Subsistence farming and factor misallocation:
Evidence from Ugandan agriculture

Bruno Morando

*Trinity College Dublin*

January 31, 2019

Abstract

Recent findings show that the inefficient distribution of land and complementary inputs across heterogeneous farmers plays a role in determining low agricultural productivity in developing countries. There is a wide consensus that resource misallocation is the outcome of poorly functioning land markets that prevent plots from being transferred from less to more capable farmers. This is at odds with the fact that relevant inefficiencies persist in countries where factor markets are well functioning. We develop a theoretical model showing that frictions in the food market resulting in incentives towards subsistence farming can cause agricultural input misallocation even when agricultural factors of production can be traded freely. This hypothesis is tested using micro data from farms in Uganda, a country where land markets are well established and historically very active. Our estimates indicate that an efficient reallocation of land and capital among existing farmers would result in a 119 percent increase in agricultural TFP. We show that in order to achieve it, inputs should be transferred from subsistence to more market oriented farmers. Consistently, misallocation is shown to be more severe where subsistence agriculture is more widespread. These findings suggest that eliminating land market restrictions represents a necessary but not sufficient condition to abate misallocation and that complementary policies aimed at reducing frictions in the food market are required.

**Keywords:** Misallocation, Productivity, Agriculture, Uganda

**JEL Classification:** Q14, O40, O13
1 Introduction

Low labour productivity in agriculture is perhaps the most distinctive feature of developing countries. Mechanically, low productivity in the primary sector implies that a larger share of the workforce needs to be employed in farming in order to produce the food necessary for subsistence needs (Gollin et al. 2007). This leads to the well known pattern of negative correlation between agricultural productivity and share of employment in farming. As shown by Restuccia et al. (2008), this fact alone can account for most of the cross country income differences. Infrastructural constraints like poor transportation networks making food imports more costly (Tombe 2015) exacerbate this problem and contribute to its persistence.

A number of concurrent explanations for low agricultural productivity have been suggested, such as negative sorting into the primary sector (Lagakos and Waugh 2013) and barriers to adoption of modern intermediate inputs (Gollin and Rogerson 2014) to name just two. A recent strand of literature, inspired by the seminal study by Hsieh and Klenow (2009), relates low total factor productivity in agriculture to misallocation in production factors. In short, the core hypothesis is that institutional barriers to land rentals/sales (that are widespread in the developing world) prevent the reallocation of factors of production towards more able individuals, ultimately resulting in low aggregate TFP.

Consistently with this theory, reforms aimed at relaxing long lasting limitations to land transactions have been shown to increase efficiency in resource allocation and foster agricultural productivity in Malawi (Restuccia and Santaeulalia-Llopis 2017), Ethiopia (Chen et al. 2017) and China (Adamopoulos et al. 2017, Chari et al. 2017). These studies have opened a debate on the actual severity of agricultural input misallocation in developing countries and its implications in terms of overall productivity losses (Restuccia and Rogerson 2017). On the one hand, Adamopoulos et al. (2017) claim that the consequences of agricultural factor misallocation are more significant than it would appear from a static analysis as it affects also the sorting of workers across sectors. On the other hand, Gollin and Udry (2017) suggest that the existing estimates of misallocation might be dramatically inflated as they fail to disentangle the actual dispersion in farmers’ productivity from other unobserved sources of output variation such as measurement errors and land quality heterogeneity.

However, there is a general consensus that the principal cause of misallocation is frictions in input (and in particular land) transactions while concomitant and potentially complementary explanations are generally overlooked.¹ Nevertheless, all the existing studies highlight that relevant inefficiencies in factor distribution persist (and in some cases increase, as in Ayerst et al. 2018) not only in the immediate aftermath of land reforms, but also in contexts where land markets are well established. Although this can be ascribed to the existence of more subtle constraints still preventing frictionless exchanges of inputs, other relevant channels might be in place.

More specifically, another distinctive fact regarding agriculture in developing countries is that a large share of the output is consumed by the individuals who cultivate it on small scale holdings (Adamopoulos and Restuccia 2014). This suggests that accessing the food

¹A significant exception is Shenoy (2017), who also considers imperfections in the credit markets.
market (as buyers or sellers) might be costly and farmers are better off producing for self consumption rather than for commercial purposes. Since the seminal paper by De Janvry et al. (1991), a number of studies have been proposed to explain the behaviour of farm households facing implicit (Fafchamps 1992) or explicit (Omamo 1998) output marketization costs. The common denominator of these models is that the producers' output shadow price is a function of their specific consumption needs and production capacity rather than being constant and equal to the market price. Namely, net food buyers attach higher value than net food sellers to their output as a result of the above mentioned frictions that create a wedge between producer and consumer prices.

In this paper, we build from this intuition and provide a model showing that resource misallocation can occur even in the presence of virtually frictionless input markets as long as subsistence farmers attribute higher value to their produce and thus are less willing to transfer their factors of production to more able market oriented farmers.\(^2\)

The empirical analysis shows that in Uganda, in spite of a remarkably active market for land, misallocation is still a preponderant issue. We find that reallocating agricultural inputs efficiently among the existing farmers would result in a 119 percent increase in the aggregate productivity. Consistently with our model, we show that in order to realize these gains, land should be transferred from subsistence (or close to subsistence) farmers to more market oriented ones. Moreover, we show that the level of village specific misallocation is a positive function of the pervasiveness of subsistence agriculture while is not related to the level of land rental/sales activity.

The most closely related work is Li (2017), which studies the production side impact of a food subsidy scheme in India aimed at lowering staples price. In particular, the author finds that the reduction in food price triggers a redistribution of land towards more productive commercial farmers and enhances agricultural productivity. The most obvious difference with respect to this study is that, having micro level farm data, we can base the theoretical model as well as the empirical analysis on the classic misallocation framework.

The paper develops as follows: section 2 presents a broad overview of the agricultural sector in Uganda, with a particular attention to the functioning of the land markets and the issue of lack of efficient transport infrastructures. Section 3 provides the theoretical framework used to measure misallocation and a model linking it to frictions in output markets and subsistence farming. Section 4 describes the data used for the empirical analysis. Section 5 is devoted to the presentation of the empirical findings corroborating the theoretical model and comments on the results. In section 6, we provide some robustness checks and discuss the findings further. Section 7 concludes with some final remarks.

\(^2\)The idea that factor misallocation might be the result of frictions in the output market is not new. However, it has never been explored fully nor tested empirically with the partial exception of Chen et al. (2017).
2 Agriculture in Uganda

Uganda is one of the poorest countries in the world, and systematically ranks at the bottom of the distribution of several economic, social and health indicators. In spite of relatively good economic performance since the 2000s, poverty remains a widespread issue. According to the most recent estimates (Ubos, 2018), the poverty rate is just below 20 percent, but the aggregate figure masks huge regional and urban/rural imbalances. In particular, the rural areas in the Northern and Eastern regions present remarkably higher poverty indicators.

Similarly to most countries in Sub Saharan Africa, the agricultural sector plays a key role in Ugandan economy, employing around 70 percent of the total workforce (Ubos, 2017). Subsistence and semi subsistence farming still represent the major source of income of most households, as farmers typically operate on a very small scale and sell a very limited share of their production.

Interestingly, as pointed out by Ulimwengu et al. (2009), although from an aggregate point of view Uganda is a net food exporter, most of the farmers are net food buyers. In particular, they find that 65 percent of the net food buyers in Uganda declare that subsistence agriculture is their primary source of livelihood, and that farmers spend a high share of their total income (45 percent) on food and beverages.

The prevalence of subsistence agriculture is typically attributed to high transaction and marketization costs, that are widely believed to represent the most serious barrier to the modernization of the primary sector in Uganda. This view is shared by farmers, scholars and policy makers (Gollin and Rogerson 2010). More specifically, high transportation costs result in wide wedges between consumer and producer prices, low spatial market integration and high price volatility. Overall, these factors are believed to hinder commercial agriculture and provide incentives towards self production of food and might explain the disproportionate share of employment in subsistence farming (Gollin and Rogerson 2010; Li 2017).

Unlike other developing countries, there seem to be virtually no barriers to land transfers in Uganda. Indeed, land markets have been very active with virtually no interruption ever since the English domination, and land rentals and sales are common throughout the whole country (Deininger and Mpuga 2002).

As reported by Baland et al. (2007), local leaders have virtually no powers when it comes to land redistribution and only intervene very rarely in case of land disputes. They also highlight the fact that land transactions happen regularly between people belonging to different ethnic groups and with low to inexistent social ties, suggesting that land transactions are generally inspired by market logic.

The overall picture describes an economy still largely dependent on subsistence farming, especially in the poor rural areas, and where a long tradition of market based land transactions has failed to foster the development of a commercial/market-oriented agricultural sector. The poor quality of transport infrastructure is often blamed for the persistence of subsis-

---

3In the early 50s, almost 60 percent of the landowners were operating plots that had been acquired through market transactions.
tence oriented farming activities. This issue is perceived as preponderant by farmers and the
government alike, and its gravity is effectively summarized by the claim that the paved road
density in Uganda in 2003 “was not much greater than [...] the one found in Britain in AD 350” (Gollin and Rogerson, 2010, p. 10).

These features make Uganda the ideal setting to study the impact of frictions in the food
market on agricultural factor misallocation in the absence of apparent barriers to land trans-
actions.

3 Theoretical framework

The theoretical framework is largely borrowed from the cutting edge literature on factor
misallocation in the agricultural sector (see Restuccia and Santaeulalia-Llopis 2017, Chen
et al. 2017, Adamopoulos et al. 2017 and Chari et al. 2017). In the first subsection, we present
a basic framework for measuring the magnitude of misallocation. In the second, we set up a
theoretical model featuring frictionless markets for inputs and where resource misalloca-
tion is presented as the equilibrium outcome of distortions in the output (food) market. A formal
derivation of the results and a brief discussion on the underlying conditions is provided in the
appendix. The predictions of the model will be then tested in the empirical section.

3.1 Measuring Misallocation

There is a fixed number of farmers $N$, indexed by $i$. They differ in productivity, which is
captured by an individual specific, labour augmenting parameter $s_i$. They all produce a
homogeneous agricultural good according to a common Cobb Douglas production function
using three inputs: labour $X$, capital $K$ and land $L$:

\[
\tilde{Y}_i = (s_i X_i)^{1-\gamma} [(q_i L_i)^{\alpha} K_i^{1-\alpha}]^\gamma
\]

where $q_i$ is a land augmenting parameter included to control for the quality of the soil. The
parameters $\alpha$ and $\gamma$ define the three factors’ shares.\(^5\)

Consistently with the literature, we will focus on misallocation of land and capital, while
labour is treated as a non tradable factor. It is therefore convenient to rearrange the pro-
duction function in labour units terms:

\[
\tilde{y}_i = s_i^{1-\gamma} [(q_i L_i)^{\alpha} K_i^{1-\alpha}]^\gamma
\]

in this setting, the parameter $\gamma$ will determine the returns to scale in the variable factors of
production.\(^6\)

\(^4\)High transportation costs and poor road networks typically rank first among the detrimental factors for
agricultural productivity reported by farmers in the World Bank’s LSMS.

\(^5\)More specifically, labour’s share is $1 - \gamma$, land’s share is $\alpha \gamma$ and capital share is $(1 - \alpha)\gamma$.

\(^6\)As $\gamma < 1$, the model features decreasing returns to scale, and the efficient input distribution will be non
degenerate.
Finally, we can control for heterogeneous land quality by normalizing the output $\bar{y}_i/q_i^{(1-\alpha)\gamma} = y_i$. By doing so, we obtain the operative form of the production function that will be used to infer farm specific productivity:

$$y_i = s_i^{1-\gamma}(l_i^\alpha k_i^{1-\alpha})\gamma. \tag{3}$$

The magnitude of resource misallocation and its impact on aggregate factor productivity can be estimated comparing the production resulting from the actual distribution of factors across heterogeneous farmers ($l_1, \ldots, l_N; k_1, \ldots, k_N$) to the one obtained in the counterfactual scenario where land and capital are distributed efficiently. The latter can be characterized as the solution of the social planner problem, who maximizes the aggregate production facing only the resource constraint. The problem can formally be described as:

$$\max_{l_i, k_i} \sum_{i}^N y_i \text{ subject to: } \sum_{i}^N l_i = L \text{ and } \sum_{i}^N k_i = K \tag{4}$$

where L and K are the total land and capital available in the economy.

Intuitively, the first order conditions imply that the the marginal products of the factors of production are equalized across all the farmers. As the production function features decreasing returns to scale, these conditions are met when each farmer operates a share of land and capital that is proportional to her productivity $s_i$:

$$l_i = s_i \frac{s_i}{\sum_j^N s_j} L \text{ and } k_i = s_i \frac{s_i}{\sum_j^N s_j} K \tag{5}$$

By plugging in these values into the production function, we can obtain the farm specific efficient output level: $y_i^e$ as $s_i^{1-\gamma}(l_i^\alpha k_i^{1-\alpha})\gamma$. In turn, this allows us to estimate the severity of factor misallocation as the ratio between the total output in the efficient counterfactual and the actual production:

$$e = \frac{\sum_{j}^N y_j^e}{\sum_{j}^N y_j} = \frac{Y^e}{Y} \tag{6}$$

Conveniently, the same procedure can be applied in order to compute the potential gain from efficient factor reallocation at lower geographical levels and/or within some categories of farmers. In the remainder of the paper, we will take advantage of this opportunity to test some predictions of the model.\footnote{More specifically, we will estimate the village level misallocation as the ratio between the total production that would be obtained if marginal products were equalized across all the farmers and the actual village level output. We will also estimate the misallocation within farmers participating and non participating to land markets and within subsistence/commercial farmers in a given location.}
3.2 Misallocation and subsistence farming

Although the debate on the magnitude of the potential gain from reallocation of agricultural inputs and its potential contribution to the reduction of the agricultural productivity gap in low income countries is very active, less attention has typically been devoted to the causes of the misallocation. Typically, frictions in land rental and sales market are indicated as the major (and sometimes unique) cause of the inefficient resource distribution.

In support of this thesis, existing findings show that more active land rental markets are correlated with lower level of factor misallocation. However, the same studies highlight that resource distribution remains far from efficient after land reforms and even when input markets appear to be well functioning. It is therefore important, especially from a policy perspective, to identify some concurrent sources of misallocation. In the following, we provide a simple theoretical framework aimed at showing that input misallocation can occur when there are no barriers to input trade as long as accessing the food market (both as a producer and as a consumer) is costly.

For the sake of tractability, we abstract from the difference in land and capital inputs and we assume that farmer $i$ in village $v$ produces a homogeneous agricultural good $y$ using only one composite input $\xi$ which can be exchanged freely among agents. The resulting production function (expressed in per labour unit terms) is:

$$y_i = s_i^{1-\gamma} \xi_i^\gamma.$$  

(7)

In line with a well established literature on production decisions of farm households (see, among others, the seminal work Finkelshtain and Chalfant 1991), farmers are not modeled as producers willing to maximize their profits, but rather as consumers and producers, who evaluate their output differently according to their net position in the food market. For simplicity, we assume that farmers consume agricultural produce (food) until the threshold $\bar{y}$ is met deriving constant marginal utility normalized to 1.\(^8\) The production exceeding $\bar{y}$ is sold, and for each unit the farmer receives a constant utility normalized to $1-\tau_v$.\(^9\) The resulting maximization problem is:

$$\max_{\xi_i} \begin{cases} y_i - r\xi_i & \text{if } y_i \leq \bar{y} \\ (1 - \tau_v)y_i - r\xi_i & \text{if } y_i > \bar{y} \end{cases}$$  

(8)

In general, $\tau$ can be thought of as a combination of more and less tangible factors (such as transaction/transportation costs, volatile market prices, thin markets) that incentivize

\(^8\)A more realistic model would feature decreasing marginal utility of food consumption. We maintain the assumption of constant marginal utility for the sake of manageability but it is worth mentioning that decreasing marginal utility would leave the main predictions of the model untouched (and possibly strengthen them).

\(^9\)Here, we are assuming that the threshold $\bar{y}$ is the same across the different households. Note that this assumption is less problematic than it might appear as output $y$ is expressed in per labour terms. Therefore, as long as we assume that all the individuals in a households are employed in agriculture and labour is not tradeable (in line with the model), the assumption could be rephrased as: “each individual farmer has equal food consumption threshold”.

7
farmers to grow their own food rather than buying it at the market and symmetrically make selling agricultural output less profitable. We assume that these output market distortions are village specific, and therefore are identified by the subscript \( v \). Intuitively, a larger \( \tau \) implies higher incentives towards food self production and lower profitability of commercial agriculture.

In order to provide a clear intuition of the basic functioning of the model, it is convenient to analyze the two extreme cases of no frictions in the output market (\( \tau = 0 \)) or unavailability of food markets (\( \tau = 1 \)).

When \( \tau = 0 \), farmers face no additional cost in buying/selling their produce and therefore the value they attach to the output is the same regardless of whether they consume or sell it. As a result, their maximization problem mirrors the one of a profit maximizing producer:

\[
\max_{\xi_i} y_i - r\xi_i \tag{9}
\]

where \( r \) is the price of the composite agricultural input \( \xi \).\(^{10}\)

Similarly to the case of the social maximizer problem depicted in equation 4, the resulting first order condition requires the marginal productivity of inputs to be equalized across farmers:

\[
\gamma s_i^{1-\gamma} \xi_i^{-\gamma-1} = MP_i^* = r \tag{10}
\]

Therefore, in equilibrium the quantity of input used will be proportional to the farmer specific productivity \( s_i \), and the factors of production will be distributed efficiently:

\[
\xi_i^* = As_i \tag{11}
\]

where \( A = \frac{r^{\frac{1}{1-\gamma}}} {\gamma} \). The equilibrium in this scenario is depicted in Figure 1.

At the opposite side of the spectrum, markets for food do not exist (or are too costly to access) as in De Janvry et al. (1991). In the model’s notation, we set \( \tau = 1 \). This implies that farmers will be unable to sell their produce and therefore do not derive any utility from output exceeding the consumption threshold \( \bar{y} \). Therefore, farmers will be maximising:

\[
y_i - r\xi_i \text{ subject to: } y_i \leq \bar{y} \tag{12}
\]

In this case, the equilibrium features two categories of farmers depending on their agricultural skills. As the constraint is not binding for low enough farming skills, less productive individuals will use a quantity of input \( \xi \) so to equalize their marginal cost \( r \) and marginal product \( MP_i \). On the other hand, farmers with higher agricultural skills will operate on a scale that is just sufficient to produce \( \bar{y} \) and in equilibrium, their marginal productivity will exceed the marginal cost of input.

\(^{10}\)The fact that there is no difference between the cost of renting the input in and the returns to rent it out (as in Chari et al. 2017) and that there is no maximum threshold of input that can be rented in (as in Shenoy 2017) reflects the assumption that input markets are perfectly functioning.
As a result, while the share of factor is proportional to productivity for the first set of (non constrained) farmers, it is inversely proportional to $s$ for the second group of individuals.\footnote{The fact that the input is efficiently distributed within non constrained farmers depends crucially on the assumption that marginal utility of food consumption is constant. Decreasing marginal utility would imply different shadow price of output within subsistence farmers (with higher marginal value when $y$ if further from $\bar{y}$) and ultimately result in factor misallocation also within less productive farmers.}

Formally:

$$\xi_i^* = \begin{cases} 
(\frac{2}{5})^{\frac{1}{1-\gamma}} s_i & \text{if } s_i \leq \bar{y} \left(\frac{r}{\gamma}\right)^{\frac{1}{1-\gamma}} \\
\left(\frac{\bar{y}}{s_i^{1-\gamma}}\right)^{\frac{1}{\gamma}} & \text{if } s_i > \bar{y} \left(\frac{r}{\gamma}\right)^{\frac{1}{1-\gamma}}
\end{cases} \quad (13)$$

Unlike the one previous one, this equilibrium outcome, depicted in Figure 2, features different marginal productivity across farmers, and therefore is inefficient as the resulting allocation differs from the one of a hypothetical social planner.\footnote{The only case in which the equilibrium would be exactly equal to the one without any friction in the food market is when the distribution of farmers’ productivity is such that the constraint $y_i \leq \bar{y}$ is not binding for any of them. While in the opposite scenario of the constraint being binding for all the farmers, the share of input operated by each individual will be inversely proportional to her skills.}

$$MP_i^* = \begin{cases} 
r & \text{if } s_i \leq \bar{y} \left(\frac{r}{\gamma}\right)^{\frac{1}{1-\gamma}} \\
Bs_i^{\frac{1-\gamma}{\gamma}} & \text{if } s_i > \bar{y} \left(\frac{r}{\gamma}\right)^{\frac{1}{1-\gamma}}
\end{cases} \quad (14)$$

where $B = \frac{\gamma y_1}{y}$.  

To sum up, the resulting distribution of resources is inefficient across the two groups as less productive farmers use a larger than efficient share of input and have lower marginal productivity than more skilled individuals in equilibrium. Additionally, the equilibrium outcome is characterized by misallocation within the group of constrained farmers as well, as the more productive will need a lower quantity of the composite factor $\xi$ to meet their consumption threshold $\bar{y}$ and thus present higher marginal productivity.

In the intermediate and more realistic case where $0 < \tau < 1$, the market for food exists but is costly to access. Therefore, farmers are better off producing the food they consume rather than buying or selling it. Borrowing from the terminology used by Li (2017), they experience consumption advantage that eventually affects the input distribution across them. In particular, farmers with low productivity relying on agriculture for subsistence purposes only, will operate a higher than efficient share of input as their output decision (or shadow) price is higher than the one of commercial farmers.

The equilibrium features three categories of farmers, depending on their productivity:

- a) those with $s_i \leq \bar{y} \left(\frac{r}{\gamma}\right)^{\frac{1}{1-\gamma}}$, who would not commercialize their produce even if there were no output market frictions as their production is lower than the consumption threshold. They would operate a share of land such that the marginal product equals the cost of input $r$. The farmer with ability $s_1$ in Figure 3 belongs to this category;
b) those with \( \bar{y} \left( \frac{r}{y} \right)^{\frac{1}{\gamma}} < s_i \leq \bar{y} \left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} \), who only hire input \( \xi \) to meet their consumption threshold \( \bar{y} \), but would produce a surplus if there were no frictions in the market for agricultural produce. The total quantity of input \( \xi_i \) used will be equal to \( \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} \) and therefore those among them with relatively higher level of productivity will cultivate less land. Accordingly, the marginal productivity will be an increasing function of \( s_i \): \( B s_i^{\frac{1-\gamma}{\gamma}} \). The farmer with productivity \( s_2 \) in figure 3 belongs to this category;

c) those with \( s_i > \bar{y} \left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} \), who will engage in commercial agriculture regardless of the frictions in output market and will hire input until the marginal productivity equals \( \frac{r}{\gamma(1-\tau_v)} \). Similarly to case a, the input distribution will reflect farmers’ productivity within the group. Nevertheless, their marginal productivity will be higher than the one in a and b. The farmer with productivity \( s_3 \) in Figure 3 belongs to this category.

Formally, in equilibrium, the quantity of input used will be:

\[
\xi^*_i = \begin{cases} 
(\frac{r}{y})^{\frac{1}{\gamma}} s_i & \text{if } s_i \leq \bar{y} \left( \frac{r}{y} \right)^{\frac{1}{\gamma}} \\
\left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} s_i & \text{if } \bar{y} \left( \frac{r}{y} \right)^{\frac{1}{\gamma}} < s_i \leq \bar{y} \left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} \\
\left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} s_i & \text{if } s_i > \left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} 
\end{cases}
\]

(15)

With a resulting marginal productivity of:

\[
MP^*_i = \begin{cases} 
\frac{r}{\gamma(1-\tau_v)} & \text{if } s_i \leq \bar{y} \left( \frac{r}{y} \right)^{\frac{1}{\gamma}} \\
B s_i^{\frac{1-\gamma}{\gamma}} & \text{if } \bar{y} \left( \frac{r}{y} \right)^{\frac{1}{\gamma}} < s_i \leq \bar{y} \left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} \\
\frac{r}{\gamma(1-\tau_v)} & \text{if } s_i > \left[ \frac{r}{\gamma(1-\tau_v)} \right]^{\frac{1}{\gamma}} 
\end{cases}
\]

(16)

A graphical representation of this equilibrium is provided in Figure 3. Similarly to the case with no food market, the final outcome displays different levels of marginal productivity across the three groups, with less able farmers having lower marginal productivity in equilibrium. Additionally, resources are inefficiently distributed within farmers belonging to the group b, whose members operate a share of land that is a negative function of their agricultural skills.\(^{13}\) Figure 4 illustrates the resulting input distribution in the three scenarios just described.

In practice, the model returns three main predictions on the magnitude and patterns of factor misallocation in a context where food markets are thin and costly to access.\(^{14}\) The first one is

\(^{13}\)Even in this case, efficient input distribution within farmers in group a is a direct consequence of constant marginal utility of food consumption. Relaxing this assumption would result in resource misallocation among them as well and in turn in higher overall inefficiencies.

\(^{14}\)These predictions are based on the most realistic scenario of \( 0 < \tau < 1 \), where the market for food exists but presents some imperfections in the form of transaction costs, thinness and high price uncertainty.
that in order to achieve gains from reallocation, land should be transferred from subsistence to market oriented farmers. Indeed, farmers who sell a positive share of their production have a higher marginal productivity than farmers who only farm for self consumption purposes.

Secondly, according to the model, the misallocation in any given village will be higher among subsistence farmers than among marked oriented farmers. Indeed, while in equilibrium commercial farmers have all the same marginal productivity, subsistence farmers are a heterogeneous group composed by those who do not meet their consumption threshold in equilibrium (category a) who have all the same marginal productivity, and those who produce for self consumption only but would exceed this quota if the marketization costs were not in place (category b). Individuals belonging to the latter group will not only have higher marginal productivity than the ones belonging to the former, but also have different marginal productivity to one another.\textsuperscript{15}

Lastly, we expect that villages with more prohibitive marketization costs $\tau$ will be characterized by higher levels of subsistence agriculture and in turn display more severe factor misallocation.\textsuperscript{16}

To sum up, the model shows in an intuitive fashion how agricultural factor misallocation might arise also when farmers face no frictions in trading inputs. It is worth to mention that by abstracting for the potential role of frictions in input markets we are not claiming that they do not represent an issue. Indeed, explicit or implicit barriers in trading land and other agricultural factors represent a massive obstacle to agricultural productivity in large part of the developing world. Instead, we want to highlight that the abatement of these barriers represents a necessary but not sufficient condition to tackle misallocation.

\section{Data}

In performing the empirical analysis, we will use data from a number of different sources, as briefly summarised in Table 1. The most relevant is the Uganda Census of Agriculture (UCA) by the Ugandan Bureau of Statistics, which collects detailed holding level information on more than 25,000 Ugandan farms and is meant to be nationally representative. The data is a cross section and refer to the 2008/09 agricultural year.\textsuperscript{17} Variables on agricultural input and output, as well as on participation on land markets and disposition of crops produced are derived from this dataset. Consequently, the farm will be our main unit of analysis as output and inputs are aggregated at the holding level.

Cultivated land is defined as the total area of the plots farmed in the two agricultural seasons under analysis. As every parcel was measured through GPS techniques, the available data are of particularly high quality and arguably less subject to measurement errors than figures based on farmers’ own estimations. The questionnaire demands also the mode of acquisition

\textsuperscript{15}In the more realistic scenario of decreasing marginal utility from food consumption, also farmers belonging to the first group would have different marginal productivities in equilibrium.

\textsuperscript{16}This relies on the assumption that marketization costs are village specific.

\textsuperscript{17}Namely, the dataset collects information regarding the second agricultural season 2008 and the first agricultural season 2009.
of each of the plots operated so that it can be assessed whether each farm included at least some land that was acquired through market transactions as opposed to inheritances or donations or other forms of informal appropriation.\textsuperscript{18}

As for the labour input, farmers were asked about the exact amount of working days provided by family members and hired labourers as well as their gender and whether they were children or adults. Unfortunately, these data were missing in 56 percent of the observations, and the only information available was whether each worker was employed full time or only temporarily for a given season. From the available observations, we infer that on average, temporary workers provided about half the working days than their full time counterparts. Therefore, in measuring the total quantity of labour input, we count all part time workers as half a unit. Additionally, we introduce different weights depending on the gender and on the age of the workers on the basis of the relative median agricultural wage of each group as measured in the 2009/10 LSMS survey.\textsuperscript{19} As a result, labour input is defined in terms of full time adult male equivalents rather than total working hours. Although this is not ideal and it would be better to have more detailed measures of labour input, we are still able to capture a lot of variation across different holdings based on realistic assumptions and reliable surveys.

Finally, capital is defined as the sum of the value of the agricultural asset used in the farming activity.\textsuperscript{20} Differently from previous studies (Restuccia and Santeulalia-Llopis 2017, Chen et al. 2017), we do not include in the measure the value of means of transportation and livestocks as we have no way to establish whether they were actually employed in agricultural production.\textsuperscript{21}

In terms of final output, farmers were asked about their total production for the 18 major crops that overall account for more than 97 percent of the total farmed area.\textsuperscript{22} Conveniently, the total production for each crop is reported in kilos of dry final produce by applying different conversion factors depending on the self reported state of the harvest and units. Similarly to the case of capital items, prices of the different crops were not available. Therefore, in order to aggregate the total production, we use a set of median prices derived from the LSMS survey.

Furthermore, the survey collects information on the destination of the final output. In

\begin{itemize}
  \item \textsuperscript{18}In the following analysis, we will only make a distinction between farmers operating at least some plots rented in and/or bought without differentiating between temporary and permanent transactions. Unfortunately, we have no information on whether farmers were renting out some of their plots and as such our figures on land market participation might be biased downward.
  \item \textsuperscript{19}More specifically, the median hourly agricultural wage of a woman was 0.8 times the one of a man, while the one of a child, regardless of the gender, was 0.55 times the one of a man.
  \item \textsuperscript{20}Conveniently, unlike other surveys, the UCA questionnaire explicitly asks about the quantities of each item actually employed in agricultural production rather than focusing on the ownership.
  \item \textsuperscript{21}As the price/value of each capital item are not available in the UCA, we derived them from the LSMS using the median self reported value in the agricultural module of the 2010 survey. This was possible as the list of assets in the two questionnaires is exactly the same.
  \item \textsuperscript{22}The list of crops is: maize, finger millet, sorghum, rice, beans, field peas, cow peas, pigeon peas, ground-nuts, sesame, soy beans, bananas (three varieties: food, brewing, and sweet), cassava, sweet potatoes, potatoes and coffee.
\end{itemize}
particular, farmers were asked, for each crop, the share of the total production that was self consumed, used for other purposes by the households, sold or stored. We define the percentage of self consumption as the fraction of the non stored outcome that was not sold.\textsuperscript{23}

The descriptive statistics for the most relevant variables are shown in Table 2. In line with the description of the Ugandan agricultural sector, we find that farmer tend to operate on a very low scale and using a rather limited amount of capital. In particular, the median farmer was cultivating 0.65 hectares and with capital equivalent to 10 US dollars.\textsuperscript{24} The distribution of both land and capital is highly skewed to the right, with median values much lower than average values, suggesting that a high share of the holdings is rather small and farmers tend to operate on a close to subsistence level.

Accordingly, it seems that a large share of the agricultural input is not commercialised but self consumed. Only less than 30 percent of the farmers in our sample were selling at least two thirds of their output and 13 percent of them were producing for self consumption only. The average and median farmers were selling 31 and 25 percent of their output respectively.

In line with the insights on Ugandan agriculture provided in Section 2, land markets are shown to be very active: indeed more than 70 percent of the farmers are operating at least one plot of land that was either acquired in the past or rented in for the current agricultural season. This figure does not come as a surprise considering the Ugandan context, characterized by the lack of barriers to land transfers and the total lack of influence exerted by traditional local authorities on land redistribution (unlike the in China, Ethiopia and Malawi).

In addition to the farm level data listed above, in the following empirical analysis we exploit also two spatial datasets in order to disentangle actual differences in farmers’ productivity from other sources of output variation. Namely, in order to capture different soil and climatic conditions across different areas, we use the GAEZ dataset developed by FAO and IIASA, that provides very precise and disaggregate (the dataset consists in $10 \times 10$ kilometers cells) information on the local agronomic conditions. Similarly, we control for rainfall levels using the estimates provided by TAMSAT referring to the two rainy seasons occurred in the period under analysis (Sept-Nov 2008 and March-May 2009). Since both datasets are realistically too coarse to capture any variation at the household level, the resulting variables are defined at the village level only.

\section{Empirical specifications and results}

This section is devoted to the presentation of the results of the empirical analysis. We will start by showing the estimates of the magnitude of misallocation and comparing them to the ones from existing studies. We will then move to some empirical tests aimed at corroborating the three main predictions offered by the model.

\textsuperscript{23}Only in around 0.1 percent of the observations the whole produce was stored, therefore the variable is available for virtually the whole sample.

\textsuperscript{24}The figure on land refers to the average land operated per season.
5.1 The magnitude of misallocation

The first step required to provide some figures on the magnitude of misallocation consists in estimating the farm specific productivity $s_i$ that can be obtained as, rearranging equation 3:

$$ s_i = \left( \frac{y_i}{l_i^\alpha k_i^{1-\alpha}} \right)^{\frac{1}{1-\gamma}} $$

As data on farm level input use and production are readily available, we only need to estimate the parameters $\alpha$ and $\gamma$, that are a function of the shares of the three inputs entering the production function: labour $X$, land $L$ and capital $K$. Following the procedure suggested by Chari et al. (2017), we obtain them by estimating a production function as follows:

$$ \log \tilde{Y}_{ivt} = \beta_1 \log X_{ivt} + \beta_2 \log q_{iv} L_{ivt} + \beta_3 \log K_{ivt} + W_{vt}' \delta + \phi_i + \tau_t + \epsilon_{ivt} $$

As our data are cross sectional, the production function is estimated using the World Bank’s LSMS data for Uganda referring to the 4 agricultural seasons between 2010 and 2011. In practice, the logarithm of the value added for each farmer $i$ and season $t$ is regressed on the logarithm of the quantities of the three inputs as well as a vector $W$ of time and village (indexed by the subscript $v$) specific shocks plus farmers’ and time fixed effects. The land input is augmented by the parameter $q_{iv}$ that is meant to capture heterogeneous land quality across farms.

According to these estimates, the shares of labour, land and capital are 0.403, 0.475 and 0.122 respectively. This implies $\alpha = 0.796$ and $\gamma = 0.597$. Reassuringly, as shown in Table 3, unlike Chari et al. (2017), we are not able to estimate a crop specific production function as the inputs are recorded at the plot rather than at the crop level and typically farmers grow more than one crop per plot. Therefore, we need to make the implicit assumption of equal input elasticities across crops. Reassuringly, in the case of Chari et al. (2017), using crop specific or general coefficients does not affect the results of the estimation significantly.

In our case, the vector $W$ contains the rainfall percentile as well as its interaction with a dummy variable taking value 1 if the district is prevalently unimodal.

Unlike in our main analysis, the land quality index is computed on the basis of the relative variables in the LSMS questionnaire (village section) rather than the GAEZ dataset. The procedure used to compute the land quality index follows quite closely the one in Restuccia and Santaeulalia-Llopis (2017) and Chen et al. (2017) and is based on the simple regression:

$$ \tilde{y}_{iv} = X_{iv}' \beta_1 + Z_{iv}' \beta_2 + \frac{k_{iv}}{l_{iv}} \beta_3 + \epsilon_{iv} $$

including the value added per labour days of the farmer $i$ in village $v$ on the left hand side and a vector $X$ of farm specific and $Z$ of village specific variables that capture the soil and terrain characteristics, as well as the ratio between capital and land to control for different capital intensities across fields. The farm specific vector $X$ includes self reported soil quality, occurrence of erosion problems, presence of irrigation facilities and prevalent soil type and topography. As the value reported were parcel rather than farm specific, they are aggregated at the farm level by considering the weighed average of the continuous variable (using the size of the cultivated parcels as weights) or the weighted mode for the categorical variables. The index will be simply defined as $X_{iv}' \beta_1 + Z_{iv}' \beta_2$.

We were unable to reject constant returns to scale at the 90% significance level and therefore we use the estimates derived by imposing $\beta_1 + \beta_2 + \beta_3 = 1$. 
our parameters are quite close to the ones estimated by similar studies involving developing countries. Consistently with the literature, the capital share is significantly lower than the one in more industrialized economies such as the US.

Once these parameters are estimated, we can plug them into equation 17 to obtain an estimate of the farm specific productivity $s_i$ for each of the $N$ observations in our sample.\footnote{In estimating $s_i$, we control for land quality creating an index based on the 10 dimensions of soil suitability available in the GAEZ dataset (soil workability, toxicity, terrain slope, rooting conditions, oxygen availability, nutrient retention and availability, altitude, excess salinity and dominant soil type) and for weather shocks using rainfall data from TAMSAT. More specifically, we follow the procedure outlined by Chen et al. (2017) and control for different rainfall patterns by regressing the output on a set of dummies, each representing a decile of total precipitations during the two rainy seasons (Sept-Nov 2008 and Mar-May 2009) and using the residuals as a measure of total production.} In order to address some potential concerns related to measurement errors, we trim the top and bottom 2.5\% of the observations in the productivity distribution. However, all of our results are robust to more and less conservative trimming strategies.

On the basis of the distribution of the farm level productivity, it is possible to estimate the total production in the full efficiency scenario, namely, when inputs are redistributed proportionally to each individual’s agricultural skills, as shown in equation 5. We find that, if factors of production were efficiently distributed among the existing farmers, the agricultural factor productivity would more than double as a result.

Table 4 provides a comparison between our figures and the ones provided by similar studies. Reassuringly, our estimates are roughly in line with the existing ones. In particular, it is worth to point out that the measure of misallocation resulting from our analysis is rather similar to the one by Gollin and Udry (2017) for Uganda, although they use different techniques and data. It is also in line with the one by Ayerst et al. (2018) for Vietnam, another country characterized by relatively well functioning and active land markets.

On the other hand, misallocation seems to be less severe than in contexts where land markets have just recently been liberalized such as Malawi and Ethiopia, suggesting that in fact, facilitating agricultural input transactions has a positive impact on the efficiency of factor distribution. However, unlike the case of the two above mentioned studies, there seems not to be any significant difference in the magnitude of within group misallocation between farmers who operate land acquired through the market and farmers who do not. Overall, these findings suggest that input distribution (even within farmers who bought and/or rent land in) can be far from efficient also in the absence of barriers to market based land transactions.

Intuitively, misallocation reflects the fact that more productive individuals do not necessarily operate larger quantities of inputs, and that consequently the marginal productivity of factors of production differs greatly across farmers and is a positive function of their productivity. Figures 5 and 6 confirm that the distribution of both land and capital seems to be virtually independent of farm specific productivity.

Furthermore, we show that a relevant productivity gain could be achieved if resources were allocated efficiently within the villages in our sample. Namely, redistributing land and capital at the village level in proportion to the farmers’ productivity would increase the value of the
agricultural output by nearly 40 percent. As land quality, cropping patterns as well as weather shocks experienced are arguably homogeneous within villages, this finding provides some additional evidence that resources are not efficiently allocated.

In the following section, we will provide some suggestive empirical evidence that the existing misallocation might be caused by frictions in the food market rather than restrictions in land or capital transactions.

5.2 Testing the model’s predictions

The key intuition conveyed by the model is that inefficient input distribution might be the result of the fact that the shadow value of the agricultural output differs depending on whether the farmer produces mostly for self consumption or for commercial purposes. Consequently, since in equilibrium farmers operate a level of input that equalises the marginal utility from an additional unit of production, the marginal productivity should be lower for subsistence farmers.

As individuals in our sample tend to cultivate only plots located in their village, it is sensible to compare marginal productivities of farmers living in the same enumeration area. Formally, we compute the regression:

$$\log(MP_{jiv}) = \beta_1 \text{subsistence index}_{iv} + \nu_v$$  \hspace{1cm} (19)

where the marginal productivity of input $j$ for farmer $i$ in the enumeration area $v$ is regressed on some variable capturing whether $i$ is a commercial or a subsistence farmer as well as a set of village level fixed effects.

As expected, subsistence farmers display significantly lower marginal productivity of both land and capital. As shown in Table 5, the magnitude of the gap is all but negligible. When considering the percentage of agricultural output self consumed, we estimate a 30 and 63 percent difference in marginal productivity of land and capital between a farmer consuming all of her output and one selling the whole produce to the market. As the specification is log-linear, this translates into an estimated 0.3 and 0.6 percent decrease of marginal productivity of labour and capital for each percentage point increase in the share of output consumed. In columns 2 and 4 we estimate an alternative regression that mirrors the model more closely where we define as subsistence farmers only those who consumed their entire production and market oriented farmers those who sold at least 66 percent of their output. The estimates show that most of the gap is actually due to the difference between purely subsistence farmers (that produce for self consumption only) and the others. Unsurprisingly, the coefficients of

\[\text{In the following, we will use the terms village and enumeration area interchangeably. The latter refers to the smallest geographical statistical unit in the UCA.}\]

\[\text{According to the model, the marginal productivity should be equalized across all the farmers who sell a non zero share of their output and so we should only observe a gap between pure subsistence farmers and the others. However, it is realistic to assume that farmers selling a high share of their produce and semi subsistence farmers who only sell a very small amount of output (but are still potentially net buyers in the market for food) attribute different shadow value to their production.}\]
the “market oriented farm” dummy are positive for both inputs, but their magnitude is much smaller and not significantly different from zero in the case of capital.

These findings indicate that factors of production should be transferred from subsistence to more market oriented farmers in order to achieve within village efficiency.\textsuperscript{32} This can be shown and tested directly by comparing the actual land distribution ($l_i$) to the one in the within village efficiency counterfactual: ($l_{ev}^i$).\textsuperscript{33} Once again, we set the analysis at the village level as the bulk of land transfers happens within enumeration areas.

We compute two regressions aimed at comparing the actual to the efficient land scale. The first one consists in a simple OLS where the dependent variable is the log difference between the actual and the efficient farm scale, formally:

$$log(l_i) - log(l_{ev}^i) = \beta_1 \text{ subsistence index}_i + \nu_v$$

(20)

where once again subsistence is captured either by the percentage of output value self consumed (entered linearly) or by the two dummies included in the previous estimation.

The second one is a probit model where the binary outcome dependent variable takes value 1 when the farmer $i$ is operating a larger than efficient share of land:

$$Pr(l_i > l_{ev}^i) = \Phi(\beta_1 \text{ subsistence index}_i + \nu_v)$$

(21)

The results of these regressions are shown in Table 6. Consistently with our hypothesis, subsistence farmers are operating on a larger than efficient scale both on the intensive and on the extensive margin.

According to the linear specifications with village fixed effects, a percentage point increase in the value of the output consumed increases the gap between the actual and efficient land by 0.7 percent. This results in a 70 percent difference between a farmer who only operate for self consumption and a fully commercial farmer. The results do not change when farmers are divided in pure subsistence, market oriented and the residual (median) category. Pointedly, subsistence (market oriented) farmers are estimated to operate at a 63 percent larger (16 percent lower) than efficient scale as compared to the reference group.

The results on the probability of operating at a larger than efficient scale point in the same direction, indicating that subsistence or close to subsistence farmers are more likely to be cultivating too large an amount of land as opposed to the efficient counterfactual. The average marginal effects are both statistically and economically significant across all the specifications. Depending on whether village fixed effects are included or not, we estimate a 0.09 or 0.17 percentage points higher probability of operating a larger than efficient share of land for each percentage point increase in output consumed. Dividing the farmers in the

\textsuperscript{32}Note that this is not true for all the farmers belonging to the group $b$ as the cost of input $r$ will be affected by the redistribution so that some of them might actually end up with lower share of land than the full efficiency counterfactual.

\textsuperscript{33}We present and discuss the findings for land as it represents by far the most important tradable factor of production, but it can be shown that the results are very similar for capital input.
usual three categories returns similar results, as we estimate that purely subsistence farmers are 18 percentage points more likely to cultivate a larger than efficient share of land than market oriented ones (when village fixed effects are included).

Estimates presented in Tables 5 and 6 show that the input distribution within villages is consistent with the patterns presented in the model, where in equilibrium farmers close to subsistence operate a higher than efficient shares of input and consequently have lower marginal productivity. It is worth highlighting that rather than being the result of a more or less arbitrary allocation procedure by the village leaders (potentially motivated by personal preferences or equality concerns), the observed factor distribution is the outcome of market based transactions, as in most of the villages (see Table 2) a relevant share of the farmers are operating land that has been acquired, either permanently or temporally, through the market. Thus, it might be argued that misallocation is not necessarily due to frictions in land markets, but rather the consequence of the different shadow values individuals attach to their produce.

In order to corroborate our model further, we test empirically whether resource misallocation is actually more severe within subsistence farmers in a given area. As pointed out in the theoretical framework, this would happen as in equilibrium subsistence farmers, unlike commercial ones, display different marginal productivities (see Equation 16).34 Although the theoretical model is framed in a way that only differences in marginal productivity dispersion between subsistence and commercial farmers within a village are captured, we generalize this conclusion and test it for different commercialization groups $g$ and broader administrative units $z$.

Formally, we estimate a number of regressions in the following form:

$$e_{zg} = \sum_g \beta_g \text{group}_{zg} + \gamma N_{zg} + \nu_z$$

where depending on the specification we will adopt different ways to define the commercialization groups $g$ and the administrative units $z$. The resulting estimates are depicted in Table 7. On the one hand, in columns 1, 3 and 5 we compare strictly subsistence farmers (with no output sold) to those who have sold a non zero quantity of their production. On the other, in columns 2, 4 and 6 farmers in each zone $z$ are divided in three quantiles depending on the percentage of farm output sold.

As for the geographical level $z$, we run the analysis at the three smallest administrative units available: villages (columns 1 and 2), parishes (columns 3 and 4) and subcounties (columns 5 and 6). Moving the analysis to a larger zone reduces the number observations but improves the credibility of the estimates as the misallocation is computed for a larger number of individuals per group.

34In the model, the misallocation arises within constrained farmers (belonging to group $b$) as well as across groups (as the more skilled constrained farmers have higher marginal productivity). By allowing for decreasing marginal utility in food consumption, there would be misallocation also within non constrained farmers. As long as the frictions in output markets have also an individual specific component, we would expect some (although less pronounced) misallocation within commercial farmers as well.
The dependent variable $e_{zg}$ is the TFP gain that would be achieved by reallocating resources efficiently among the farmers belonging to the commercialization group $g$ in zone $z$. The explanatory variables of interest are the dummies indicating each group (subsistence versus non subsistence for columns 1, 3 and 5 or the consumption tertiles for columns 2, 4 and 6). We follow Shenoy (2017) and include the number of farmers per group $N_{zg}$ as a control along with a set of zone specific fixed effects.

Consistently with our expectations, we find that for each of the administrative units considered, misallocation within subsistence farmers is higher than within commercial ones. The relative magnitude of the difference between the two groups is relatively constant for each geographical unit as larger absolute values of the coefficients for broader administrative areas are associated to higher average levels of within group misallocation. Pointedly, subsistence farmers are estimated to have 14.1, 14.2 and 18.9 percent larger gains from efficient factor reallocation than commercial farmers within villages, parishes and subcounties respectively.

The main message remains the same when the comparison is made across consumption tertiles. Indeed, we find significantly higher misallocation within farmers who belong to the third one, namely those who self consume the highest share of their production. This suggests that factors of production are in fact less efficiently distributed among subsistence or close to subsistence farmers and that differences in output shadow values originated by frictions in the output markets are a good candidate to explain misallocation in a context where land markets seem to operate with virtually no frictions.\(^{35}\)

Finally, we want to test whether the severity of input misallocation at the village level is positively correlated to the predominance of subsistence agriculture. More specifically, according to our theoretical framework, more pronounced output market frictions $\tau$ would incentivise subsistence farming and in turn result in factor misallocation across farmers.

In practice, we will estimate some regressions taking the general form:

$$e_v = \beta_1 \text{ subsistence index}_v + \beta_2 \% \text{ land market participation}_v + \gamma \log [\text{sd}(\text{TFP})_v] + \delta_d \quad (23)$$

where the village level severity of input misallocation (or potential gain from efficient reallocation) is regressed on different indexes meant to measure the pervasiveness of subsistence farming in the village. Additionally, we include the percentage of farmers participating in land markets to check whether, in line with Restuccia and Santaeulalia-Llopis (2017), misallocation is less pronounced where land transactions are more widespread. Following Chen et al. (2017), we control for the village specific dispersion in total factor productivity to make sure that the estimates are not merely capturing different productivity distributions across villages. Finally, a set of district or county fixed effects are added to control for location specific factors that might affect the level of misallocation and/or subsistence agriculture.

\(^{35}\)As individuals belonging to first and second groups typically sell a non trivial fraction of their total production, the fact that they do not present different level of inefficiency is consistent with the theoretical framework since, unlike subsistence or close to subsistence farmers, they would all make production decisions based on the same shadow value.
The findings reported in Table 8 corroborate our hypothesis. Villages where the median farmer consumes a higher share of her produce, that present a higher percentage of purely subsistence farmers and where a larger fraction of the total output is self consumed present lower efficiency in agricultural factor distribution. Unlike Restuccia and Santaeulalia-Llopis (2017), we do not find any significant negative correlation between misallocation and land market participation at the village level. A possible interpretation of this finding is that the positive impact on efficiency of land market liberalization is only captured in the immediate aftermath of the reforms when previously constrained farmers are allowed to exchange their land and make the most of the existing, yet unrealized, gains from trade. However, in the long run and in presence of other frictions like the ones described in the model, there might no longer be any gain from trade left and therefore land transactions would not necessarily result in more efficient input distribution.36

It is worth remarking that the coefficients of the subsistence indexes are all significant at least at the 95 percent confidence level across all the specifications, suggesting that misallocation can indeed be the result of frictions in the food markets incentivising subsistence farming. Their magnitude is however quite low: a standard deviation increase in the percentage of consumed output by the median farmer or in the share of the total village output self consumed increases the gravity of misallocation of only one tenth of standard deviation. This does not necessarily come as a surprise as it seems that most of the variation in the dependent variable is captured by the underlying village level productivity distribution.

To sum up, the empirical analysis shows that even though land markets are very active in Uganda, agricultural factors of production are far from being efficiently distributed across farmers. Namely, after controlling for land quality and weather shocks, we find that reallocating efficiently the existing inputs among the existing farmers would more than double total factor productivity.

We find that farmers operating in subsistence or close to subsistence present systematically lower marginal productivity than those who commercialise a non trivial share of their produce. This implies that in order to achieve the potential efficiency gains mentioned above, land and capital should be transferred from individuals who cultivate mostly for self consumption to more market oriented producers. As land markets are very active and virtually frictionless, this suggests that efficiency improving transactions do not take place because subsistence farmers attribute higher value to their output and are not willing to separate from their land and agricultural capital. Consistently with this hypothesis, we show that village level resource misallocation is uncorrelated to land market activity but is increasing in the prevalence of subsistence farming.

These results shed new light on the real causes of misallocation in Sub Saharan Africa and provide some relevant insights for policy makers. Indeed, they suggest that although the abatement of barriers to agricultural factors (and crucially land) transactions is a necessary step in order to tackle input misallocation, it might not be sufficient to guarantee an efficient

---

36In the model, although the equilibrium features different marginal productivity across farmers, there are no possible gains from trade as less productive individuals attribute higher value to their output and so are not willing to separate from their factors of production.
resource distribution among farmers. In particular, we highlight that reducing misallocation requires a redistribution of land and capital from subsistence to commercial farmers and therefore would entail the development of an effective strategy to tackle the existing incentives towards subsistence agriculture.

6  Robustness checks and further discussion

This section aims at providing some additional evidence to corroborate the empirical findings listed so far. In the first subsection, we disentangle the component of misallocation due to inefficient scale of the different holdings from the one due to inefficient input mix within farms. Intuitively, if inputs markets are actually efficient as assumed in the theoretical framework, farmers should operate with the efficient mix of land and capital input and misallocation should only reflect inefficiencies in the scale of the holdings.

In the second subsection, we perform the same regressions on the subsample of farmers who do not cultivate coffee or devote just a very small share of their land to that crop. The rationale for this is that while other crops included in the analysis are relatively homogeneous in terms of market value, production process and level of marketization, coffee is characterized by a much less labour intensive production process, higher value added and high level of marketization. For these reasons, there might be concerns that our results might be capturing differences between coffee growers and the other farmers, rather than between subsistence and market oriented farms.

6.1  Mix vs scale misallocation

In developing the theoretical framework, we abstracted from the difference between land and capital and combined them in the fictional composite input $\xi$. This choice is consistent with the assumption that factors of production can be exchanged freely and that consequently farmers can and will operate with the most efficient mix of the two tradeable inputs. As a result, according to our model, input misallocation is the result of inefficient scale distribution of the different holdings (with subsistence farmers operating too high a fraction of the inputs) rather than of inefficient mix of land and capital within farms.

To a certain degree, it is possible to test whether this assumption is credible, following the methodology proposed by Shenoy (2017) to disentangle resource misallocation deriving from inefficient input mix and from inefficient scale distribution of farms. More specifically, the share of misallocation attributable to an inefficient choice of input mix can be computed by comparing the actual production ($Y$) to the one achieved by the existing farmers if they were operating an output maximizing mix of factors ($Y^*$).

The distribution of land ($l_1^*, \ldots, l_N^*$) and capital ($k_1^*, \ldots, l_N^*$) in the counterfactual where farms have the same scale but adopt the efficient mix of factors of production can be characterised
as the solution to the following maximization problem:

$$\max \sum_{i}^{N} y_i \text{ subject to } pl_i + ck_i = p\bar{l}_i + c\bar{k}_i,$$

$$\sum_{i}^{N} l_i = \sum_{i}^{N} \bar{l}_i = L \text{ and } \sum_{i}^{N} k_i = \sum_{i}^{N} \bar{k}_i = K$$  \hspace{1cm} (24)

where, $\bar{l}_i$ and $\bar{k}_i$ represents the amount of land and capital actually used by each farmer $i$.

Differently from the case of the social planner maximization problem depicted in equation 4, the “scale” constraint $pl_i + ck_i = p\bar{l}_i + c\bar{k}_i$ is added to the resource constraint. Intuitively, it imposes that the value of the optimal $l$ and $k$ must be the same as the one of the inputs actually utilized in the agricultural production at the market clearing prices $p$ and $c$ so that the resulting allocation leaves each farm scale unaffected. The first order conditions imply:

$$\frac{l_i^*}{l} = \frac{k^*}{k} = \left[ \frac{\bar{l}_i}{L} + \frac{1 - \alpha}{\bar{k}_i} \right]$$  \hspace{1cm} (25)

The resulting output for each farmer $i$ is $y_i^* = s_i^{1-\gamma}(l_i^*\alpha k_i^*(1-\alpha))^{\gamma}$. Therefore, the misallocation caused by suboptimal input mix within farms can be computed as:

$$\frac{\sum_{i}^{N} y_i^*}{\sum_{i}^{N} y_i} = \frac{Y^*}{Y}$$  \hspace{1cm} (26)

Our estimate of the ratio is 1.16, indicating that by reallocating inputs optimally across farmers maintaining their operational scale unaffected, the overall productivity would increase by a mere 16 percent.

It can be shown that the total gain from reallocation $G = \frac{Y^* - Y}{Y}$ can be decomposed as the sum between the gain from redistributing factors to guarantee an optimal mix of land and capital $G^{MIX}$ plus the gains from redistributing the inputs (in their optimal combination) proportionally to the farmers’ agricultural skills $G^{SCALE}$, where $G^{MIX} = \frac{Y^* - Y}{Y}$ and $G^{SCALE} = \frac{Y^* - Y^*}{Y}$. This implies that the total misallocation estimated in Section 5.1 is predominantly caused by inefficiencies in the distribution of farms’ scale, as postulated by the model. In particular, in our sample the gains from an efficient redistribution of factor mix within farms would reduce the total misallocation only by 14 percent. The predominance of scale misallocation is even more pronounced when considering smaller administrative units. At the village level, the median contribution of inefficiencies in input mix distribution to the total misallocation is only 3 percent (mean = 5 percent), the relative figure at the subcounty and district level is of 5 (7) and 9 (11) percent respectively.

It is important to mention that this methodology presents some relevant flaws. Most remarkably, as explained in Shenoy (2017), the results should be interpreted as a lower bound of the relative importance of factor market misallocation as imperfections in land and capital
markets might affect the scale and not only the mix of the input used. Intuitively, perfectly functioning input markets should at least guarantee a perfect combination of the tradable factor at the farm level, and this procedure only allows us to assess to which extent this is true. Whether the remaining amount of misallocation is actually due to other factors affecting the scale of the farms or to frictions in the input markets is a question that cannot be answered by this decomposition. Nevertheless, the fact that the input combination at the farm level seems to be almost fully optimal (especially when considering small geographical areas) represents a valuable argument in support of the assumption of well functioning factor markets.\textsuperscript{37}

6.2 Subsample analysis: non coffee farmers

A potential concern is that the results presented might be driven by different crop choices rather than reflecting real differences in farmer specific productivity. In particular, it might be the case that farmers who are labeled as commercial in the empirical analysis present higher level of productivity because they are growing cash crops with higher market value (or different input intensities in the production process) rather than having better agricultural skills. Although this is an equally interesting issue that deserves more in depth research, it would lead to a fundamentally different interpretation of the findings and would ultimately be inconsistent with the model and the assumption of homogeneity in the output.

Among the 18 crops listed in the survey, coffee is by far the most problematic from this point of view. More specifically, being an permanent crop it arguably presents very different input intensities and is realistically less labour intensive than most of the other agricultural outputs.\textsuperscript{38} Additionally, it has very high market value and is predominantly marketed by the farmers, as the consumption levels are mostly trivial: in the dataset, more than 98 percent of the coffee produced was either stored or sold. It follows that farmers who devote a high share of their inputs to coffee production might mechanically display better agricultural skills and higher output marketization.

In order to make sure that the results are not only capturing different mix of crop across different holdings, we re estimate all the regressions after dropping the observations for these farmers who devoted at least one third of their land to coffee cultivation, that represent slightly more than 10 percent of the sample.\textsuperscript{39}

First of all, it is important to point out that the resulting estimates of misallocation obtained are perfectly in line with the ones obtained with considering the full sample. The total TFP gain from an efficient reallocation of resources is now 117 percent, virtually identical to the

\textsuperscript{37}In the only comparable study, Shenoy (2017) finds a considerably more relevant contribution of misallocation in input combination, ranging from 36% in 1996 to 75% in 2008 in the case of Thailand.

\textsuperscript{38}As we do not have crop specific input use but only holding level data, we are not able to test this hypothesis directly.

\textsuperscript{39}In the average farm, 7 percent of the land is cultivated with coffee. However, the distribution is very skewed with many holdings almost exclusively specialized in coffee and a high share of farmers growing just very small amount of that crop. The results are robust to the exclusion of all the farmers who grew a non zero quantity of coffee, but this would imply a consistent reduction of the sample size (around 30 percent).
previous one. This indicates that a large part of the inefficiency in resource distribution takes place within farmers who do not predominantly grow coffee and therefore we can exclude that our results are driven by structural differences between coffee farmers and the others.

Furthermore, all the findings listed in Tables 5 to 8 are left virtually unchanged in their magnitude and significance levels, as can be checked in Tables 9 to 12.

7 Conclusion

The wide gap in agricultural productivity between poor and rich countries is one of the most debated and crucial topics in the development literature. As shown by Adamopoulos and Restuccia (2018), the huge disparities observed cannot be attributed to systematic differences in land endowments but rather must be the outcome of sub optimal production patterns. A particularly appealing explanation for them is that, as markets for agricultural inputs and output fail systematically in low income countries (see Dillon and Barrett 2017 for a recent study on Sub Saharan Africa), resources are not allocated efficiently among existing farmers. In a similar fashion to Hsieh and Klenow (2009), the main hypothesis is that consistent productivity gains could be achieved by redistributing the existent factors of production from less to more productive individuals.

Recent studies have corroborated this theory, showing that in a number of developing countries input distribution is virtually uncorrelated to farmer specific productivity, resulting in significant aggregate productivity losses that could be avoided if more (less) skilled producers operated on a larger (smaller) scale. With some relevant but scarce exceptions, the existing literature has imputed this misallocation to distortions in the market for agricultural inputs, and in particular to legal or informal restrictions in land transactions, that are widespread in the developing world.

It is difficult to overstate the importance of liberalizing land markets in order to achieve a more efficient allocation of resources, and the evidence provided in that sense are robust and persuasive. However, as stressed by virtually all existing studies, high levels of misallocation seem to persist also in countries where land markets appear to be active and well functioning. This is typically attributed to the occurrence of more subtle failures and imperfections in land markets that are more difficult to observe. Although this view has some merit, it is not totally convincing.

This paper aims at addressing this gap by considering the concurrent role of distortions in output markets in determining resource misallocation in agriculture. In particular we explore the possibility that, even in the presence of frictionless input markets, severe misallocation can arise when, due to marketization costs and other similar frictions, farmers find it more convenient to grow the food they consume rather than buying it at the market. This idea derives from a well established strand of literature on farm household choices dating back to De Janvry et al. (1991) and recently reconsidered by Li (2017).

We provide a simple theoretical model showing that, even when capital and land can be traded freely, distortions in the food markets providing incentives towards subsistence farming
might ultimately result in inefficient resource distribution. The hypotheses of the model are tested using data from Uganda, a country where land markets are well established and very active.

The estimated magnitude of the productivity gain from reallocation is around 120 percent, somewhat lower than the figures provided for other African countries with more recently established and less active land markets, but still far from insignificant. This is consistent with a scenario where functioning inputs markets are a necessary, yet not sufficient condition to achieve an efficient distribution of agricultural inputs across farmers.

In line with the model proposed, we find that subsistence farmers operate systematically on a larger than efficient scale. This suggests that they attribute a higher value to their output and therefore are reluctant to separate from their land and agricultural capital. This prevents more productive commercial farmers from increasing their holdings’ size. Misallocation therefore seems to be caused by frictions in food markets that induce a wedge between subsistence and commercial farmers’ shadow prices rather than explicit barriers to input transactions. Consistently with our hypothesis, we find a positive correlation between the village level magnitude of misallocation and the prevalence of subsistence farming. In contrast to to previous studies, frequency of land transactions does not seem to play any role.

To our knowledge, this is the first paper that establishes a link between subsistence farming and agricultural factor misallocation. This represents a relevant contribution as it identifies a new channel that might explain persisting inefficiencies in resource distribution even in contexts where input markets appear to be functioning. Additionally, it provides some valuable insights to policy makers by questioning the simplistic view according to which liberalizing land markets is a sufficient condition to abate misallocation. More specifically, we argue that complementary policies/investment aimed at reducing frictions in food markets are needed to achieve a efficient input distribution across farmers and enhance agricultural productivity.

Appendix

Social planner problem

The solution for the social planner problem can be obtained by plugging in the production function (equation 3) in equation 4, and obtaining the Lagrangian:

$$L = \sum_{i} s_{i}^{1-\gamma}(l_{i}^{\alpha}k_{i}^{1-\alpha})^{\gamma} - \lambda_{1} \left( \sum_{i} l_{i} - L \right) - \lambda_{2} \left( \sum_{i} k_{i} - K \right)$$

(A1)
with first order conditions:

\[
\alpha \gamma s_i^{1-\gamma} l_i^{\alpha \gamma - 1} k_i^{(1-\alpha)\gamma} = \lambda_1 \\
(1 - \alpha) \gamma s_i^{1-\gamma} l_i^{\alpha \gamma} k_i^{(1-\alpha)\gamma - 1} = \lambda_2
\]

\[
\sum_i^N l_i = L \\
\sum_i^N k_i = K
\]

By taking the ratio of the first two conditions for a generic farmer \(i\), we obtain that:

\[
\frac{k_i}{l_i} = \frac{1 - \alpha}{\alpha} \frac{\lambda_1}{\lambda_2} l_i
\]

that states that the mix of capital and land inputs must be the same for each the farmer \(i\).

By plugging this into equation \(A2a\) and solving for \(\frac{k_i}{l_j}\) for any two farmers \(i\) and \(j\), it can be found that:

\[
\frac{k_i}{l_j} = \frac{s_i}{s_j}
\]

this implies that land (and consequently capital, given equation \(A3\)) input is distributed across farmers proportionally to their agricultural skill \(s\). In turn, given the resource constraints (equations \(A2c\) and \(A2d\)), this results in:

\[
l_i = \frac{s_i}{\sum_j^N s_j} L \quad \text{and} \quad k_i = \frac{s_i}{\sum_j^N s_j} K
\]

that is the final result presented in Section 3.

**Equilibrium with \(\tau = 0\)**

The maximisation problem of farmers who do not face any market frictions is trivial and boils down to:

\[
\max s_i^{1-\gamma} \xi_i^{\gamma} - r \xi_i
\]

with first order condition:

\[
\gamma s_i^{1-\gamma} \xi_i^{\gamma - 1} = r
\]

Intuitively, this means that the the marginal productivity is constant across farmers. As a result, the quantity of output operated in equilibrium will be proportional to the farmers’ ability:

\[
\xi_i = \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i
\]
The cost of input $r$ can be obtained as the market clearing price given the exogenous total input available $\Xi$ and plugging $A8$ into the resource constraint:

$$\sum_{i}^{N} \left( \frac{\gamma}{r} \right)^{1-\gamma} s_i = \Xi$$  \hspace{1cm} (A9)

Solving for $r$ leads to:

$$r = \gamma \left( \frac{\sum_{i} s_i}{\Xi} \right)^{1-\gamma}$$  \hspace{1cm} (A10)

This expression represents a useful benchmark to assess how changes in the output frictions $\tau$ affects $r$ and in turn the distribution of the input across the farmers. Moreover, this can be plugged into equation $A8$ obtaining:

$$\xi_i = \left( \frac{s_i}{\sum_{i} s_i} \right) \Xi$$  \hspace{1cm} (A11)

that shows explicitly that, when $\tau = 0$, the share of inputs operated in equilibrium by each farmer $i$ is directly proportional to their ability, as in the social planner case.

**Equilibrium with $\tau = 1$**

In this case, farmers do not gain any utility from exceeding the consumption threshold $\overline{y}$, therefore, after plugging in the production function, the maximisation problem becomes:

$$\max s_i^{1-\gamma} \xi_i^\gamma - r \xi_i \text{ subject to } s_i^{1-\gamma} \xi_i^\gamma \leq \overline{y}$$  \hspace{1cm} (A12)

and the corresponding Lagrangian is:

$$\mathcal{L} = s_i^{1-\gamma} \xi_i^\gamma - r \xi_i - \mu(s_i^{1-\gamma} \xi_i^\gamma - \overline{y})$$  \hspace{1cm} (A13)

whose first order conditions are:

$$(1 - \mu) \gamma s_i^{1-\gamma} \xi_i^{\gamma-1} = r$$  \hspace{1cm} (A14a)

$$s_i^{1-\gamma} \xi_i^\gamma \leq \overline{y}$$  \hspace{1cm} (A14b)

$$\mu \geq 0$$  \hspace{1cm} (A14c)

$$\mu(s_i^{1-\gamma} \xi_i^\gamma - \overline{y}) = 0$$  \hspace{1cm} (A14d)

It follows that in equilibrium there will be two different types of farmers, those for whom the constraint is binding ($\mu > 0$) who will use just enough input $\xi$ to fulfil their consumption need, and the ones for whom the constraint is not binding, who will rent input $\xi$ until the marginal productivity equals the marginal cost $r$.

The former will therefore operate a share of input that is negatively related to their productivity, as more able farmers will need less composite output to produce $\overline{y}$. Formally:

$$\xi_i = \left( \frac{\overline{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}}$$  \hspace{1cm} (A15)
On the other hand, unconstrained farmers will operate a level of input such that the marginal productivity equals the marginal cost $r$, and so:

$$\xi_i = \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i$$  \hspace{1cm} (A16)

although this condition is exactly the same as the one for each farmer in the case of $\tau = 0$ depicted in equation A8, it is worth to point out that it underlies a different quantity and share of inputs since the frictions in the output market may affect the equilibrium price $r$ as it will be made explicit in the following.

Farmers belong to the first or the second group depending on their innate productivity $s$. More specifically, the threshold above which the constraint can be derived by plugging equation A15 in the marginal productivity formula and setting it $= r$. By doing so, we will find the productivity level at which producing exactly $y$ results in a marginal productivity that is equal to the cost of input $r$. Formally:

$$\gamma s_i^{1-\gamma} \left(\frac{y}{s_i^{1-\gamma}}\right)^{\gamma-1} = r$$  \hspace{1cm} (A17)

Solving for $s_i$ leads to:

$$s_i = y \left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}}$$  \hspace{1cm} (A18)

thus, all farmers with lower productivity, (identified as $i^-$ in the following) will not be constrained, while farmers with higher agricultural skills ($i^+$) will.

The new equilibrium price of input $r$ can be found by imposing the resource constraint. The condition implies that the total input used by unconstrained and constrained farmers must add up to $\Xi$, that is to say the total quantity of input available in the village. In formal terms:

$$\sum_{i^-} \xi_i + \sum_{i^+} \xi_i = \sum_{i^-} \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i + \sum_{i^+} \left(\frac{y}{s_i^{1-\gamma}}\right)^{\frac{1}{\gamma}} = \Xi$$  \hspace{1cm} (A19)

where the first equality is obtained by plugging in the equilibrium value of input used by non constrained (equation A16) and constrained (equation A15) farmers.

Solving for $r$ leads to:

$$r = \gamma \left(\frac{\sum_{i^-} s_i}{\Xi - \bar{y}^{1/\gamma} \sum_{i^+} \frac{1}{s_i^{(1-\gamma)/\gamma}}}\right)^{1-\gamma}$$  \hspace{1cm} (A20)

where the numerator of the ratio between brackets represents the sum of the ability of the unconstrained farmers, while the denominator is the total land used by unconstrained farmers (that is to say the difference between the total input $\Xi$ and the the fraction used by constrained farmers to produce their consumption thresholds $\bar{y}$). The difference with respect
to the formula for $r$ when $\tau = 0$ is immediate and can be appreciated comparing the last finding to equation A10. Depending on the quantity of total input available $\Xi$ and the farmers’ ability distribution, different scenarios might occur.

In the extreme case where:

$$\Xi \geq \sum_i \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}}$$  \hspace{1cm} (A21)

there is more than sufficient land available for all the farmers to rely on self production to meet their consumption needs $\bar{y}$. This implies that in equilibrium all the farmers will be constrained, or more formally that $i^- = \emptyset$ and therefore the resulting cost of input $r$ will be equal to zero. This reflects the fact that the marginal value every farmer attaches to a marginal unit of input is null as by assumption when $\tau = 1$ they are not able to sell their excess production. In the resulting equilibrium, the share of input operated by each farmer will be a negative function of their agricultural ability, as less skilled farmers need more input to produce the same amount of food.

When instead $\Xi < \sum_i \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}}$, it can be shown that the resulting equilibrium price $r$ will be lower than the one in the case with $\tau = 0$ as long as some very reasonable conditions hold, Namely, from equations A10 and A20:

$$\gamma \left( \frac{\sum i^{-}s_i}{\Xi - \bar{y}^{1/\gamma} \sum i^{+} 1/s_i^{(1-\gamma)/\gamma}} \right)^{1-\gamma} < \gamma \left( \frac{\sum i s_i}{\Xi} \right)^{1-\gamma}$$  \hspace{1cm} (A22)

whenever, for at least one farmer $i$:

$$\frac{s_i}{\sum i s_i} \Xi < \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}}$$  \hspace{1cm} (A23)

or, solving for $s_i$:

$$s_i > \bar{y} \left( \frac{\sum i s_i}{\Xi} \right)^{\gamma}$$  \hspace{1cm} (A24)

this condition implies that for at least one farmer, the ability $s$ is high enough for the total input operated in the case of $\tau = 0$ to be higher than the one necessary to produce $\bar{y}$. Intuitively, this brings about a reduction of the input price because the factors of production would be transferred from more productive (constrained) farmers, to less productive ones who are willing to pay less. The resulting allocation of input is the one depicted by the dashed green line in Figure 4.

In particular, in equilibrium all the farmers with $s_i \leq \bar{y} \left( \frac{\sum i^{-}s_i}{\Xi - \bar{y}^{1/\gamma} \sum i^{+} 1/s_i^{(1-\gamma)/\gamma}} \right)^{\gamma}$ will not be constrained and operate a share of land proportional to their ability, while the others will operate just enough input to meet their subsistence threshold $\bar{y}$. As a result, all farmers with ability higher (lower) than $\bar{y} \left( \frac{\sum i s_i}{\Xi} \right)^{\gamma}$ will be operating a lower (higher) than efficient
quantity of inputs. Note that as \( s_i \leq \frac{\sum_{i=1}^{s_i} - s_i}{\Xi} \frac{\sum_{i=1}^{s_i}}{1 - \gamma} \), there will be some constrained farmers (who only operate a amount of input sufficient to produce \( \bar{y} \)) who are nevertheless operating more input than in the \( \tau = 0 \) counterfactual, this is a result of the reduction in the input price \( r \).

If on the other hand, the condition A24 does not hold for any of the farmers, none of them would be constrained as they would all be producing less than what their consumption threshold \( \bar{y} \) and, as a result, \( i^+ = \emptyset \). Therefore, the cost of input would be the same as in the case of \( \tau = 0 \). Indeed, when none of the farmers is constrained, \( \sum s_i = \sum_i s_i \) and \( \sum (\frac{1}{s_i})^{1-\gamma} = 0 \), equation A20 becomes identical to equation A10. Thus, the resulting input distribution would be exactly the same and fully efficient.

**Equilibrium with \( 0 < \tau < 1 \)**

In the intermediate case of some frictions in the output market, every farmer is facing the same maximisation problem (after plugging in the production function):

\[
\max_{\xi_i} \begin{cases} 
 s_i^{1-\gamma} - r \xi_i & \text{if } s_i^{1-\gamma} \xi_i \leq \bar{y} \\
 (1 - \tau)(s_i^{1-\gamma} \xi_i) - r \xi_i & \text{if } s_i^{1-\gamma} \xi_i > \bar{y}
\end{cases}
\]  

(A25)

Intuitively, in equilibrium commercial farmers who obtain a lower marginal utility (captured by \( 1 - \tau \)) for their production, will have a higher marginal productivity than subsistence farmers. Formally, in equilibrium commercial farmers (who produce more than \( \bar{y} \)) will have a marginal productivity equal to \( \frac{r}{1 - \tau} \) while those producing \( y_i \leq \bar{y} \) will have a marginal productivity equal to \( r \). There will be a residual category of farmers with intermediate level of ability who will produce a quantity of output exactly equal to their consumption needs \( \bar{y} \), but who would produce more if it was not for the marketization costs \( \tau \).

The last group can be characterised as those farmers whose marginal productivity is included between \( r \) and \( \frac{r}{1 - \tau} \) whenever they are producing \( \bar{y} \) using a input equal to \( \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} \), plugging this into the marginal productivity formula, the condition becomes:

\[
\bar{y} \left( \frac{r}{\gamma} \right)^{\frac{1}{\gamma}} < s_i < \bar{y} \left( \frac{r}{\gamma(1 - \tau)} \right)^{\frac{1}{\gamma}}
\]  

(A26)

In line with the terminology used in the main section, we will refer to farmers with lower productivity as belonging to group a, to farmers with productivity in the interval as in equation A26 as belonging to group b and to farmers with higher productivity as group c.

The quantity of input used can be inferred from the marginal productivity conditions in the
case of group a and c and from \( y_i = \bar{y} \) in the case of group b. By doing so, we obtain:

Group a: \( \xi_i = \left( \frac{\gamma}{r} \right)^{\frac{1}{1-\gamma}} s_i \)  
(A27a)

Group b: \( \xi_i = \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} \)  
(A27b)

Group c: \( \xi_i = \left[ \frac{\gamma(1-\tau)}{r} \right]^{\frac{1}{1-\gamma}} s_i \)  
(A27c)

In turn, we can find the formula for the cost of input \( r \) in this intermediate case just by setting the total input used by each of the three categories equal to the total input available \( \Xi \). Formally:

\[
\sum_{i^a} \xi_i + \sum_{i^b} \xi_i + \sum_{i^c} \xi_i = \Xi 
\]

(A28)

By plugging in the values in equations A27a to A27c, and solving for \( r \):

\[
r = \gamma \left[ \frac{\sum_{i^a} s_i + (1-\tau)^{1/(1-\gamma)} \sum_{i^c} s_i}{\Xi - \sum_{i^b} \left( \frac{\bar{y}}{s_i^{1-\gamma}} \right)^{1/\gamma}} \right]^{1-\gamma} 
\]

(A29)

It is easy to show that when \( \tau = 0 \) the formula collapses to equation A10 as all the farmers would belong to either group a or c and they would be equally weighted at the numerator. If instead \( \tau = 1 \) the formula would mirror to equation A20 as group c would not be populated and group a (b) would be composed by the same elements of group \( i^- \) (\( i^+ \)). Although the impact of an change in \( \tau \) depends on the productivity distribution, from a qualitative point of view a positive \( \tau \) will always result in a decrease in \( r \) as long as there is at least one farmer with productivity higher than \( \bar{y} \left( \frac{\sum_{i^c} s_i}{\Xi} \right)^{\gamma} \). More specifically, for sufficiently low \( \tau \), group b will not be populated and the lower \( r \) would be the result of a lower weighting of the more productive farmers in category c in the numerator. Higher values of \( \tau \) will also affect the groups’ composition, as more productive farmers in group a and less productive farmers in group c will move to group b. For sufficiently high \( \tau \), \( r \) as well as the resulting input distribution would mirror the situation described when \( \tau = 1 \).

Finally, In the case where for all the farmers \( s_i < \bar{y} \left( \frac{\sum_{i^c} s_i}{\Xi} \right)^{\gamma} \), the cost of input \( r \) as well as the input distribution would be independent on \( \tau \) as none of the farmers would obtain a excess production in any case.
References


The figure represents the equilibrium rental cost $r$ of the tradeable input $\xi$, as well as the marginal revenue productivity (that in this case equals the marginal productivity as the utility is normalized to 1 regardless of whether the produce is sold or consumed) for three farmers with increasing agricultural skills ($s_3 > s_2 > s_1$). In equilibrium, they will operate a share of the composite input $\xi$ that is proportional to their productivity $s$. 
Figure 2: Equilibrium with no food market

The red solid lines represent the marginal revenue productivity (in this case, the product of the marginal productivity and the marginal utility, that is 1 for \( y \leq \bar{y} \) and 0 otherwise) of the input \( \xi \) for farmers with increasing level of agricultural skills \( (s_3 > s_2 > s_1) \) and the equilibrium rental cost of the composite input \( r \). The black dotted lines represent their marginal productivity (that in the case of no frictions in the output market equals their marginal revenue productivity) and the equilibrium rental cost of capital in the frictionless scenario. Note that including frictions in the food markets always reduces the capital rental cost unless the new constraints are not binding for any of the farmers. In the figure, the constraint is binding for farmer 2 and 3 while it is not for farmer 1. As a result, the share of input operated by each of them is a negative function of their productivity.
Figure 3: Equilibrium with frictions in the food market

The red solid lines represent the marginal revenue productivity (in this case, the product of the marginal productivity and the marginal utility, that is 1 for \( y \leq \bar{y} \) and \( 1 - \tau \) otherwise) of the input \( \xi \) for farmers with increasing level of agricultural skills (\( s_3 > s_2 > s_1 \)) and the equilibrium rental cost of the composite input \( r \). The black dotted lines represent their marginal productivity (that in the case of no frictions in the output market equals their marginal revenue productivity) and the equilibrium rental cost of capital in the frictionless scenario. Note that including frictions in the food markets always reduces the capital rental cost unless the new constraints are not binding for any of the farmers (and therefore they all belong to group \( a \)).

In the figure, farmer with ability \( s_1 \) belongs to category \( a \) and in equilibrium operates a higher than efficient share of land, farmer with ability \( s_2 \) belongs to group \( b \) and is operating a lower than efficient share of land. Finally, farmer with skill level \( s_3 \) belongs to the group of commercial farmers \( s_3 \) and operates on a lower than efficient scale.
The red solid line represents the input operated in equilibrium per for each level of skills when there are no frictions in the food market. The blue pointed line represents the intermediate case with $0 < \tau < 1$ while the green dotted line the scenario where food markets do not exist and $\tau = 1$. The lines depict the “normal” situation in which the total input $\Xi$ is lower than $\overline{y}^\gamma \sum_i \left( \frac{1}{s_i} \right)^{\frac{1}{\gamma-1}}$, $\gamma = 0.597$ and there is at least one farmer $i$ such that $s_i > \overline{y} \left( \frac{\sum_i s_i}{\Xi} \right)^\gamma$. See the appendix for an explanation of these conditions.
Figure 5: Efficient vs actual land input distribution

Note: In both graphs, the black line represents the linear fit of the actual observations while the red dashed line represents the efficient distribution of input (proportional to the farmer specific productivity) and the resulting marginal productivity (equalized across agents). The land input is defined as the operational scale, that is to say the land actually used for agricultural production (regardless of whether it is owned or rented in). The correlation between land input and ability (in logs) is 0.00.

Figure 6: Efficient vs actual capital input distribution

Note: In both graphics, the black line represents the linear fit of the actual observations while the red dashed line represents the efficient distribution of input (proportional to the farmer specific productivity) and the resulting marginal productivity (equalized across agents). The correlation between capital input and ability (in logs) is -0.03 and is not statistically different from zero at any conventional level.
Table 1: Datasets used

<table>
<thead>
<tr>
<th>Main analysis</th>
<th>Name</th>
<th>Source</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input level, production and output disposition</td>
<td>UCA</td>
<td>Ubos</td>
<td>household</td>
</tr>
<tr>
<td>Estimation of $\alpha$ and $\gamma$ and output and capital value</td>
<td>LSMS</td>
<td>World Bank</td>
<td>household</td>
</tr>
</tbody>
</table>

* Adjustments in productivity*

<table>
<thead>
<tr>
<th>Rainfall</th>
<th>Quarterly rainfall</th>
<th>Source</th>
<th>4 km² resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land quality</td>
<td>GAEZ</td>
<td>FAO &amp; IIASA</td>
<td>10 km² resolution</td>
</tr>
</tbody>
</table>

* Available at the village level and only for the geolocalized subset of the sample (around 93% of the villages).

Table 2: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour (full time adult male equivalent)</td>
<td>2.92</td>
<td>2.30</td>
<td>2.08</td>
<td>24,431</td>
</tr>
<tr>
<td>Land (hectares, 2 seasons)</td>
<td>2.00</td>
<td>1.30</td>
<td>2.40</td>
<td>24,431</td>
</tr>
<tr>
<td>Capital (1000 Shillings)</td>
<td>219.722</td>
<td>21.00</td>
<td>1590.4</td>
<td>24,431</td>
</tr>
<tr>
<td>Value added per labour unit (1000 Shillings)</td>
<td>4,675.36</td>
<td>255.23</td>
<td>663.72</td>
<td>24,431</td>
</tr>
<tr>
<td>Participation to land markets (1 = yes)</td>
<td>0.70</td>
<td>1.00</td>
<td>0.46</td>
<td>24,431</td>
</tr>
<tr>
<td>% produce marketed</td>
<td>0.31</td>
<td>0.25</td>
<td>0.30</td>
<td>24,431</td>
</tr>
<tr>
<td>Subsistence farmer*</td>
<td>0.13</td>
<td>0</td>
<td>0.34</td>
<td>24,431</td>
</tr>
<tr>
<td>Market oriented farmer†</td>
<td>0.28</td>
<td>0</td>
<td>0.45</td>
<td>24,431</td>
</tr>
<tr>
<td><strong>Village level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of farmers</td>
<td>7.68</td>
<td>8</td>
<td>2.31</td>
<td>3,181</td>
</tr>
<tr>
<td>% of farmers participating in land markets</td>
<td>0.70</td>
<td>0.8</td>
<td>0.32</td>
<td>3,181</td>
</tr>
<tr>
<td>% produce marketed</td>
<td>0.40</td>
<td>0.33</td>
<td>0.23</td>
<td>3,181</td>
</tr>
</tbody>
</table>

* Subsistence farmer is a dummy that takes value 1 when the farmer does not sell any of her produce. † Market oriented farmer is a binary variable taking value 1 when the farmer sells at least two thirds of her output. The figures refer to the observations left once outliers have been dropped following the trimming procedure described in the next section.
### Table 3: Shares of agricultural inputs in selected countries

<table>
<thead>
<tr>
<th></th>
<th>Uganda</th>
<th>Malawi</th>
<th>China</th>
<th>Uganda*</th>
<th>Ghana</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour</td>
<td>0.403</td>
<td>0.42</td>
<td>0.243</td>
<td>0.389</td>
<td>0.212</td>
<td>0.46</td>
</tr>
<tr>
<td>Land</td>
<td>0.475</td>
<td>0.39</td>
<td>0.417</td>
<td>0.473</td>
<td>0.536</td>
<td>0.18</td>
</tr>
<tr>
<td>Capital</td>
<td>0.122</td>
<td>0.19</td>
<td>0.088</td>
<td>-</td>
<td>0.181</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*The alternative figures from Uganda come from Gollin and Udry (2017), who do not include capital input in the production function. The figures presented come from Restuccia and Santaeulalia-Llopis (2017) for Malawi, Chari et al. (2017) for China, Gollin and Udry (2017) for Uganda (with no capital input) and Ghana and Valentinyi and Herrendorf (2008) for the US. In the case of China, the total estimates (without allowing from crop specific shares) are presented.

### Table 4: Comparison to existing studies

<table>
<thead>
<tr>
<th></th>
<th>Uganda</th>
<th>Malawi</th>
<th>Ethiopia</th>
<th>China</th>
<th>Vietnam</th>
<th>Ghana</th>
<th>Uganda</th>
<th>Thailand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UCA</td>
<td>Malawi</td>
<td>Ethiopia</td>
<td>China</td>
<td>Vietnam</td>
<td>Ghana</td>
<td>LSMS</td>
<td></td>
</tr>
<tr>
<td>Nationwide analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total gain</td>
<td>2.19</td>
<td>3.59</td>
<td>3.07</td>
<td>1.57</td>
<td>1.78</td>
<td>1.16</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<td>Operating marketed land</td>
<td>2.22</td>
<td>3.00</td>
<td>2.61</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Non operating marketed land</td>
<td>2.11</td>
<td>4.60</td>
<td>3.18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>% operating market land</td>
<td>70.1</td>
<td>16.6</td>
<td>24.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Village level analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total gain*</td>
<td>1.38</td>
<td>1.65</td>
<td>-</td>
<td>1.27</td>
<td>-</td>
<td>-</td>
<td>1.50</td>
<td>1.19</td>
</tr>
<tr>
<td>% Contribution</td>
<td>41</td>
<td>39</td>
<td>-</td>
<td>53</td>
<td>-</td>
<td>-</td>
<td>60</td>
<td>-</td>
</tr>
</tbody>
</table>

* In this case, total gain refers to the aggregate TFP gain that would be obtained by reallocating resources efficiently within villages. The estimates are derived using different methodologies and as such they are not directly comparable. The figures from Malawi come from Restuccia and Santaeulalia-Llopis (2017), for Ethiopia from Chen et al. (2017), for China from Adamopoulos et al. (2017), for Vietnam from Ayerst et al. (2018), for Ghana and Uganda (using LSMS) from Gollin and Udry (2017) and from ? for Thailand.
Table 5: Marginal productivity of factors

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>MPL (log)</th>
<th>MPK (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>% output consumed</td>
<td>-0.266***</td>
<td>-0.486***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Subsistence farm</td>
<td>-0.224***</td>
<td>-0.472***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Market oriented farm</td>
<td>0.067**</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>N</td>
<td>24,292</td>
<td>24,292</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.562</td>
<td>0.565</td>
</tr>
</tbody>
</table>

Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Village FE are included in each specification. Subsistence farm is a dummy taking value 1 when the whole production is consumed by the household while market oriented farm takes value 1 when at least two thirds of the farm’s output (value) is sold.

Table 6: Actual vs efficient farm scale

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>log($l_i$) - log($l^v$)</th>
<th>Pr($l_i$ &gt; $l^v$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS coefficients</td>
<td>Average Marginal Effects (Probit)</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>% output consumed</td>
<td>0.447***</td>
<td>0.726***</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Subsistence farm</td>
<td>0.514***</td>
<td>0.632***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Market oriented farm</td>
<td>0.012</td>
<td>-0.162**</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Village FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Unconditional mean</td>
<td>0.830</td>
<td>0.830</td>
</tr>
<tr>
<td>of dependent variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.004</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In the first four columns, the dependent variable is the difference between the logarithm of the area of land farmed by the holding and the logarithm of the efficient farm scale that would be implied by a efficient redistribution of land within village. In last four columns, the dependent variable is a dummy taking value 1 when the amount of land farmed by household $i$, $l_i$, is greater than the one than would be implied if the land input were redistributed across farmers residing in the village proportionally to their agricultural skills $s$ so to equalize marginal productivity of land. Subsistence farm is a dummy taking value 1 when the whole production is consumed by the household while market oriented farm takes value 1 when at least two thirds of the farm’s output (value) is sold.
Table 7: Commercialization group specific gains from reallocation

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Group specific gain from reallocation</th>
<th>( \frac{Y_{gs}}{Y_g} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Village level</td>
<td>Parish level</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Subsistence farms</td>
<td>(1) 0.038*</td>
<td>(2) 0.051**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Consumption tertile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2(^{nd})</td>
<td>0.012</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>3(^{rd})</td>
<td>0.020**</td>
<td>0.039***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

Unconditional mean of dependent variable

|                     | 1.27                                 | 1.22             | 1.36             | 1.27             | 1.54             | 1.47             |

| N                   | 5,231                                | 7,931            | 4,097            | 6,057            | 1,603            | 2,166            |
| Clusters            | 80                                    | 80               | 80               | 80               | 80               | 80               |
| adj. \( R^2 \)      | 0.231                                 | 0.238            | 0.229            | 0.254            | 0.275            | 0.280            |

Standard errors clustered at the district level in parentheses: * \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \). The dependent variable is the gain from reallocation within location (village, parish or subcounty) and commercialization group. In particular, in the first, third and fifth columns farmers with no output sold are compared to the others, while in the remaining specifications they are divided by consumption tertile (where the lowest tertile is the ones with the highest level of marketization). Location and number of farmers per group fixed effects are included in all the specifications.
Table 8: Village level gains from reallocation, land market participation and subsistence farming

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Village level gain from reallocation $e_v = \frac{Y_{v^*}}{Y_v}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% consumed output (median farmer)</td>
<td>0.075***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>% non commercial farms</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>% total output consumed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>% participants to land markets</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

Fixed effects

<table>
<thead>
<tr>
<th></th>
<th>District</th>
<th>County</th>
<th>District</th>
<th>County</th>
<th>District</th>
<th>County</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional mean of dependent variable</td>
<td>1.31</td>
<td>1.31</td>
<td>1.31</td>
<td>1.31</td>
<td>1.31</td>
<td>1.31</td>
</tr>
<tr>
<td>N</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
</tr>
<tr>
<td>Clusters</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.308</td>
<td>0.315</td>
<td>0.307</td>
<td>0.314</td>
<td>0.307</td>
<td>0.315</td>
</tr>
</tbody>
</table>

Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable represents the total TFP gain from reallocations that would be obtained in a village if the inputs were redistributed across farmers in proportion to their productivity. In all the specifications, we control for the logarithm of the standard deviation in farmers’ productivity. % consumed output refers to the median percentage of agricultural output consumed by farmers in the village, the mean is 0.65. % of non commercial farms indicates the percentage of farms operating in the village whose output is entirely used by the household. The average of this variable is 0.29. % total output consumed is the share of the agricultural output produced in the village that is consumed by the farmers. Finally, % of participants to land markets indicates the share of farmers in the village operating at least one plot of land that has been either purchased or rented in. The average is 0.69.
Table 9: Marginal productivity of factors (no coffee farmers)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>MPL (log)</th>
<th>MPK (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% output consumed</td>
<td>-0.317***</td>
<td>-0.520***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Subsistence farm</td>
<td>-0.227***</td>
<td>-0.454***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Market oriented farm</td>
<td>0.098***</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

| N                   | 21,622    | 21,622    |
| adj. $R^2$          | 0.580     | 0.583     |

These results refer to the subsample of farmers who devoted less than one third of their land to the production of coffee. Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Village FE are included in each specification. Subsistence farm is a dummy taking value 1 when the whole production is consumed by the household while market oriented farm takes value 1 when at least two thirds of the farm’s output (value) is sold.

Table 10: Actual vs efficient farm scale (no coffee farmers)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>$log(l_i) - log(l_{ei})$</th>
<th>Pr($l_i &gt; l_{ei}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS coefficients</td>
<td>(1) (2) (3) (4)</td>
<td>(5) (6) (7) (8)</td>
</tr>
<tr>
<td>% output consumed</td>
<td>0.546*** 0.847***</td>
<td>0.110*** 0.199***</td>
</tr>
<tr>
<td></td>
<td>(0.06) (0.09)</td>
<td>(0.01) (0.02)</td>
</tr>
<tr>
<td>Subsistence farm</td>
<td>0.517*** 0.633***</td>
<td>0.083*** 0.123***</td>
</tr>
<tr>
<td></td>
<td>(0.04) (0.05)</td>
<td>(0.01) (0.01)</td>
</tr>
<tr>
<td>Market oriented farm</td>
<td>-0.047 -0.232***</td>
<td>-0.035*** -0.077***</td>
</tr>
<tr>
<td></td>
<td>(0.05) (0.06)</td>
<td>(0.01) (0.01)</td>
</tr>
<tr>
<td>Village FE</td>
<td>No Yes No Yes</td>
<td>No Yes No Yes</td>
</tr>
<tr>
<td>N</td>
<td>21,622 21,622 21,622 21,622</td>
<td>21,622 21,615 21,622 21,615</td>
</tr>
<tr>
<td>Clusters</td>
<td>80 80 80 80</td>
<td>80 80 80 80</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.004 0.048 0.012 0.056</td>
<td>0.002 0.006 0.003 0.011</td>
</tr>
<tr>
<td>pseudo $R^2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results refer to the subsample of farmers who devoted less than one third of their land to the production of coffee. Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In the first four columns, the dependent variable is the difference between the logarithm of the area of land farmed by the holding and the logarithm of the efficient farm scale that would be implied by an efficient redistribution of land within village. In last four columns, the dependent variable is a dummy taking value 1 when the amount of land farmed by household $i$, $l_i$, is greater than the one that would be implied if the land input were redistributed across farmers residing in the village proportionally to their agricultural skills $s$ so to equalize marginal productivity of land. Subsistence farm is a dummy taking value 1 when the whole production is consumed by the household while market oriented farm takes value 1 when at least two thirds of the farm’s output (value) is sold.
Table 11: Commercialization group specific gains from reallocation (no coffee farmers)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Group specific gain from reallocation $\frac{Y_{gs} - Y_{gs}}{Y_{gs}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Village level (1)</td>
</tr>
<tr>
<td>Subsistence farms</td>
<td>0.036*</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>Consumption tertile (Reference group: 1st)</td>
<td>0.008</td>
</tr>
<tr>
<td>2nd</td>
<td>(0.01)</td>
</tr>
<tr>
<td>3rd</td>
<td>0.019**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>N</td>
<td>5,108</td>
</tr>
<tr>
<td>Clusters</td>
<td>80</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.228</td>
</tr>
</tbody>
</table>

These results refer to the subsample of farmers who devoted less than one third of their land to the production of coffee. Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the gain from reallocation within location (village, parish or subcounty) and commercialization group. In particular, in the first, third and fifth columns farmers with no output sold are compared to the others, while in the remaining specifications they are divided by consumption tertile (where the lowest tertile is the ones with the highest level of marketization). Location and number of farmers per group fixed effects are included in all the specifications.
Table 12: Village level gains from reallocation, land market participation and subsistence farming (no coffee farmers)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Village level gain from reallocation $e_v = \frac{Y_v^*}{Y_v}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% consumed output (median farmer)</td>
<td>0.064*** (0.02)</td>
</tr>
<tr>
<td>% non commercial farms</td>
<td></td>
</tr>
<tr>
<td>% total output consumed</td>
<td></td>
</tr>
<tr>
<td>% participants to land markets</td>
<td>-0.021 (0.03)</td>
</tr>
</tbody>
</table>

Fixed effects

<table>
<thead>
<tr>
<th></th>
<th>District</th>
<th>County</th>
<th>District</th>
<th>County</th>
<th>District</th>
<th>County</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
<td>3,097</td>
</tr>
<tr>
<td>Clusters</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.308</td>
<td>0.315</td>
<td>0.307</td>
<td>0.314</td>
<td>0.307</td>
<td>0.315</td>
</tr>
</tbody>
</table>

These results refer to the subsample of farmers who devoted less than none third of their land to production of coffee. Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable represents the total TFP gain from reallocations that would be obtained in a village if the inputs were redistributed across farmers in proportion to their productivity. In all the specifications, we control for the logarithm of the standard deviation in farmers’ productivity. % consumed output refers to the median percentage of agricultural output consumed by farmers in the village, the mean is 0.65. % of non commercial farms indicates the percentage of farms operating in the village whose output is entirely used by the household. The average of this variable is 0.29. % total output consumed is the share of the agricultural output produced in the village that is consumed by the farmers. Finally, % of participants to land markets indicates the share of farmers in the village operating at least one plot of land that has been either purchased or rented in. The average is 0.69.