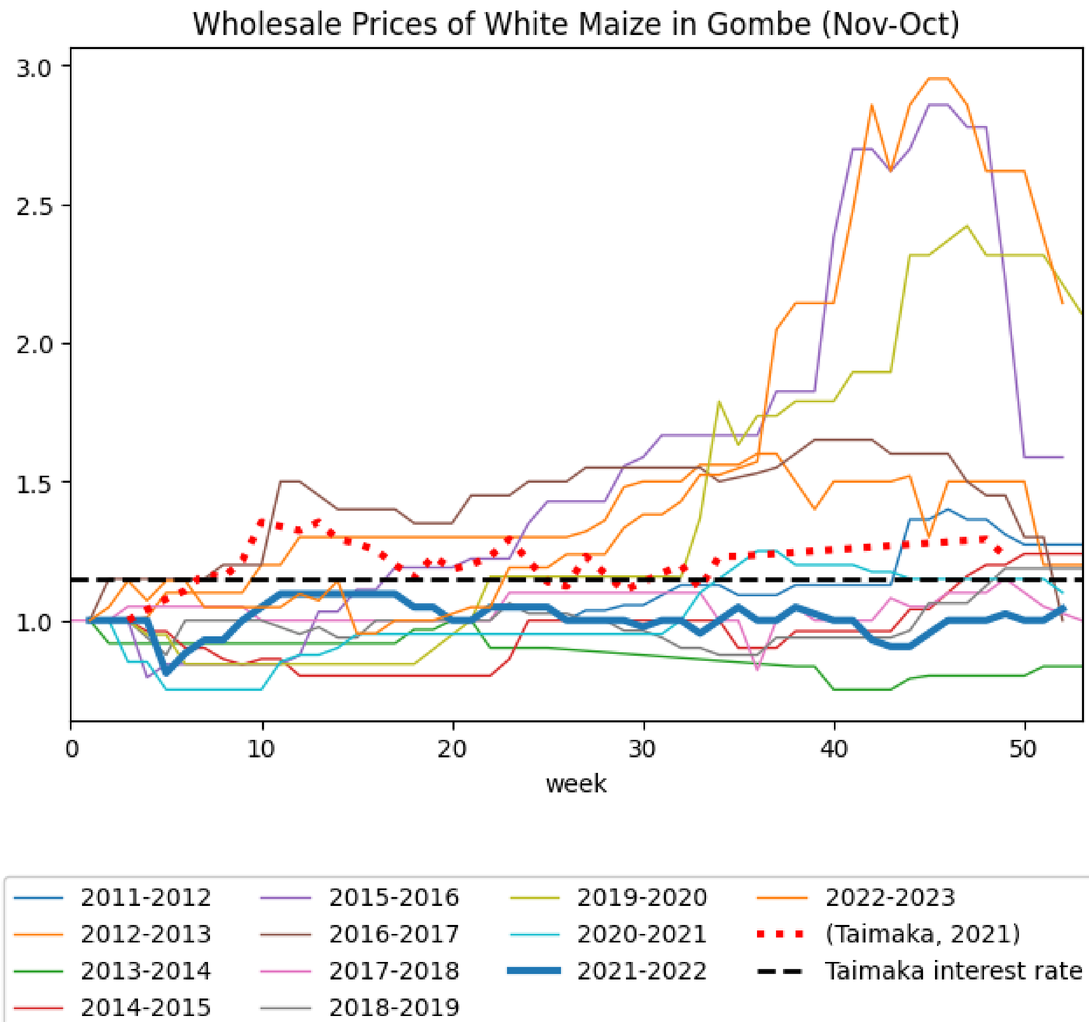
Our results from an RCT of a PHL in Gombe State Nigeria, in a year in which prices did not rise, show that the loan induced households offered a cash loan to store more grain for longer, but those offered an in-kind loan. However, neither treatment led to large increases in profits or welfare according to most measures – apart from a marginally significant reduction in the experimentally elicited IMRS. Over the course of the season, we also see small increases in business investment and livestock holdings in the cash loan group. This is consistent with relaxing credit constraints allowing households to store more, but these investments not paying off.

We attempt to use our model of intertemporal arbitrage to say something more broadly about the *expected* returns to storing grain, particularly whether households could be better off *ex ante* by holding different portfolios. We do so by testing whether variation in households' realized IMRS is correlated with variables in their prior period's information set. We fail to reject the null that treatment assignment, lagged (estimated) IMRS, stocks, and experimentally elicited IMRS are uncorrelated with this variation, although we are nearly able to for the latter case. This suggests that while the integration of Gombe with the broader economy is poor, we find little evidence that farmers are failing to respond optimally to local prices and returns. Nevertheless, this by no means rules out that significant market failures are at play.

Taken together, the unfavorable (for arbitrageurs) realizations of prices, the lack of evidence against the *ex ante* optimality of portfolios, and the potential adverse selection channels suggest that PHLs may not be the optimal policy to improve seasonal welfare. As a practical matter, further research could consider alternatives such as forward contracts, which would provide farmers liquidity without exposing them to price risk later in the season. More broadly, There is poor evidence against the efficiency of allocation and production within Gombe *given local prices*. However, there is solid evidence that the integration of Gombe with the broader economy seems to be poor. On the consumption side, this is supported by the evidence of a "lean season" and seasonal variation in average IMRSs, indicating that Gombe is poorly integrated with broader credit markets. On the production side though, there's considerable uncertainty regarding local grain prices. This would be fine if variation in these prices was mirrored by world prices in these commodities, but the correlation here, while significant, is quite weak. The consequence is that local supply shocks affect prices more than they would were the economy better integrated; these highly variable prices lead to high variation in IMRSs and limit incentives for investment.

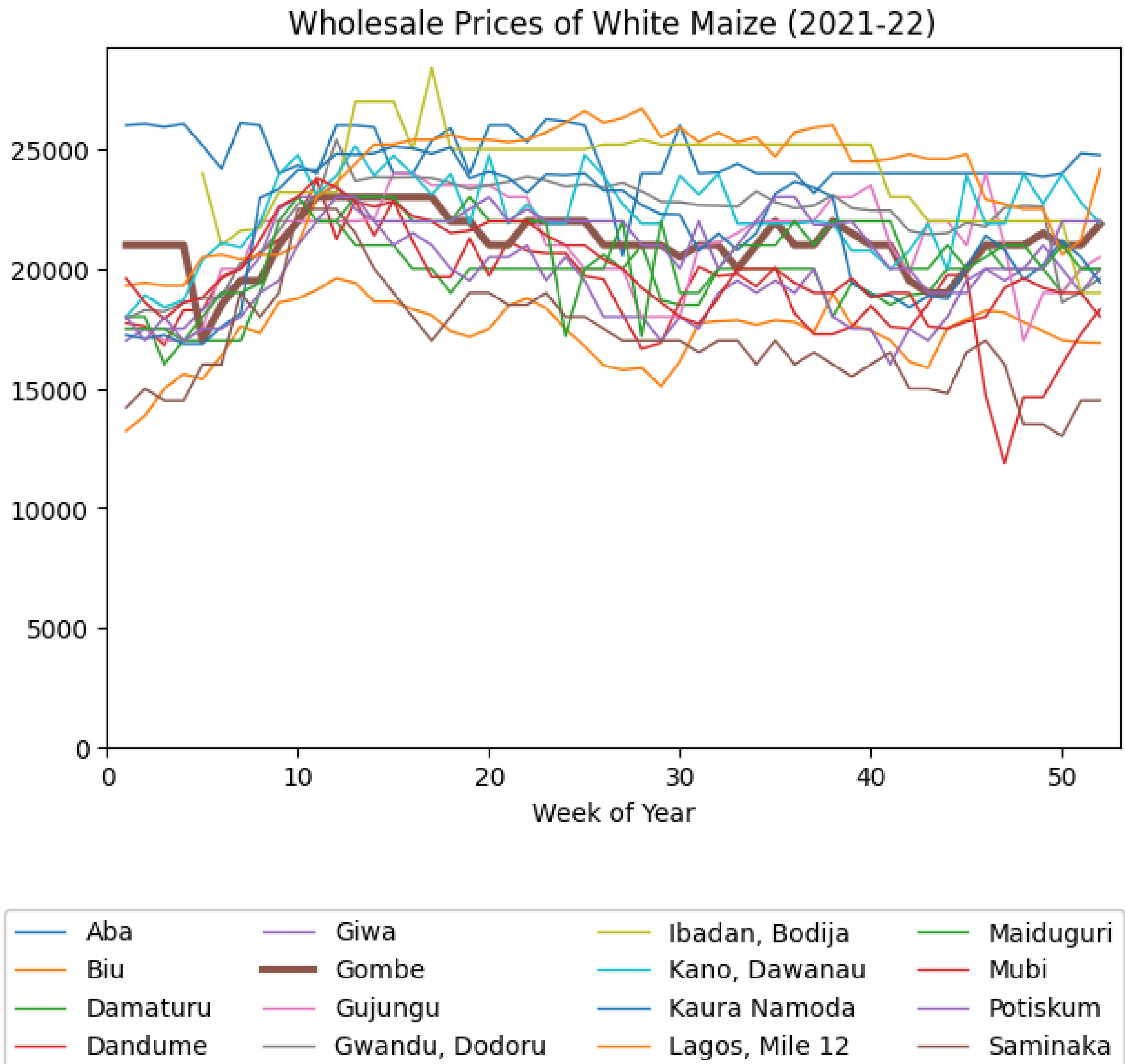
2.8 Figures and Tables

Figure 2.1: Maize price increases in Gombe, by year



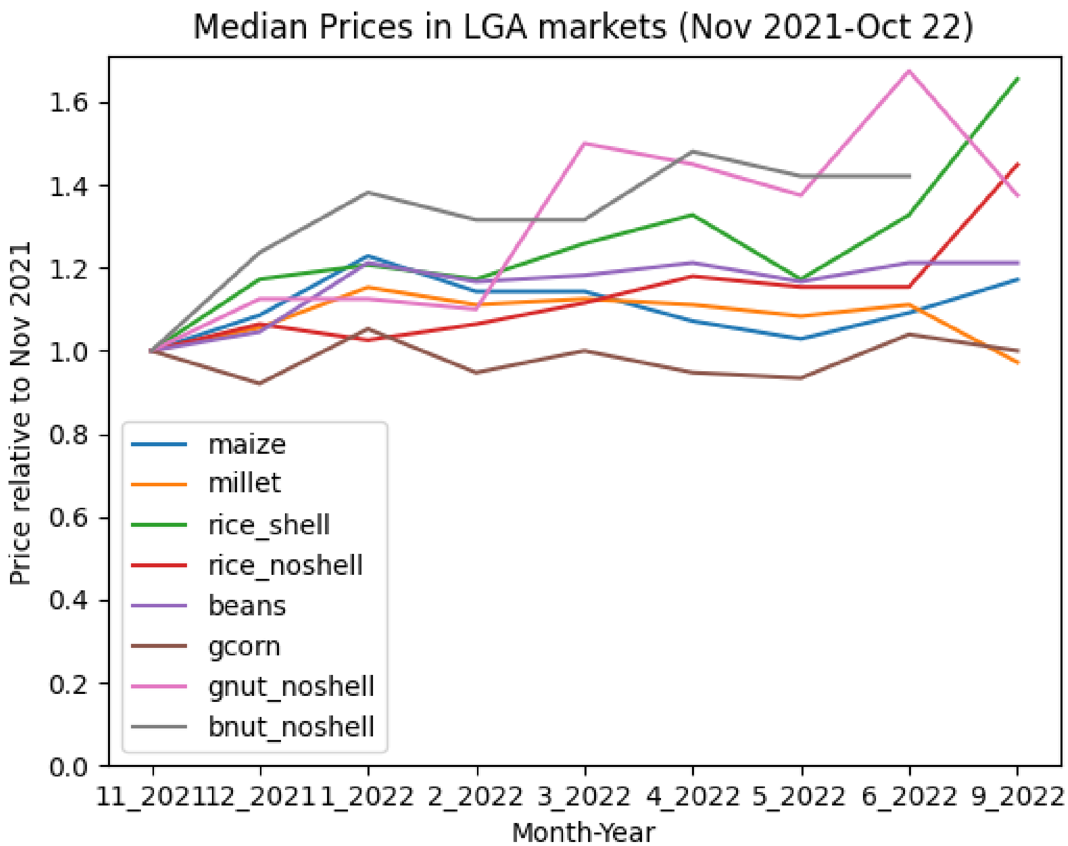
This figure contains data from FEWS-Net 2023 on the prices of white maize in Gombe market relative to November 1st in each 12-month period at a weekly frequency. The bold line corresponds to prices during the study period while the dotted line shows the evolution of the median price collected by Taimaka from markets across Gombe state. The dashed line shows Taimaka's interest rate of 15%.

Figure 2.2: Maize prices in Nigeria, 2021-22



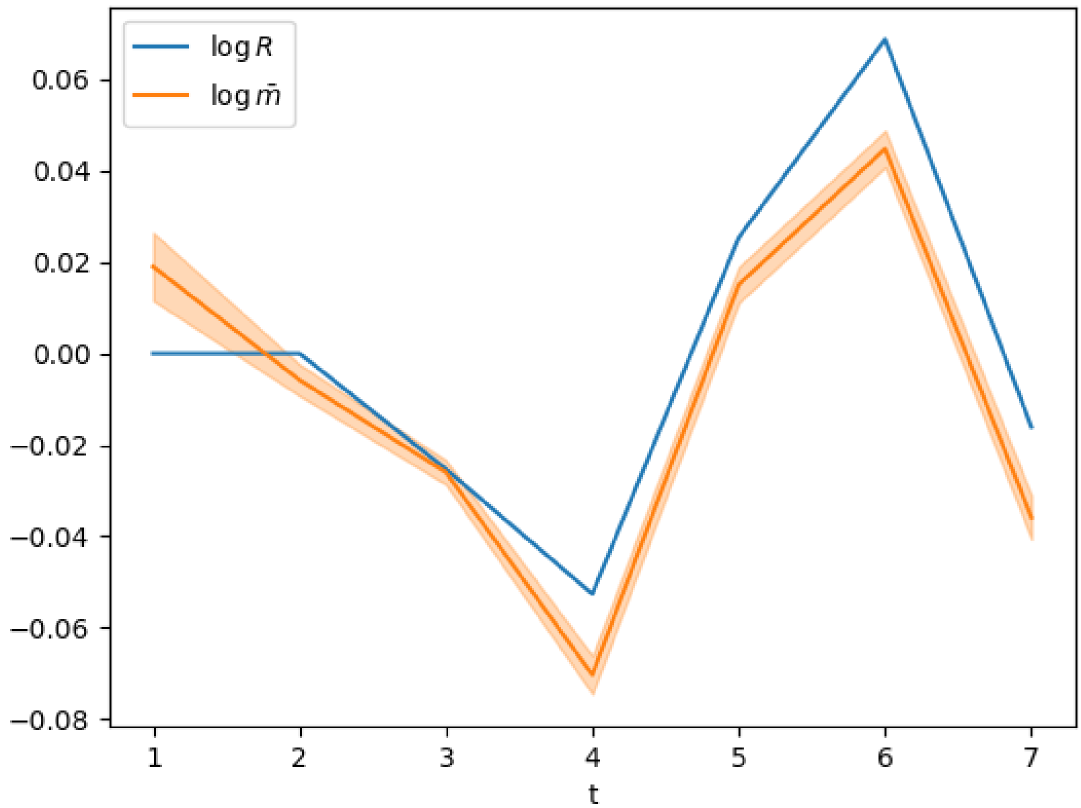
This figure contains data from FEWS-Net 2023 on the prices of white maize in markets across Nigeria from November 1st, 2021 to October 31st, 2022 at a weekly frequency. Prices for Gombe are bolded.

Figure 2.3: Prices in local markets collected by Taimaka, 2021-22



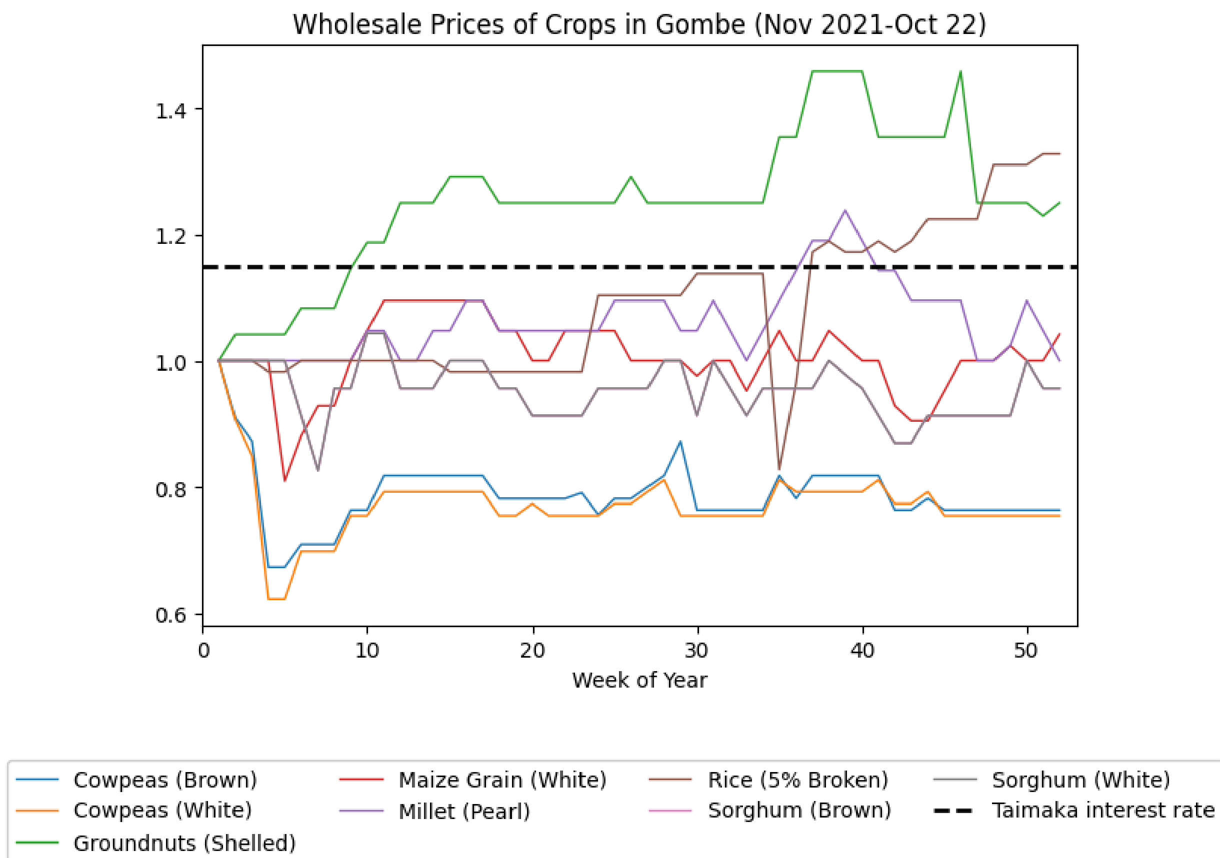
This figure contains data from Taimaka on the median prices of various crops in markets across Gombe from November 1st, 2021 to October 31st, 2022 at a weekly frequency.

Figure 2.4: Maize returns and IMRS



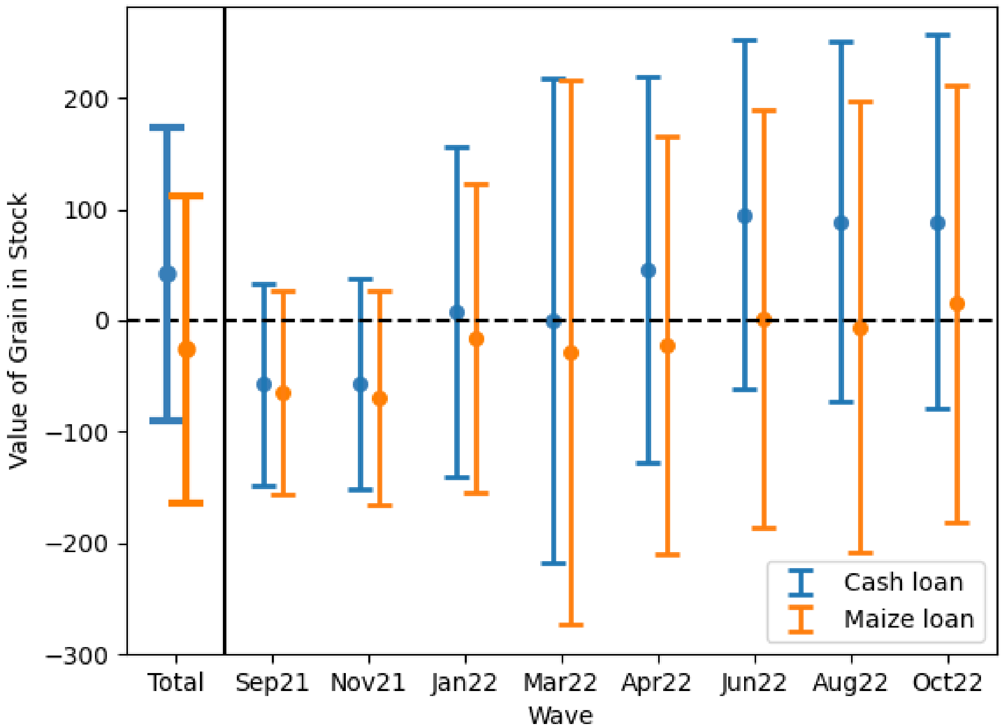
Time series of logarithms of maize returns in Gombe and the average intertemporal marginal rate of substitution $\bar{m}$. The latter series is scaled to have the same standard deviation as the former. The region around $\log \bar{m}$ series indicates standard errors of the estimated mean.

Figure 2.5: Prices of major crops in Gombe, 2021-22, relative to November 1st 20221



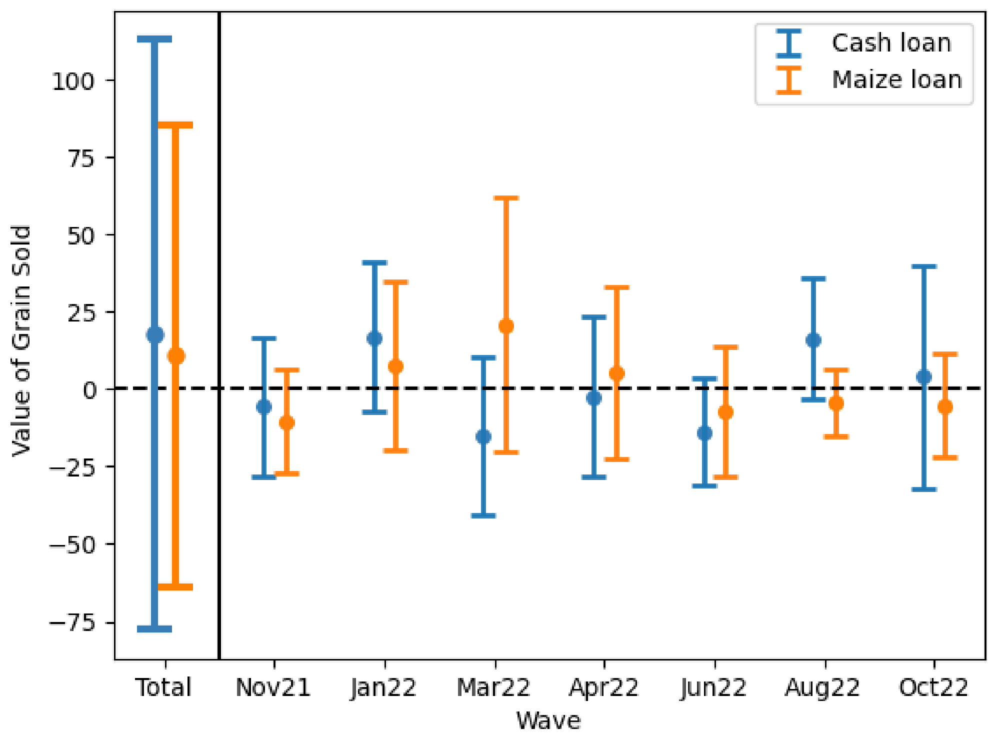
This figure contains data from FEWS-Net 2023 on the prices of various crops in Gombe from November 1st 2021 to October 31st 2022 at a weekly frequency, relative to November 1st.

Figure 2.6: Treatment effects on value of grain in stock by wave



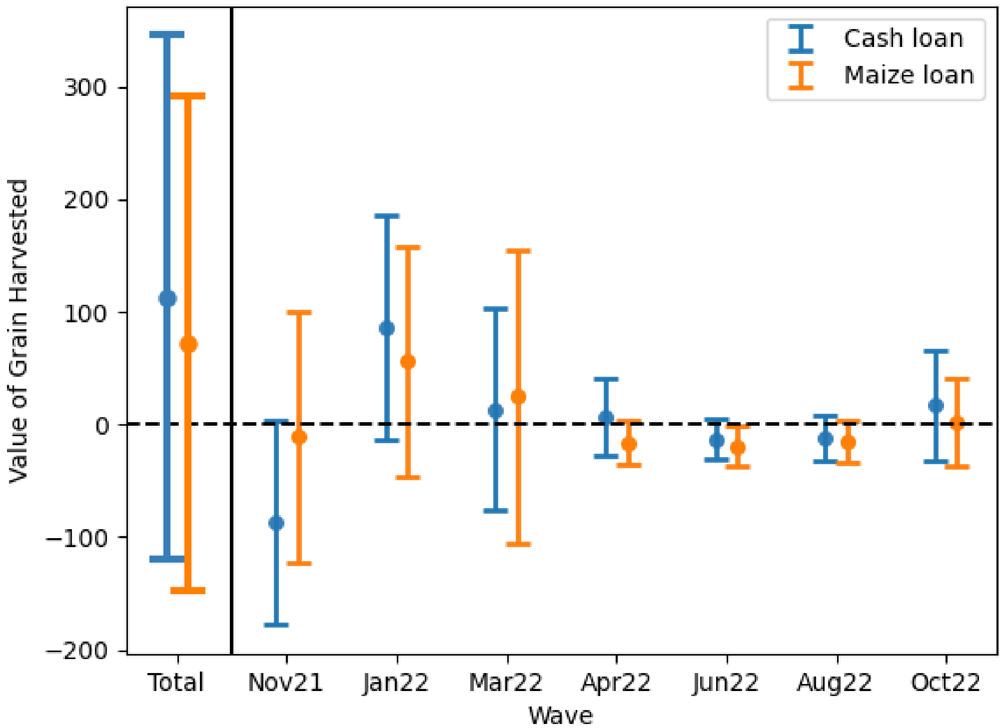
This figure shows the average treatment effects of the cash and maize loans on the value of households' stocks of all crops in 000's of Naira. To the left of the solid vertical line is the average effect on stocks, controlling for wave and strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level. Stocks are imputed as the cumulative sum of grain flows when not directly reported.

Figure 2.7: Treatment effects on value of grain sold by wave



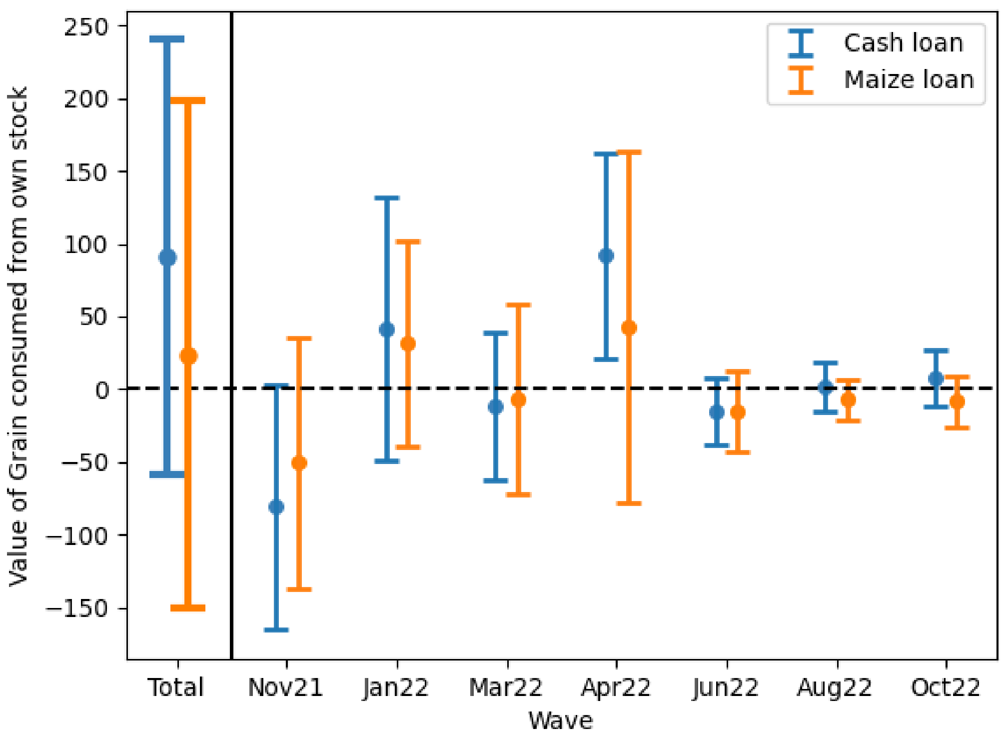
This figure shows the average treatment effects of the cash and maize loans on the value of households' sales of all crops in thousands of Naira. To the left of the solid vertical line is the effect on total sales over the season, controlling for strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level. Sales are imputed following the procedure in Appendix B.1.

Figure 2.8: Treatment effects on value of grain harvested by wave



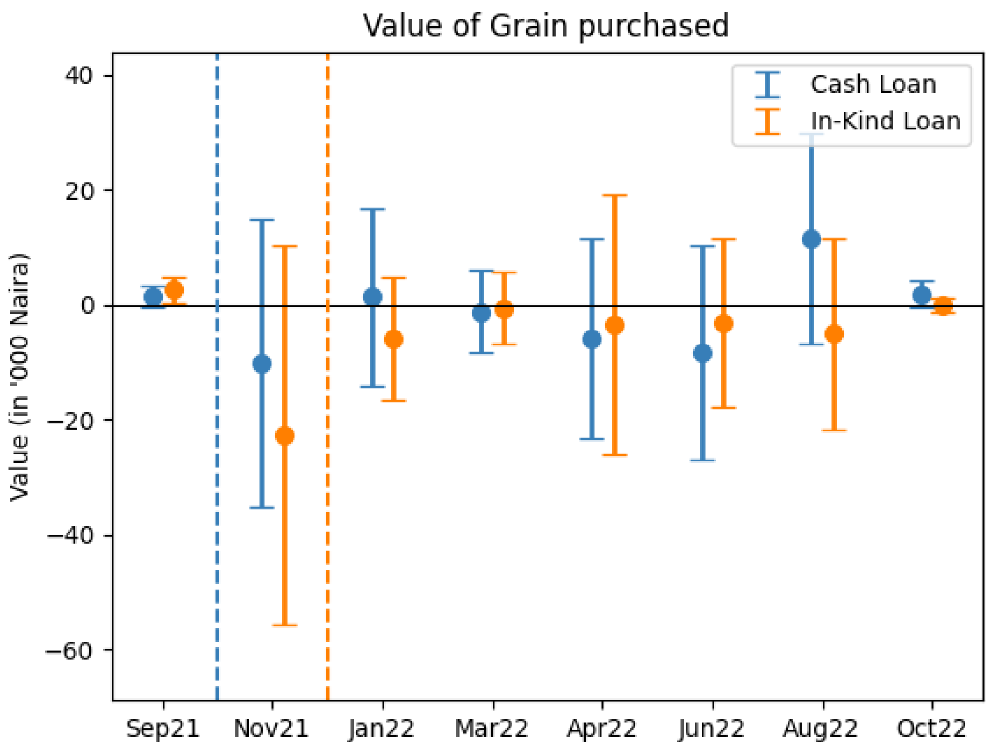
This figure shows the average treatment effects of the cash and maize loans on the value of households' harvests of all crops in 000's of Naira. Note that in the survey, harvests were asked as the amount of grain brought back to the homestead from the field, which excludes any sales that may have taken place directly from the field. To the left of the solid vertical line is the effect on total harvests over the season, controlling for strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level. Harvests are imputed following the procedure in Appendix B.1.

Figure 2.9: Treatment effects on value of grain consumed from own stock, by wave



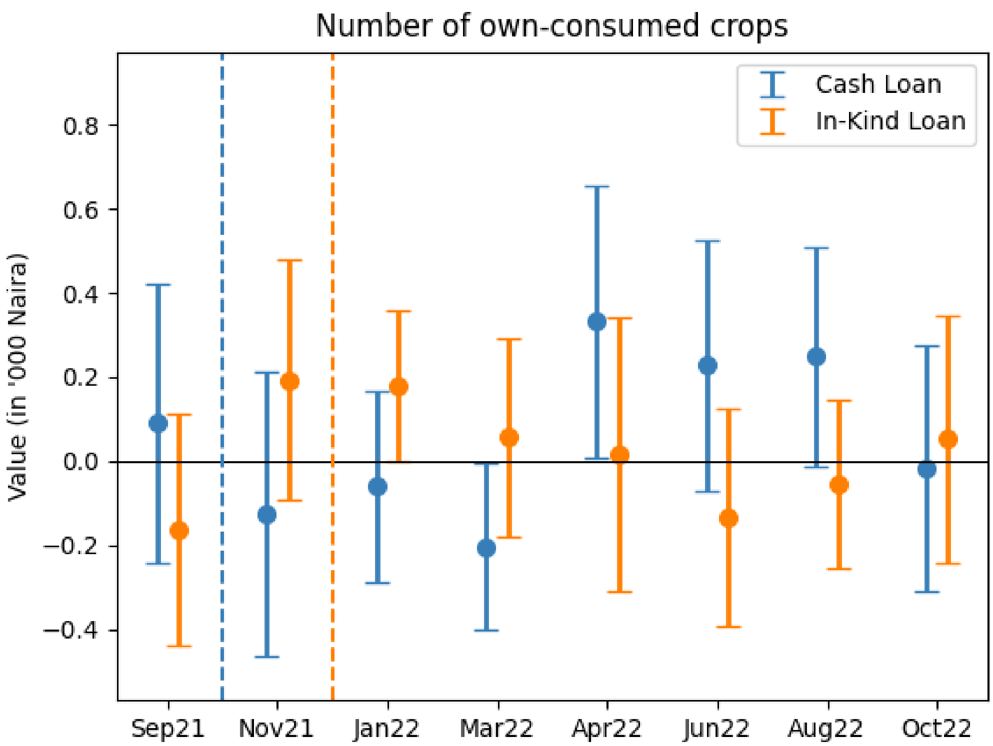
This figure shows the average treatment effects of the cash and maize loans on the value of households' consumption of their stored crops in 000's of Naira. Note that in the survey, consumption from own stocks was only directly elicited after the third survey wave, and is otherwise imputed following the procedure in Appendix B.1. To the left of the solid vertical line is the effect on total consumption over the season, controlling for strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level.

Figure 2.10: Treatment effects on grain purchased to store, by wave



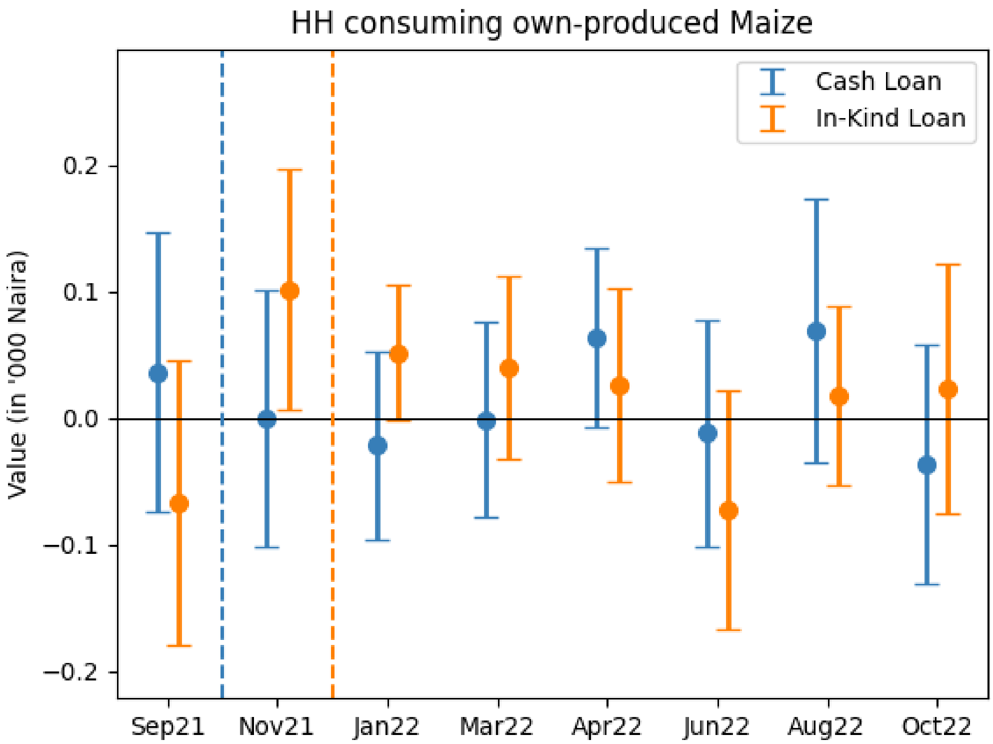
This figure shows the average treatment effects of the cash and maize loans on the value of households' purchases of crops for the purposes of storage. To the left of the solid vertical line is the effect on total purchases over the season, controlling for strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level. Sales are imputed following the procedure in Appendix B.1.

Figure 2.11: Treatment effects on number of crops consumed from own stock, by wave



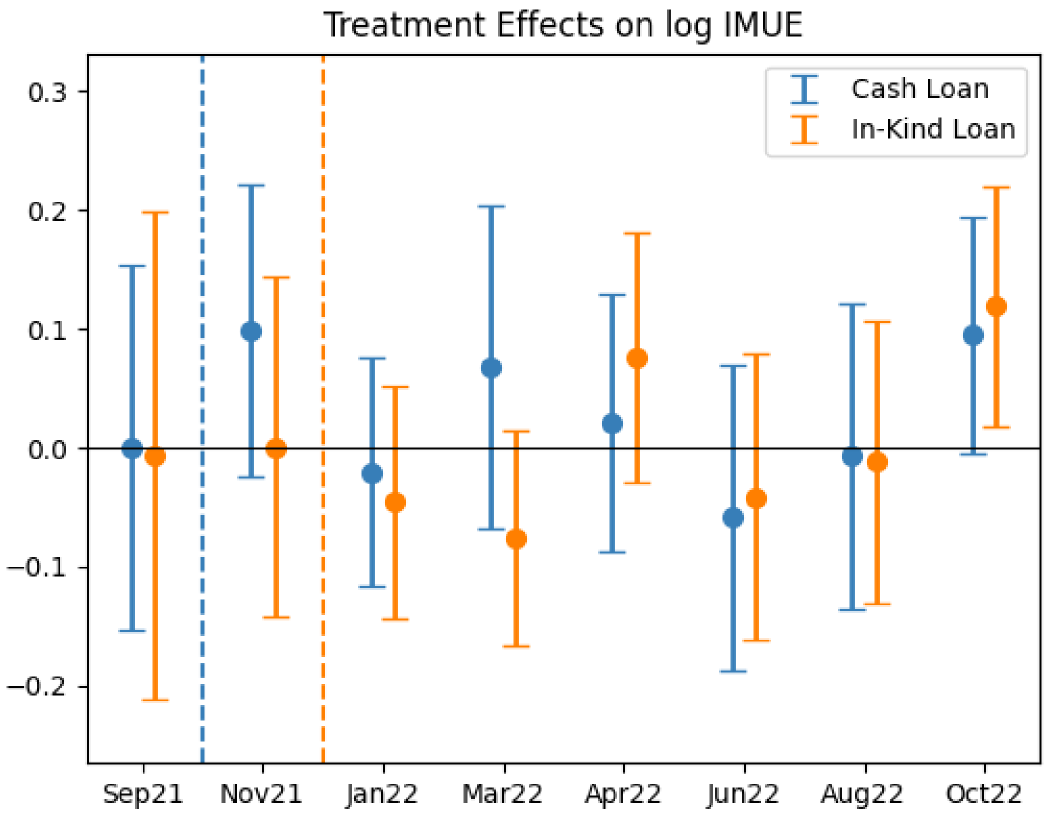
This figure shows the average treatment effects of the cash and maize loans on the number of crops households consumed from their own stocks in a given month. To the left of the solid vertical line is the effect on the average number of crops consumed over the season, controlling for wave and strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level.

Figure 2.12: Treatment effects on dummy for consuming own-produced maize, by wave



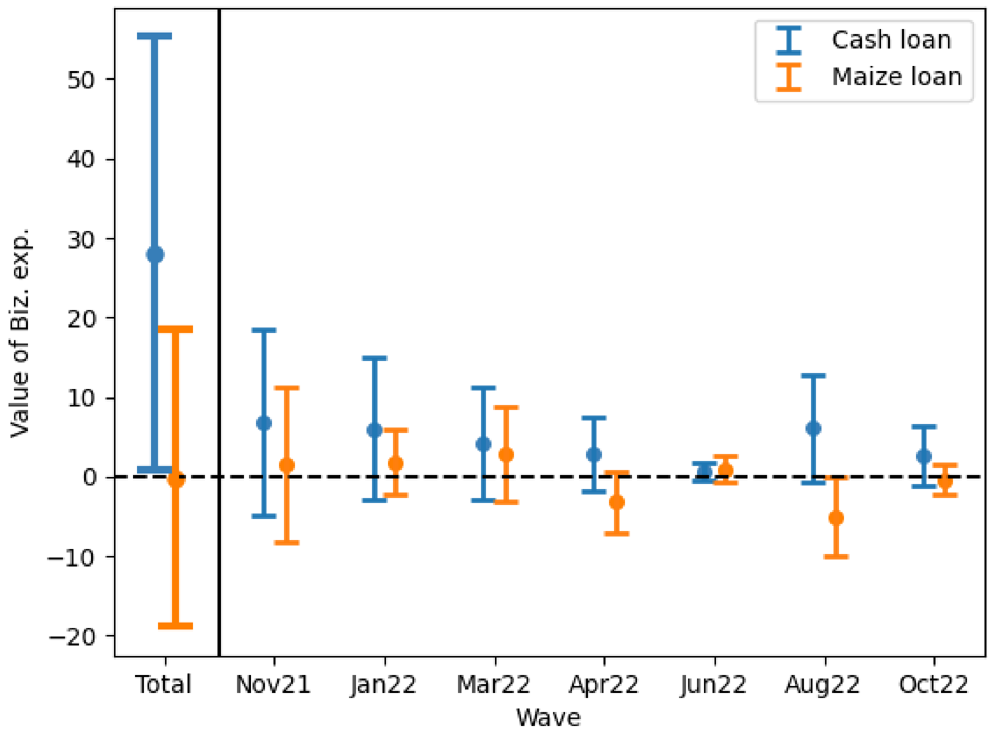
This figure shows the average treatment effects of the cash and maize loans on whether households consumed own-produced maize in a given month. To the left of the solid vertical line is the average effect over the course of the season, controlling for wave and strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level.

Figure 2.13: Treatment effects on log IMUE, by wave



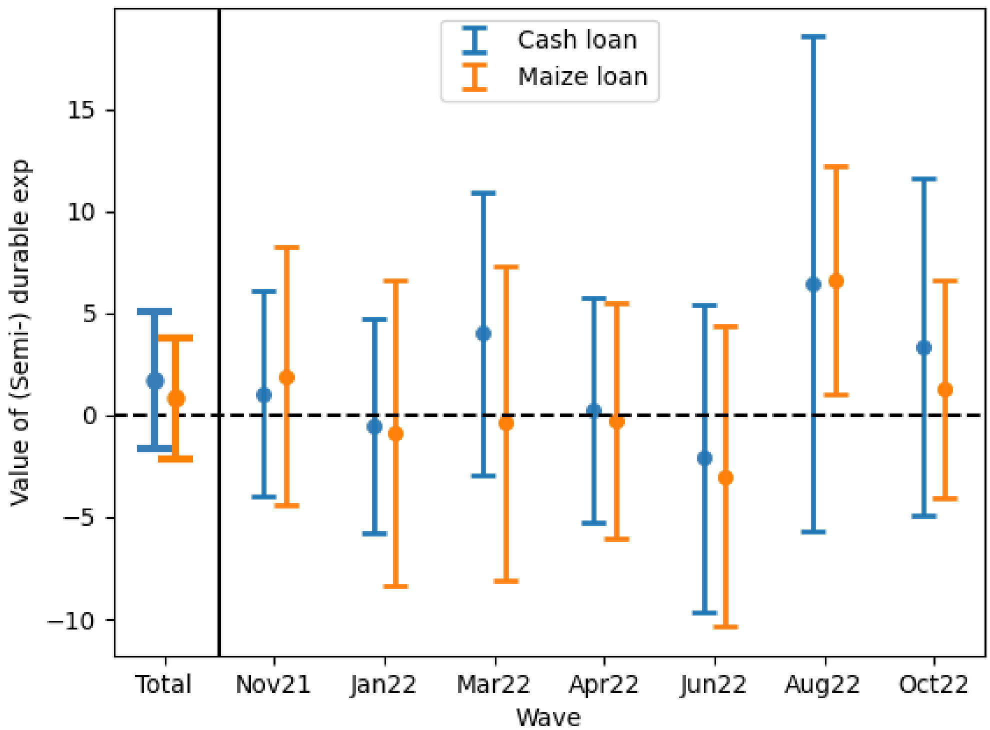
This figure shows the average treatment effects of the cash and maize loans on our estimate of the marginal utility of expenditure, log λ, following Ligon (2020). To the left of the solid vertical line is the average effect over the course of the season, controlling for wave and strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level.

Figure 2.14: Effects on business expenditure, by wave



This figure shows the average treatment effects of the cash and maize loans on business expenditures. To the left of the solid vertical line is the effect on total expenditures over the course of the season, controlling for wave and strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level.

Figure 2.15: Effects on semi-durable expenditure, by wave



This figure shows the average treatment effects of the cash and maize loans on durable and semi-durable nonfood expenditures. To the left of the solid vertical line is the effect on total expenditures over the course of the season, controlling for strata fixed effects. To the right, coefficients are broken out by survey wave, where September 2021 is the baseline period. Standard errors are clustered at the stratum level.

Table 2.1: Sales value by crop

	Sales Value ('000s Naira)					
	Total	Maize	Millet	Beans	Guinea Corn	Rice
Cash Loan	18.255	7.652	14.220	-0.234	-9.650	4.389
	(49.521)	(13.339)	(35.016)	(25.278)	(6.863)	(4.937)
Maize Loan	1.725	5.791	0.004	-10.152	-4.615	3.021
	(36.602)	(9.950)	(11.025)	(18.878)	(6.762)	(7.076)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.329	0.219	0.163	0.375	0.050	0.129
Control mean	323.698	54.631	71.143	147.828	14.436	15.473
Observations	930	930	930	930	930	930

This table contains estimates of the treatment effects of cash and maize loans on the total sales of each crop. Sales values are imputed following the procedure in Appendix B.1. Standard errors are clustered at the stratum level.

Table 2.2: Consumption of own stocks value by crop

	Value of Stock Consumed ('000s Naira)					
	Total	Maize	Millet	Beans	Guinea Corn	Rice
Cash Loan	92.490	6.279	44.007*	48.228***	3.275	-15.513
	(78.486)	(35.889)	(24.756)	(17.032)	(11.653)	(14.510)
Maize Loan	22.825	3.765	9.105	32.867	-1.815	-18.590
	(85.620)	(36.671)	(22.567)	(21.075)	(6.990)	(20.425)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.303	0.299	0.339	0.222	0.111	0.443
Control mean	673.146	305.729	133.176	94.621	46.285	86.109
Observations	930	930	930	930	930	930

This table contains estimates of the treatment effects of cash and maize loans on the consumption of own grain stocks by crop. Consumption values are imputed following the procedure in Appendix B.1. Standard errors are clustered at the stratum level.

Table 2.3: Harvest value by crop

	Value of Harvest ('000s Naira)					
	Total	Maize	Millet	Beans	Guinea Corn	Rice
Cash Loan	112.548	13.535	46.186	43.521	5.570	-7.531
	(117.218)	(43.409)	(53.608)	(31.572)	(8.571)	(15.084)
Maize Loan	47.392	20.536	3.908	16.533	4.042	-1.748
	(105.324)	(41.252)	(34.433)	(26.685)	(6.906)	(22.641)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.309	0.266	0.338	0.382	0.188	0.299
Control mean	949.036	320.456	255.228	228.499	43.415	74.778
Observations	930	930	930	930	930	930

This table contains estimates of the treatment effects of cash and maize loans on the amounts of each crop brought home from the field ("harvested"). Harvest values are imputed following the procedure in Appendix B.1. Standard errors are clustered at the stratum level.

Table 2.4: Treatment effects on household consumption

	Log exp.	Log non-. storable exp.	log λ	Elicited IMRS	Seasonal hunger index
Cash Loan	-0.028	-0.004	0.016	-0.035	0.009
	(0.084)	(0.055)	(0.024)	(0.022)	(0.019)
Maize Loan	0.000	0.020	0.000	-0.041*	0.020
	(0.084)	(0.066)	(0.028)	(0.020)	(0.031)
Fixed Effects	Strat+Wv	Strat+Wv	Strat+Wv	Strat+Wv	Strat+Wv
R^2	0.045	0.129	0.243	0.191	0.006
Control mean	7.750	6.619	0.324	1.691	-0.010
Observations	6404	6404	6404	6404	6404

This table contains estimates of the treatment effects of cash and maize loans on measures of household consumption and welfare, pooled across periods. The outcome in the first column is the log of households' most recent food expenditures. The second is the log of expenditure on foods that are not commonly stored. Column 3 is the estimated (log) marginal utility of expenditure, following Ligon (2020). The final column is the Reduced Coping Strategies Index for seasonal hunger developed by Maxwell et al. (2014). All specifications contain stratum + survey wave fixed effects. Standard errors are clustered at the stratum level.

Table 2.5: Effects on agricultural outcomes

	Ag. exp . ('000 Naira)	Planted area (ha)	Dry season ag. exp.	Dry season area (ha)	Harvest value
Cash Loan	13.51	0.124	0.125	0.001	163.7
	(20.82)	(0.608)	(0.138)	(0.008)	(152.9)
Maize Loan	-0.043	-0.374	-0.147	-0.012	72.87
	(22.24)	(0.576)	(0.838)	(0.019)	(112.6)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.247	0.170	0.070	0.071	0.137
Control mean	144.122	5.482	.894	0.024	784.949
Observations	829	829	829	829	808

This table contains estimates of the treatment effects of cash and maize loans on measures of agricultural investment in the 2022 planting season. The outcome in the first column is total expenditure, including land, labor, fertilizer, seed, and equipment during the 2022 rainy season in 000's of Naira. The second column is the total area planted in hectares. The next two columns repeat the same outcomes for the 2022 dry season. The fifth column reports the value of the 2022 harvest in 000's of Naira, which includes actual and anticipated harvests as not all households had completed harvesting at endline. All specifications contain stratum fixed effects. Standard errors are clustered at the stratum level.

Table 2.6: Effects on business and borrowing

	(Semi-) Durable Exp.	Biz. Exp.	Amount Borrowed	Any Biz Activity	Any Borrowing
Cash Loan	1.047	4.181**	0.866	0.005	0.028*
	(1.692)	(1.970)	(0.901)	(0.015)	(0.015)
Maize Loan	0.213	-0.042	-0.434	-0.004	0.017
	(1.457)	(1.401)	(0.614)	(0.019)	(0.012)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.027	0.030	0.023	0.057	0.065
Control mean	13.810	5.396	2.540	0.121	0.099
Observations	6404	6404	6404	6404	6404

This table contains estimates of the treatment effects of cash and maize loans on measures of monthly business expenditure and borrowing. The outcome in the first column is total expenditure, including land, labor, fertilizer, seed, and equipment during the 2022 rainy season in 000's of Naira. The second column is the total area planted in hectares. The next two columns repeat the same outcomes for the 2022 dry season. The fifth column reports the value of the 2022 harvest in 000's of Naira, which includes actual and anticipated harvests as not all households had completed harvesting at endline. All specifications contain stratum + survey wave fixed effects. Standard errors are clustered at the stratum level.

Table 2.7: Effects on livestock holdings

	Cows	Goats	Sheep	Chickens	Donkeys
Cash Loan	0.538	0.947	0.415	1.710*	-0.004
	(0.441)	(0.581)	(0.538)	(0.885)	(0.012)
Maize Loan	-0.305	-0.527	-0.564	0.119	-0.005
	(0.281)	(0.403)	(0.424)	(0.689)	(0.011)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.134	0.230	0.127	0.162	0.035
Control mean	1.528	3.843	2.472	2.262	0.010
Observations	829	829	829	829	829

This table contains estimates of the treatment effects of cash and maize loans on livestock holdings at endline. All specifications contain stratum fixed effects. Standard errors are clustered at the stratum level.

Table 2.8: Effects on land holdings

	Land owned (ha)	Land rented out (ha)	Land rented in (ha)
Cash Loan	2.871	0.161	-0.340
	(2.634)	(0.209)	(0.422)
Maize Loan	-0.597	-0.015	-0.558
	(0.577)	(0.117)	(0.413)
Fixed Effects	Strat	Strat	Strat
R^2	0.060	0.086	0.331
Control mean	5.785	0.187	2.395
Observations	829	829	829

This table contains estimates of the treatment effects of cash and maize loans on landholdings at endline. All specifications contain stratum fixed effects. Standard errors are clustered at the stratum level.

Table 2.9: Tests of the Euler asset pricing equations.

	Treatment	L^2Stocks	L^2m x loan	IMRS	All
χ^2	2.19	6.97	9.83	8.06	30.59
df	11.00	17.00	17.00	5.00	35.00
p-value	1.00	0.98	0.91	0.15	0.68

Different columns involve tests of the orthogonality of errors to different sets of variables. Column 1 uses a dummy for assignment to either the cash or maize treatment arms. L^2 denotes the second lag of the IMRS, m_{it-2}, which is interacted with the value of stocks and the treatment dummy. Column 4 uses the elicited IMRS and Column 5 tests these restrictions jointly.

Chapter 3

Enter Sandmo: Production Function Estimation for Firms that Consume

3.1 Introduction

Small family farms employ and feed a massive share of the world's poor. Therefore, understanding productivity and welfare requires reliable estimates of their production functions. However, production function estimation is always wrought with identification challenges, which can be especially problematic when producers are not profit-maximizing firms but expected utility-maximizing households. A classic result by Sandmo (1971) posits that if firms are risk-averse, (expected) output will be lower under risk than under certainty. Yet, the implications of risk aversion for production function estimation are still not completely understood. In this chapter, I introduce a class of novel production function estimators for farm households and other enterprises undertaken by risk-averse producers. This includes the homothetic and generalized Cobb-Douglas estimators used in Chapter 1, as well as a multi-stage production function in which shocks are realized sequentially within a season. I apply these approaches to data from farm households in rural Thailand.

The main challenge to production function identification in general is that observed input choices are likely to be correlated with unobservable factors, such as (anticipated) productivity, unobserved inputs, measurement error, and perhaps most importantly in the farm household context, distortions in input and financial markets. This is particularly problematic for approaches that involve regressing output on inputs, given the challenges of finding a suitable instrument for input use that's uncorrelated with these unobservables. Structural approaches in the spirit of Olley and Pakes (1996) offer an appealing alternative, using producers' optimal behavior to essentially proxy for unobservables known to or anticipated by the firm. In practice, this is done by inverting demand for a flexible input (Levinsohn and Petrin, 2003). However, this requires the firm's optimization problem to be well-specified. In particular, productivity must be a scalar unobservable in firms' input demands. The logic is that firms take all available information into account when choosing their inputs, including information unobservable to the econometrician.[1] In the most common applications (e.g. Ackerberg et al., 2015; Gandhi et al., 2020), this requires assuming that firms are risk-neutral profit maximizers facing competitive markets. While this may or may not be a reasonable approximation for manufacturing firms in advanced economies, it is almost certainly not for farm households in developing countries. Absent a full set of complete markets for inputs, credit, and risk, the household's consumption and production problems are non-separable (Benjamin, 1992), meaning that the marginal utility of expenditure, along with input wedges, enters input demands. In such settings, making productivity a scalar unobservable requires explicitly accounting for these deviations from the benchmark of profit maximization under complete markets.

This has been done in previous work to address adjustment costs (Asker et al., 2014), input price dispersion (De Loecker et al., 2016; Grieco et al., 2016), and markups (De Loecker and Warzynski, 2012; Asker et al., 2019; Cairncross et al., 2023), but not the types of distortions that farm households are likely to face, such as uninsured risk. My approach builds on the simple theory-consistent model of households' constrained optimal behavior laid out in Chapter 1 to identify the production function given how input and financial frictions enter first-order conditions. Doing so ensures unobserved shocks' effects on input demands are subsumed by households' constrained-optimal choices of consumption and investment.

Expanding on this theoretical framework, I develop a novel method to structurally estimate the production function from households' first-order conditions, even when frictions are present. My estimates of input and financial distortions, as described in Chapter 1, account for exactly *how* input and financial distortions affect input demands, making structural approaches valid again. Estimating the production function then amounts to identifying the parameters that rationalize these constrained optimal choices, as in a portfolio choice problem. To do so, I develop a linear Generalized Method of Moments (GMM) estimator in the spirit of Hansen and Singleton (1982) under the assumption of rational expectations.[2] To my knowledge, this is the first use of moments in consumption data to estimate a physical production function.

In contrast, the agricultural misallocation literature has typically calibrated the production function using input shares from settings where markets are assumed to function well, (Chen et al., 2023; Adamopoulos and Restuccia, 2020; Adamopoulos et al., 2022b), or used lagged instruments to estimate the production function in-sample (Shenoy, 2017; Manysheva, 2021). The issues with the former approach are that the underlying production function may be different in the U.S. and Canada than in Sub-Saharan Africa and Southeast Asia. The latter approach is valid in theory (Shenoy, 2021) but relies on strong assumptions about the

[1] The simplest example of this is calibrating Cobb-Douglas coefficients to observed revenue shares. However, these are not valid under imperfect markets because firms do not maximize expected profits and do not face common prices.

[2] Much like a Consumption Capital Asset Pricing Model (C-CAPM) problem, I treat inputs as risky assets whose (marginal) returns covary with the return to a household's overall portfolio, captured by the marginal utility of expenditure. However, in my case, the returns rather than marginal utilities (which have been estimated in the previous step) are the estimands of interest.

nature of unobserved shocks (i.e. autoregressivity).[3] In Table C.1.1, I replicate the Anderson and Hsiao (1981) estimation strategy used by Shenoy (2017) with the monthly (rather than annual) Townsend Thai dataset. First, the coefficients are implausibly low for labor and high for land. Second, I reject the overidentifying restrictions of the model when exploiting the panel structure of the data and the assumption that shocks are AR(1). This suggests that an alternative approach to estimating the production function may be required. Instead of trying to work around the endogeneity of inputs through instrumentation, I attempt to directly model it using the household's consumption problem. The main difference between my estimator and the dynamic panel estimators used elsewhere in the literature (e.g. Shenoy, 2017; Manysheva, 2021) is that the bulk of my assumptions rests on household optimization rather than the dynamics of unobserved shocks.

I also go beyond the homothetic CES specifications, Cobb-Douglas in particular, that remain the workhorse in many strands of literature, including the modern misallocation literature (Hsieh and Klenow, 2009; Adamopoulos et al., 2022b). The multiplicatively hetereoskedastic generalization that I consider, first introduced by Just and Pope (1978, 1979) in the spirit of Harvey (1976), simply relaxes the common restriction that the elasticity of the standard deviation of output equal the elasticity of the mean of output with respect to each input. In other words, this allows for some inputs to be "riskier" than others, allowing for risk to affect both the scale and composition of production in a tractable manner. This helps fill the gap is ruled out by construction in much of the empirical literature, despite the evidence on the importance of uninsured risk (Karlan et al., 2014; Emerick et al., 2016; Donovan, 2021).[4] While empirical work (e.g. Just and Pope, 1979; Antle, 1983; Di Falco and Chavas, 2009) has estimated production functions that allow for inputs to contribute differentially to higher moments of output, these analyses remain subject to the usual endogeneity concerns. This is particularly concerning if the input demands of the most risk-averse producers are differentially sensitive to shocks. In contrast, this analysis is the first, to our knowledge, to address these concerns in a theory-consistent manner, using the structure of input demands under risk aversion.

The third part of this chapter, which focuses on multi-stage production functions with sequential shocks, captures how farmers make input decisions at different points in the season

[3]Shenoy (2017)'s application of the Anderson and Hsiao (1981) estimator assumes that productivity shocks are AR(1) and that, when estimated with 2SLS, the first-stage coefficients on lagged inputs are homogeneous (Heckman and Vytlacil, 1998). Gollin and Udry (2021) accommodate heterogeneous first stages across plots within households by instrumenting for inputs with shocks to other plots and households in a correlated random coefficient model following Masten and Torgovitsky (2016). Note that by inverting input demands in the first stage, this approach also invokes the scalar unobservable, which does not theoretically hold but appears to generate negligible bias in their application. However, their results rely on differences across simultaneously cultivated plots by the same producer for identification; see Aragón et al. (2022) for a discussion of some of the pitfalls of using plot-level data, including fixed factors at the household level and increased susceptibility to measurement error.

[4]This differs from the analysis by Donovan (2021), in which the composition of production changes through supply-side channels in general equilibrium, rather than through inputs contributing differentially to risk.

with different information sets. In particular, inputs chosen later in the season are made after more of the uncertainty in production has been resolved. While a similarly dynamic production function has been introduced by Felkner et al. (2012), we are the first to our knowledge to estimate it allowing for risk-averse producers, capturing the importance of both overall and within-season risk on productivity and misallocation.

I find that this broad class of estimators performs well, both in Monte Carlo simulations and when applied to the Thai data. In the baseline Cobb-Douglas case it produces more plausible, precise, and robust estimates than the dynamic panel method used by Shenoy (2017) and Manysheva (2021) in similar contexts. There are much fewer applications of the generalized Cobb-Douglas specification based on Just and Pope (1978, 1979) in the modern empirical literature to compare against. However, comparing my estimates to those from the homothetic Cobb-Douglas specification highlights important takeaways. I find that while the elasticities of the mean of output with respect to each input are similar across both specifications, inputs applied earlier in the season are in fact riskier than those applied later.

While even a static non-homothetic framework captures this notion, the dynamic specification extends it to a much more realistic setting.

The rest of this chapter is organized as follows.

3.2 The Farm-Household Problem

We first introduce the basic setup of production featuring a general production function F and preferences U. I then derive estimators for three specifications of F: homothetic, heteroskedastic, and sequential Cobb-Douglas.

3.2.1 Setup

Time consists of discrete periods indexed by t and households indexed by j are infinitely-lived. Production takes place over S distinct stages indexed by s, after which output is harvested. In each stage, farmers apply K_s inputs q_{ks} prior to the realization of a shock φ_s. As such, final output $Y_{t+S} \in X$ is given by

$$Y_{t+S} = F(q, \varphi) \tag{3.1}$$

Note that we assume that technology F is common across households but that households may differ in productivity (captured by anticipated components of φ.

We assume households have time-separable, von-Neumann-Morgenstern preferences, which may depend on characteristics z, over a vector $c \in C \subseteq X$ of consumption goods and discount factor δ (note that leisure may be an element of c). As such, they maximize

$$\sum_{s=t}^{\infty} \delta^{s-t} u(c_s, z)$$

where u is continuously differentiable, strictly increasing and concave in c,[5] subject to the following budget constraint.

$$k_{t+1} = k_t + p_t(H_t Y_t - c_t) - w_{st} q_{st} + R_t(B_t - B_{t-1}) + x_t \tag{3.2}$$

Here, k is total assets, p_t is the price of the vector of consumption goods c_t, and w_{st} are input prices. B_t is an asset that households may borrow and save at interest rate R, although they may face a borrowing constraint $B_t \leq \bar{B}$ and x_t is an endowment that may yield uncertain (net) income in each state.

As a slight modification of Equation 1.7 in Chapter 1, the general first order condition for investment under uncertainty is

$$\lambda_{jr} w_{jkr} = \delta^R \mathbb{E}_t \left[\lambda_{jR} F_{kt}(q_j, \varphi_j) \right] \tag{3.3}$$

where q_j is a vector of K inputs in each of R periods

t denotes the current period, R denotes the harvest period, and

Production

Suppose that a farmer begins cropping operations at time t, and chooses to harvest S periods later. We assume that final output Y_{t+S} is given by

$$Y_{t+S} = \prod_{s=0}^{S} F_s(q_s, \varphi_s)$$

where F_s is the stage s-specific production function, q is a vector of K_s inputs applied at each stage s, and φ is an S vector of stage-specific shocks realized after q_s is determined.[6,7]

Since the farmer has to make decisions in a given stage in a state of ignorance regarding subsequent shocks, it's convenient to give state the problem recursively. Given any vectors of inputs q_s and any realized sequence of shocks ϵ_s, we have

$$A_{s+1} = A_s F_s(q_s, \varphi_s),$$

so that A_s summarizes the influence of all earlier shocks and inputs on production.

[5] The empirical implementation will require that preferences can be represented by the broad class of constant Frisch elasticity demands (Ligon, 2020).

[6] The assumption of multiplicative separability across stages is hardly general, but trivially holds for single stage production functions and will be assumed for the dynamic Cobb-Douglas specification discussed later.

[7] Allowing for different crops and plots is conceptually simple, but notationally complex, and we avoid doing this here. But the idea would be to treat A as a vector with each element corresponding to a single plot, with corresponding changes made to inputs q and shocks ϵ. Differences in stages across plots could be accommodated by similarly treating s as a vector, and the analysis below would go through.

Budget constraints

The farm-household enters the period with liquid assets k, crop progress A, and observes a stage/period shock ϵ. Out of the stock of assets the farm-household must purchase any consumption goods or services, as well as finance any crop operations.

Within any given period t the following happens. First, the farmer can purchase some costly vector of input q_s suitable to the current stage of production s at a price w_s, and also purchase a vector of consumption goods c. Note that prices of both inputs and consumption may also be random variables, in which case assets in the subsequent period are given by

$$k' = k - p_t^\top c - w_s^\top q_s.$$

Alternatively, the farmer can decide that it's time to harvest the crop. In this case, the farmer realizes harvest revenue $y_t(\omega, A) = p_t(\omega)A$. This leaves the decision of how much land q_0 to rent or set aside for production in the next season, at a price $w_0(\omega)$.

In this case, assets in the next period are given by

$$k' = k - p_t(\omega)^\top c + y_t(\omega, A) - w_0(\omega)q_0.$$

These two cases can be combined by defining a harvest decision variable $H \in \{0, 1\}$, where $H = 1$ indicates a decision to harvest, with

$$k' = k - p_t(\omega)^\top c + Hy_t(A, \omega) - w_s(\omega)^\top q_s. \tag{3.4}$$

Note here that prices for consumption goods depend on the month t as well as the aggregate shock ω, while the cost of inputs $w_s(\omega)$ depends on the aggregate shock and the stage of production.

Preferences

The preferences of the farm household may depend on a set of characteristics z, and are assumed to be both time-separable and von Neumann-Morgenstern. Then given z, the within-period utility function for the household is denoted $U(c; z)$; the function U is assumed to be continuously differentiable, strictly increasing, and concave in c.

Dynamic Program

At the beginning of any period, the relevant state for the household is given by the tuple (s, t, A, k, φ), where s is the current stage of production, t is the name of the month, A is the current crop state, k is the stock of liquid assets available to the household, and φ is a shock, which may contain both aggregate and idiosyncratic components.

$$V(s, t, A, k, \varphi) = \max_{c,q,H} U(c; z) + \delta \begin{cases} \mathbb{E}_t \left[V(s+1, t', AF_s(q, \varphi), k', \varphi')) \,|\varphi \right] & \text{if } H = 0 \\ \mathbb{E}_t \left[V(0, t', A_0, k', \varphi') \,|\varphi \right] & \text{if } H = 1 \end{cases} \tag{3.5}$$

such that seasons advance in their periodic fashion, with $t' = (t+1) \mod 12$, and such that the budget constraint is satisfied. Note that A_0 denotes the initial condition at which production begins at stage 0.

First order conditions

First, consider the first-order conditions with respect to the choice of consumption good i:

$$u_i(c; z) = p_i \lambda, \tag{3.6}$$

where u_i denotes the marginal utility of good i, and where λ is the multiplier associated with the budget constraint.

Second, the remaining first-order conditions correspond to different input quantities q_{ks}, chosen at period t for stage s:

$$F_{sk}(q_s, \varphi_s)\delta\mathbb{E}_t\left[\frac{\partial V}{\partial A}(s', t', A', \varphi')\right] = w_{ls}\lambda, \tag{3.7}$$

where it's understood that if this is a harvest period then $s' = 0$ and $A' = A_0$, and otherwise $s' = s + 1$.

Envelope condition

The partial derivative $\partial V/\partial A$ is of particular interest, given the role it plays in the first-order conditions for productive inputs in Equation (3.7). Applying the envelope theorem to the dynamic program Equation (3.5), we obtain, again conditional on the discrete harvest decision H:

$$\frac{\partial V}{\partial A}(s, t, A, k, \varphi) = \lambda H p_t + (1 - H)\delta\mathbb{E}_t\left[\frac{A'}{A}\frac{\partial V}{\partial A}(s', t', A', \varphi')\right]. \tag{3.8}$$

Production Result

We wish to obtain a condition that allows us to relate current input decisions to future harvest revenue, even if we don't observe *all* inputs or stages of production. The following result provides this condition, relating current expenditures on productive inputs to the expected product of the household's intertemporal marginal rate of substitution and eventual crop revenue.

Proposition 2. *For any input k at any stage s in period t, the farm-household will choose expenditures on the input to satisfy*

$$w_{ks} = \mathbb{E}_t\left[\left(\delta^{S-t}\frac{\lambda_{t+S}}{\lambda_t}\right)\frac{F_{sk}(q_s, \varphi_s)}{F_s(q_s, \varphi_s)}Y_{t+S}\right]. \tag{3.9}$$

Proof. Equation Equation (3.7) relates the expected marginal benefit of supplying an input k in stage s in month t to the marginal cost, expressing the marginal benefit in terms of the partial derivative of the value function with respect to A in the subsequent period. But Equation (3.8) gives gives us an expression of this partial derivative in terms of either the value marginal revenue of the harvest, or in terms of future partial derivatives.

Evaluating the left hand side of Equation (3.8) at $(s+1, t+1)$, substituting the right-hand-side into Equation (3.7), and noting that $\partial A'/\partial A = A'/A$ yields

$$A_s F_{sk}(q_s, \varphi_s) \frac{A_{s+2}}{A_{s+1}} \delta^2 \mathbb{E}_t \left[\frac{\partial V}{\partial A}(s+2, (t+2) \mod 12, A_{s+2}, \varphi_{t+2}) \right] = w_{ls}\lambda_t.$$

Further forward-substitution for $\partial V/\partial A$ proceeds until the farmer chooses to harvest, say S periods hence. Though r is a random variable, taking $H = 1$ implies from Equation (3.8) that $\partial V/\partial A = \lambda p$. Noting that $A_{s+1} = A_s F_s(q_s, \varphi_s)$,

$$w_{ks}\lambda_t = \mathbb{E}_t \left[\frac{F_{sk}(q_s, \varphi_s) A_{t+S}}{F_s(q_s, \varphi_s)} \delta^S \lambda_{t+S} p_{t+S} \right]. \tag{3.10}$$

Rearranging this equation and noting that $p_{t+S}A_{t+S} = Y_{t+S}$ yields the result. □

In this chapter, we consider three functional forms for F_s

$$F(q, \varphi) \equiv A\varphi \prod_k q_k^{\alpha_k} \tag{3.11}$$

$$F(q_s, \varphi_s) \equiv A \prod_k^K q_k^{\alpha_k} + \varphi B \prod_k^K q_k^{\beta_k} \tag{3.12}$$

$$F(q_s, \varphi_s) \equiv A_s \varphi_s \prod_k q_{ks}^{\alpha_{ks}} \tag{3.13}$$

in which $S = 1$ for Equations (3.11) and (3.12). Equation 3.11 is simply the static, homothetic Cobb-Douglas with Hicks-neutral shock φ.[8] Equation 3.12 is the generalized Cobb-Douglas introduced by Just and Pope (1979), in which α_k is the elasticity of the mean of output with respect to q_k and β_k is the elasticity of the standard deviation of output with respect to q_k. Finally, Equation 3.13 is the dynamic extension of Equation 3.11 in which production is Cobb-Douglas both within and across stages

Substituting for F in these respective cases, Equation (3.10) becomes

$$\lambda_t w_{kt} q_{kt} = \delta\alpha_k \mathbb{E}_t [\lambda_{t+1} Y_{t+1}] \tag{3.14}$$

$$\lambda_t w_{kt} q_{kt} = \delta\alpha_k \mathbb{E}_t [\lambda_{t+1}] \mathbb{E}_t [Y_{t+1}] + \beta_k \text{cov}(\lambda_{t+1} Y_{t+1}) \tag{3.15}$$

$$\lambda_t w_{ks} q_{ks} = \alpha_{ks} \mathbb{E}_t \left[\delta^S \lambda_{t+S} Y_{t+S} \right] \tag{3.16}$$

Each of these provides a set of moment conditions that can be exploited to estimate the parameters α and β.

[8] Note that in this case, φ can be thought of as $\frac{e^\phi}{\mathbb{E}[e^\phi]}$ such that $\mathbb{E}[\varphi] = 1$

3.3 Estimation Framework

Estimating these moment conditions requires knowledge of the (shadow) prices households face for each input w_{jkt}, which may be household-specific, and the marginal utilities of expenditure λ_{jt}. I describe how these can be estimated in the Thai data in Chapter 1.

The first-order conditions for input demands provide moment conditions that can be exploited to recover the production function parameters under rational expectations using linear GMM in the spirit of Hansen and Singleton (1982). In a sense, I treat inputs as assets in a C-CAPM problem whose returns covary with a household's overall portfolio captured by the marginal utility of expenditure, λ. The intuition behind this approach is simple. If all markets are perfect, then all households maximize expected profits and choose inputs to equate marginal revenue products with the common input price. However, households generally maximize expected utility rather than expected profits and may not face common (shadow) prices for all inputs. However, Equations 3.14-3.16 capture how households' optimal choices differ from those under the perfect markets benchmark — through the marginal utility of expenditure λ_{jt} and possibly household-specific (shadow) prices of each input, w_{jkt}. As in Chapter 1, I write these prices without loss of generality as $\bar{w}_{kvt}\tau_{jkt}$, where $\bar{w}_{kvt}$ is the average price of input k in village v at time t and τ_{jkt} the wedge between household j's shadow price and the market price.[9] Estimates of λ and τ account for *how* input choices are distorted and allow the production coefficients to be identified from the correctly-specified first-order conditions for input demands (1.12).

Estimating λ and τ are both non-trivial. The former requires assuming a utility function — in Chapter 1 I show how it can be estimated for both CRRA and the more flexible constant Frisch elasticity functional forms using disaggregated expenditure data. This does not depend on any assumptions about the production function, as it is simply a way of mapping consumption expenditures into a scalar welfare measure.

In contrast, estimating τ does depend on the properties of the production function. In particular, I show in Chapter 1 how if F is homothetic and at least one input (say, K) is perfectly tradable such that $\tau_K = 1$ for all households, then the remaining $K - 1$ τs can be inferred from input ratios, even without knowledge of the production function. I discuss alternatives for estimating τ in the nonhomothetic static and dynamic cases below. For what follows, I assume that one has obtained consistent estimates of λ and τ (in the homothetic case) following the procedures in Chapter 1.

For all three specifications considered, the additional assumption of rational expectations is required. The intuition is that since Equations (3.12) and (3.16) hold by virtue of optimization, any deviations between expected and realized $\lambda_{jt+1}Y_{jt+1}$ are mean-zero forecast errors. The catch is that households' subjective expectations over λ and Y, based on their time t information sets, at harvest aren't observed. However, as I show in Proposition 3, the key to estimation is to be able to replace these subjective expectations with data. If expecta-

[9] Note that this makes no assumptions about the source of these distortions or whether they act as a tax or a ration. Likewise, absent full insurance, λ_{jt} is a random variable.

tions are rational, then on average, they equal the observed data as the sample grows large. Replacing these subjective expectations with data makes it simple to recover the parameters α (and β).

3.3.1 Homothetic Cobb-Douglas

Let $x_{jkt} \equiv \bar{w}_{kvt}\tau_{kt}q_{jkt}$. x_{jkt} can be interpreted as household j's "shadow" expenditure on input k at time t. This can either represent actual expenditure under possibly household-specific prices or as the cost of input k such that the household would choose q_{jkt} under perfect markets. Let $\mathcal{I}_{jt}$ denote household j's information set at time t. Rearranging constrained-optimal input demands (1.12) and making the dependence on households' time t information sets explicit yields the moment condition

$$\delta\alpha_k\mathbb{E}[\lambda_{j,t+1}Y_{j,t+1}|\mathcal{I}_{jt}] - \lambda_{jt}x_{jkt} = 0 \tag{3.17}$$

for each input k where input $x_{jkt} = \bar{w}_{vt}\hat{\tau}_{jkt}q_{jkt}$ is (shadow) expenditure on input k is applied at time t and $\hat{\tau}$ is estimated as described in Section 1.4.2. Note that both λ_{t+1} and Y_{t+1} are unknown as of time t, as they both depend on the yet-to-be-realized φ_{t+1}.

While $x_{jkt}, \lambda_{jt}, \lambda_{jt+1}$, and Y_{jt+1} are all either observed or estimated, using (3.17) to identify the α_k requires mapping the unobserved *subjective* expectation $\mathbb{E}[\lambda_{jt+1}Y_{jt+1}|\mathcal{I}_{jt}]$ to data. Proposition 3 states that α can be estimated from (3.17) (up to the time-preference discount factor δ with a simple linear GMM procedure under rational expectations. The intuition is that if expectations are rational, then subjective expectations $\mathbb{E}[\lambda_{jt+1}Y_{jt+1}|\mathcal{I}_{jt}]$ will *on average* equal the observed $\lambda_{jt+1}Y_{jt+1}$. Substituting realized $\lambda_{jt+1}Y_{jt+1}$ into (3.17) identifies the α_k up to the time-preference discount factor δ. Moreover, optimization implies that any element of $\mathcal{I}_{jt}$ should be mean-independent of forecast errors, creating a large set of potential overidentifying instruments. In particular, lagged values of λ_{jt} are natural candidates.

Proposition 3. *Assume production is given by Equation* (3.11) *and that households have rational expectations and let $h(z_{jt})$ be a measurable function of variables $z_{jt} \in \mathcal{I}_{jt}$. Then the estimator defined by*

$$\arg\min_a J(a) \equiv g_{NTK}(a)'Wg_{NTK}(a)$$

where

$$g_{NTK}(a) \equiv \frac{1}{NT}\sum_t\sum_j\sum_k \delta a(\lambda_{j,t+1}Y_{j,t+1} - \lambda_{jt}x_{jkt} - \psi_{kvt}) \otimes h(z_{jt}),$$

where ψ_{kvt} is an input by village by year fixed-effect, is a consistent estimator of the vector of coefficients α up to the time-preference discount factor δ for a symmetric and positive-definite weighting matrix W, for large N and T.

Proof. The proof is an application of Hansen and Singleton (1982) with a few modifications.

Let $\zeta_{jt+1} \equiv \mathbb{E}[\lambda_{jt+1}Y_{t+1}|\mathcal{I}_{jt}]$, which is the difference between household j's subjective expectation of $\lambda_{jt+1}Y_{jt+1}$ conditional on time t information $\mathcal{I}_{jt}$. Under rational expectations, differences between expectations and realizations of random variables are mean 0 forecast errors. Therefore $\mathbf{E}[\zeta_{jt+1}] = 0$, where $\mathbf{E}$ denotes unconditional population expectations. Furthermore, let $z_{jt} \in \mathcal{I}_{jt}$ be a vector of observed elements of household j's time t information set with finite second moments and let $h(z_{jt})$ be a measurable function of z. Rational expectations then implies that $\mathbf{E}[\zeta_{jt+1}] \otimes h(z_{jt}) = 0$, where $\otimes$ is the Kronecker product. Substituting $\zeta_{jt+1} + \lambda_{j,t+1}Y_{t+1}$ for $\mathbb{E}_t[\lambda_{jt+1}Y_{jt+1}]$ implies

$$\mathbf{E}[(\delta\alpha\lambda_{j,t+1}Y_{j,t+1} + \zeta_{jt+1} - \lambda_{jt}x_{jkt}) \otimes h(z_{jt})] \tag{3.18}$$

The sample counterpart of is

$$g_{NTK}(a) \equiv \frac{1}{NTK}\sum_j\sum_t\sum_k \delta a(\lambda_{j,t+1}Y_{j,t+1} + \zeta_{jt+1} - \lambda_{jt}x_{jkt} - \psi_{kvt+1}) \otimes h(z_{jt}) \tag{3.19}$$

where $\psi_{jkt+1}\frac{1}{N}\sum_{j=1}^{N}\zeta_{jt+1}$ itself can be thought of as the aggregate shock within each period. Let $\psi_{t+1} \equiv \frac{1}{N}\sum_{j=1}^{N}\zeta_{jt+1} \otimes h(z_{jt})$, which is the sample covariance of unanticipated shocks in each period with the lagged instruments in each period.

Since (by definition) idiosyncratic forecast errors by household are on average equal to the common forecast error, $g_{NTK}(a) \rightarrow \frac{1}{T}\sum_{t=0}^{T}\psi_{t+1}$ as $N \rightarrow \infty$. If shocks are purely idiosyncratic, then the average forecast error is zero in *each* period $\psi_{t+1} \rightarrow 0\ \forall t$ as $N \rightarrow \infty$. However, even there are aggregate shocks within each period, rational expectations still imply that are they are mean-zero. Therefore $\frac{1}{T}\sum_{t=0}^{T}\psi_{t+1} \rightarrow 0$ as $T \rightarrow \infty$. In this case, the GMM estimate of α is

$$\arg\min_a J(a) \equiv g_{NTK}(a)'Wg_{NTK}(a) = 0$$

where W is a symmetric and positive-definite weighting matrix. The efficient choice of W is $\mathbf{E}[g_{NTK}(a)g_{NTK}(a)']^{-1}$. □

In practice, including ψ_{kvt} as a fixed effect directly imposes the weaker restriction that the average forecast error in the (marginal utility-weighted) returns to each input in each village-year is a mean-zero random variable as opposed to 0 itself. This random variable can be thought of as capturing the effects of aggregate unanticipated shocks.

3.3.2 Generalized Cobb-Douglas

Applying similar logic as in Section 3.3.1 to the moment condition 3.15, where x_{jkt} again captures shadow input expenditure $\bar{w}_{kvt}\tau_{jkt}$ and $\mathcal{I}_{jt}$ again denotes time t information sets, yields.

$$\alpha_k\mathbb{E}[\lambda_{j,t+1}|\mathcal{I}_{jt}]\mathbb{E}[Y_{j,t+1}|\mathcal{I}_{jt}] + \beta_k cov(\lambda_{j,t+1,}Y_{j,t+1}|\mathcal{I}_{jt}) - \lambda_{jt}x_{jkt} = 0 \tag{3.20}$$

where $cov_t(\lambda_{j,t+1}Y_{j,t+1}) = \mathbb{E}_t[\lambda_{j,t+1}Y_{j,t+1} - \mathbb{E}_t[\lambda_{j,t+1}]\mathbb{E}_t[Y_{j,t+1}]]$ can be thought of as a measure of how households expect their utility at harvest to depend on the realizations of production shocks, conditional on their time t information.

The main difference between Equation (3.20) and Equation (3.17) is that estimation now requires distinguishing between $\mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}]\mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}]$ and $E[\lambda_{jt+1}Y_{jt+1}|\mathcal{I}_{jt}]$ before applying the logic of Proposition 3. More formally, Differences between the expected and realized products of output and marginal utilities can be expressed as:

$$\lambda_{j,t+1}Y_{t+1} - \mathbb{E}[\lambda_{j,t+1}Y_{t+1}|\mathcal{I}_{jt}] = \zeta_{j,t+1} \tag{3.21}$$

One approach is to projecting realizations of λ_{jt+1} and Y_{jt+1} on to functions of $\mathcal{I}_{jt}$, say $l(\mathcal{I}_{jt})$ and $y(\mathcal{I}_{jt})$, and using the predicted values, $\hat{l}(\mathcal{I}_{jt})$ and $\hat{y}(\mathcal{I}_{jt})$,to substitute for $\mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}]$ and $\mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}]$, respectively. In this case

$$\begin{aligned}
\lambda_{jt+1} &= \mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}] + \pi^L_{jt+1}\\
Y_{jt+1} &= \mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}] + \pi^Y_{jt+1}\\
\lambda_{jt+1} &= \hat{l}(\mathcal{I}_{jt}) = +\upsilon^L_{jt+1}\\
Y_{jt+1} &= \hat{y}(\mathcal{I}_{jt}) + \upsilon^Y_{jt+1}
\end{aligned} \tag{3.22}$$

The household's prediction errors π are mean zero by rational expectations and the estimation errors υ are mean 0 by construction. This means that the difference these two errors $\psi^Y_{jt} \equiv \pi^L_{jt} - \upsilon^L_{jt}$ and $\psi^Y_{jt} \equiv \pi^L Yjt - \upsilon^Y_{jt}$ are each mean zero by linearity of expectations. However substituting the *product* of subjective $\mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}]\mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}]$ for realizations implies:

$$\begin{aligned}
&\mathbb{E}[\delta\alpha_k(\hat{l}(\mathcal{I}_{jt}) + \upsilon^L_{jt+1})(\hat{y}(I) + \upsilon^Y_{jt+1})\\
&+ \delta\beta_k(\lambda_{jt+1}Y_{jt+1} - (\hat{l}(\mathcal{I}_{jt}) + \upsilon^L_{jt+1})(\hat{y}(I) + \upsilon^Y_{jt+1}) - \lambda_{jt}x_{jkt}) \otimes h(\mathcal{I}_{jt})]\\
&= (\alpha_k - \beta_k)(\psi^L_{jt+1}\hat{y}(\mathcal{I}_{jt}) + \psi^Y_{jt+1}\hat{l}(\mathcal{I}_{jt}) + \psi^L_{jt+1}\psi^Y_{jt+1} \otimes h(\mathcal{I}_{jt})) = 0
\end{aligned} \tag{3.23}$$

Assuming $\hat{l}(\mathcal{I}_{jt})$ and $\hat{y}(\mathcal{I}_{jt})$ provide accurate predictions of the true subjective expectations, $\mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}]$ and $\mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}]$, α is identified as the coefficient on the product of these estimated expectations and β is identified as the coefficient on the difference between these expectations and realizations.

Lemma 1. *Assume production is given by Equation* (3.12) *and that households have rational expectations and let $h(z_{jt})$ be a measurable function of variables $z_{jt} \in \mathcal{I}_{jt}$. Further assume that $\hat{l}(\mathcal{I}_{jt})$ and $\hat{y}(\mathcal{I}_{jt})$ are consistent estimators of $\mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}]$ and $\mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}]$, respectively, with associated errors terms υ^L_{jt+1} and υ^Y_{jt+1}. If the differences between these error terms and the household's corresponding forecast errors, then the estimator defined by*

$$\arg\min_a J(a) \equiv g_{NTK}(a,b)'Wg_{NTK}(a,b)$$

where

$$g_{NTK}(a,b) \equiv \frac{1}{NTK}\left(\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\delta(a_k - b_k)\hat{l}(\mathcal{I}_{jt})\hat{y}(\mathcal{I}_{jt}) + \delta b_k(\lambda_{j,t+1}Y_{j,t+1}) - \lambda_{jt}x_{jkt}\right)\otimes h(\mathcal{I}_{jt}),$$

is a consistent estimator of the vectors of coefficients α and β up to the time-preference discount factor δ for a symmetric and positive-definite weighting matrix W, for large N and T.

Proof. The proof is extremely similar to that of Proposition 3, with additional housekeeping required due to the error terms of $\hat{l}(\mathcal{I}_{jt})$ and $\hat{y}(\mathcal{I}_{jt})$.

Let υ^Y_{jt+1} and υ^L_{jt+1} be the error terms associated with the regressions of Y_{jt+1} and λ_{jt+1} on $Y(\mathcal{I}_{jt})$ and $l(\mathcal{I}_{jt})$, respectively

$$\begin{aligned} \lambda_{jt+1} &= \hat{l}(\mathcal{I}_{jt}) + \upsilon^L_{jt+1} \\ Y_{jt+1} &= \hat{y}(\mathcal{I}_{jt}) + \upsilon^Y_{jt+1} \end{aligned} \tag{3.24}$$

$\mathbb{E}[\upsilon^L_{jt+1}] = \mathbb{E}[\upsilon^Y_{jt+1}] = 0$ by construction. Likewise, realized λ_{jt+1} and Y_{jt+1} equal households' time t conditional expectations plus forecast errors π^L_{jt+1} and π^L_{jt+1}, respectively.

$$\begin{aligned} \lambda_{jt+1} &= \mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}] + \pi^L_{jt+1} \\ Y_{jt+1} &= \mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}] + \pi^Y_{jt+1} \end{aligned} \tag{3.25}$$

Under rational expectations $\mathbb{E}[\pi^L_{jt+1}] = \mathbb{E}[\pi^Y_{jt+1}] = 0$. This means that the difference these sets of two errors $\psi^Y_{jt} \equiv \upsilon^L_{jt} - \pi^L_{jt}$ and $\psi^Y_{jt} \equiv \upsilon^Y_{jt} - \pi^Y_{jt}$ are each mean zero by linearity of expectations. Substituting (3.25) into (3.20) yields:

$$\begin{aligned} &\mathbb{E}[(\alpha_k(\lambda_{jt+1} + \pi^L_{jt+1})(y + \pi^Y_{jt+1}) + \beta(\lambda_{jt+1}Y_{jt+1} - (\lambda_{jt+1} + \pi^L_{jt+1})(Y_{jt+1} + \pi^Y_{jt+1})) \\ &- \lambda_{jt+}x_{jkt}) \otimes h(\mathcal{I}_{jt})] = 0 \end{aligned} \tag{3.26}$$

Under what conditions will using $\hat{l}(\mathcal{I}_{jt})$ and $\hat{y}(\mathcal{I}_{jt})$ to proxy for $\mathbb{E}[\lambda_{jt+1}|\mathcal{I}_{jt}]$ and $\mathbb{E}[Y_{jt+1}|\mathcal{I}_{jt}]$ produce consistent estimates of α and β?

$$\begin{aligned} &\mathbb{E}\left[\left(\alpha_k\hat{l}(\mathcal{I}_{jt})\hat{y}(I) + \beta_k(\lambda_{jt+1}Y_{jt+1} - \hat{l}(\mathcal{I}_{jt})(\hat{y}(\mathcal{I}_{jt}))) - \lambda_{jt}x_{jkt}\right) \otimes h(\mathcal{I}_{jt})\right] \\ &= \mathbb{E}\left[(\alpha_k - \beta_k)\left(\lambda_{jt+1} + \upsilon^L_{jt+1})(Y_{jt+1} + \upsilon^Y_{jt+1}) + \beta_k\lambda_{jt+1}Y_{jt+1} - \lambda_{jt}x_{jkt}\right) \otimes h(\mathcal{I}_{jt})\right] \end{aligned} \tag{3.27}$$

Since Equation 3.26 equals 0 by optimality, simple subtraction implies that if

$$\mathbb{E}\left[\left((\alpha_k - \beta_k)(\psi^L_{jt+1}Y_{jt+1} + \psi^Y_{jt+1}\lambda_{jt+1} + \pi^L_{jt+1}\pi^Y_{jt+1} - \upsilon^L_{jt+1}\upsilon^Y_{jt+1})\right) \otimes h(\mathcal{I}_{jt}) = 0\right] \tag{3.28}$$

then Equation 3.27 also equals 0. Essentially this amounts to assuming that the differences between the household's forecast errors and the econometrician's estimation errors are mean independent of the instrument set.

In this case, taking sample averages,

$$g_{NTK}(a,b) \equiv \frac{1}{NTK}\left(\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\delta(a_k - b_k)\hat{l}(\mathcal{I}_{jt})\hat{y}(\mathcal{I}_{jt}) + \delta b_k(\lambda_{j,t+1}Y_{j,t+1}) - \lambda_{jt}x_{jkt}\right) \otimes h(\mathcal{I}_{jt}), \tag{3.29}$$

converges to 0 with large NT under similar conditions as in Proposition 3. Thus, the GMM estimate of α is

$$\arg\min_{a,b} J(a,b) \equiv g_{NTK}(a,b)' W g_{NTK}(a,b) \tag{3.30}$$

where W is the standard optimal weighting matrix. □

A second challenge is separately identifying τ, since households facing common technology and prices will no longer necessarily have the same input ratios. To make progress, I draw on empirical IO methods to estimate product-level production functions with unobserved input prices. In the case of De Loecker et al. (2016), they observe single- and multi-product firms producing the same goods but only observe inputs at the firm level. Their solution is to estimate the production function restricting the sample to single-product firms, and then apply a selection correction to control for unobservable differences between these two types of firms.

The problem in my case is that τ is not necessarily observed. Depending on the nature of input distortions, τ may correspond to the difference between the market price of an input and the price actually paid by a household that purchases this input, or it may be a shadow price that a household faces when rationed. I observe both input expenditures and quantities in the data. I assume that when households hire labor or rent land, any distortion is reflected in the observed price they pay. In this case, τ_{jkt} is included in the x_{jkt} I observe, which is the appropriate variable for (3.29). Thus I restrict the sample to transacted inputs when estimating α and δ, which I then use to recover τs for the households that do not transact these inputs. Note that I do not have to make such assumptions about the nature of τs when production is homothetic, as I can estimate these directly from factor ratios.[10]

3.3.3 Dynamic Production with Sequential Shocks

In the dynamic Cobb-Douglas case, the production function can be estimated consistently in a similar matter to the static case, essentially treating each inputs at different stages as separate inputs. This is because Equation (3.16) hholds for any input at any stage, only that households have different information sets and face different prices at each stage. Rational expectations still implies that households' forecast errors are mean-zero, only that their variance decreases as the season progresses.[11]

[10]This approach relies on some strong assumptions — namely that there is no selection into hiring inputs, that transacted inputs have the same returns as those owned by the household, and that households who purchase positive amounts of inputs do not come up against a ration. To provide support for the first assumption, I can apply the control function approach in De Loecker et al. (2016). I can also restrict the sample to households that use their own inputs in some seasons and purchase inputs in others. To address the second, I observe individual laborer and plot identifiers and can test whether their observed productivity differs when they are used by their respective households or hired. The third assumption is more difficult to test, but I can attempt to restrict the sample to households that appear less likely to face a binding ration.

[11]The proof follows that of Proposition 3 only that inputs are indexed by ks instead of ss. This may nevertheless pose some challenges for efficiency that I do not address in this chapter.

Some practical challenges include dealing with zeroes. Naturally, inputs such as land and seed can be assumed to only be relevant to the first stage. However, inputs such as labor, fertilizer and equipment may be applied at different times throughout the season.

Relatedly, the length of the season, S, is endogenous. Since harvests are determined by crop maturity, it may be inappropriate to assume that households who harvest at stage s would have obtained the same returns at stage $s+1$ as a household that chooses not to harvest at s.

A reasonable solution may be to coarsen stages of production into planting midseason and harvest. Nevertheless, the bias from naively estimating Equation (3.16) without λ_{jt} and λ_{jt+1} for any (non-zero) input is informative about the effects of financial constraints on production. The ratio of these two estimates would be $\frac{\mathbb{E}_t[\lambda_{jt+1}\varphi_{jt+1}]}{\lambda_{jt}}$. While this bias term would equal 1 under complete markets, under imperfect insurance we would expect it to shrink towards 1 over the course of the season as uncertainty over λ_{jt+1} and φ_{jt+1} is sequentially realized.

3.4 Results

I apply these three approaches to the Townsend Thai Data, discussed in detail in Section 1.3.

Table 3.1 reproduces the main results of the two static specifications from Chapter 1, with standard errors from 200 bootstrap replications. Column 1 produces estimates that are in line with the rest of the literature, implying returns to scale of 0.82. In Table A.2.1, I show robustness to various alternative cuts of the sample and specifications of λ and τ. Columns 2 and 3 of Table 3.1 relax the restriction that α_k, the elasticity of the mean of output with respect to input k, is equal to the elasticity of the standard deviation of output with respect to k, β_k. While the αs and βs are similar for most inputs, they are much higher for the three inputs applied only at the beginning of the seasons — land, planting labor, and seed. This is consistent with the idea that uninsured risk causes households to underinvest in production, but this channel weakens as shocks are realized over the course of the season.

The dynamic specification in Equation (3.16) provides a more direct micro-foundation for this phenomenon. While production is homothetic within stages, the covariance between marginal products of each input and the IMRS is larger (in absolute value) at earlier stages of production. This suggests that the bias from naively calibrating coefficients from expenditure shares should be decreasing over time. In addition, if risk causes households to underinvest in production, then the naive calibration should be biased downwards. Intuitively, households have inefficiently low expenditures on each input so observed expenditure is lower than the coefficient.

As discussed in Section 3.3.3, estimating this specification requires some non-trivial assumptions about the nature of observed 0 expenditures, and whether the elasticity of substitution of each input across different stages is truly 1. However, the bias from the naive calibration vs. the estimates using the IMRS is unlikely to be dependent on what stance one takes on these issues. Therefore Table 3.3, which presents estimates of Equation (3.16) with and without weighting input expenditures and output by λ_{jt} and λ_{jt+S}, respectively, should be interpreted as a purely descriptive exercise. Nevertheless, the results confirm the intuition that the "raw" estimates of α_{ks} are downward biased, but decreasingly so over the course of the season.

3.5 Conclusion

This section introduces new methods for estimating the production function for firms that are also consumers — most commonly farm households. The advantage of the estimator we develop is that it uses information from the household's consumer problem to identify how financial frictions affect input demands. This allows us to identify the production function as that which rationalizes constrained optimal input choices and lends itself to a straightforward linear GMM procedure.

We apply these methods to data from farm households in rural Thailand using three specifications of the production function: the workhorse Cobb-Douglas with Hicks-Neutral shocks, a generalized Cobb-Douglas with differentially risky inputs, and a multi-stage Cobb-Douglas with shocks realized sequentially. In addition to producing reasonable estimates, these specifications highlight the importance of accounting for the effects of risk on input demands throughout the season. In particular, the results suggest that the ubiquitous calibration of Cobb-Douglas coefficients from input expenditure shares would produce downward-biased estimates when risk reduces input demands. Moreover, both the generalized static and the dynamic specifications highlight that households face greater uncertainty when applying inputs earlier in the seasons. In the former, we find that land, planting labor and seed disproportionately contribute to the variance of output, which reduces their demand. In the latter case, we show suggestive evidence that the downward bias from naively assuming away risk decreases over the course of the season.

Together, these results show the importance — and usefulness — of accounting for households' dual roles as producers and consumers. As shown in Chapter 1, this has important implications for understanding micro- and macro-level outcomes. Nevertheless, relaxing the functional form and rational expectations assumptions required for such procedures may be a fruitful avenue for future research.

3.6 Tables

Table 3.1: Static production function estimates reproduced from Chapter 1

	α CD	α NH	β NH
Equip.	0.084	0.161	0.144
	(0.005)	(0.013)	(0.048)
Fert.	0.089	0.103	0.110
	(0.002)	(0.004)	(0.016)
Harv. Labor	0.225	0.175	0.181
	(0.006)	(0.028)	(0.077)
Land	0.208	0.219	0.362
	(0.004)	(0.069)	(0.208)
Plant. Labor	0.117	0.120	0.210
	(0.004)	(0.045)	(0.430)
Seed	0.092	0.087	0.130
	(0.002)	(0.005)	(0.028)
Weed. Labor	0.013	0.041	0.050
	(0.001)	(0.017)	(0.029)
J-stat	35.06	36.41	
p-val	0.465	0.132	
γ	0.828	0.906	
s.e.	(0.01)	(0.09)	

This table presents results from the main GMM specifications used to estimate the production function under both the Hicks-neutral Cobb-Douglas specification in the main text and the generalized Cobb-Douglas in Appendix A.2. Column 1 shows the estimates of the Cobb-Douglas coefficients α from (3.17) The second and third columns show estimates of α and β from (3.20), which are the elasticities of the mean and standard deviation of output with respect to each input. All specifications use tambon dummies and lags of λ_{jt} from the 5 months before input k is first applied as instruments. An annual discount factor of $\delta = .95$ is assumed. Results are computed using fertilizer and seed as the reference input for the estimation of τ from (1.25) (only relevant for Column 1), using rice plots only and CFE λs at the farm level. The J-statistic and p-values reported are from a test of the model with the full instrument set against one with only tambon dummies and a single lag of λ_{jt}. γ is the returns to scale parameter implied by the sum of the production coefficients. Standard errors are computed from 234 bootstraps of the full estimation procedure at the household level.

Table 3.2: Static production function estimates reproduced from Chapter 1

	Fert τ	Seed τ	CRRA	Rice only
Equip.	0.084	0.080	0.165	0.094
	(0.005)	(0.004)	(0.005)	(0.005)
Fert.	0.089	0.089	0.100	0.084
	(0.002)	(0.002)	(0.002)	(0.002)
Harv. Labor	0.225	0.255	0.124	0.243
	(0.006)	(0.017)	(0.006)	(0.006)
Land	0.208	0.208	0.190	0.222
	(0.004)	(0.004)	(0.004)	(0.004)
Plant. Labor	0.117	0.125	0.050	0.121
	(0.004)	(0.003)	(0.004)	(0.004)
Seed	0.092	0.092	0.080	0.100
	(0.002)	(0.002)	(0.002)	(0.002)
Weed. Labor	0.013	0.014	0.016	0.019
	(0.001)	(0.001)	(0.001)	(0.001)
J-stat	35.06	45.53	36.64	37.93
p-val	0.465	0.11	0.393	0.337
γ	0.828	0.864	0.724	0.882
s.e.	(0.010)	(0.019)	(0.010)	(0.010)

This table presents results from the main GMM specifications used to estimate the production function, reproduced from Chapter 1. An annual discount factor of $\delta = .95$ is assumed. Columns (1) and (2) present results using fertilizer and seed as the reference input for the estimation of τ from (1.25), using rice plots only and CFE λs at the farm level. Column (3) presents results under CRRA preferences with a coefficient of relative risk aversion equal to 1.5. Column (4) includes all upland crops in the sample. Column (5) presents results using the plot rather than the farm level as the unit of aggregation. All specifications use tambon dummies and lags of λ_{jt} from the 5 months before input k is first applied. The J-statistic and p-values reported are from a test of the model with the full instrument set against one with only tambon dummies and a single lag of λ_{jt}. γ is the returns to scale parameter implied by the sum of the production coefficients. Standard errors are computed from 234 bootstraps of the full estimation procedure at the household level.

Table 3.3: GMM estimates of Dynamic Cobb-Douglas Coefficients

Input	Raw $\hat{\alpha}$	s.e.	IMRS-weighted $\hat{\alpha}$	s.e.
Equip: Stage 0	0.061	(0.000)	0.068	(0.001)
Equip: Stage 1	0.007	(0.001)	0.007	(0.001)
Equip: Stage 2	0.003	(0.001)	0.003	(0.001)
Equip: Stage 3	0.009	(0.001)	0.009	(0.001)
Equip: Stage 4	0.021	(0.001)	0.022	(0.001)
Equip: Stage 5	0.019	(0.002)	0.021	(0.002)
Equip: Stage 6	0.011	(0.002)	0.011	(0.003)
Equip: Stage 7	0.001	(0.005)	0.001	(0.005)
Fert: Stage 0	0.014	(0.001)	0.014	(0.001)
Fert: Stage 1	0.023	(0.001)	0.031	(0.001)
Fert: Stage 2	0.016	(0.001)	0.023	(0.002)
Fert: Stage 3	0.008	(0.002)	0.010	(0.003)
Fert: Stage 4	0.002	(0.004)	0.002	(0.005)
Fert: Stage 5	0.000	(0.003)	0.000	(0.003)
Fert: Stage 6	0.000	(0.002)	0.000	(0.000)
Labor: Stage 0	0.040	(0.001)	0.059	(0.001)
Labor: Stage 1	0.012	(0.000)	0.015	(0.001)
Labor: Stage 2	0.017	(0.001)	0.016	(0.001)
Labor: Stage 3	0.033	(0.002)	0.032	(0.002)
Labor: Stage 4	0.042	(0.002)	0.047	(0.003)
Labor: Stage 5	0.036	(0.003)	0.032	(0.003)
Labor: Stage 6	0.018	(0.004)	0.017	(0.004)
Labor: Stage 7	0.002	(0.006)	0.002	(0.008)
Land: Stage 0	0.227	(0.002)	0.264	(0.003)
Seed: Stage 0	0.087	(0.001)	0.093	(0.001)
Seed: Stage 1	0.001	(0.003)	0.001	(0.006)
Seed: Stage 2	0.000	(0.004)	0.000	(0.002)

This table contains estimates of the dynamic Cobb-Douglas specification in Equation 3.16. The first two columns contain the coefficients and standard errors using raw input expenditures and output values instead of weighting them by λ_{jt} and λ_{jt+S}, respectively. The second two columns use the λs corresponding to each month of production. Both specifications use two-step linear GMM with 5 monthly lags of λ_{jt} and a constant as instruments. Data are disaggregated at the crop-plot level and all inputs are valued at the market price. Coefficients on inputs at each stage are weighted by proportion of crop-plots using non-zero amount of the input at that stage. No standard error adjustments are made.

Bibliography

Abay, K., Abate, G., Barrett, C., and Bernard, T. (2019). Correlated non-classical measurement errors, 'second best' policy inference, and the inverse size-productivity relationship in agriculture. *Journal of Development Economics*, 139:171–184.

Abay, K., Bevis, L., and Barrett, C. (2021). Measurement error mechanisms matter: Agricultural intensification with farmer misperceptions and misreporting. *American Journal of Agricultural Economics*, 103(2):498–522.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Adamopoulos, T., Brandt, L., Chen, C., Restuccia, D., and Wei, X. (2022a). Land security and mobility frictions. Technical report, National Bureau of Economic Research.

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2022b). 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):1667–97.

Adamopoulos, T. and Restuccia, D. (2020). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics*, 12(3):1–39.

Aggarwal, S., Francis, E., and Robinson, J. (2018). Grain today, gain tomorrow: Evidence from a storage experiment with savings clubs in kenya. *Journal of Development Economics*, 134:1–15.

Agness, D. J., Baseler, T., Chassang, S., Dupas, P., and Snowberg, E. (2022). Valuing the time of the self-employed. Technical report, National Bureau of Economic Research.

Aiken, E. L., Bedoya, G., Coville, A., and Blumenstock, J. E. (2020). Targeting development aid with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 310–311.

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

Antle, J. M. (1983). Testing the stochastic structure of production: a flexible moment-based approach. *Journal of Business & Economic Statistics*, 1(3):192–201.

Aragon, F. M., Restuccia, D., and Rud, J. P. (2022). Are small farms really more productive than large farms? *Food Policy*, 106:102168.

Aragón, F. M., Restuccia, D., and Rud, J. P. (2022). Assessing misallocation in agriculture: plots versus farms. Technical report, National Bureau of Economic Research.

Arthi, V., Beegle, K., De Weerdt, J., and Palacios-López, A. (2018). Not your average job: Measuring farm labor in tanzania. *Journal of Development Economics*, 130:160–172.

Asker, J., Collard-Wexler, A., and De Loecker, J. (2014). Dynamic inputs and resource (mis) allocation. *Journal of Political Economy*, 122(5):1013–1063.

Asker, J., Collard-Wexler, A., and De Loecker, J. (2019). (mis) allocation, market power, and global oil extraction. *American Economic Review*, 109(4):1568–1615.

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments*, volume 1, pages 73–140. Elsevier.

Basu, K. and Wong, M. (2015). Evaluating seasonal food storage and credit programs in east indonesia. *Journal of Development Economics*, 115:200–216.

Beegle, K., Carletto, C., and Himelein, K. (2012). Reliability of recall in agricultural data. *Journal of Development Economics*, 98(1):34–41.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

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

Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics*, 46(1):51–84.

Bils, M., Klenow, P. J., and Ruane, C. (2021). Misallocation or mismeasurement? *Journal of Monetary Economics*, 124:S39–S56.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Bohr, C., Mestieri, M., and Robert-Nicoud, F. (2023). Heterothetic cobb douglas. Technical report, Centre for Economy Policy Research.

Bold, T., Kaizzi, K. C., Svensson, J., and Yanagizawa-Drott, D. (2017). Lemon technologies and adoption: measurement, theory and evidence from agricultural markets in uganda. *The Quarterly Journal of Economics*, 132(3):1055–1100.

Breza, E., Kaur, S., and Shamdasani, Y. (2021). Labor rationing. *American Economic Review*, 111(10):3184–3224.

Bryan, G., de Quidt, J., Silva-Vargas, M., Wilkening, T., and Yadav, N. (2022). Market design for land trade: Evidence from uganda and kenya. Technical report, London School of Economics, London.

Burchardi, K. B., Gulesci, S., Lerva, B., and Sulaiman, M. (2019). Moral hazard: Experimental evidence from tenancy contracts. *The Quarterly Journal of Economics*, 134(1):281–347.

Burke, M., Bergquist, L. F., and Miguel, E. (2019). Sell low and buy high: arbitrage and local price effects in kenyan markets. *The Quarterly Journal of Economics*, 134(2):785–842.

Cairncross, J., Morrow, P., Orr, S., and Rachapalli, S. (2023). Robust markups. *Unpublished manuscript*.

Cardell, L. and Michelson, H. (2022). Price risk and small farmer maize storage in sub-saharan africa: New insights into a long-standing puzzle.

Carletto, C., Gourlay, S., and Winters, P. (2015). From guesstimates to gpstimates: Land area measurement and implications for agricultural analysis. *Journal of African Economies*, 24(5):593–628.

Carletto, C., Savastano, S., and Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, 103:254–261.

Carrillo, P., Donaldson, D., Pomeranz, D., and Singhal, M. (2023). Misallocation in firm production: A nonparametric analysis using procurement lotteries. Technical report, National Bureau of Economic Research.

Caunedo, J. and Kala, N. (2021). Mechanizing agriculture. Technical report, National B ureau of Economic Research.

Channa, H., Ricker-Gilbert, J., Feleke, S., and Abdoulaye, T. (2022). Overcoming smallholder farmers' post-harvest constraints through harvest loans and storage technology: Insights from a randomized controlled trial in tanzania. 157:102851.

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 Santaeulalia-Llopis, R. (2017). Land misallocation and productivity. Technical report, National Bureau of Economic Research.

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.

Christian, P. and Dillon, B. (2018). Growing and learning when consumption is seasonal: long-term evidence from tanzania. *Demography*, 55(3):1091–1118.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84(2):445–510.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–2471.

Desiere, S. and Jolliffe, D. (2018). Land productivity and plot size: Is measurement error driving the inverse relationship? *Journal of Development Economics*, 130:84–98.

Di Falco, S. and Chavas, J.-P. (2009). On crop biodiversity, risk exposure, and food security in the highlands of ethiopia. *American Journal of Agricultural Economics*, 91(3):599–611.

Dillon, B., Brummund, P., and Mwabu, G. (2019). Asymmetric non-separation and rural labor markets. *Journal of Development Economics*, 139:78–96.

Diop, B. Z. (2023). Upgrade or migrate: The consequences of input subsidies on household labor allocation.

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):1537–61.

Famine Early Warning Systems Network (2023). Staple food price data nigeria november 2008 - october 2023. Distributed by Famine Early Warning Systems Network.

Felkner, J., Tazhibayeva, K., and Townsend, R. (2012). The impact of climate change on rice yields: the importance of heterogeneity and family networks.

Fink, G., Jack, B. K., and Masiye, F. (2020). Seasonal liquidity, rural labor markets, and agricultural production. *American Economic Review*, 110(11):3351–92.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016.

Goldstein, M. and Udry, C. (2008). The profits of power: Land rights and agricultural investment in ghana. *Journal of Political Economy*, 116(6):981–1022.

Gollin, D. and Udry, C. (2021). Heterogeneity, measurement error, and misallocation: Evidence from african agriculture. *Journal of Political Economy*, 129(1):1–80.

Gordeev, S. and Singh, S. (2023). Misallocation and product choice.

Gottlieb, C. and Grobovšek, J. (2019). Communal land and agricultural productivity. *Journal of Development Economics*, 138:135–152.

Grieco, P. L., Li, S., and Zhang, H. (2016). Production function estimation with unobserved input price dispersion. *International Economic Review*, 57(2):665–690.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280.

Hansen, L. P. and Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, pages 1269–1286.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica: journal of the Econometric Society*, pages 461–465.

Hayashi, F. (2011). *Econometrics*. Princeton University Press.

Heckman, J. and Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, pages 974–987.

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.

Hughes, D. and Majerovitz, J. (2023). Measuring misallocation with experiments.

Jones, M., Kondylis, F., Loeser, J., and Magruder, J. (2022). Factor market failures and the adoption of irrigation in rwanda. *American Economic Review*, 112(7):2316–52.

Just, R. E. and Pope, R. D. (1978). Stochastic specification of production functions and economic implications. *Journal of econometrics*, 7(1):67–86.

Just, R. E. and Pope, R. D. (1979). Production function estimation and related risk considerations. *American Journal of Agricultural Economics*, 61(2):276–284.

Kaboski, J. P. and Townsend, R. M. (2011). A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, 79(5):1357–1406.

Kaboski, J. P. and Townsend, R. M. (2012). The impact of credit on village economies. *American Economic Journal: Applied Economics*, 4(2):98–133.

Karaivanov, A. and Townsend, R. M. (2014). Dynamic financial constraints: Distinguishing mechanism design from exogenously incomplete regimes. *Econometrica*, 82(3):887–959.

Karlan, D., Kutsoati, E., McMillan, M., and Udry, C. (2011). Crop price indemnified loans for farmers: A pilot experiment in rural ghana. *Journal of Risk and Insurance*, 78(1):37–55.

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., Samphantharak, K., Townsend, R., and Cossio, D. A. V. (2020). Propagation and insurance in village networks. Technical report, National Bureau of Economic Research.

Kinnan, C., Samphantharak, K., Townsend, R., and Vera-Cossio, D. (2024). Propagation and insurance in village networks. *American Economic Review*, 114(1):252–284.

Kinnan, C. and Townsend, R. (2012). Kinship and financial networks, formal financial access, and risk reduction. *American Economic Review*, 102(3):289–93.

Kochar, A. (1999). Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural india. *Review of Economics and Statistics*, 81(1):50–61.

LaFave, D. and Thomas, D. (2016). Farms, families, and markets: New evidence on completeness of markets in agricultural settings. *Econometrica*, 84(5):1917–1960.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Ligon, E. (2020). Estimating household welfare from disaggregate expenditures.

Lipsey, R. G. and Lancaster, K. (1956). The general theory of second best. *The Review of Economic Studies*, 24(1):11–32.

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

Manysheva, K. (2021). Land property rights, financial frictions, and resource allocation in developing countries. Technical report, mimeo.

Masten, M. A. and Torgovitsky, A. (2016). Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics*, 98(5):1001–1005.

Maxwell, D., Vaitla, B., and Coates, J. (2014). How do indicators of household food insecurity measure up? an empirical comparison from Ethiopia. *Food policy*, 47:107–116.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of development Economics*, 99(2):210–221.

Merfeld, J. D. and Morduch, J. (2023). Poverty at higher frequency. *KDI School of Pub Policy & Management Paper No. DP23-03*.

Mobarak, A. M. and Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review*, 103(3):375–380.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Omotilewa, O. J., Ricker-Gilbert, J., Ainembabazi, J. H., and Shively, G. E. (2018). Does improved storage technology promote modern input use and food security? evidence from a randomized trial in uganda. *Journal of Development Economics*, 135:176–198.

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

Rosenzweig, M. R. and Udry, C. (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies*, 87(1):343–381.

Rotemberg, M. and White, T. K. (2021). Plant-to-table (s and figures): Processed manufacturing data and measured misallocation.

Samphantharak, K. and Townsend, R. M. (2018). Risk and return in village economies. *American Economic Journal: Microeconomics*, 10(1):1–40.

Sandmo, A. (1971). On the theory of the competitive firm under price uncertainty. *The American Economic Review*, 61(1):65–73.

Shenoy, A. (2017). Market failures and misallocation. *Journal of Development Economics*, 128:65–80.

Shenoy, A. (2021). Estimating the production function under input market frictions. *Review of Economics and Statistics*, 103(4):666–679.

Singh, I., Squire, L., and Strauss, J. (1986). Agricultural household models: Extensions applications and policy.

Sraer, D. and Thesmar, D. (2023). How to use natural experiments to estimate misallocation. *American Economic Review*, 113(4):906–938.

Suri, T. and Udry, C. (2022). Agricultural technology in africa. *Journal of Economic Perspectives*, 36(1):33–56.

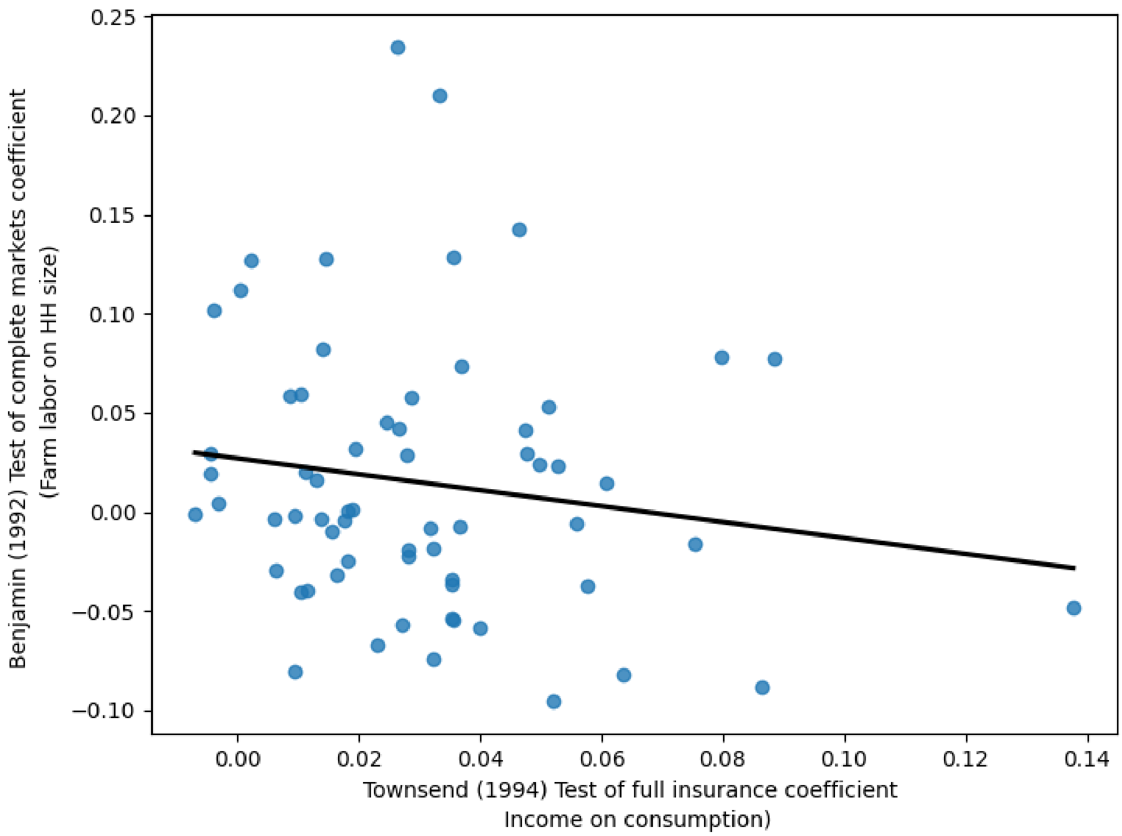
Townsend, R. M. (1994). Risk and insurance in village india. *Econometrica*.

Appendix A

Farm Household Misallocation — Appendix

A.1 Appendix to Section 1.3

Figure A.1.1: Comparison of Test Coefficients Across Villages



This figure contains a scatter plot of the coefficients from the Townsend (1994) and Benjamin (1992) tests, run separately for each village in each 48-month block of the full panel. For the Townsend coefficients on the x-axis, the full monthly sample of households (agricultural and non-agricultural) and monthly measures of total income and consumption are used. Data from the full sample of producers aggregated to the household-year level are used to estimate the Benjamin coefficients on the y-axis.

Table A.1.1: Summary statistics for agricultural households by township

	All	Chachoengsao	Buriram	Lopburi	Sisaket
HH Size	5.564	5.827	5.622	5.03	5.923
	(2.333)	(2.857)	(2.214)	(2.018)	(2.389)
Age Head	56.037	59.792	53.295	53.756	59.597
	(13.259)	(13.515)	(13.275)	(12.387)	(12.745)
Sex Head	0.804	0.757	0.821	0.842	0.769
	(0.397)	(0.429)	(0.383)	(0.365)	(0.422)
Head Primary Educ	0.87	0.951	0.699	0.948	0.938
	(0.337)	(0.215)	(0.459)	(0.223)	(0.241)
Head Secondary Educ	0.1	0.07	0.08	0.121	0.115
	(0.3)	(0.255)	(0.271)	(0.326)	(0.319)
Formal Loan	0.341	0.149	0.432	0.368	0.307
	(0.519)	(0.361)	(0.573)	(0.493)	(0.519)
Any Loan	0.733	0.566	0.716	0.77	0.788
	(0.442)	(0.496)	(0.451)	(0.421)	(0.409)
Years in Ag	10.535	8.798	9.672	10.199	12.507
	(5.514)	(6.438)	(5.4)	(5.081)	(5.026)
N Households	568	71	174	161	162

This table shows summary statistics for agricultural households by township. The table displays means and standard deviations for each variable averaged across household-year observations.

Table A.1.2: Summary statistics for agricultural households by township

	All	Chachoengsao	Buriram	Lopburi	Sisaket
Rice	0.691	0.884	0.966	0.007	0.937
	(0.462)	(0.32)	(0.182)	(0.081)	(0.243)
Maize	0.09	0.009	0.004	0.328	0.001
	(0.286)	(0.097)	(0.059)	(0.47)	(0.03)
Farm size	4.797	6.837	2.293	9.663	2.489
	(7.892)	(5.602)	(1.631)	(13.237)	(1.836)
# plots	3.227	3.078	2.097	4.704	3.026
	(2.787)	(2.424)	(1.28)	(4.069)	(1.944)
Any plot rented	0.16	0.395	0.144	0.267	0.025
	(0.367)	(0.489)	(0.351)	(0.443)	(0.155)
Any labor hired	0.682	0.76	0.781	0.849	0.461
	(0.466)	(0.427)	(0.414)	(0.358)	(0.499)
% labor hired	0.287	0.194	0.284	0.539	0.127
	(0.318)	(0.194)	(0.268)	(0.362)	(0.211)
Any fert.	0.89	0.929	0.92	0.803	0.92
	(0.313)	(0.256)	(0.271)	(0.398)	(0.271)
Any equip.	0.907	0.904	0.939	0.923	0.873
	(0.29)	(0.294)	(0.239)	(0.267)	(0.333)
Profit share	0.228	1.056	0.176	0.039	0.172
	(0.688)	(0.905)	(0.564)	(0.606)	(0.585)
N Households	578	73	177	165	163

This table shows summary statistics for agricultural households by township. The table displays means and standard deviations for each variable averaged across household-year observations.

Table A.1.3: Coefficients of variation in factor and output prices by township

	Chachoengsao	Lopburi	Srisaket
Land rent (per rai)	0.5197	0.4376	0.4552
Wage (hourly)	0.7179	0.5652	0.9919
Planting wage (hourly)	0.6822	0.4718	0.8543
Weeding wage (hourly)	0.5899	0.5312	0.5830
Harvest wage (hourly)	0.6151	0.5480	0.9213
Price of rice seed (per kg)	0.2663	0.2069	0.1096
Price of chem. fert. (per kg)	0.1780	0.1413	0.0946
Power tiller rental (per rai)	0.2749	0.4121	0.6040
Large tractor rental (per rai)	0.2093	0.3669	0.2870
Output price of rice (per kg)	0.0944	0.1148	0.0853

This table shows the coefficients of variation of input and output prices within each township averaged across years. The top panel shows the inputs that I assume are distorted, while the bottom panel shows those that I assume are freely traded. The coefficients of variation are computed at the township-year level after trimming outlier per-unit plot-level expenditures at the upper and lower 2.5% tails and restricting the sample to inputs/outputs with at least 20 observations within a township-year. The three townships shown are those that nearly universally produce rice. The data do not contain the number of days that tractors or power tillers are used — therefore the unit prices I compute are the total expenditure for each type of machinery at the plot level divided by the plot area. Therefore, much of the price dispersion depicted is likely to result from number of days used, machine sizes, or measurement error. Since a more diverse range of crops is grown in Buriram, there is additional heterogeneity due to varieties of seed and fertilizers used for different crops (which I observe). When accounting for this heterogeneity, similar patterns of high price dispersion in land and labor but low price dispersion for traded inputs and outputs emerge.

Table A.1.4: Diagnostic Tests for Market Failures

	log Consumption Val.	log Labor Hrs.	
	(1)	(2)	(3)
log Income	0.0547***		
	(0.0037)		
HH Size		0.0211*	
		(0.0112)	
Male adults			0.0257
			(0.0258)
Female Adults			0.0269
			(0.0253)
Male children			0.0121
			(0.0217)
Female Children			0.0165
			(0.0210)
Household FE	Yes		
Village-month FE	Yes		
Village-year FE		Yes	Yes
F-stat			11.61**
p-val			0.0205
Observations	83,384	5,689	5,689

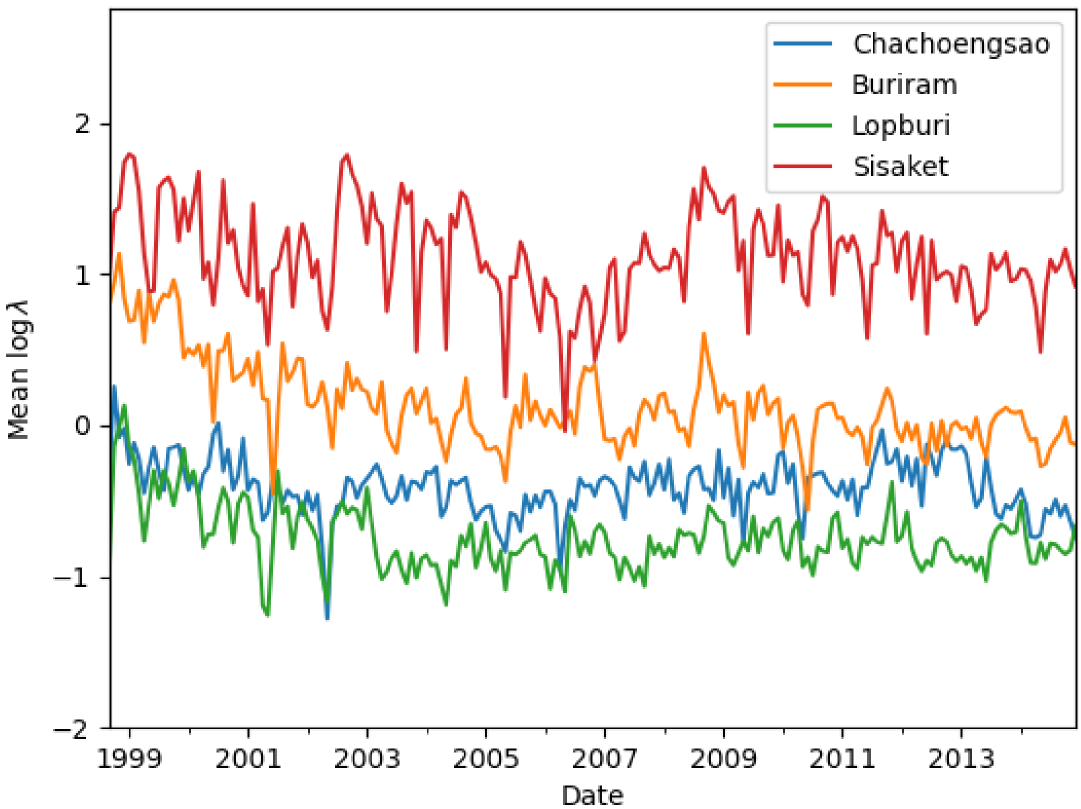
This table presents the results for two of the canonical tests of incomplete markets in the literature. Column (1) shows the results of a regression of (log) consumption on income with household and village-month fixed effects as in Townsend (1994). The full monthly sample of households (agricultural and non-agricultural) and monthly measures of total income and consumption are used. Column (2) shows the results of the Benjamin (1992) test of separability, which regresses (log) household labor hours on household characteristics, controlling for farm size. For simplicity, household size is the only measured included. Data from the full sample of producers aggregated to the household-year level are used.

A.2 Appendix to Section 1.4

I provide additional details on the CFE demand system of Ligon (2020) used for the main results. CFE demands satisfy the condition that $\log p_i c_i = a_i(p) + b_i(z) - \beta_i \log \lambda$, where expenditures on good i depend on functions of the price vector p and household characteristics z and are log-linear in λ. β_i is the eponymous constant elasticity, which imposes that the elasticity of expenditure on good i with respect to the marginal utility of expenditure (as opposed to total expenditure) is a constant. This allows for highly non-linear Engel curves and an unrestricted rank of the demand system. Ligon (2020) shows that CFE is the only globally regular demand system in which identical households with different budgets' demands for goods differ only through a common aggregator. The paper also derives an estimator for the MUE that uses disaggregated consumption data. The key assumption for estimation is that observed 0 expenditures can essentially be treated as a missing data problem. While this may appear strong, the assumption essentially requires that welfare can be inferred from observed expenditures and the Frisch elasticities of those goods. See Ligon (2020) for more detail.

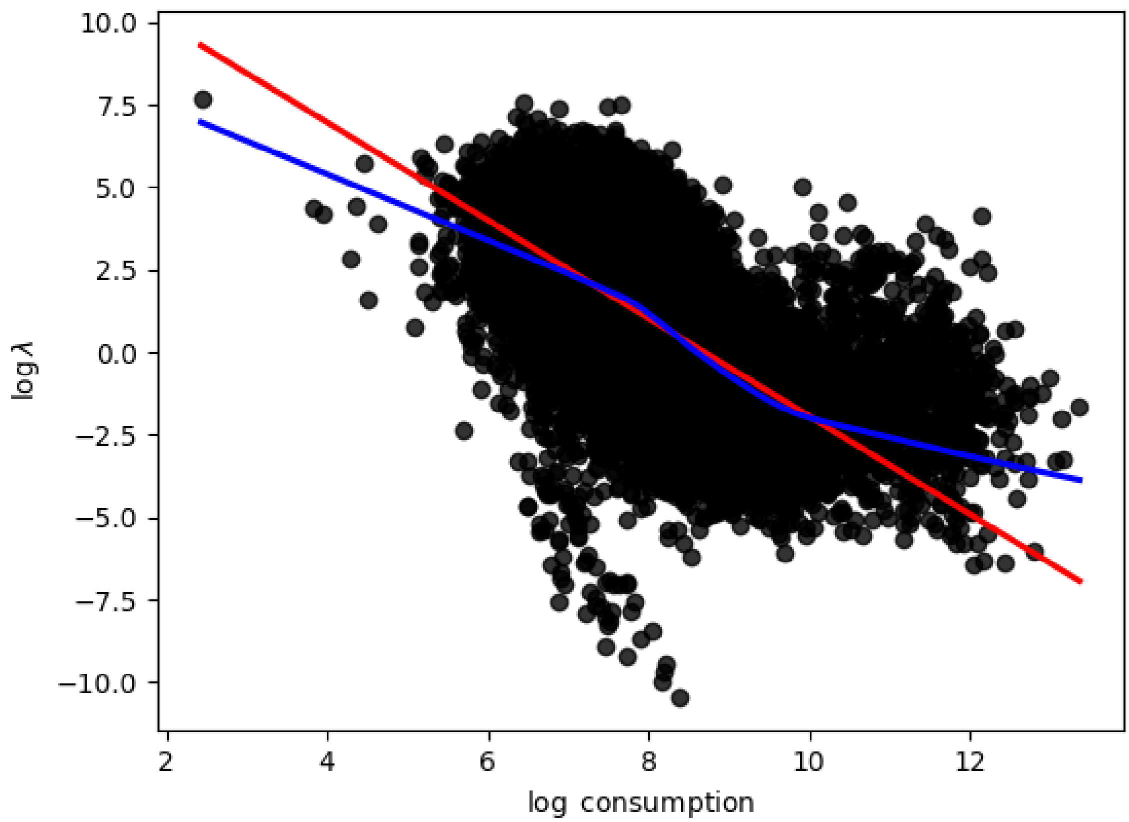
What matters for the model in Section 1.4 is the curvature of utility. The elasticity of λ with respect to total consumption is (minus) the coefficient of relative risk aversion. If this elasticity is constant, then CFE reduces to the nested CRRA case. The slope of Figure A.2.2 shows that while there does appear to be some curvature in relative risk aversion, there is not a huge difference from CRRA. Accordingly, the results in Table A.2.1 are similar across specifications.

Figure A.2.1: Time series plots of $\log \lambda$ by tambon



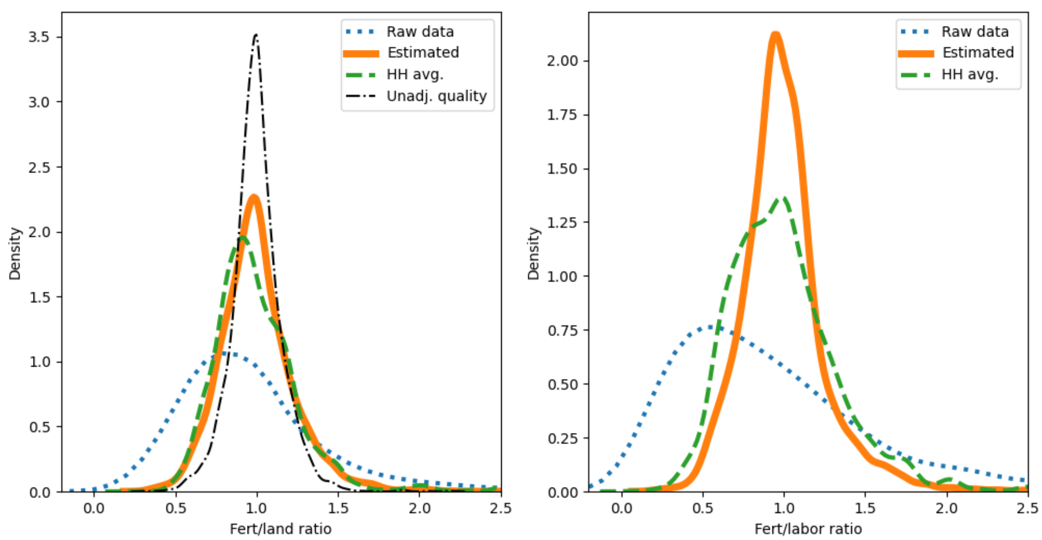
This figure plots the time series of the mean $\log \lambda$, estimated from the CFE demand system of Ligon (2020) over the 196-month sample period in each tambon (township).

Figure A.2.2: Relative risk aversion under CFE demands



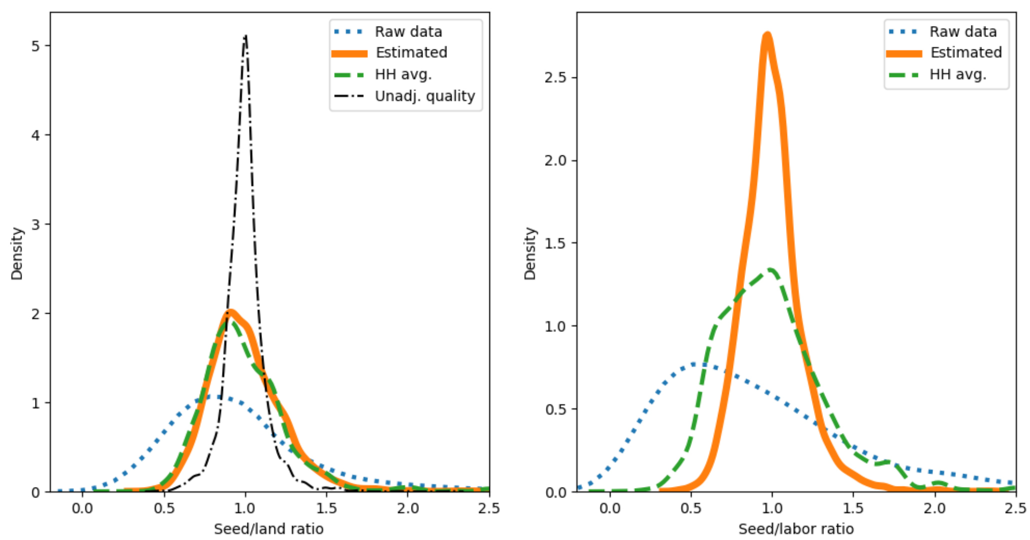
The figure plots estimated log λs against the log of consumption after partialing out month fixed-effects. The slope of the graph at any point is (minus) the coefficient of relative risk aversion under von Neumann-Morgenstern preferences. The red line is the estimate of relative risk aversion when imposing CRRA preferences, while the blue line is a Lowess fit of the relative risk aversion implied by CFE demands.

Figure A.2.3: Kernel density estimation of τ by input (fertilizer)



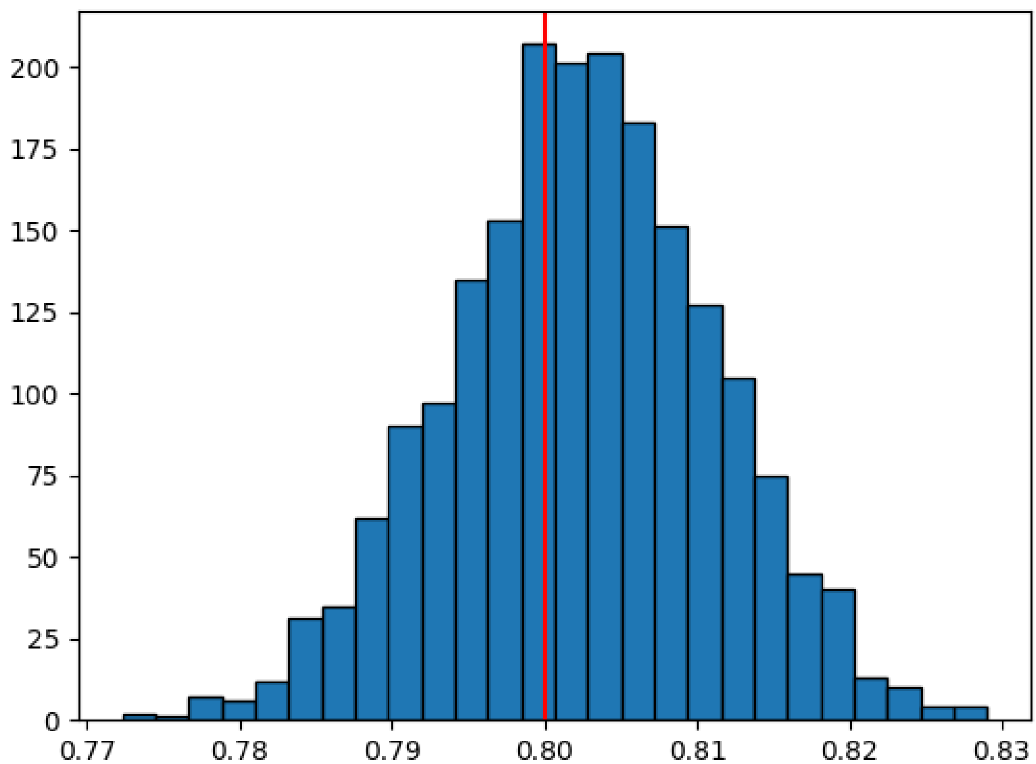
This figure plots kernel density estimates of τ for land and each labor input using fertilizer as the normalizing input. The blue lines show the density of raw input ratios relative to the township-year mean, the green lines show the density of household average input ratio relative to the township means and the orange lines show the estimated τs following (1.25). The black line in the left panel shows the density for τ_{LAND} when not adjusting for land quality. An Epanechnikov kernel is used.

Figure A.2.4: Kernel density estimation of τ by input (seed)



This figure plots kernel density estimates of τ for land and total labor input using seed as the normalizing input. The blue lines show the density of raw input ratios relative to the township-year mean, the green lines show the density of household average input ratio relative to the township means and the orange lines show the estimated τs following (1.25). The black line in the left panel shows the density for τ_{LAND} when not adjusting for land quality. An Epanechnikov kernel is used.

Figure A.2.5: Monte Carlo Simulations of Estimation with Aggregate Shocks



This figure presents a histogram of the regression coefficients of 1,000 Monte Carlo simulations of the GMM estimator. I develop a simulated data-generating process under a single-input production function with $\alpha = 0.8$ and CRRA preferences with $\theta = 1.5$. I simulate an $N = 1,000$ by $T = 16$ year panel. For each t, I draw $\phi_{jt} \sim \mathcal{N}(\mu_t, \sigma)$ where the μ_t's themselves are drawn from a $\mathcal{N}(0, \sigma)$ distribution. In the main simulations, I choose $\sigma = 0.4$ (to match the variance of the residuals in Section 1.4.3.1). I then apply the GMM estimator to each simulated dataset. The distribution of coefficients is centered near the true value of 0.8 (indicated by the red line in the figure) with a mean of 0.8024 and standard error of 0.0087.

Table A.2.1: GMM results

	Fert τ	Seed τ	CRRA	Rice only
Equip.	0.084	0.080	0.165	0.094
	(0.005)	(0.004)	(0.005)	(0.005)
Fert.	0.089	0.089	0.100	0.084
	(0.002)	(0.002)	(0.002)	(0.002)
Harv. Labor	0.225	0.255	0.124	0.243
	(0.006)	(0.017)	(0.006)	(0.006)
Land	0.208	0.208	0.190	0.222
	(0.004)	(0.004)	(0.004)	(0.004)
Plant. Labor	0.117	0.125	0.050	0.121
	(0.004)	(0.003)	(0.004)	(0.004)
Seed	0.092	0.092	0.080	0.100
	(0.002)	(0.002)	(0.002)	(0.002)
Weed. Labor	0.013	0.014	0.016	0.019
	(0.001)	(0.001)	(0.001)	(0.001)
J-stat	35.06	45.53	36.64	37.93
p-val	0.465	0.11	0.393	0.337
γ	0.828	0.864	0.724	0.882
s.e.	(0.010)	(0.019)	(0.010)	(0.010)

This table presents results from the main GMM specifications used to estimate the production function. An annual discount factor of $\delta = .95$ is assumed. Columns (1) and (2) present results using fertilizer and seed as the reference input for the estimation of τ from (1.25), using rice plots only and CFE λs at the farm level. Column (3) presents results under CRRA preferences with a coefficient of relative risk aversion equal to 1.5. Column (4) includes all upland crops in the sample. Column (5) presents results using the plot rather than the farm level as the unit of aggregation. All specifications use tambon dummies and lags of λ_{jt} from the 5 months before input k is first applied. The J-statistic and p-values reported are from a test of the model with the full instrument set against one with only tambon dummies and a single lag of λ_{jt}. γ is the returns to scale parameter implied by the sum of the production coefficients. Standard errors are computed from 234 bootstraps of the full estimation procedure at the household level.

Table A.2.2: Correlation between time-varying financial wedges and financial participation

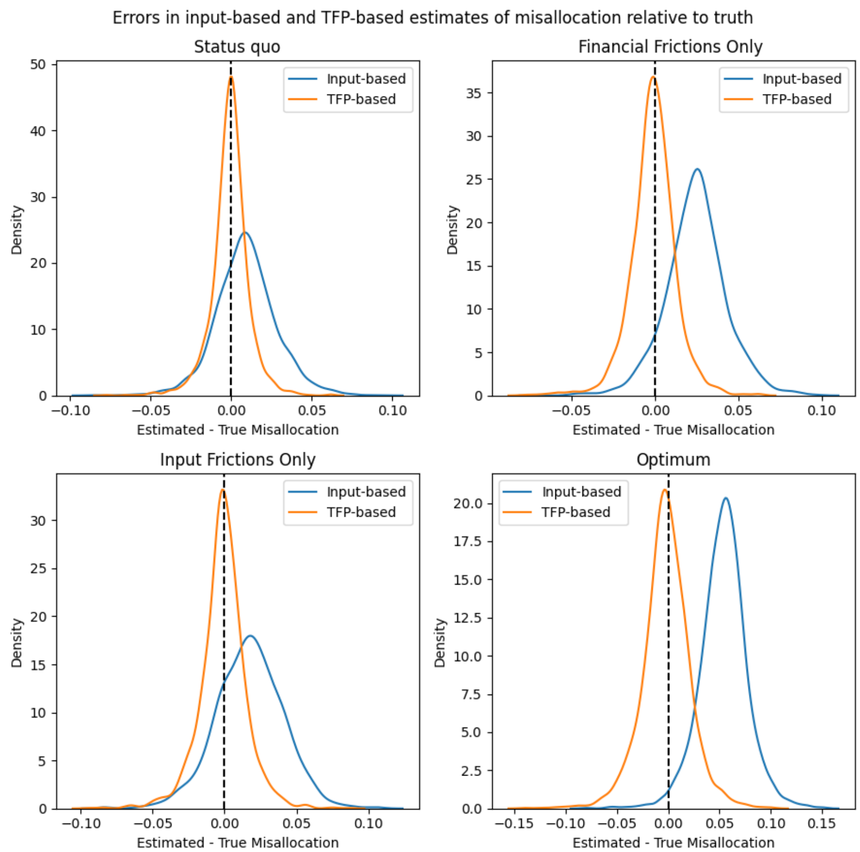
	Dependent variable:					
	Savings bal.	Debt bal	Credit bal.	Gifts made	Gifts rec'd.	Net gifts
	(1)	(2)	(3)	(4)	(5)	(6)
$\log \Lambda$	0.33***	0.11*	0.12	0.29***	0.16***	
	(0.09)	(0.07)	(0.20)	(0.10)	(0.05)	
Λ						−10,425.33
						(9,330.07)
Vil.+ Time FE	Yes	Yes	Yes	Yes	Yes	
Obs.	5,442	4,951	561	4,966	5,808	5,830
Adj. R^2	0.17	0.20	0.19	0.03	0.27	0.02
Note:						*p<0.1; **p<0.05; ***p<0.01

Table A.2.3: Production shocks' effect on interhousehold transfers

	Dependent variable:				
	Gifts made	log gifts made	Gifts recieved	log gifts recieved	Net gifts
	(1)	(2)	(3)	(4)	(5)
Shock (s.d)	1,504.34		−326.86***		−1,831.20
	(1,883.43)		(88.54)		(1,871.36)
log shock		0.003		−0.10***	
		(0.06)		(0.02)	
Vil. + Time FE	Yes	Yes	Yes	Yes	Yes
Obs.	5,830	4,966	5,830	5,808	5,830
Adj. R^2	0.02	0.14	0.09	0.27	0.02
Note:					*p<0.1; **p<0.05; ***p<0.01

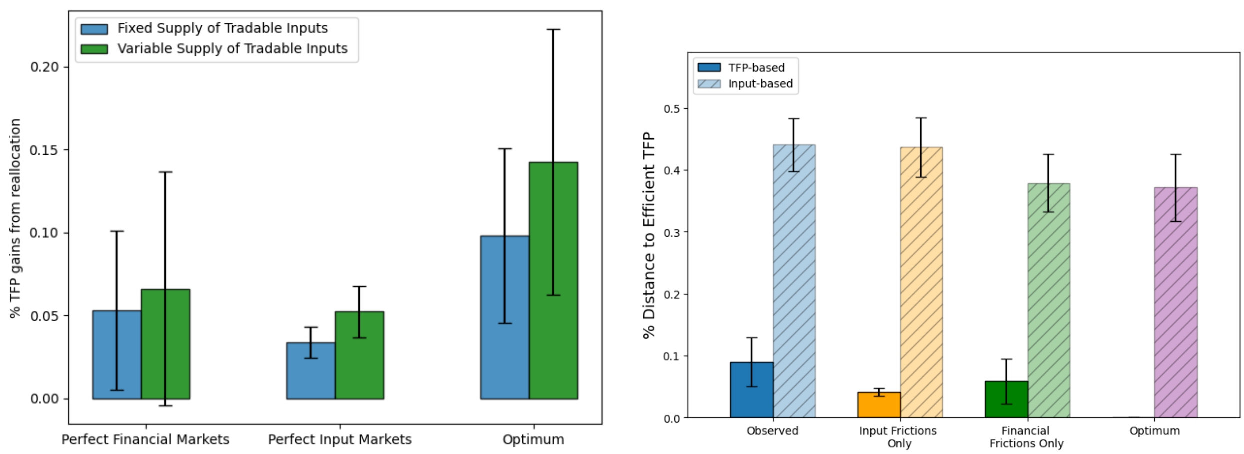
A.3 Appendix to Section 1.5

Figure A.3.1: Comparison of errors from input- and TFP-based estimates



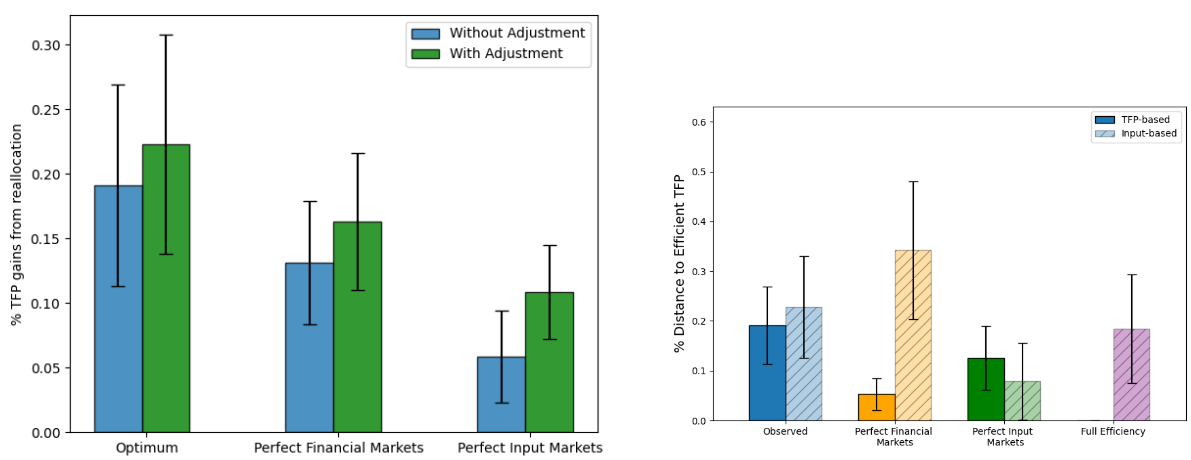
These figures show the distribution of estimates of misallocation from 2,000 Monte Carlo simulations of the model. The model consists of 500 households observed for 16 years using a two-input production function with $\gamma = 0.7$. TFP, Λ and τ are drawn from a multivariate lognormal distortion with $\mu = 0$ and positively correlated distortions. Measurement error in inputs and production shocks are drawn from log-normal distributions with $\sigma = .5$. The blue lines show the densities of estimates using the input-based measure from (1.26) and the orange lines show the densities using the TFP-based measure from (1.18). In all four scenarios, the TFP-based estimates have negligible bias while the input-based estimates are biased upwards and have larger variance. Similar patterns hold for other distributions of shocks and distortions.

Figure A.3.2: Plot-level estimates of misallocation



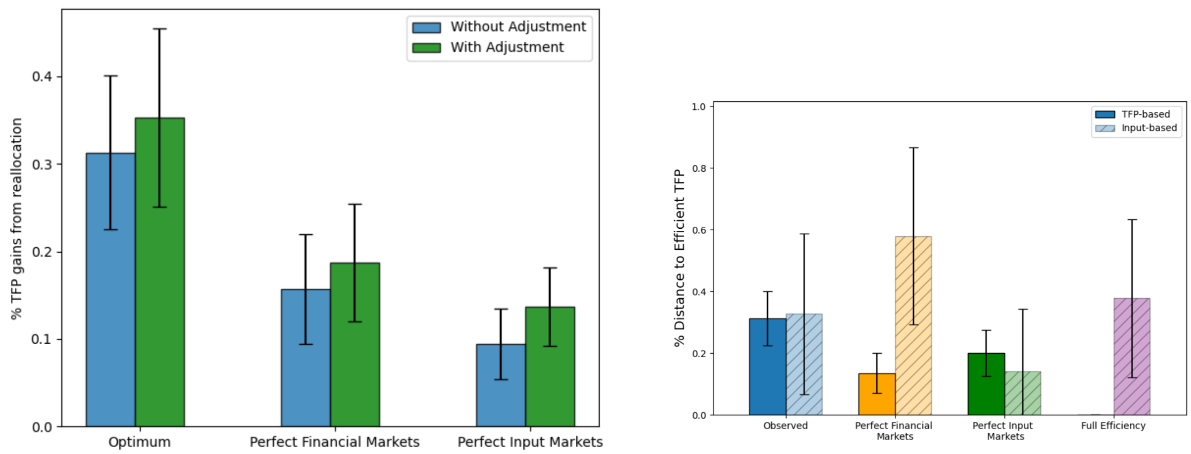
The figure shows results from the main counterfactuals in Figure 1.3 and Figure 1.5 in panels (a) and (b), respectively, using plot-level rather than farm-level data. Results are computed using CFE demands, and fertilizer as the normalizing input for τs, restricting the sample to rice plots. The measure of misallocation is the difference between aggregate TFP under a given allocation and the efficient one, expressed as a percent of modeled TFP. The solid bars compute these using the TFP-based measure of misallocation, using (1.17). The shaded bars are calculated by taking raw input observed in the data and augmenting them by the estimated τ and Λ, where relevant. 95% confidence intervals from 200 bootstrap replications are plotted.

Figure A.3.3: Main results with CRRA preferences



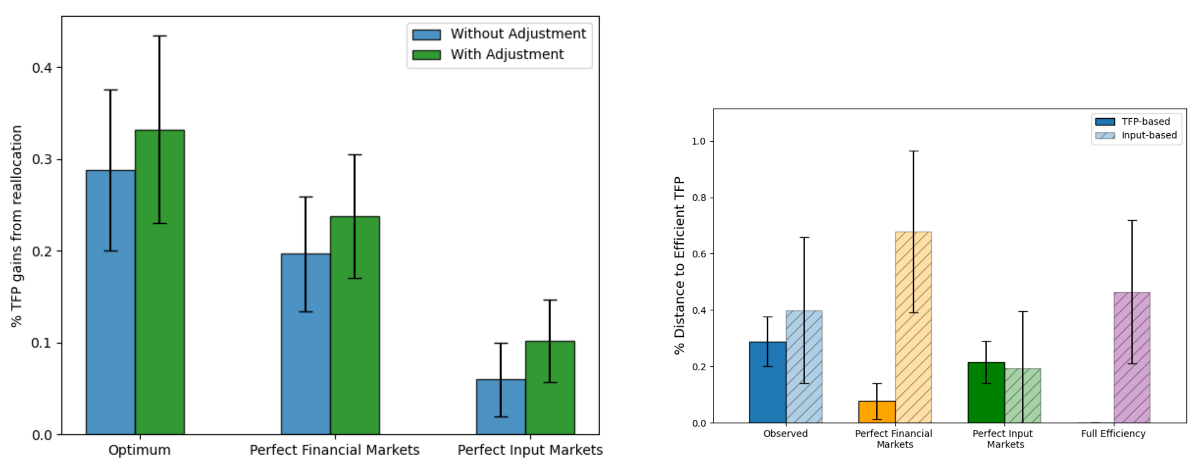
The figure shows results from the main counterfactuals in Figure 1.3 and Figure 1.5 in panels (a) and (b). Results are computed using CFE demands, fertilizer as the normalizing input for τs, only rice plots, and aggregating to the farm level. The measure of misallocation is the difference between aggregate TFP under a given allocation and the efficient one, expressed as a percent of modeled TFP. The solid bars compute these using the TFP-based measure of misallocation, using (1.17). The shaded bars are calculated by taking raw input observed in the data and augmenting them by the estimated τ and Λ, where relevant. 95% confidence intervals from 200 bootstrap replications are plotted.

Figure A.3.4: Results using only rice



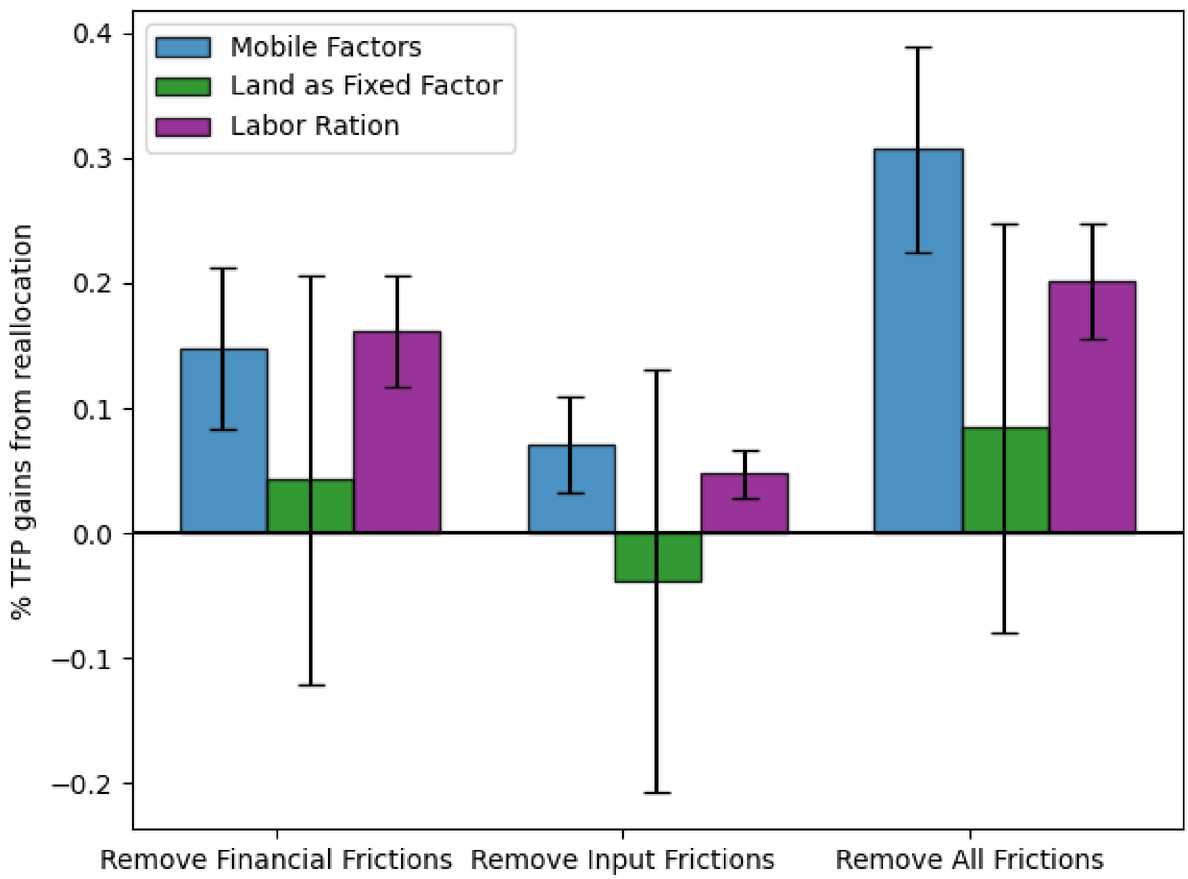
The figure shows results from the main counterfactuals in Figure 1.3 and Figure 1.5 in panels (a) and (b). Results are computed using CFE demands, fertilizer as the normalizing input for τs, restricting the sample to rice plots, and aggregating to the farm level. The measure of misallocation is the difference between aggregate TFP under a given allocation and the efficient one, expressed as a percent of modeled TFP. The solid bars compute these using the TFP-based measure of misallocation, using (1.17). The shaded bars are calculated by taking raw input observed in the data and augmenting them by the estimated τ and Λ, where relevant. 95% confidence intervals from 200 bootstrap replications are plotted.

Figure A.3.5: Main results using seed as the reference input



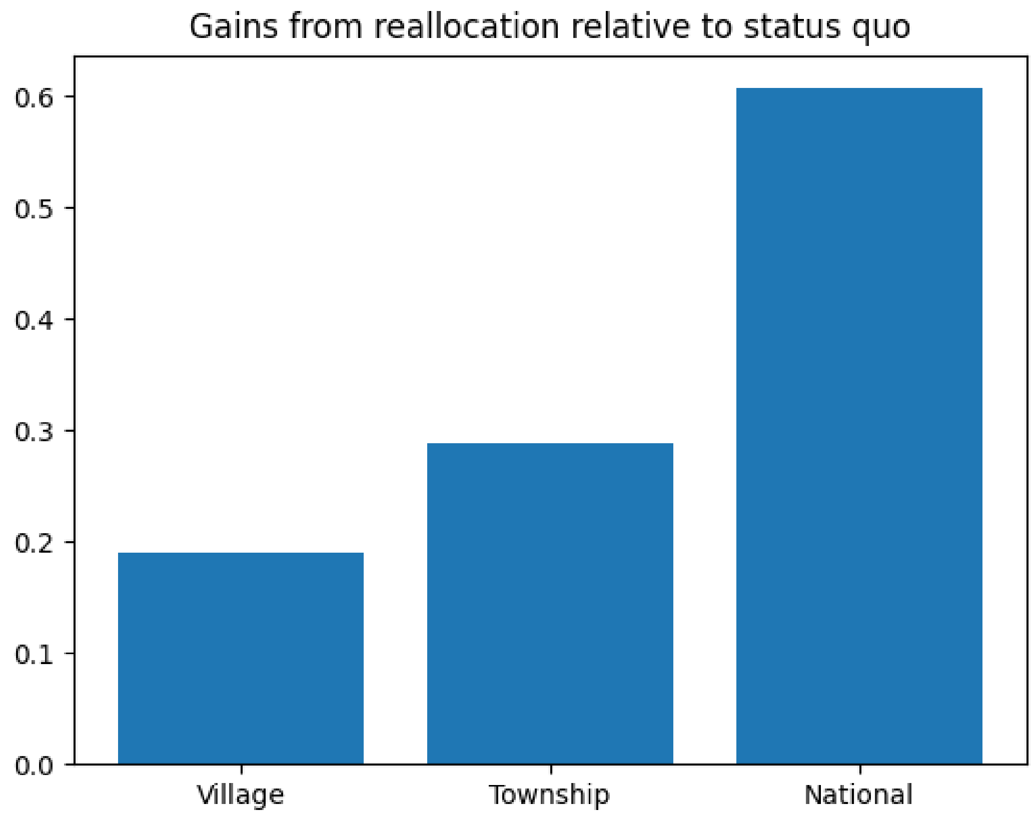
The figure shows results from the main counterfactuals in Figure 1.3 and Figure 1.5 in panels (a) and (b). Results are computed using CFE demands, seed as the normalizing input for τs, only rice plots, and aggregating to the farm level. The measure of misallocation is the difference between aggregate TFP under a given allocation and the efficient one, expressed as a percent of modeled TFP. The solid bars compute these using the TFP-based measure of misallocation, using (1.17). The shaded bars are calculated by taking raw input observed in the data and augmenting them by the estimated τ and Λ, where relevant. 95% confidence intervals from 200 bootstrap replications are plotted.

Figure A.3.6: Counterfactual gains from reallocation under input rations



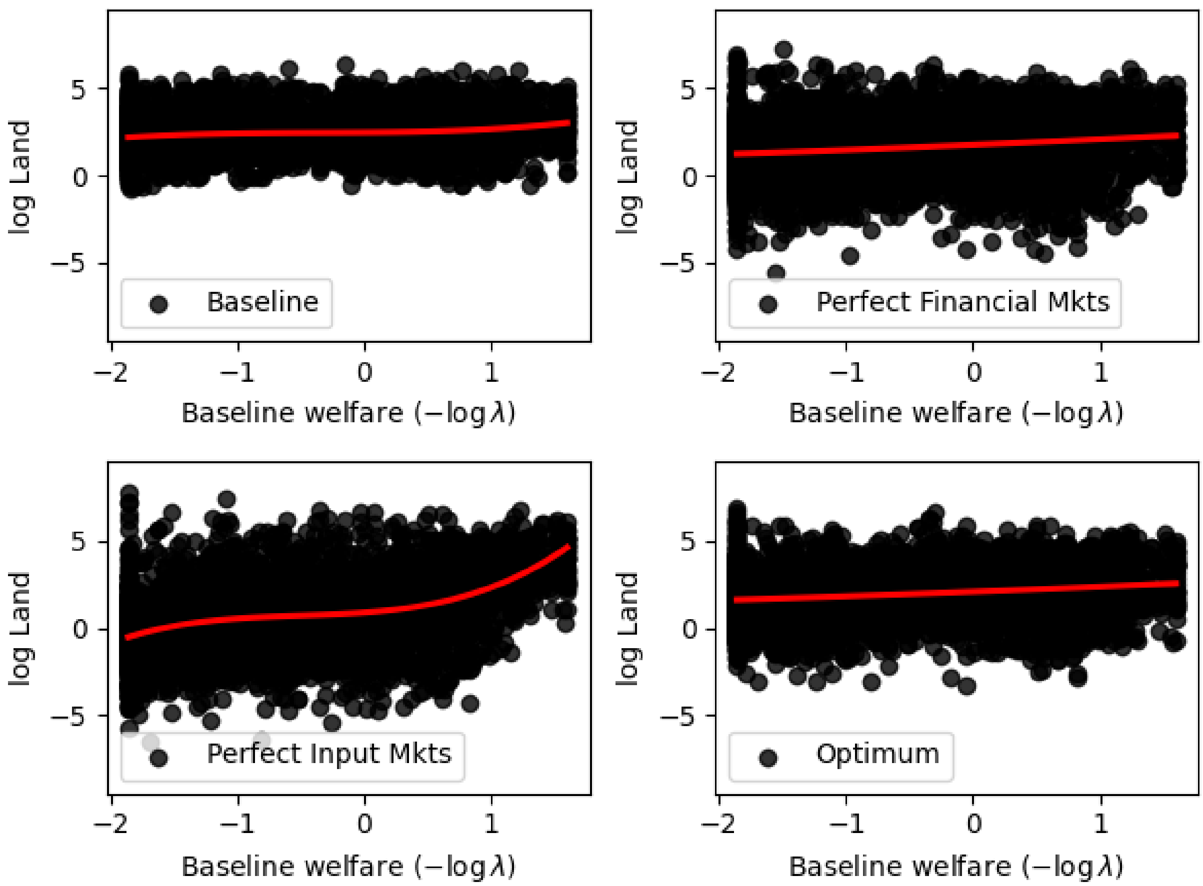
The figure shows the gains from reallocation under the main counterfactuals depending on which factors are mobile within townships. The blue (left) bars reproduce the baseline scenario, in which all factors are mobile and can be reallocated. The green (middle) bars show results holding land fixed at observed levels in all three scenarios, even when relaxing other input frictions. The purple (right) bars show results assuming households with $\tau < 1$ for each labor input face a binding downward ration. Results are computed using CFE demands, fertilizer as the normalizing input for τs, all crops, and aggregating to the farm level. 95% confidence intervals from 200 bootstrap replications are plotted.

Figure A.3.7: Potential gains from full reallocation



This figure shows the total gains from the efficient allocation as a percent of status quo aggregate TFP when aggregating at the village, township, and national levels.

Figure A.3.8: Land Distribution



This figure shows the distribution of land under the baseline and main counterfactuals as a function of baseline welfare, which is the negative of the log MUE. The scatter plots are shown with a lowess fit. This is shown for the closed economy case, using fertilizer as the normalizing input, CFE demands, and restricting the sample to rice crops at the farm level.

Appendix B

The Welfare Effects of Postharvest Loans Under Price Risk — Appendix

B.1 Appendix to Section 2.3

B.1.1 Algorithm for imputing stocks and flows

It is very clear that the raw data is not always an accurate accounting of farmers' inventory. Common inconsistencies include:

- Households will report selling or consuming goods that we don't see in previous stocks, purchases, or harvests
- Households will report purchasing or harvesting goods that we never see them dispose of, but at endline they report having less than the acquired amounts in stock
- In later rounds, households' answers to whether they have consumed a good from their own stock in the inventory module do not match their answers to whether they have acquired the good from their own production or purchased it.

To resolve these issues, we apply the following algorithm, which allows us to compute a number of different indicators of stocks, with higher indices corresponding to more imputation. We proceed until baseline stocks plus the cumulative sum of flows matches endline stocks, with the restriction that stocks never fall below 0 at any point in the period. We use responses from the consumption module as an additional source of information about whether households had positive stocks at a given period.

1. Save reported stocks (which are only available in waves 1-3 and at baseline). Assign dummy value of 1 if positive
2. Then compute stocks as the cumulative sum of report flows (purchases + harvest - sales - consumption). Replace 0 dummy with 2 if this is positive. Add reported initial stocks (positive for only a few HH) and replace 0 dummy with 3 if this is positive
3. Replace 0 dummy with 4 if the household reports consuming crop from own stock in consecutive waves.
4. For waves 1-3, impute consumption as the difference between the change in stocks between the previous and current period net of flows (harvest + purchases for storage - sales). Update flows accordingly (updating stocks is trivial)
5. If the household has harvested a positive amount between waves 1-3, add the residual between reported stocks and harvests to the harvest. Assign dummy value of 6 when this turns stocks positive. Update flows and stocks accordingly. Note that we need to do this iteratively (because if a household meets this condition for multiple waves but has activity in the interim, one iteration won't account for this.)

6. Still have some people who sell/consume crops they've harvested after period 3 that are unaccounted for. For example, one household claims to harvest and sell 400kg of beans in period 2 but then sells another 200 in period 4 and 100 in period 5. It seems more likely that these came from the harvest rather than purchases. Likewise for consumption that's reported as being own-produced. We are going to want to attribute these to the latest feasible harvest (between waves 1-3). Assign dummy value of 7 when this turns stocks positive. Update stocks and flows again.

7. Where there is unaccounted-for sales/consumption in later waves for crops that a household never harvests, attribute these to purchases. Assign dummy value of 8

8. When households have excess positive stocks in excess of the endline stocks they report, attribute the maximum amount that will not cause subsequent stocks to go negative at any point to current period consumption. This again needs to be done iteratively. Assign dummy value of 9 and update accordingly.

9. When households don't report consuming these crops from own production, attribute to sales and assign dummy value of 10.

10. Assume any remaining discrepancy is stock carried over to endline.

B.2 Appendix to Section 2.4

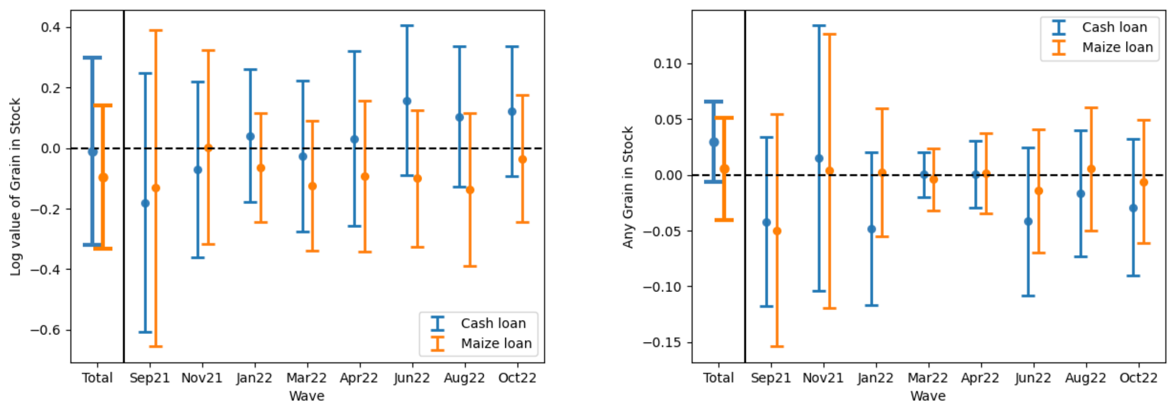
This table shows baseline summary statistics by treatment arm and reports p-values for the test of equality of means across arms.

Table B.2.1: Treatment Balance

Variable	Control	Cash	Kind	p-value Cash-Control	p-value Kind-Control
Age Head	39.24	38.84	37.96	0.73	0.22
Rainy Seed Exp	11.43	10.09	11.26	0.53	0.93
Rainy Fert Exp	62.26	58.14	67.58	0.43	0.35
Rainy Chem Exp	6.76	4.87	5.78	0.04	0.30
Rainy Mech Exp	4.15	4.27	4.64	0.90	0.62
Rainy Labor Exp	18.40	15.93	17.95	0.43	0.88
Any Dry	0.01	0.00	0.01	0.23	0.94
Dry Labor Exp	0.33	0.30	0.22	0.88	0.58
Chickens	7.68	8.43	7.42	0.43	0.77
Goats	5.31	5.63	5.25	0.45	0.89
Sheep	3.52	3.97	3.80	0.25	0.45
Cows	2.35	2.75	2.69	0.34	0.42
Donkeys	0.03	0.07	0.03	0.27	0.77
Large Item Exp	23.97	21.38	22.05	0.43	0.55
Biz Exp	17.37	39.46	23.05	0.01	0.17
Biz Rev	10.46	17.13	13.34	0.01	0.09
Borrow Amt	10.54	19.98	10.33	0.34	0.81
Land Own Ha	5.37	6.20	5.25	0.13	0.79
Land In Ha	2.51	2.11	3.17	0.18	0.62
Land Out Ha	0.28	0.43	0.22	0.26	0.46
Rainy Area Ha	5.53	5.59	6.14	0.89	0.42
Dry Area Ha	0.07	0.03	0.10	0.13	0.53
Largest Rainy Beans	0.03	0.02	0.02	0.44	0.16
Largest Rainy Maize	0.43	0.55	0.50	0.00	0.08
Largest Rainy Millet	0.27	0.19	0.19	0.02	0.02
Educ Head Comppri	0.17	0.11	0.12	0.05	0.10
Educ Head Compsec	0.12	0.07	0.10	0.05	0.54
Educ Head Somepri	0.10	0.10	0.10	0.94	0.94
Educ Head Tert	0.01	0.06	0.07	0.00	0.00
Female Head	0.08	0.06	0.06	0.39	0.26
Any Biz	0.18	0.19	0.20	0.80	0.29
Any Borrow	0.19	0.27	0.25	0.24	0.29
Loglambda	0.63	0.70	0.66	0.63	0.72
Males 0-5	0.46	0.45	0.42	0.82	0.54
Females 0-5	0.41	0.34	0.42	0.28	0.85
Males 5-10	0.67	0.66	0.60	0.83	0.34
Females 5-10	0.54	0.55	0.47	0.92	0.25
Males 10-15	0.47	0.44	0.36	0.67	0.07
Females 10-15	0.42	0.44	0.38	0.73	0.52
Males 15-20	0.45	0.38	0.44	0.24	0.79
Females 15-20	0.35	0.35	0.36	0.97	0.89
Males 20-30	0.61	0.56	0.63	0.48	0.76
Females 20-30	0.52	0.51	0.48	0.86	0.48
Males 30-50	0.56	0.63	0.60	0.13	0.33
Females 30-50	0.37	0.40	0.30	0.62	0.15
Males 50-60	0.19	0.13	0.14	0.08	0.11
Females 50-60	0.06	0.06	0.05	0.86	0.88
Males 60-100	0.12	0.13	0.10	0.67	0.46
Females 60-100	0.06	0.04	0.04	0.50	0.44

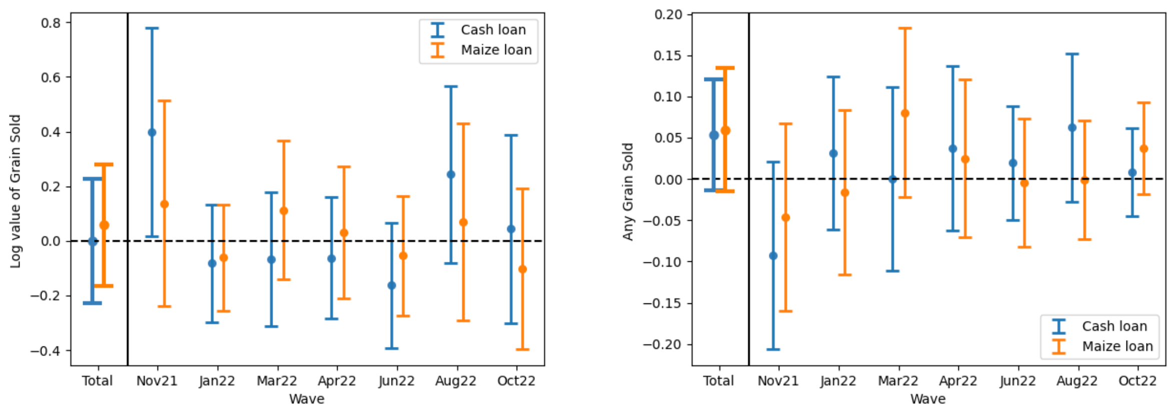
B.3 Appendix to Section 2.5

Figure B.3.1: Extensive and intensive margin effects on stocks



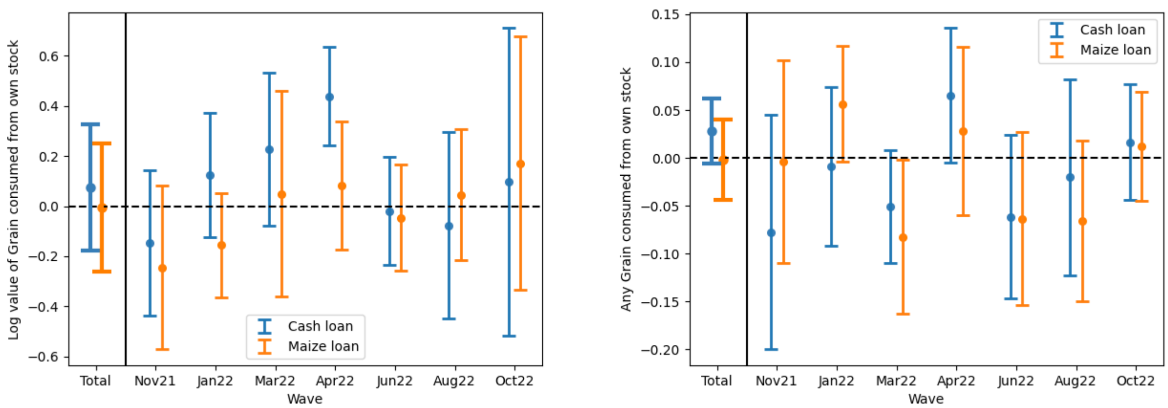
This figure supplements Figure 2.6, showing the extensive and intensive margin effects of the cash and maize treatment on overall grain stocks. The left panel uses the log of stock values (conditional on being positive) as the dependent variable while the right panel uses a dummy variable for whether a household had positive stocks in the given wave

Figure B.3.2: Extensive and intensive margin effects on sales



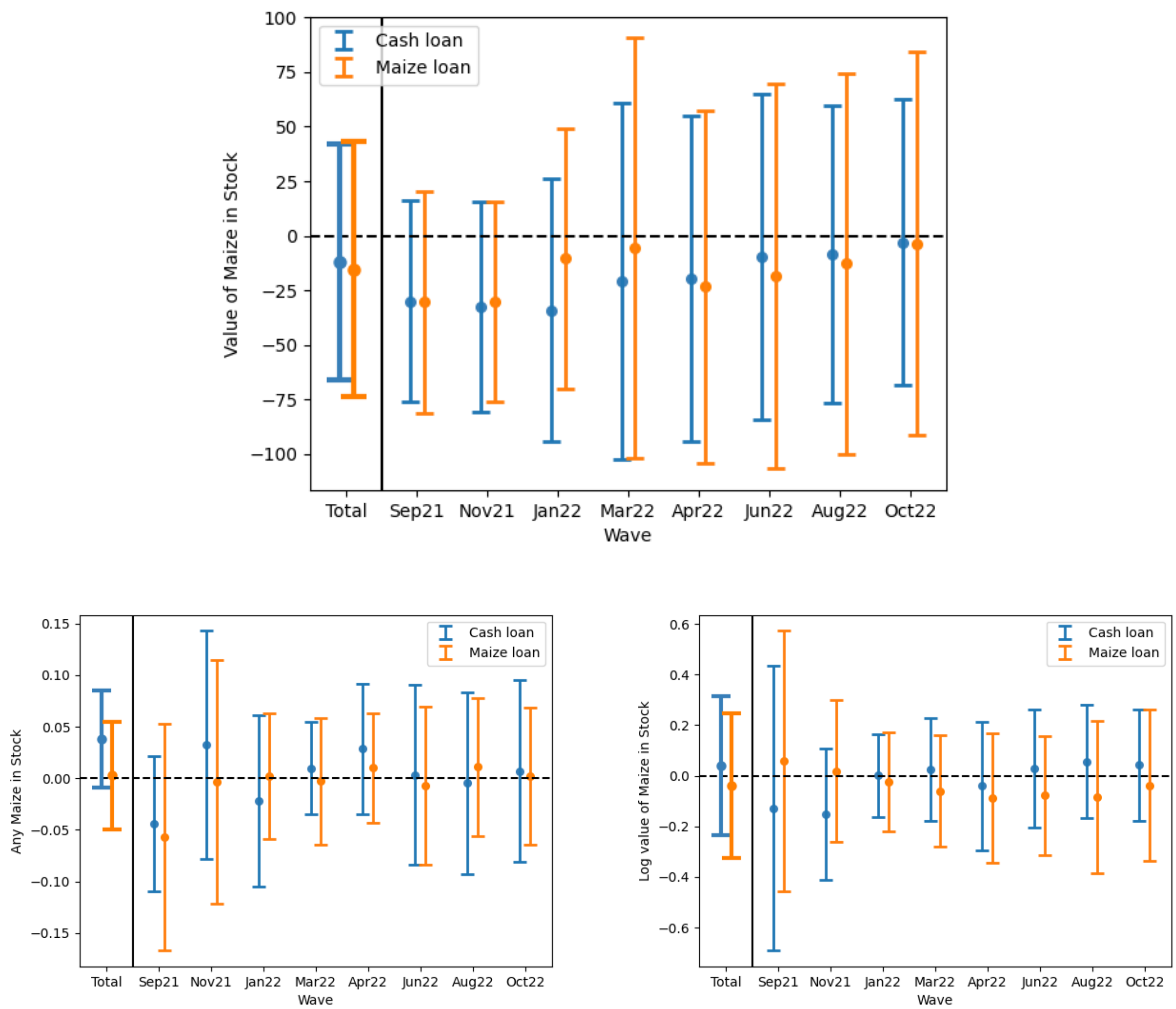
This figure supplements Figure **??**, showing the extensive and intensive margin effects of the cash and maize treatment on overall grain sales. The left panel uses the log of sales values (conditional on being positive) as the dependent variable while the right panel uses a dummy variable for whether a household had positive sales in the given wave.

Figure B.3.3: Extensive and intensive on own stock consumption



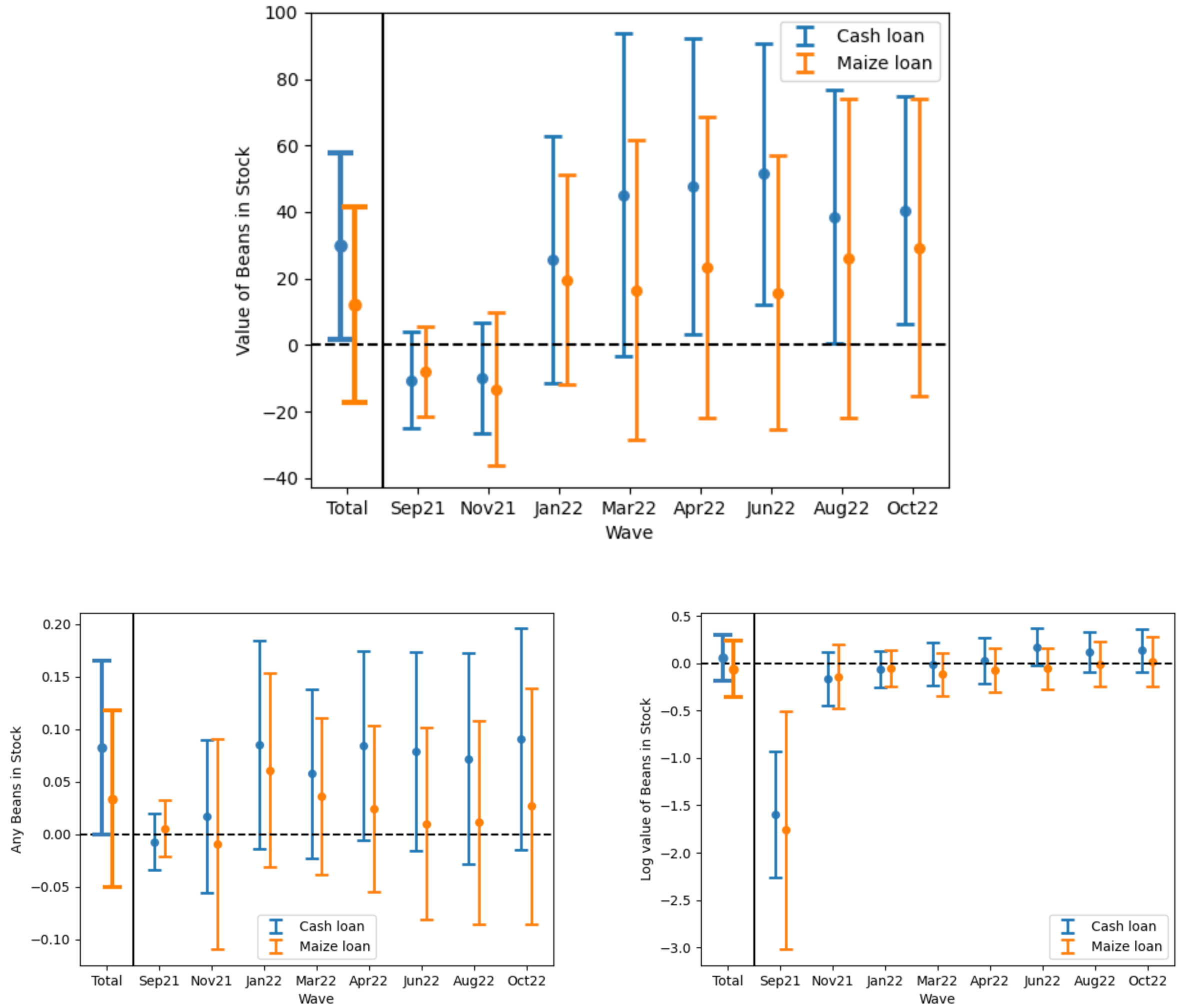
This figure supplements Figure 2.9, showing the extensive and intensive margin effects of the cash and maize treatment on the value of grain consumed from own stocks. The left panel uses the log consumption value (conditional on being positive) as the dependent variable while the right panel uses a dummy variable for whether a household had positive consumption. This figure supplements Figure 2.6, showing the effect

Figure B.3.4: Effects on maize stocks



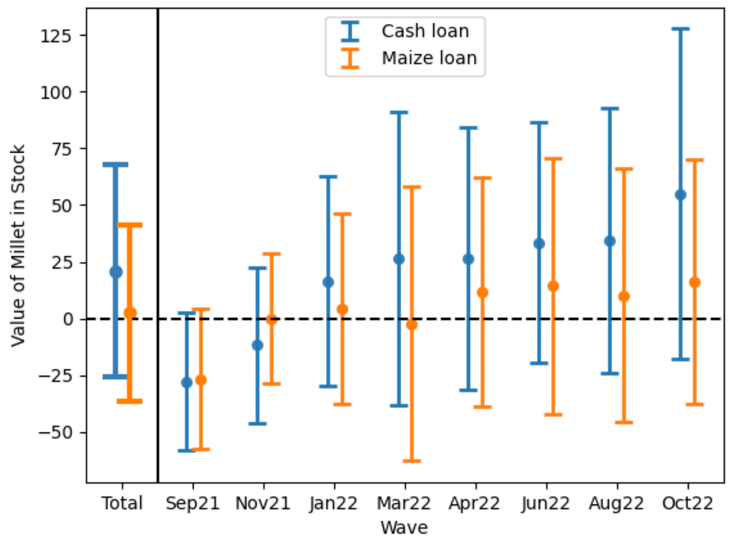
of both treatment arms on the value of maize stocks. The top panel shows the effects on the level value of stocks, measured in thousands of Naira. The left panel uses the log of stock values (conditional on being positive) as the dependent variable while the right panel uses a dummy variable for whether a household had positive stocks in the given wave.

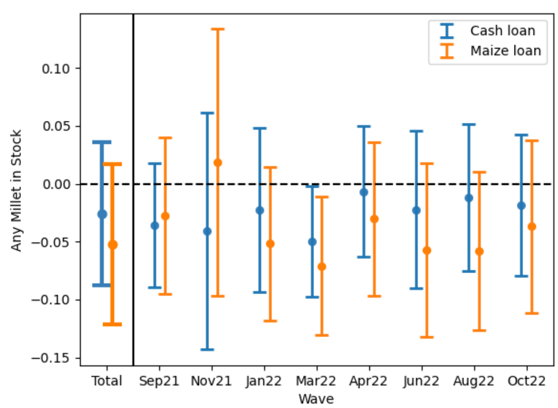
Figure B.3.5: Effects on bean stocks

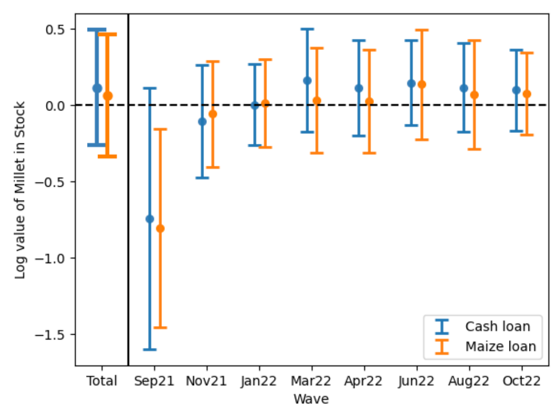


This figure supplements Figure 2.6, showing the effect of both treatment arms on the value of bean stocks. The top panel shows the effects on the level value of stocks, measured in thousands of Naira. The left panel uses the log of stock values (conditional on being positive) as the dependent variable while the right panel uses a dummy variable for whether a household had positive stocks in the given wave.

Figure B.3.6: Effects on Millet stocks







This figure supplements Figure 2.6, showing the effect of both treatment arms on the value of millet stocks. The top panel shows the effects on the level value of stocks, measured in thousands of Naira. The left panel uses the log of stock values (conditional on being positive) as the dependent variable while the right panel uses a dummy variable for whether a household had positive stocks in the given wave.

Table B.3.1: Agricultural outcomes: Ancova

	Ag. exp . ('000 Naira)	Planted area (ha)	Dry season ag. exp.	Dry season area (ha)
Cash Loan	16.007	0.180	0.119	0.001
	(19.709)	(0.577)	(0.138)	(0.008)
Maize Loan	-1.043	-0.442	-0.152	-0.011
	(21.529)	(0.564)	(0.841)	(0.019)
Baseline Value	0.610***	0.132	-0.010	0.001
	(0.163)	(0.090)	(0.010)	(0.001)
Fixed Effects	Strat	Strat	Strat	Strat
R^2	0.306	0.207	0.070	0.072
Control mean	144.122	5.482	0.894	0.024
Observations	829	829	829	829

This figure replicates Table 2.5 controlling for the baseline values of each outcome.

Table B.3.2: Agricultural outcomes: Double post LASSO

	Ag. exp . ('000 Naira)	Planted area (ha)	Dry season ag. exp.	Dry season area (ha)	Harvest value
Cash Loan	13.511	0.124	0.126	0.001	1.637E+05
	(20.817)	(0.608)	(0.139)	(0.008)	(1.529E+05)
Maize Loan	-0.044	-0.374	-0.148	-0.012	7.287E+04
	(22.242)	(0.576)	(0.838)	(0.019)	(1.126E+05)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.247	0.170	0.070	0.071	0.137
Control mean	146.166	5.559	0.907	0.025	796.083
Observations	829	829	829	829	808

This figure replicates Table 2.5 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.3: Consumption outcomes: Ancova

	Log exp.	Log non-. storable exp.	log λ	Elicited IMRS	Seasonal hunger index
Cash Loan	0.011	0.063	0.004	-0.035	0.011
	(0.069)	(0.046)	(0.027)	(0.022)	(0.022)
Maize Loan	0.019	0.021	-0.003	-0.041*	0.023
	(0.072)	(0.055)	(0.026)	(0.020)	(0.037)
Baseline Value	0.060***	0.104***	0.092***	0.030	0.918
	(0.011)	(0.021)	(0.014)	(0.020)	(0.565)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.061	0.061	0.278	0.138	0.007
Control mean	7.750	6.619	0.324	1.691	-0.010
Observations	5474	5474	5474	5474	5474

This figure replicates **??** controlling for the baseline values of each outcome.

Table B.3.4: Consumption outcomes: Double post LASSO

	Log exp.	Log non-. storable exp.	log λ	Elicited IMRS	Seasonal hunger index
Cash Loan	0.002	0.031	0.016	-0.038	0.013
	(0.071)	(0.041)	(0.024)	(0.024)	(0.020)
Maize Loan	0.013	0.029	0.004	-0.044*	0.023
	(0.078)	(0.066)	(0.026)	(0.022)	(0.034)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.063	0.068	0.281	0.145	0.007
Control mean	7.750	6.619	0.324	1.691	-0.010
Observations	5809	5809	5809	5809	5809

This figure replicates Table 2.4 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.5: Business and financial outcomes: Ancova

	(Semi-) Durable Exp.	Biz. Exp.	Amount Borrowed	Any Biz Activity	Any Borrowing
Cash Loan	1.748	-0.898	-0.318	0.003	0.019
	(1.719)	(1.003)	(0.423)	(0.014)	(0.014)
Maize Loan	0.678	-1.192	-0.395	-0.011	0.012
	(1.364)	(0.923)	(0.353)	(0.016)	(0.011)
Baseline Value	0.132***	0.213***	0.158***	0.119***	0.036**
	(0.024)	(0.032)	(0.007)	(0.018)	(0.015)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.052	0.196	0.154	0.078	0.053
Control mean	13.810	5.396	2.540	0.121	0.099
Observations	5474	5474	5474	5474	5474

This figure replicates Table 2.6 controlling for the baseline values of each outcome.

Table B.3.6: Business and financial outcomes: Double post LASSO

	(Semi-) Durable Exp.	Biz. Exp.	Amount Borrowed	Any Biz Activity	Any Borrowing
Cash Loan	1.777	-0.595	-0.005	0.000	0.024*
	(1.540)	(1.203)	(0.336)	(0.016)	(0.014)
Maize Loan	0.928	-1.503	-0.322	-0.013	0.015
	(1.284)	(0.895)	(0.296)	(0.019)	(0.011)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.065	0.206	0.145	0.079	0.053
Control mean	13.810	5.396	2.540	0.121	0.099
Observations	5809	5809	5809	5809	5809

This figure replicates Table 2.6 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.7: Land outcomes: Ancova

	Land owned (ha)	Land rented out (ha)	Land rented in (ha)
Cash Loan	2.776	0.078	-0.344
	(2.602)	(0.107)	(0.423)
Maize Loan	-0.554	0.004	-0.563
	(0.562)	(0.106)	(0.414)
Baseline Value	0.182	0.546**	0.011
	(0.112)	(0.235)	(0.011)
Fixed Effects	Strat	Strat	Strat
R^2	0.062	0.340	0.332
Control mean	5.785	0.187	2.395
Observations	829	829	829

This figure replicates Table 2.8 controlling for the baseline values of each outcome.

Table B.3.8: Land outcomes: Double post LASSO

	Land owned (ha)	Land rented out (ha)	Land rented in (ha)
Cash Loan	2.485	0.218	-0.290
	(1.956)	(0.213)	(0.369)
Maize Loan	-0.140	0.023	-0.516
	(0.638)	(0.103)	(0.363)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.142	0.111	0.387
Control mean	1.639	0.059	0.733
Observations	829	829	829

This figure replicates Table 2.8 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.9: Livestock outcomes: Ancova

	Cows	Goats	Sheep	Chickens	Donkeys
Cash Loan	0.503	0.924*	0.404	1.890**	-0.007
	(0.377)	(0.506)	(0.399)	(0.760)	(0.012)
Maize Loan	-0.349	-0.387	-0.532	0.528	-0.005
	(0.251)	(0.392)	(0.374)	(0.674)	(0.010)
Baseline Value	0.302***	0.377***	0.485***	0.268***	0.089
	(0.079)	(0.069)	(0.073)	(0.081)	(0.069)
Fixed Effects	Strat	Strat	Strat	Strat	Strat
R^2	0.277	0.340	0.336	0.288	0.107
Control mean	1.528	3.843	2.472	2.262	0.010
Observations	829	829	829	829	829

This figure replicates Table 2.7 controlling for the baseline values of each outcome.

Table B.3.10: Livestock outcomes: Double post LASSO

	Cows	Goats	Sheep	Chickens
Cash Loan	0.564*	0.997*	0.352	1.906**
	(0.312)	(0.510)	(0.374)	(0.772)
Maize Loan	-0.269	-0.303	-0.534	0.572
	(0.239)	(0.407)	(0.396)	(0.638)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.321	0.369	0.364	0.297
Control mean	1.550	3.894	2.496	2.294
Observations	829	829	829	829

This figure replicates Table 2.7 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.11: Sales: Double post LASSO

	Sales Value ('000s Naira)					
	Total	Maize	Millet	Beans	Guinea Corn	Rice
Cash Loan	1.926	3.721	13.180	-4.399	-9.924	5.007
	(36.504)	(12.038)	(28.645)	(19.446)	(6.879)	(4.732)
Maize Loan	9.364	5.578	5.915	-6.203	-4.780	2.427
	(32.866)	(9.156)	(13.169)	(17.911)	(7.020)	(7.124)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.472	0.302	0.265	0.495	0.064	0.142
Control mean	323.698	54.631	71.143	147.828	14.436	15.473
Observations	922	922	922	922	922	922

This figure replicates Table 2.1 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.12: Consumption Double post LASSO

	Value of Stock Consumed ('000s Naira)					
	Total	Maize	Millet	Beans	Guinea Corn	Rice
Cash Loan	83.775	0.237	36.181*	45.471***	0.569	-16.635
	(63.283)	(33.219)	(19.105)	(14.161)	(11.357)	(14.817)
Maize Loan	15.814	0.542	8.233	31.023	-2.480	-20.008
	(72.105)	(32.729)	(18.846)	(19.612)	(6.450)	(20.420)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.495	0.441	0.508	0.334	0.175	0.470
Control mean	673.146	305.729	133.176	94.621	46.285	86.109
Observations	922	922	922	922	922	922

This figure replicates Table 2.2 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.13: Harvest: Double post LASSO

	Value of Harvest ('000s Naira)					
	Total	Maize	Millet	Beans	Guinea Corn	Rice
Cash Loan	79.920	4.855	44.531	40.084*	3.117	-0.201
	(84.860)	(37.413)	(37.390)	(23.634)	(6.591)	(11.827)
Maize Loan	47.582	17.818	12.253	20.255	4.498	1.082
	(90.404)	(36.529)	(28.850)	(22.991)	(6.720)	(19.914)
Fixed Effects	Stratum	Stratum	Stratum	Stratum	Stratum	Stratum
R^2	0.532	0.430	0.522	0.541	0.285	0.377
Control mean	949.036	320.456	255.228	228.499	43.415	74.778
Observations	922	922	922	922	922	922

This figure replicates Table 2.3 controlling for covariates selected by the double post LASSO method of Belloni et al. (2012)

Table B.3.14: Heterogeneity by Gender: Grain Flows

	Sales Value	Harvest Value	Value of Stock Consumed
Cash Loan	13.334	103.192	90.476
	(53.364)	(125.587)	(84.089)
Cash Loan×Female Head	59.408	90.665	-7.259
	(83.697)	(256.772)	(159.497)
Maize Loan	3.572	50.379	21.192
	(38.572)	(112.634)	(92.345)
Maize Loan×Female Head	-55.731	-126.199	-23.386
	(62.323)	(148.111)	(140.835)
Female Head	-56.859	-173.678	-114.275
	(41.646)	(106.185)	(81.534)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.330	0.311	0.304
Control mean	323.698	949.036	673.146
Observations	930	930	930

This figure replicates outcomes for grain flows interacting each treatment with a dummy for whether the household head is female

Table B.3.15: Heterogeneity by Gender: Consumption

	Log exp.	Log non-. storable exp.	log λ	Elicited IMRS	Seasonal hunger index
Cash Loan	-0.045	-0.017	0.017	-0.032	0.010
	(0.087)	(0.053)	(0.027)	(0.020)	(0.020)
Cash Loan×Female Head	-0.056	-0.090	0.013	-0.008	-0.027
	(0.165)	(0.221)	(0.094)	(0.057)	(0.029)
Maize Loan	-0.063	-0.042	-0.003	-0.033	0.019
	(0.077)	(0.057)	(0.029)	(0.020)	(0.033)
Maize Loan×Female Head	0.743***	0.733***	0.097	-0.087*	0.006
	(0.211)	(0.149)	(0.067)	(0.051)	(0.040)
Female Head	-0.686***	-0.645***	0.064**	0.102**	-0.020
	(0.113)	(0.093)	(0.026)	(0.047)	(0.029)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.050	0.135	0.244	0.194	0.006
Control mean	7.750	6.619	0.324	1.691	-0.010
Observations	6404	6404	6404	6404	6404

This figure replicates Table 2.4, interacting each treatment with a dummy for whether the household head is female

Table B.3.16: Heterogeneity by Gender: Business and Financial Outcomes

	(Semi-) Durable Exp.	Biz. Exp.	Amount Borrowed	Any Biz Activity	Any Borrowing
Cash Loan	0.766	4.583**	0.596	0.002	0.025
	(1.833)	(2.131)	(0.977)	(0.015)	(0.016)
Cash Loan×Female Head	3.258	-5.963**	3.506*	0.031	0.037
	(4.075)	(2.922)	(1.942)	(0.040)	(0.035)
Maize Loan	0.238	0.079	-0.520	-0.007	0.019
	(1.525)	(1.554)	(0.677)	(0.019)	(0.013)
Maize Loan×Female Head	-2.409	-2.114	0.590	0.016	-0.047
	(2.504)	(2.925)	(1.255)	(0.045)	(0.031)
Female Head	-2.696	0.083	-1.402	-0.040	-0.028*
	(2.062)	(1.557)	(1.044)	(0.031)	(0.015)
Fixed Effects	Strat	Strat	Strat	Strat	Strat
R^2	0.028	0.031	0.023	0.057	0.066
Control mean	18.484	3.252	4.659	0.087	0.098
Observations	6404	6404	6404	6404	6404

This figure replicates Table 2.6, interacting each treatment with a dummy for whether the household head is female

Table B.3.17: Heterogeneity by Gender: Agricultural Outcomes

	Ag. exp . ('000 Naira)	Planted area (ha)	Dry season ag. exp.	Dry season area (ha)	Harvest value
Cash Loan	16.291	0.215	0.135	0.000	170.1
	(21.695)	(0.637)	(0.159)	(0.009)	(162.2)
Cash Loan×Female Head	-65.591**	-2.053	-0.152	-0.001	-321.046*
	(29.538)	(1.807)	(0.162)	(0.009)	(179.1)
Maize Loan	-0.757	-0.358	-0.153	-0.013	73.16
	(22.895)	(0.543)	(0.901)	(0.021)	(116.7)
Maize Loan×Female Head	-39.130	-2.013*	0.120	0.010	-303.440
	(42.439)	(1.058)	(0.990)	(0.022)	(202.3)
Female Head	-59.251**	-1.747**	-0.016	-0.005	-266.772**
	(28.784)	(0.739)	(0.406)	(0.011)	(123.3)
Fixed Effects	Strat	Strat	Strat	Strat	Strat
R^2	0.256	0.184	0.070	0.071	0.141
Control mean	144.122	5.482	0.894	0.024	784.949
Observations	829	829	829	829	808

This figure replicates Table 2.5, interacting each treatment with a dummy for whether the household head is female

Table B.3.18: Heterogeneity by Gender: Livestock Outcomes

	Cows	Goats	Sheep	Chickens
Cash Loan	0.547	1.004	0.468	1.791*
	(0.482)	(0.612)	(0.578)	(0.962)
Cash Loan×Female Head	-0.490	-1.205*	-0.850	-1.052
	(0.672)	(0.679)	(0.854)	(1.896)
Maize Loan	-0.359	-0.593	-0.581	0.049
	(0.287)	(0.430)	(0.447)	(0.719)
Maize Loan×Female Head	0.458	0.822	0.349	2.183
	(0.463)	(0.965)	(1.023)	(1.468)
Female Head	-0.812**	-0.764	-0.089	0.511
	(0.386)	(0.745)	(0.589)	(0.736)
Fixed Effects	Strat	Strat	Strat	Strat
R^2	0.137	0.232	0.128	0.163
Control mean	1.528	3.843	2.472	2.262
Observations	829	829	829	829

This figure replicates Table 2.7, interacting each treatment with a dummy for whether the household head is female

Table B.3.19: Heterogeneity by Gender: Land Outcomes

	Land owned (ha)	Land rented out (ha)	Land rented in (ha)
Cash Loan	2.941	0.140	-0.486
	(2.861)	(0.211)	(0.448)
Cash Loan×Female Head	-1.401	0.218	1.708
	(3.526)	(0.351)	(1.625)
Maize Loan	-0.506	-0.030	-0.553
	(0.547)	(0.123)	(0.375)
Maize Loan×Female Head	-3.081	0.141	-1.481*
	(1.896)	(0.250)	(0.872)
Female Head	-1.077	-0.198	-1.236*
	(1.352)	(0.198)	(0.657)
Fixed Effects	Strat	Strat	Strat
R^2	0.060	0.086	0.337
Control mean	5.785	0.187	2.395
Observations	829	829	829

This figure replicates Table 2.8, interacting each treatment with a dummy for whether the household head is female

Table B.3.20: Heterogeneity by Baseline Wealth: Grain Flows

	Sales Value	Harvest Value	Value of Stock Consumed
Cash Loan	6.392	65.010	68.138
	(48.641)	(121.755)	(78.282)
Cash Loan×BL Asset Index	-96.338**	-94.210	9.820
	(45.022)	(114.952)	(100.867)
Maize Loan	10.773	47.475	25.328
	(35.758)	(99.358)	(81.479)
Maize Loan×BL Asset Index	-3.696	-10.820	18.193
	(48.793)	(137.014)	(107.839)
BL Asset Index	57.139*	63.697	109.647**
	(28.165)	(64.035)	(41.568)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.362	0.362	0.337
Control mean	323.698	949.036	673.146
Observations	930	930	930

This figure replicates outcomes for grain flows, interacting each treatment with an index comprising the first principal component of baseline assets.

Table B.3.21: Heterogeneity by Baseline Wealth: Consumption

	Log exp.	Log non-. storable exp.	log λ	Elicited IMRS	Hunger index
Cash Loan	-0.013	0.013	0.021	-0.036	0.016
	(0.087)	(0.057)	(0.024)	(0.022)	(0.020)
Cash Loan×BL Asset Index	0.117*	0.053	0.052**	-0.023	0.001
	(0.059)	(0.050)	(0.025)	(0.016)	(0.029)
Maize Loan	0.017	0.040	-0.002	-0.043**	0.018
	(0.080)	(0.062)	(0.029)	(0.020)	(0.029)
Maize Loan×BL Asset Index	0.003	-0.027	0.014	-0.014	-0.052*
	(0.059)	(0.056)	(0.027)	(0.015)	(0.026)
BL Asset Index	0.085	0.128***	-0.023	-0.007	0.023
	(0.061)	(0.046)	(0.019)	(0.012)	(0.018)
Fixed Effects	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr	Strat-Yr
R^2	0.048	0.134	0.245	0.194	0.008
Control mean	7.750	6.619	0.324	1.691	-0.010
Observations	6404	6404	6404	6404	6404

This figure replicates Table 2.4, interacting each treatment with an index comprising the first principal component of baseline assets.

Table B.3.22: Heterogeneity by Baseline Wealth: Business and Financial Outcomes

	(Semi-) Durable Exp.	Biz. Exp.	Amount Borrowed	Any Biz Activity	Any Borrowing
Cash Loan	0.832	4.428**	0.833	0.008	0.026*
	(1.645)	(2.080)	(0.840)	(0.016)	(0.015)
Cash Loan×BL Asset Index	-0.047	4.057	-1.840	0.011	0.007
	(1.577)	(3.887)	(1.184)	(0.016)	(0.013)
Maize Loan	0.561	0.072	-0.392	-0.005	0.016
	(1.453)	(1.347)	(0.570)	(0.018)	(0.011)
Maize Loan×BL Asset Index	-0.210	-2.329	-1.254	-0.023*	-0.003
	(1.323)	(1.508)	(0.880)	(0.012)	(0.011)
BL Asset Index	1.016	1.224	1.050	0.012	-0.011
	(1.041)	(0.765)	(0.875)	(0.011)	(0.010)
Fixed Effects	Strat	Strat	Strat	Strat	Strat
R^2	0.031	0.035	0.024	0.060	0.067
Control mean	18.484	3.252	4.659	0.087	0.098
Observations	6404	6404	6404	6404	6404

This figure replicates Table 2.6, interacting each treatment with an index comprising the first principal component of baseline assets.

Table B.3.23: Heterogeneity by Baseline Wealth: Agricultural Outcomes

	Ag. exp .	Planted area (ha)	Dry season ag. exp.	Dry season area (ha)	Harvest value
Cash Loan	11.617	0.081	0.175	0.001	170.1
	(21.111)	(0.631)	(0.227)	(0.008)	(150.3)
Cash Loan×BL Asset Index	7.771	0.013	0.692	0.015	-146.101
	(19.900)	(0.632)	(1.127)	(0.030)	(129.1)
Maize Loan	1.407	-0.325	-0.248	-0.014	85.32
	(22.856)	(0.571)	(0.897)	(0.020)	(121.7)
Maize Loan×BL	-7.998	-0.219	0.303	0.005	54.04
	(24.390)	(0.515)	(0.347)	(0.011)	(228.9)
BL Asset Index	14.862	0.532**	-0.627	-0.003	71.60
	(11.502)	(0.258)	(0.714)	(0.017)	(64.42)
Fixed Effects	Strat	Strat	Strat	Strat	Strat
R^2	0.260	0.183	0.076	0.075	0.141
Control mean	144.122	5.482	0.894	0.024	784.949
Observations	829	829	829	829	808

This figure replicates Table 2.5, interacting each treatment with an index comprising the first principal component of baseline assets.

Table B.3.24: Heterogeneity by Baseline Wealth: Livestock Outcomes

	Cows	Goats	Sheep	Chickens
Cash Loan	0.509	0.940	0.366	2.081**
	(0.428)	(0.609)	(0.538)	(0.822)
Cash Loan×Baseline Asset Index	0.092	0.323	0.056	0.319
	(0.508)	(0.543)	(0.507)	(1.061)
Maize Loan	-0.268	-0.505	-0.554	0.511
	(0.262)	(0.397)	(0.437)	(0.533)
Maize Loan×Baseline Asset Index	0.089	0.267	-0.259	-0.860
	(0.480)	(0.481)	(0.390)	(0.630)
Baseline Asset Index	0.113	-0.024	0.165	0.824
	(0.277)	(0.270)	(0.362)	(0.567)
Fixed Effects	Strat	Strat	Strat	Strat
R^2	0.139	0.235	0.133	0.206
Control mean	1.528	3.843	2.472	2.262
Observations	829	829	829	829

This figure replicates Table 2.7, interacting each treatment with an index comprising the first principal component of baseline assets.

Table B.3.25: Heterogeneity by Baseline Wealth: Land Outcomes

	Land owned (ha)	Land rented out (ha)	Land rented in (ha)
Cash Loan	2.975	0.176	-0.373
	(2.677)	(0.209)	(0.441)
Cash Loan×BL Asset Index	3.637	0.062	-0.174
	(4.082)	(0.256)	(0.342)
Maize Loan	-0.404	0.019	-0.578
	(0.538)	(0.086)	(0.403)
Maize Loan×BL Asset Index	0.118	-0.227	0.135
	(0.879)	(0.195)	(0.307)
BL Asset Index	0.556	0.134	0.071
	(0.479)	(0.144)	(0.250)
Fixed Effects	Strat	Strat	Strat
R^2	0.068	0.099	0.334
Control mean	5.785	0.187	2.395
Observations	829	829	829

This figure replicates Table 2.8, interacting each treatment with an index comprising the first principal component of baseline assets.

Appendix C

Enter Sandmo: Production Function Estimation for Firms that Consume — Appendix

C.1 Appendix to Section 3.1

Table C.1.1: Dynamic Panel Production Estimates

Dependent Variable:	Δlog Ouptut		
Model:	Just IDed	OverIDed 2SLS	OverIDed GMM
Variables			
Δlog Land	0.4641***	0.5239***	0.4314***
	(0.0399)	(0.0471)	(0.0480)
Δlog Labor	0.1033***	0.0604***	0.0839***
	(0.0225)	(0.0225)	(0.0217)
Δlog Equipment	0.0854***	0.1044***	0.1291***
	(0.0222)	(0.0265)	(0.0239)
Δlog Fertilizer	0.0561***	0.0452**	0.0273
	(0.0176)	(0.0182)	(0.0225)
Δlog Seed	0.0934***	0.1178***	0.1216***
	(0.0238)	(0.0253)	(0.0314)
Lagged instruments	1st	1st and 2nd	1st and 2nd
Observations	3,289	2,937	3,209
Within R^2	0.4579	0.4715	
Sargan test, p-value		0.0122	0.0027
AR(2) test, p-value			0.0001

Clustered (j) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

This table provides estimates of α following the Anderson and Hsiao (1981) (AH) procedure used by Shenoy (2017). To be consistent with Shenoy (2017), I group inputs into land, labor, and materials, where materials are the sum of expenditures on fertilizer, seed, and equipment. The first column shows the just-identified AH specification, in which the log differences in inputs are instrumented with their lagged values. The second shows the same specification with two first and second lags of inputs as instruments, estimated using two-stage least squares. The third estimates the same specification with GMM. The Sargan test rejects the null that both sets of lags are exogenous with p-values of 0.0122 and 0.0027, respectively and the Arellano-Bond test rejects the null of no second-order autocorrelation with a p-value of 0.0001 .