The World Bank Economic Review, 37(4), 2023, 570–598
https://doi.org10.1093/wber/lhad017
Article

Subsistence Farming and Factor Misallocation: Evidence from Ugandan Agriculture

Bruno Morando

Abstract

This paper presents a model where misallocation in the agricultural factors of production is caused by transportation costs to and from local markets, which result in an inefficiently large share of inputs operated by less productive subsistence farmers. The model derives some testable predictions which are verified in the empirical analysis, based on a representative census of Ugandan farms. Specifically, subsistence farmers operate inefficiently high shares of land and capital and the efficiency losses are more severe in areas where subsistence farming is more widespread, due to lower connectivity with local markets. Conversely, there is no relationship between the level of misallocation and credit access and/or land-market activity. These findings suggest that transportation costs play a key role in determining the efficiency of agricultural input distribution and that land-market liberalization is a necessary but not sufficient condition to tackle misallocation.

JEL classification: Q14, O40, O13

Keywords: misallocation, productivity, agriculture, Uganda

1. Introduction

A growing body of literature (Restuccia and Santaeulalia-Llopis 2017; Chen et al. 2017; Ayerst et al. 2020; Chari et al. 2021; Chen et al. 2021; Adamopoulos et al. 2022) attributes low agricultural productivity in developing countries to factor misallocation. The core hypothesis is that institutional barriers to land rentals/sales (which are common in the developing world) prevent the reallocation of resources from less to more able producers and ultimately result in sizeable efficiency losses.

Given the existing evidence of widespread failures in rural markets (Dillon and Barrett 2017), it is plausible that formal restrictions to input transactions represent only one of the determinants of resource misallocation and that other factors contribute to the observed inefficiencies in input distribution. However, with the exception of the work on credit markets by Shenoy (2017) and the recent paper by Britos et al. (2021) looking also at road connectivity, ethnicity, and education, the role of concurrent factors has not been explored by the existing literature.

This paper aims to fill this gap by considering the role of transportation costs to and from local markets in determining resource misallocation in agriculture. This is motivated by the fact that, especially in

Bruno Morando is a lecturer (assistant professor) at Maynooth University, Maynooth, Ireland; his email address is bruno.morando@mu.ie. The author would like to thank Salvatore Di Falco, Douglas Gollin, Tara Mitchell, and Carol Newman for their valuable comments and support. The author is also grateful to the editor Norman Loayza and two anonymous referees for their insights and feedback. A supplementary online appendix is available with this article at *The World Bank Economic Review* website.

Sub-Saharan Africa, farmers typically produce mostly for self-consumption and sell a very small fraction of their output (Christiaensen and Demery 2017), suggesting that they face significant barriers in accessing food markets. This insight is formally developed in a theoretical model that shows that, even when inputs can be traded freely, costly access to food markets (captured by a wedge between consumer and producer prices) results in an equilibrium featuring an inefficiently large subsistence sector and resource misallocation among farmers.

The empirical analysis supports the main predictions of the model. Specifically, it shows that efficiency-enhancing transactions would entail transfers of land and agricultural capital from subsistence to commercial farmers. Moreover, there is a strong positive relationship between village-specific estimated travel time to the closest urban center, the size of the subsistence sector, and the severity of factor misallocation. Conversely, there is no empirical link between the magnitude of the efficiency losses and the percentage of farmers with access to land or credit markets.

This work complements the existing studies on agricultural factor misallocation (Restuccia and Santaeulalia-Llopis 2017; Chen et al. 2017; Ayerst et al. 2020; Chari et al. 2021; Adamopoulos et al. 2022) by identifying a concurrent cause leading to deviations from the full efficiency counterfactual. Moreover, it contributes to the debate on the actual magnitude of the potential efficiency gains in the context of farming in developing countries recently raised by Gollin and Udry (2019) and Maue et al. (2020). Specifically, the most relevant empirical results are corroborated at a very granular level (within farmers in the same village) and are possibly less subject to misspecification of the production function (due to different cropping patterns and farming techniques) and unobserved heterogeneity in land quality. Additionally, the empirical analysis confirms a number of theoretically informed predictions that go beyond the variance of marginal productivities (e.g. the relative scale of productivity dispersion among subpopulations and the nature of the implied efficiency-enhancing transactions) that are possibly less sensitive to measurement errors and misspecifications.

More broadly, this paper relates to the literature on farming productivity and cross-country income differences (Gollin et al. 2007; Restuccia et al. 2008; Lagakos and Waugh 2013) as it provides a micro-level perspective on the food problem and the negative relationship between agricultural productivity and the share of labor force in agriculture. In particular, the results presented echo the findings of a well-established literature on the detrimental impact of transportation costs on the productivity and size of the farming sectors in developing countries (Adamopoulos 2011; Gollin and Rogerson 2014; Tombe 2015; Shamdasani 2021). Crucially, while existing studies are concerned with the impact of transportation frictions on the distribution of inputs between the agricultural and non-agricultural sectors, this paper shows that transportation costs have a detrimental effect also on the efficiency of the distribution of factors of production among existing farmers.[1]

Also, in looking at how internal trade barriers shape agricultural activities, this work complements a recent strand of literature aiming to quantify the productivity gains achievable by reducing internal trade frictions (Costinot and Donaldson 2016; Adamopoulos 2019; Sotelo 2020). Unlike these studies, the main focus is rural/urban connectivity rather than regional market integration.

Finally, the basic insight of the paper, as well as some modeling choices, is derived from the seminal literature on farm households' behavior in the presence of market constraints (De Janvry et al. 1991; Fafchamps 1992; Omamo 1998). These early papers introduced the concept of interdependence of consumption and production decisions to explain some apparently suboptimal behaviors of farmers in the developing world (like underinvestment in revenue-maximizing cash crops). This paper builds on these insights to study their implications in terms of the resource distribution among heterogeneous farmers.

1 Supplementary online appendix S5 provides a unifying framework where transportation costs and reduced market access affect both the distribution of inputs within farmers and between the agricultural and non-agricultural sectors.

Recent works showing the impact of reduced market access on farmers production choices and the resulting productivity loss include Qin and Zhang (2016), Li (2023), and Haque (2022).

The paper develops as follows: Section 2 introduces the theoretical framework and shows how factor misallocation can also arise in the presence of frictionless input markets when accessing the food market is made costly by transportation costs. Section 3 provides some background and illustrates some features of the agricultural sector in Uganda. In section 4, I describe the data used in the empirical analysis. Section 5 presents results corroborating the theoretical model and discusses their implications. Section 6 concludes with some final remarks.

2. Theoretical Framework

This section presents the qualitative model and its most relevant predictions. A set of core assumptions from the related literature (Restuccia and Santaeulalia-Llopis 2017; Chen et al. 2017; Ayerst et al. 2020) is maintained; namely, a common production function, decreasing returns to scale in the tradable inputs (land and capital), labor-augmenting productivity, and perfect output substitutability (i.e. flat individual demand curve). This implies that the counterfactual allocation maximizing the total production is non-degenerate, which allows one to compute the achievable gains from efficient reallocation of existing agricultural inputs among existing farmers i.e. the so-called static misallocation.

The model departs from the existing ones as it introduces transportation costs which are modeled as frictions in the output market (i.e. a wedge between producers and consumers prices), following Omamo (1998). This allows one to examine how the resulting interdependence between consumption and production decisions affects the distribution of inputs across farmers who are heterogeneous in their agricultural skills and in turn the aggregate productivity.

The next section describes the basic components of the model and the resulting maximization problem farm households face. Then, it derives the social-planner solution which represents the full efficiency counterfactual and will serve as a benchmark for the computation of the magnitude of the efficiency losses.[2] I show that the same input allocation is achieved by the markets where there are no frictions. Finally, the social-planner outcome is compared with the one obtained for different levels of frictions to study their impact on aggregate productivity and derive some testable predictions. An analysis of the impact of such frictions on farmers' welfare along their productivity distribution is provided in the supplementary online appendix S1.

2.1. The Model Setup

There is a continuum of farmers (each indexed by i) of dimension M all operating in the same area (village).[3] They are heterogeneous in productivity a, which has a density function $f(\cdot)$ with lower and upper bounds $\underline{a}$ and $\bar{a}$. All farmers produce a homogeneous agricultural good Y according to the same Cobb–Douglas production function that combines productivity a, labor X, land L, and capital K:

$$Y_i = (a_i X_i)^{1-\gamma} (L_i^{\alpha} K_i^{1-\alpha})^{\gamma}.$$

Labor is inelastically supplied to agriculture and cannot be traded. The production function can thus conveniently be expressed as

$$y_i = a_i^{1-\gamma} (l_i^{\alpha} k_i^{1-\alpha})^{\gamma}, \tag{1}$$

2 In the case of the social planner, production and consumption decisions are independent (i.e. aggregate production can be freely distributed across agents so as to maximize total welfare) and as such the optimal allocation of resources (where the amount of resources devoted to farming is fixed) is always the one that maximizes production.

3 This assumption is relaxed in supplementary online appendix S5, where the model allows for households' self-selection into farming depending on their agricultural productivity and the transaction costs.

where the lower case refers to output and inputs in per labor unit. It is assumed that $0 < \gamma < 1$, which implies that the production function features decreasing returns to scale in the tradable inputs. In particular, γ is the span-of-control parameter that governs the returns to scale and the relationship between farm size and productivity in the frictionless scenario.

For the sake of tractability, there are no limitations to input transactions and both land and capital can be exchanged in a frictionless, perfectly competitive market.[4] In light of this, the distinction between the two tradable inputs is irrelevant for the purpose of the model, and land and capital can be aggregated into a composite agricultural input $\xi_i = l_i^{\alpha} k_i^{1-\alpha}$. The resulting production function is

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

Farmers maximize their utility U, which depends on their consumption of food c_i^f and of another generic good c_i^o, which is also the numeraire. The utility function is non-homothetic and features a subsistence threshold $\bar{c}$[5]

$$U_i = (c_i^f - \bar{c})^{\beta} (c_i^o)^{1-\beta}.$$

Food can either be produced or purchased at the local market at the exogenous price p.[6] Farmers also have the option of selling their output at the market at the same price. However, when buying food, all farmers in the village face a proportional tax τ which captures the transportation costs they incur and it is a negative function of market access.[7]

Farm households can also buy the numeraire good c_i^o and trade the composite input ξ among each other at the endogenously determined price r. Finally, each household has the same initial endowment of fixed income w and composite agricultural input $\bar{\xi}$ per unit of labor.

Technically, farmers face two separate constraints:

$$c_i^f = y_i + b_i - s_i = a_i^{1-\gamma} \xi^{\gamma} + b_i - s_i \tag{2}$$

and

$$r(\xi_i - \bar{\xi}) + p\tau b_i + c_i^o = w + ps_i. \tag{3}$$

In particular, equation (2) states that the household's food consumption is equal to the amount produced y_i plus the net quantity bought $(b_i - s_i)$, while equation (3) represents the budget constraint, where the total amount spent to buy excess input, food, or the numeraire good is equal to the sum of the fixed income w and the revenue obtained selling food. The household maximization problem can thus be described by

4 This assumption is not a necessary feature of the model and can be relaxed. A model with input-specific transactions is presented in supplementary online appendix S3.

5 Note that the threshold $\bar{c}$ is constant across farm households. However, $\bar{c}$ is expressed in per labor-unit terms, therefore it is higher for households with more members (in line with the assumption that labor is inelastically supplied to agriculture and cannot be hired).

6 The production dynamics taking place in the village are assumed to be irrelevant in determining the local market price p, as the market serves many different villages in the area.

7 In practice, this implies that farmers cannot avoid transportation costs by trading food directly or by offering in-kind compensation in exchange for agricultural inputs (i.e. where the more productive farmers pay less productive farmers in kind in exchange for their land/labor). However, neither of these practices is particularly widespread in Uganda. According to the 2009/10 Ugandan Living Standard Measurement Survey (LSMS), more than 75 percent of the sales of agricultural goods involved a trader rather than a consumer and in less than 7 percent of the cases the reported buyers were neighbors or relatives. Similarly, only 6 percent of the total farm households' crop consumption derived from donations and/or in-kind transfers, while they relied heavily on self-production (56 percent) or purchases (38 percent). Additionally, only 5 percent of the farm households derived at least one-third of their crop consumption from in-kind transfers. These figures indicate that direct food transfers among farmers in the same village (either sales or in exchange for labor or other agricultural inputs), although not uncommon, do not play a key role.

the Lagrangian,

$$\mathcal{L} = U_i(c_i^f, c_i^o) - \lambda\big[c_i^f - a_i^{1-\gamma}\xi_i^\gamma - b_i + s_i\big]$$
$$-\mu[r(\xi_i - \bar{\xi}) + p\tau b_i + c_i^o - w - ps_i] + \nu_1 b_i + \nu_2 s_i,$$

where b_i, s_i, ν_1, and ν_2 are non-negative.

Intuitively, the fact that there are two separate constraints for food consumption and for income implies that the production and consumption decisions are interdependent where $\tau > 1$, and that the utility-maximizing choice is not the one maximizing production.

The price of the composite agricultural factor r is village specific and can be identified using the market-clearing condition, which states that the total amount of input employed by households is equal to the initial endowment:[8]

$$\int_{\underline{a}}^{\bar{a}} \xi_i f(a)\, da = \int_{\underline{a}}^{\bar{a}} \bar{\xi} f(a)\, da = M\bar{\xi}.$$

The equilibrium distribution of the composite input across the heterogeneous agents can be characterized as the solution of the farmers' maximization problem. In turn, this distribution can be used to compute the resulting aggregate production and farmer-specific welfare. It is possible to generate such equilibrium for different levels of τ to study how costly market access affects aggregate productivity and farmer-specific welfare. In the next section, the social-planner solution is derived. The resulting input allocation represents the full efficiency benchmark used to compute efficiency losses caused by transaction costs τ.

2.2. Social-Planner Problem

Since the social planner can redistribute aggregate output among agents freely after production takes place, its problem can be characterized as the maximization of aggregate production given the available inputs and the existing farmers' productivity distribution. Formally,

$$\max_{\xi_i} \int_{\underline{a}}^{\bar{a}} a_i^{1-\gamma}\xi_i^\gamma f(a)\, da \quad \text{subject to} \int_{\underline{a}}^{\bar{a}} \xi_i f(a)\, da = M\bar{\xi}. \tag{4}$$

The first-order condition implies that production is maximized where marginal productivity is equalized among farmers. Formally,

$$\gamma a_i^{1-\gamma}\xi_i^{\gamma-1} = \lambda \quad \forall\, i,$$

where λ is the Lagrangian multiplier for the resource constraints in equation (4). Solving for ξ_i leads to

$$\xi_i = \left(\frac{\gamma}{\lambda}\right)^{\frac{1}{1-\gamma}} a_i,$$

which indicates that each farmer operates a quantity of input ξ which is proportional to their productivity a. The exact amount of composite input used by each agent can be obtained by imposing the market-clearing condition $\int_{\underline{a}}^{\bar{a}} \xi_i f(a)\, da = M\bar{\xi}$, which leads to

$$\xi_i = \frac{a_i}{a_M}\bar{\xi}, \tag{5}$$

where a_M is the average productivity. It follows that the production of each farm household is also proportional to their agricultural skills and is equal to $a_i(\bar{\xi}/a_M)^\gamma$ and their marginal productivity is $\gamma(a_M/\bar{\xi})^{1-\gamma}$. The resulting total production is

$$\int_{\underline{a}}^{\bar{a}} a_i^{1-\gamma}\left(\frac{a_i}{a_M}\bar{\xi}\right)^\gamma f(a)\, da = M a_M^{1-\gamma}\bar{\xi}^\gamma. \tag{6}$$

8 Since the model is concerned with static misallocation only, it is assumed that there is a fixed quantity of inputs which are traded among existing producers.

Equation (6) describes the maximum achievable aggregate production given the distribution of farmers' productivity and the total amount of factors of production available in the economy, which will be used as the full efficiency counterfactual benchmark.

In the next sections, the model is solved for different levels of frictions τ, showing how the input distribution and the resulting aggregate production departs from the fully efficient resource allocation characterized above.

2.3. The Equilibrium with Frictionless Markets ($\tau = 1$)

It can be shown that, as far as input distribution and aggregate production are involved, the market equilibrium with free access to food markets ($\tau = 1$) is equivalent to the social-planner solution and as such maximizes total output given the available factors of production and the productivity distribution of farmers.[9]

Notably, when $\tau = 1$, farmers have free access to food markets (formally, there is no wedge between the producer's and the consumer's price) and as such their consumption and production decisions are independent. Substituting for $\tau = 1$, and given the production function and the credit constraint, the farmer's problem can be expressed (after taking logs) as

$$\max_{b_i, s_i, \xi_i} \beta \log(a_i^{1-\gamma} \xi^{\gamma} - \bar{c} + b_i - s_i) + (1-\beta) \log[w + r(\xi - \bar{\xi}) + p(s_i - b_i)].$$

The optimization can be solved separately for food buyers, sellers, and autarkic farmers ($b_i > 0$, $s_i > 0$, and $b_i = s_i = 0$). In either case, the first-order conditions for ξ_i imply that

$$\gamma a_i^{1-\gamma} \xi_i^{\gamma-1} = \frac{r}{p} \quad \rightarrow \quad \xi_i = a_i \left(\frac{p\gamma}{r}\right)^{\frac{1}{1-\gamma}} \quad \rightarrow \quad \xi_i = A a_i,$$

where $A = (p\gamma/r)^{\frac{1}{1-\gamma}}$ is common across farmers. Thus, farmers operate a quantity of input which is proportional to their productivity, and have the same marginal productivity in equilibrium as a result. The final allocation of resources and aggregate production is therefore equivalent to the social-planner solution outlined in equation (5). This hardly comes as a surprise, since the economy features perfectly competitive markets and therefore allocates the factors of production efficiently. Thus, as far as production efficiency is concerned, the social-planner allocation is equivalent to the one where private agents face no frictions.

This condition can in turn be plugged into the maximization problem of net buyers and net sellers to determine the quantity of output bought/sold and the productivity threshold where farmers exit subsistence (i.e. sell a non-zero share of their production). In the case of food buyers, the objective is

$$\max_{b_i} \beta \log(A^{\gamma} a_i - \bar{c} + b_i) + (1-\beta) \log[w - r(\xi - \bar{\xi}) - p b_i],$$

and solving for b_i,

$$b_i = B - C a_i, \tag{7}$$

where B and C are constant terms. Thus, the quantity of food bought is a negative, linear function of farmer's productivity.[10] Similarly, the first-order condition with respect to s_i for food sellers implies that

$$s_i = -B + C a_i. \tag{8}$$

Equations (7) and (8) show that the equilibrium quantity of excess food production (or food bought in the case of buyers) is a continuous and linear function of a farmer's productivity and that, as such, all

9 Farmers' welfare might differ in these two counterfactuals as the social planner can redistribute existing output to maximize any aggregate utility function, while in the case of $\tau = 1$, output redistribution after production does not occur. This is not concerning, as the main scope for the paper is to look at efficiency/productivity losses.

10 The constants are $B = ((1-\beta)\bar{c}p + \beta w + \beta r\bar{\xi})/p$ and $C = ((1-\beta)A^{\gamma}p + A\beta r)/p$.

farmers will be either net sellers or net buyers, as there is only one exact level of productivity a^S where the amount of food bought and sold is equal to zero. This value represents the productivity threshold at which farmers become commercial and can be found as:

$$a^S = \frac{B}{C}.$$

This threshold can be used as a benchmark to compare the relative size of the subsistence sector in the frictionless scenario to the case where $\tau > 1$, examined in the next section.

2.4. The Equilibrium with Costly Market Access ($\tau > 1$)

Where $\tau > 1$, food markets are costly to access, and there is a wedge between the buying and selling prices of the agricultural output.[11] In this instance, the optimization problem faced by farmers depends on their net position in the food market, as this affects their shadow price for the agricultural output. In particular, the decision value for food buyers is higher than that for food sellers. Formally, substituting the food consumption and the budget constraints in the utility function and taking the logarithm leads to

$$\max_{\xi_i, b_i, s_i} \beta \log(a_i^{1-\gamma}\xi_i^{\gamma} + b_i - s_i - \bar{c}) + (1-\beta)\log[w - r(\xi_i - \bar{\xi}) + ps_i - p\tau b].$$

In this case, the solution differs depending on whether the farmer is a food buyer, a food seller, or autarkic. The first-order conditions for net food buyers with respect to the composite input ξ and the quantity of food bought b_i are

$$\frac{\beta\gamma a_i^{1-\gamma}\xi^{\gamma-1}}{a_i^{1-\gamma}\xi^{\gamma} + b_i - \bar{c}} = \frac{r(1-\beta)}{w - r(\xi - \bar{\xi}) - p\tau b_i},$$
$$\frac{\beta}{a_i^{1-\gamma}\xi_i^{\gamma} + b_i - \bar{c}} = \frac{p\tau(1-\beta)}{w - r(\xi - \bar{\xi}) - p\tau b_i},$$

and solving for ξ,

$$\xi_i = \left(\frac{p\tau\gamma}{r}\right)^{\frac{1}{1-\gamma}} a_i = \tilde{\tau} A a_i,$$

where $\tilde{\tau} = \tau^{\frac{1}{1-\gamma}}$. Thus, food buyers operate a quantity of input which is proportional to their productivity, and as a result have the same marginal productivity. This condition can be used to obtain the amount of food bought, showing that it is still a negative linear function of productivity a

$$b_i = \widetilde{B} - \widetilde{C} a_i.$$

However, in this case the parameters differ with respect to where $\tau = 1$.[12] In this instance, the productivity level at which farmers stop buying food is

$$a^{S_1} = \frac{\widetilde{B}}{\widetilde{C}}.$$

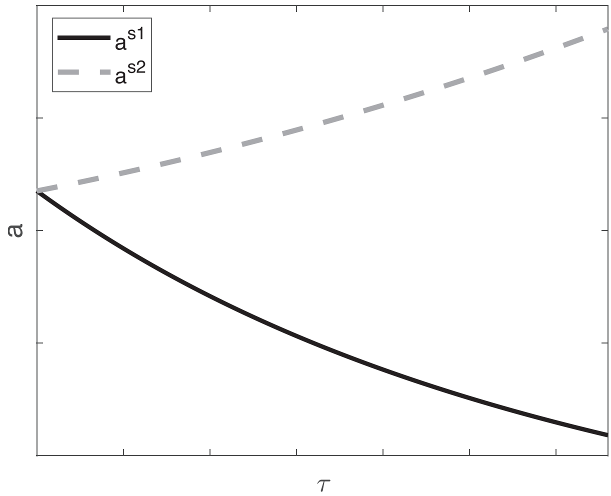
Since the wedge between buying and selling prices is modeled as a multiplicative factor to buying price, the first-order conditions for net sellers are the same as the ones obtained in the case where $\tau = 1$. Formally,

$$\xi_i = \left(\frac{p\gamma}{r}\right)^{\frac{1}{1-\gamma}} a_i = A a_i \tag{9}$$

11 Although I focus on transportation costs (like in Jayne (1994) and Omamo (1998)), the model is also consistent with a scenario where there is no perfect information on market price realizations and risk-averse farmers prefer to rely on self-production to minimize price risk (like in Fafchamps (1992)), or where there is any other factor preventing households from sourcing food from markets.

12 Specifically, $\widetilde{B} = (1-\beta)\bar{c} + \beta((w + r\bar{\xi})/p\tau)$ and $\widetilde{C} = (((1-\beta)A^{\gamma}p + A\beta r)/p)\tau^{\frac{\gamma}{1-\gamma}}$.

Figure 1. Productivity Thresholds in Equilibrium.



Source: Author's own computation.
Note: $a^{S}1$ represents the level of productivity where farmers stop being food buyers, while $a^{S}2$ is the level of productivity where farmers become food sellers.

and

$$s_i = -B + Ca_i.$$

Equation (9) shows that also net sellers operate a quantity of input proportional to their productivity. However, the multiplicative factor (A) is lower than in the case of net buyers ($\tilde{\tau}A$). This implies that (less productive) food buyers operate a comparatively higher share of inputs and as such have lower marginal productivities in equilibrium.

In this case, the productivity threshold at which farmers become commercial (sell a non-zero quantity of output) is

$$a^{S_2} = \frac{B}{C}.$$

It can be shown that, for $\tau > 1, a^{S_2} > a^{S_1}$.[13] Thus, the equilibrium features a non-null share of farmers that become autarkic (do not sell or buy any food) in order to avoid the costs associated with accessing the markets. Intuitively, larger transportation costs τ discourage an increasing fraction of households from participating in the food markets (either as buyers or as sellers). Figure 1 plots the switching thresholds as a function of τ and shows that for higher transaction costs fewer households buy (a^{S_1} goes down) or sell (a^{S_2} goes up) food in equilibrium and autarky becomes the optimal choice for a larger share of farmers.

The optimization problem for subsistence farmers can be obtained setting $b_i = s_i = 0$:

$$\max_{\xi_i} \beta \log(a_i^{1-\gamma}\xi_i^{\gamma} - \bar{c}) + (1-\beta)\log[w - r(\xi_i - \bar{\xi})].$$

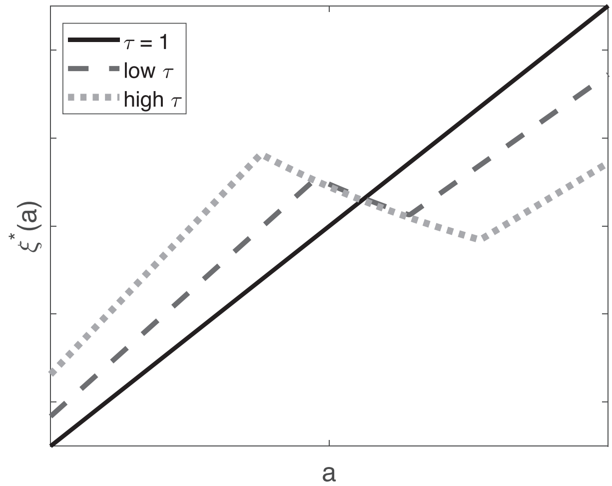
The resulting first-order condition does not have a closed-form solution, but can be expressed as an implicit function F of a_i and ξ_i:

$$F(\xi_i, a_i) = \frac{\beta a_i^{1-\gamma}\xi_i^{\gamma-1}}{a_i^{1-\gamma}\xi_i^{\gamma} - \bar{c}} - \frac{r(1-\beta)}{w - r(\xi_i - \bar{\xi})} = 0.$$

The implicit function theorem can be used to derive the sign of $d\xi_i^*(a)/da$, which indicates whether the amount of input operated by subsistence farmers is increasing or decreasing in their productivity. In particular, the sign must be the same as $-F_a/F_\xi$, where the subscripts denote partial differentiation. It can be

13 This is due to the fact that for $\tau > 1$, $\widetilde{B} < B$ and $\widetilde{C} > C$ and so $\widetilde{B}/\widetilde{C} < B/C$.

Figure 2. Distribution of Composite Input.



Source: Author's own computation.
Note: $\xi^*(a)$ denoted the equilibrium level of composite agricultural input operated by farmers with productivity a.

proven that the sign is negative conditional on $a_i^{1-\gamma}\xi^{\gamma} - \bar{c} > 0$, i.e. in the relevant scenario where farmers' consumption exceeds their subsistence threshold.

It is thus clear that the resulting input allocation deviates from the frictionless scenario/the social-planner solution. Crucially, the quantity of agricultural input operated by farmers is not proportional to their productivity; thus, marginal productivity is not equalized across farmers and the equilibrium with $\tau > 1$ features input misallocation. Intuitively, higher levels of τ increase misallocation across two dimensions. First, they widen the gap in marginal productivities between food buyers and food sellers, and second, they increase the share of autarkic farmers, which increases the severity of inefficiencies as they operate a quantity of input which is a negative function of their productivity.

The model can be closed by imposing the market-clearing condition stating that the total demand for the composite agricultural input must be equal to the existing supply. Formally,

$$\int_{\underline{a}}^{a^{S_1}} A\tilde{\tau}a_i f(a)\, da + \int_{a^{S_1}}^{a^{S_2}} \xi_i^{\mathrm{AUT}}(a) f(a)\, da + \int_{a^{S_2}}^{\bar{a}} Aa_i f(a)\, da = M\bar{\xi},$$

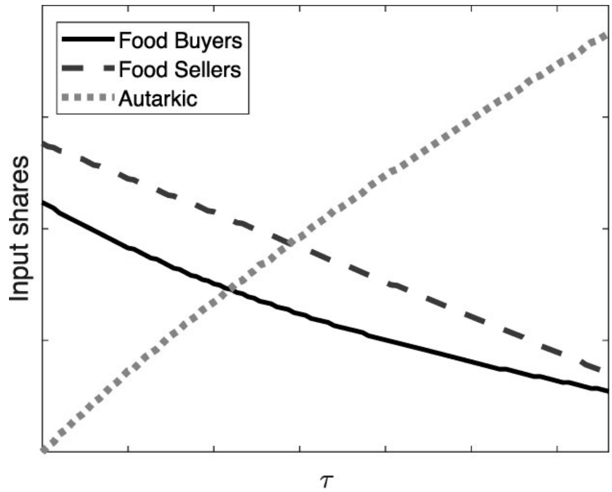
where $\xi_i^{\mathrm{AUT}}(a)$ is the demand for the composite input by subsistence farmers, which is a not-defined, negative function of a. Since the autarkic-farmer optimization has no closed-form solution, the model is estimated numerically.

The resulting distribution of the composite input in equilibrium is shown for different levels of τ in fig. 2. In the absence of frictions ($\tau = 1$), farmers operate a share of land and capital proportional to their productivity, maximizing aggregate production. However, for $\tau > 1$, the allocation of input deviates from the efficient distribution. In particular, larger values of τ increase the gap between the actual and the optimal amounts of inputs for food buyers (who operate a larger than efficient quantity) and sellers (who instead use a lower than efficient quantity), as well as the share of autarkic farmers.

As shown in fig. 3, this also implies that the share of input operated by subsistence (net buyers and autarkic) farmers is an increasing function of τ, indicating that higher transportation costs increase the size of the subsistence sector.[14]

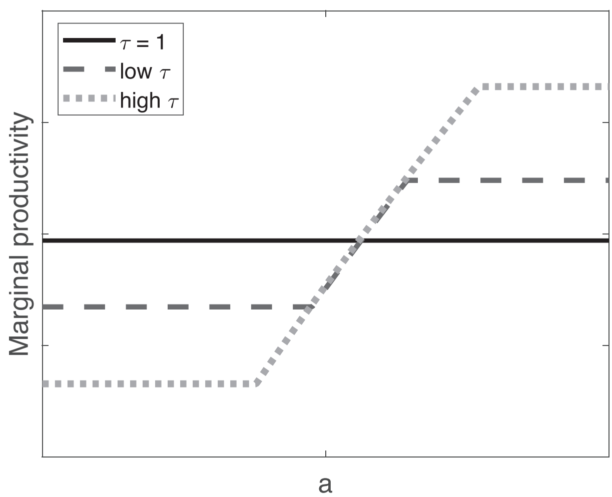
14 In this case, the size of the agricultural sector is assumed to stay constant and as such I am only considering the changes in the composition of the farming sector. In supplementary online appendix S5, I allow workers to move across sectors

Figure 3. Input Shares by Position in the Food Market.



Source: Author's own computation.

Figure 4. Distribution of Marginal Productivity.

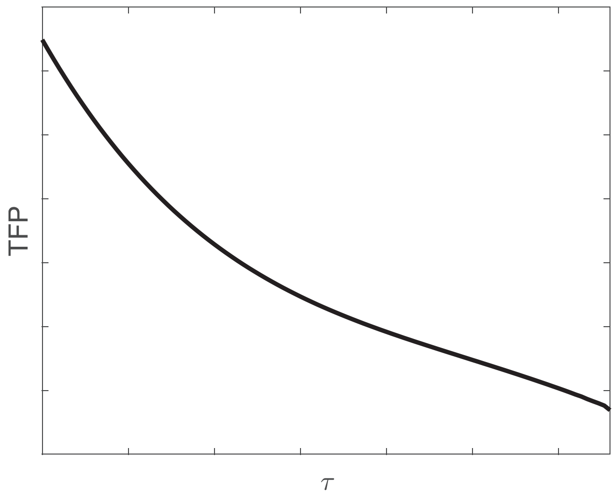


Source: Author's own computation.

As pointed out above, when there are no frictions, farmers operate a quantity of input proportional to their productivity. Intuitively, this allocation maximizes aggregate production as the marginal productivity of each farmer is equalized (i.e. it is not possible to increase production by reallocating inputs from one farmer to another). Figure 4 plots the equilibrium marginal productivity for $\tau > 1$ and compares it to the frictionless case. Also in this case, the departure from the frictionless case is evident since the marginal productivity is not equalized across farmers, but only within net buyers and net sellers. The figure also highlights the different components of input misallocation in this model. Specifically, marginal productivities are not equalized across categories (buyers, autarkic, and sellers) nor among autarkic farmers.

and show that in this case, similarly to Gollin and Rogerson (2010, 2014), both the size of the agricultural sector and the share of subsistence farmers increase in the magnitude of transportation costs.

Figure 5. Aggregate Productivity and τ.



Source: Author's own computation.

Intuitively, higher levels of τ increase misallocation by widening the gap between the marginal productivity of net buyers and net sellers and by increasing the size of the autarkic sector.

Finally, it is possible to examine the impact of transportation costs on total factor productivity by plotting the total production achieved for different levels of τ.[15] The results are presented in fig. 5, which shows that TFP decreases monotonically in τ. This implies that the larger the transportation costs are, the larger the resulting efficiency losses, measured as the gap in production with respect to the social-planner/frictionless scenario. In the empirical analysis, the qualitative predictions of this model will be tested using comprehensive micro-data on Ugandan farms.

3. Background: 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 some recent estimates (UBOS 2019), 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.

Similar to most countries in Sub-Saharan Africa, the agricultural sector plays a key role in the Ugandan economy, employing around 70 percent of the total workforce (UBOS 2019). Subsistence and semi-subsistence farming still represent the major source of livelihood of most households, as farmers typically operate on a very small scale and sell a very limited share, if any, of their production. According to UBOS (2019), about two-thirds of the workers in agriculture are subsistence farmers, and nearly 45 percent of the working population in 2017 indicated subsistence agriculture as their only form of economic activity.

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 do not sell any of their output. In particular, they find that 65 percent of the net food buyers in Uganda declare that subsistence agriculture is their pri-

15 Since total inputs and the productivity distribution are left unchanged, all the differences in production can be attributed to lower total factor productivity (TFP) caused by inefficiencies in the distribution of factors of production.

mary economic activity, and that farmers spend a high share of their total income (45 percent) on food and beverages. Along the same lines, Nabwire (2015) claims that 75 percent of the farmers surveyed reported that their main reason for being employed in farming was the necessity to obtain food for their household.

The prevalence of subsistence agriculture is typically attributed to high transportation and marketization costs, that are widely believed to represent the most serious barrier to the modernization of the farming 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).

Unlike other developing countries considered in similar papers, there seem to be no systematic and formal barriers to land transfers in Uganda. Indeed, land markets have been very active with virtually no interruption ever since the British domination, and land rentals and sales are common throughout the country: according to Deininger and Mpuga (2003), as far back as the early 1950s, almost 60 percent of the landowners were operating plots that had been acquired through market transactions. 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 that land transactions happen regularly between people from different ethnic groups and/or with weak or no social ties, suggesting that these transfers are generally inspired by market logic. In line with this, Dillon and Barrett (2017) show that land-market participation is more widespread in Uganda than in other Sub-Saharan countries.[16]

Although land transactions seem to be relatively common in Uganda, it is worth stressing that customary land rights are still quite widespread and the freehold system is still far from prevalent. In particular, according to Perego (2019), only 15 to 20 percent of the land was titled in 2010 (due to a cumbersome and costly administrative procedure) and most farmers had no formal claims on their lands. Farmers can obtain certificates of customary occupancy and convert them into freehold/leasehold contracts. However, there are sometimes conflicts and overlapping rights and disputes are not infrequent (Hunt 2004). It would therefore be misleading to consider the Ugandan land markets as completely frictionless and land transactions to be exclusively driven by market considerations.[17]

The overall picture describes an economy still largely dependent on subsistence farming, especially in the poor rural areas, and where the lack of explicit barriers to 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 subsistence-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).[18]

4. Data and Descriptive Statistics

In order to study the link between transportation costs, subsistence farming, and aggregate productivity, this paper uses data from a number of different sources, which are listed in table 1. The most relevant is the

16 The findings are based on the World Bank’s LSMS, according to which 38.7 percent of farmers in Uganda were involved in the land market, as compared to 32.7 in Ethiopia, 24.9 in Malawi, 30.9 in Niger, and 19.6 in Tanzania.

17 The assumption of frictionless input market in the benchmark model is relaxed in supplementary online appendix S3, where I show that allowing for input-specific transaction costs does not affect the main predictions.

18 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.

Table 1. Data Sets Used

	Name	Source	Level
Main analysis			
Input use, production and output disposition	UCA	UBOS	Household
Estimation of α and γ and output and capital value	LSMS	World Bank	Household
Adjustments in productivity			
Rainfalls	Quarterly rainfall	TAMSAT	$\approx$ 4 km^2 resolution
Land quality	GAEZ	FAO & IIASA	$\approx$ 10 km^2 resolution
Natural shocks	UCA	UBOS	Household
Transportation costs			
Cells crossing time	Ugandan Road Network	UBOS	1 km^2 resolution
Urban conglomerates	AFRICOVER	FAO	1 km^2 resolution
Geolocation*	UCA	UBOS	Village

Source: Author's own computation.

Note: *Village-level coordinates are obtained as the median coordinates reported by farmers in the survey. After dropping missing data or coordinates outside the relative districts, nearly 90 percent of villages can be geolocated (2,584 out of 2,918). UCA refers to the 2009 Census of Ugandan Agriculture, collected by the Ugandan Bureau of Statistics (UBOS). AFRICOVER is a spatial dataset containing information on land coverage and developed by the Food and Agriculture Organization (FAO). FAO also developed the GAEZ (Global Agro-Ecological Zones) spatial dataset on land quality and crop suitability with IIASA (International Institute for Applied Systems Analysis). Quarterly rainfall estimates are taken from the database developed by TAMSAT (Tropical Applications of Meteorology using SATellite data). LSMS refers to the World Bank's Living Standard Measurement Surveys.

2009 Uganda Census of Agriculture (UCA; UBOS (2010)), which collects detailed information on more than 25,000 Ugandan farms. The data set is a cross section which refers to the 2008/09 agricultural year and is nationally representative. After dropping observations with missing variables, from enumeration areas with less than 40 percent response rate, and in the extreme 5 percent of the productivity distribution, I obtain a final sample of 23,179 holdings.[19] The farms are distributed across 2,917 enumeration areas (villages) in 79 districts.[20]

Variables on agricultural input and output, as well as on participation in land markets and the marketization of crops produced, are derived from this data set. In particular, data on production and inputs are required to estimate farm-level productivity, while the final use of crops harvested (i.e. whether they are consumed, stored, or sold) is used to identify the subsistence and the commercial farmers. These variables are generated at the holding level.[21]

Cultivated land is defined as the total area of the plots farmed in the two agricultural seasons under analysis, which is measured using GPS techniques. The questionnaire also asks about the mode of acquisition of each of these plots, so that it can be assessed whether each farm included at least some land

19 For consistency purposes, I only keep observations of farms that operated both in the second agricultural season of 2008 and in the first agricultural season of 2009.

20 The sample can be further stratified into 154 counties or 827 subcounties.

21 Recent studies on agricultural misallocation in Uganda (Gollin and Udry 2019; Maue et al. 2020) use the World Bank's LSMS as the main source of data on agricultural production. The main advantages of the LSMS with respect to the UCA are that it is a panel data set and that it contains information on households' welfare and non-farming activities. However, the UCA is much more suitable for this analysis as (a) it collects information on more villages (the sample is roughly 10 times larger), which allows one to run more credible village-level regressions, (b) every village in the sample has a sufficiently high number of farms (since the survey is meant to be representative of smallholder farmers rather than of the whole population) to run a meaningful within village analysis, (c) the coordinates of the enumeration areas are not altered, and (d) farms' output is expressed in homogeneous crop units that can be easily and consistently aggregated. In supplementary online appendix S6, I replicate the analysis using LSMS data showing that the results are largely comparable.

that was acquired through market transactions as opposed to inheritances, donations, or other forms of informal appropriation.[22]

As for the labor input, farmers were asked about the exact number of working days provided by family members and hired laborers, as well as their gender and whether they were children or adults. Unfortunately, the data are missing for 56 percent of the observations, and the only information available is whether each worker was employed full time or temporarily. From the available observations, I infer that, on average, temporary workers provided about half the working days of their full-time counterparts. Therefore, in measuring the total quantity of labor input, all part-time workers are considered as half a unit. Additionally, I introduce different weights depending on the gender and the age of the workers on the basis of the relative median agricultural wage of each group as measured in the 2010 LSMS survey.[23] As a result, labor 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 labor input, this measure still captures a lot of variation across different holdings based on realistic assumptions and reliable entries of the survey.

Capital is defined as the sum of the value of the agricultural assets used for farming activities.[24] Differently from previous studies (Restuccia and Santaeulalia-Llopis 2017; Chen et al. 2017), means of transportation and livestock are not included as there is no way to establish whether they were actually employed in agricultural production.[25]

In terms of final output, farmers were asked to indicate their total production for the 18 major crops that overall account for more than 97 percent of the total farmed area.[26] 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. Similar to the case of capital items, the prices of the different crops were not available. Therefore, in order to aggregate the total production, a set of median prices derived from the 2010 LSMS survey is used.

Furthermore, for each crop, farmers were asked the percentage they consumed, sold, or stored. This information is used to generate a household-specific variable indicating the percentage of total output being self-consumed and identify as subsistence farmers those who did not sell any of their production.[27] Overall, the survey indicates that the average farmer consumes 68 percent of their output (the median 75 percent), and that nearly 30 percent of the farmers did not sell any of their production. Unfortunately, the survey does not collect any information on households' consumption. This implies that it is not possible to disentangle food buyers from purely autarkic farmers. For this reason, I will instead focus on the differences between subsistence (including food buyers and autarkic farmers) and commercial (food sellers) farmers.

22 In the following analysis, I 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, there is no information on whether farmers were renting out some of their plots and as such the figures on land-market participation might be biased downward.

23 More specifically, the median hourly agricultural wage of a woman was 0.8 times that of a man, while that of a child, regardless of the gender, was 0.55 times that of a man.

24 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.

25 As the values of capital items are not available in the UCA, they are derived from the LSMS using the median self-reported value in the agricultural module of the 2010 survey. This was possible as the two surveys ask about exactly the same assets.

26 The list of crops is maize, finger millet, sorghum, rice, beans, field peas, cow peas, pigeon peas, groundnuts, sesame, soy beans, bananas (three varieties: food, brewing, and sweet), cassava, sweet potatoes, potatoes, and coffee.

27 In particular, for each crop I assume the quantity stored is consumed and sold in the same proportion as at the time of the survey, and aggregate across crops using their value. Less than 0.1 percent of the farmers reported storing the whole production and were dropped from the sample.

Table 2. Descriptive Statistics by Subsistence Status

	All	Commercial	Subsistence	*p*-value
Output	118.11	138.04	64.37	0.00
Land (ha)	2.00	2.20	1.48	0.00
Capital	21.55	22.74	18.33	0.10
Labor	2.94	3.04	2.67	0.00
Productivity (log)	− 0.01	0.30	− 0.83	0.00
HH size	5.58	5.66	5.38	0.00
% dependent	47.36	47.28	47.57	0.44
% female head	20.33	19.08	23.71	0.00
Age head	44.77	44.91	44.37	0.31
% illiterate head	47.48	46.30	50.67	0.01
Food insecurity (months)	1.63	1.13	2.98	0.00
% obtained credit	11.43	12.19	9.35	0.00
% land market	42.74	44.76	37.27	0.00
Slight shocks	0.39	0.39	0.39	0.66
Moderate shocks	0.47	0.46	0.51	0.00
Severe shocks	0.52	0.51	0.55	0.04

Source: Author's own computation based on the Ugandan Census of Agriculture.
Note: The figures represent the average of each variable for the whole sample and by subsistence status. Statistics are based on the sample of 23,179 households, 6,416 of whom are in subsistence (i.e. do not sell any of their production). Output and capital are expressed in 10,000 Ugandan shillings (10,000 USh ≈ $5 in 2009). Land represents the total hectares operated in the two agricultural seasons under analysis (and not the total size of the holding). Percent land market indicates the percentage of farmers who operate at least one parcel that was either bought or rented in. The *p*-values refer to the null hypothesis of the average being equal between subsistence and non subsistence farmers.

Table 2 presents the descriptive statistics for the most important variables derived from the UCA for the whole sample and separately for subsistence and commercial farmers. As expected, subsistence producers operate, on average, on a smaller scale (though the difference when it comes to agricultural capital is only significant at the 90 percent confidence level) and have lower levels of production.[28] Reassuringly, in line with the theoretical framework, they are also significantly less productive. Household's composition also tends to differ between subsistence and non-subsistence farms: unsurprisingly, women and illiterate individuals are more likely to be the head of subsistence farm households, reflecting a worse socioeconomic status. There are no significant differences in terms of dependency ratio or age of the head.

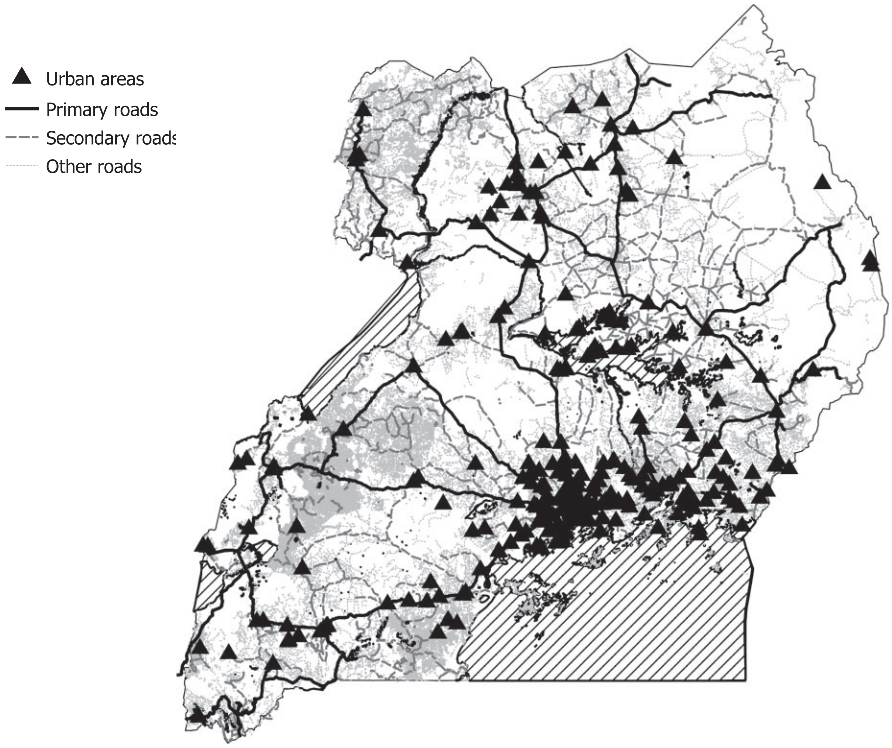
Subsistence households are also more food insecure, as indicated by the self-reported number of months in the previous year they were not able to consume enough food. They also seem to be more affected by other constraints as they are less likely to operate plots of land which were either rented in or bought or to have access to credit.[29] Finally, subsistence farmers are also more likely to have experienced a natural shock affecting the agricultural production (the difference is significant for moderate and severe shocks). Although unsurprising, this fact is somewhat problematic to reconcile with the theoretical framework as the model treats productivity as fixed and deterministic. As explained in more detail in the next section, the measure of productivity used in the empirical analysis (and presented in table 2) is obtained after adjusting for natural shocks and rainfalls, as well as land quality and intermediate inputs used.

Data on land quality are obtained from the GAEZ data set (FAO and IIASA 2015), which provides information on a number of metrics at a very granular level (about 10 square kilometers at the equator).

28 Although commercial farms are on average bigger, as shown in supplementary online appendix S4, there is some significant overlap in the size distribution of commercial and subsistence farms, which is consistent with the theoretical framework and reduces potential collinearity concerns for the empirical analysis.

29 Both these variables are derived from the UCA survey. Farmers were asked, for each plot, how it was obtained and whether they have obtained credit in the previous five years. It is worth stressing that a negative answer to these questions might either indicate lack of access to the market or the mere fact that the household did not need credit or more land.

Figure 6. Roads Network and Urban Areas.



Source: Author's own computation based on Ugandan roads data from UBOS and AFRICOVER.

Using this data set allows one to at least partly disentangle the impact of farmer-specific productivity from differences in land and soil characteristics. However, since observations are geolocated at the village level (i.e. all the holdings in the same village have the same coordinates), this does not allow one to capture differences in land quality within holdings in the same enumeration areas. A similar consideration holds for the rainfall level, obtained from TAMSAT (TAMSAT 2017), which provides information on estimated rainfalls and is used to disentangle the impact of village-specific weather shocks on production.

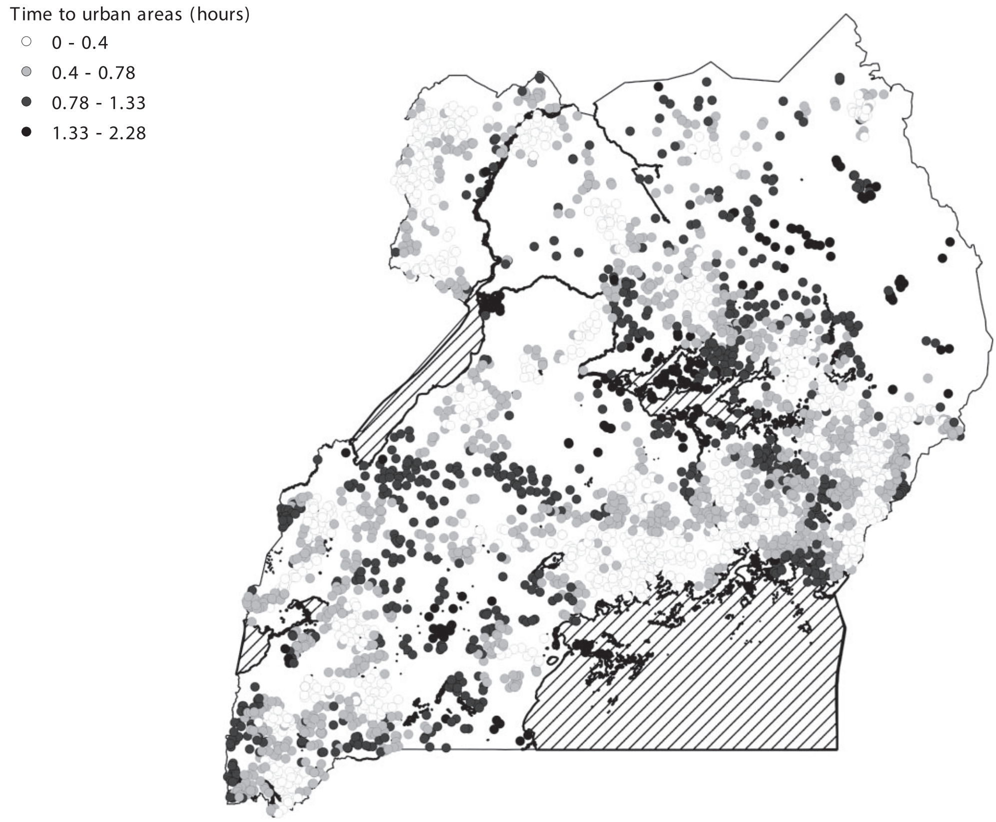
Spatial data sets also play a key role in computing the village-specific estimated time to reach the closest urban area, which is used in the empirical analysis as a proxy for τ. First of all, I use the AFRICOVER data set (FAO 2003), which maps the most important urban conglomerates on the basis of satellite data.[30] To identify the travel time to these urban areas, I employ a data set describing the Ugandan road network in 2008–2009 (UBOS 2012). In practice, following the procedure of Adamopoulos (2019), the country is divided into 1×1 kilometer cells, then each of them is assigned a crossing time on the basis of the terrain roughness and whether the cell is crossed by a road.[31] Once this grid is created, I compute an algorithm which finds the quickest path and the estimated travel time to the closest urban area for each cell (and, in turn, each village). Figures 6 and 7 depict the data used and the final outcome.[32]

30 The use of this data set allows one to avoid arbitrary choice of city size to identify urban areas. The data set provides urban conglomerates as polygons, but in the analysis they are converted to points by considering their centroids.

31 The crossing time depends on the type of road, assuming a speed of 60 km/h for primary roads, 50 for secondary roads, 30 for tertiary roads, and 10 for rural roads.

32 Although such methodology allows one to generate a consistent index of market access across different villages, the resulting estimates in terms of time (though not the relative ranking of different villages) depends crucially on the

Figure 7. Enumeration Areas and Distance from Markets.



Source: Author's own computation based on Ugandan roads data from UBOS and AFRICOVER.

Table 3. Descriptive Statistics, Village Level

	Mean	SD	Median	q25	q75
% subsistence farmers	27.71	27.68	20.00	0.00	44.44
% operating hired or bought land	41.95	32.88	40.00	11.11	70.00
% obtained credit	10.76	18.76	0.00	0.00	11.11
Time to urban area (hours)	0.60	0.48	0.55	0.49	0.78

Source: Author's own computation based on the Ugandan Census of Agriculture, Ugandan roads network, and AFRICOVER.
Note: Subsistence farmers are defined as those who did not sell any of their production.

To conclude, table 3 presents a number of village-level descriptive statistics that provide some context for empirical analysis. It is clear that there is a lot of heterogeneity in terms of credit and land-market activities, as well as in access to urban areas and prevalence of subsistence farming.

In the following section, these microdata are used to compute farm-level productivity and in turn estimate the efficiency losses. Then the patterns of within and across villages misallocation are studied to assess whether they are in line with the qualitative insights of the model and in particular whether there

assumptions made on the time to cross different types of terrains/roads. For this reason, the resulting variable is better interpreted as an index of market accessibility rather than a reliable estimate of the time it takes to reach the closest market.

is an empirical link between the prevalence of subsistence farming and the severity of agricultural factor misallocation.

5. Empirical Specification and Results

The first step required to carry out the empirical analysis is to estimate the farm-specific productivity. This is necessary to generate the full efficiency allocation, where inputs are distributed proportionally to the farmers' agricultural skills, and compare it with the actual distribution and the resulting production in the data. Finally, this section tests the qualitative predictions of the model. In particular,

(1) subsistence farmers have systematically lower marginal productivity (fig. 4);
(2) efficiency-enhancing transactions entail transfers of inputs from subsistence to commercial farmers (fig. 2);
(3) misallocation is more pronounced within subsistence (net buyer and autarkic) farmers in a given area (fig. 4);
(4) transportation costs increase the size of the subsistence sector (figs 1 and 3) and in turn misallocation (fig. 5).

5.1. Productivity Estimation

By rearranging equation (1), productivity can be defined as

$$a_i = \left[\frac{y_i}{(l_i^{\alpha} k_i^{1-\alpha})^{\gamma}}\right]^{\frac{1}{1-\gamma}}. \tag{10}$$

The elasticities are obtained through the estimation of a common production function. Following the procedure suggested by Chari et al. (2021), the function is estimated as[33]

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

As the UCA data are cross sectional, the production function is estimated using the World Bank's LSMS data for Uganda, referring to the four agricultural seasons between 2010 and 2011.[34] 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-specific (indexed by the subscript v) controls, plus farmers' and time fixed effects.[35] The land input is augmented by the parameter q_{iv} which is meant to capture heterogeneous land quality across farms.[36]

33 Unlike Chari et al. (2021), I am 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, I need to make the assumption of equal input elasticities across crops. Reassuringly, in the case of Chari et al. (2021), using crop-specific or general coefficients does not affect the results of the estimation significantly.

34 Unlike Gollin and Udry (2019), I do not use the 2009/10 wave of LSMS data as it does not contain any information on the agricultural capital. They can add it to their analysis since they do not include capital in the Ugandan production function.

35 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.

36 The land quality index is computed on the basis of the soil and land-quality variables in the village section of the LSMS questionnaire, rather than the GAEZ data set. The procedure used to compute the land-quality index follows quite closely that in Restuccia and Santaeulalia-Llopis (2017) and Chen et al. (2017) and is based on a simple regression:

$$\widetilde{y_{iv}} = X_{iv}'\beta_1 + Z_v'\beta_2 + \frac{k_{iv}}{l_{iv}}\beta_3 + \epsilon_{iv},$$

including the value added per labor day of 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

According to these estimates, the shares of labor, land, and capital are 0.403, 0.475, and 0.122 respectively.[37] This implies $\alpha = 0.796$ and $\gamma = 0.597$. Supplementary online appendix S2 compares these estimates with those deriving from other studies dealing with agricultural factor misallocation.

Farm-specific productivity can then be computed using equation (10). However, estimates obtained this way are unlikely to reflect solely idiosyncratic differences in farmers' agricultural skills, but are realistically affected by a number of other factors, such as heterogeneous land quality and natural shocks. In order to alleviate these concerns and obtain more accurate estimates of the actual magnitude of input misallocation, I regress the (log) of productivity estimates obtained using equation (10) on a number of farmer- and village-specific controls capturing these factors, and use the (exponential) residuals to capture productivity. Formally,

$$\log(a_{iv}) = \beta_0 + W_v'\delta + X_{iv}' + \epsilon_{iv},$$

and I use $\exp(\epsilon_{iv})$ as the updated estimate for farm-specific productivity. The village-specific covariates in vector W include land characteristics from the GAEZ data set and rainfall levels from TAMSAT.[38] The household-specific controls instead are the number of natural shocks experienced during the agricultural year (the occurrence and the severity of the shock were self-reported), as well as the type of intermediate inputs used.[39] Supplementary online appendix S2 comments on the resulting distribution in total factor revenue productivity and compares it with the figures from related existing studies.

5.2. Magnitude of Misallocation

Following the seminal paper by Hsieh and Klenow (2009), the magnitude of factor misallocation can be obtained by comparing the actual production, Y^A, with the one that could be obtained in the fully efficient counterfactual where inputs are redistributed proportionally to farmers' productivity, Y^E. Table 4 shows the estimated efficiency losses obtained by adopting different specifications. In particular, the table compares the results obtained without applying any adjustment to the productivity estimates (first column), adding correction to account for differences in land quality and rainfalls (second column), and holding specific natural shocks experienced and intermediate inputs used (third column). In the first row, the counterfactual is generated redistributing inputs across all farmers, while in the following ones transactions are only allowed within farmers in progressively narrower administrative areas.

According to these figures, redistributing land and capital across existing farmers in the whole sample could result in an increase in production of nearly 140 percent. The figure is reduced to 100 percent when

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 values reported were parcel rather than farm specific, they are aggregated at the farm level by considering the weighted 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 $X_{iv}'\beta_1 + Z_v'\beta_2$.

37 Since I was unable to reject the null hypothesis of constant returns to scale, the estimates presented are obtained by imposing $\beta_1 + \beta_2 + \beta_3 = 1$.

38 In particular, land quality is captured by 10 variables from the GAEZ data set, namely, dominant soil, excess salinity, altitude, nutrient availability, nutrient retention, oxygen availability, rooting conditions, slope, toxicity, and workability. Excluding slope and altitude, the variables are categorical and capture the extent to which crops' production is constrained by each factor. Rainfalls are obtained from the TAMSAT data set and enter the specification in percentile as well as their interaction with a dummy variable taking value 1 if the district is prevalently unimodal. When the data are not available (due either to missing observations or to lack of coordinates), I consider the median within the parish where available, or within the subcounty.

39 The vector X includes only a series of dummy variables indicating whether or not the household used organic/inorganic fertilizer, any pesticide, or improved seeds in the production process. Indeed, the exact quantity utilized is not reported in the data. According to Gollin and Udry (2019) the use of intermediate inputs in Uganda is very marginal and does not play a big role in productivity estimation.

Table 4. Estimated Efficiency Losses Y^E/Y^A

	No adjustments	Land quality & rainfalls	Shocks & intermediate inputs
All	2.850	2.392	2.382
District	2.114	2.047	2.044
County	1.996	1.962	1.961
Subcounty	1.634	1.630	1.630
Village	1.386	1.386	1.385

Source: Author's own computation based on the Ugandan Census of Agriculture, GAEZ and TAMSAT.
Note: The efficient counterfactual is obtained as the production achieved when reallocating land and capital such that marginal productivities are equalized across farmers. In the first row, inputs are reallocated across all the farmers in the sample, while in the following rows only among producers in the same district/county/subcounty/village.

the inputs are reallocated efficiently within the 79 districts and to about 40 percent when the counterfactual is generated allowing only for village-level transactions. Although it is helpful to provide such figures for the sake of comparison with existing studies, it is important to point out that given the importance of land in agricultural production (highlighted by the large corresponding elasticity), the most realistic estimates of the actual productivity losses are the ones at the village level, since reallocating land among farmers residing in different areas of the country might not be feasible.

Even in the most conservative estimate, the potential gains for reallocation are sizeable and indicate substantial inefficiencies in the agricultural factor distribution in Uganda. Importantly, the measurement of the magnitude of misallocation depends crucially on the observed dispersion in total factor revenue productivity which a priori can be caused either by input-specific wedges or by constraints which affect the scale of the production rather than the input mix (Hsieh and Klenow 2009). Supplementary online appendix S3 provides a more in-depth discussion on this and performs a decomposition (following Shenoy (2017)) suggesting that "scale" factors like transportation costs seem to be the most important source of inefficiencies in factor distribution.

These findings, although informative, need to be interpreted with caution. Indeed, this methodology implies that the dispersion in marginal productivities necessarily reflects inefficiencies and frictions. As shown by a number of studies on the manufacturing sector (see David and Venkateswaran (2019) and David et al. (2021) for an unified framework and a review of the related literature), this is not necessarily the case, as dispersion in marginal productivities can be the efficient outcome where agents face adjustment costs and uncertainty. As discussed by Gollin and Udry (2019) and Maue et al. (2020), this consideration is even more relevant in the context of data from farm households where there is scope for measurement error when estimating productivities.

In light of this, the following analysis also tests a number of theoretically informed predictions of the distribution of marginal productivities and the relative size of the estimated gains from reallocation with the two-fold objective to validate the hypothesis that transportation costs represent a relevant source of factor misallocation and to show that the observed dispersion in marginal productivity presents some regular patterns consistent with a theoretical framework rather than the mere result of measurement errors.

5.3. Marginal Productivities and Efficiency-Enhancing Transactions

The main intuition underlying the model is that whenever food markets are costly to access due to prohibitive transportation costs, food buyers and autarkic farmers attribute a higher value to their output and as such, in equilibrium, operate an inefficiently high quantity of inputs with respect to commercial farmers. Figures 2 and 4 illustrate this insight. In particular, where $\tau > 1$, marginal productivities are not equalized, but are higher for commercial farmers than for subsistence producers. Also, commercial farmers operate

Table 5. Marginal Productivities and Subsistence Status

	Marginal productivity of land (log)			Marginal productivity of capital (log)		
	(1)	(2)	(3)	(4)	(5)	(6)
% self-consumed	− 0.264***	–	–	− 0.489***	–	–
	(0.04)			(0.06)		
Subsistence farm	–	− 0.232***	− 0.220***	–	− 0.482***	− 0.473***
		(0.02)	(0.02)		(0.03)	(0.03)
Commercial farm	–	–	0.073***	–	–	0.056
			(0.03)			(0.04)
N	23,719	23,719	23,719	23,719	23,719	23,719
Adj. R^2	0.492	0.495	0.495	0.426	0.434	0.434

Source: Author's own computation based on the Ugandan Census of Agriculture.

Note: All regressions include village-level fixed effects. Subsistence farm is a dummy taking value 1 where the all the production was self-consumed. Commercial farm takes value 1 when at least two-thirds of the total production was sold. Standard errors clustered at the district level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

at a smaller-than-efficient scale (and as such produce less than in the full efficiency counterfactual) while food buyers and autarkic producers generally operate too much input.[40]

The first prediction is tested through an OLS regression where the (log) marginal productivity of land and capital is regressed on the type of farmer (subsistence or commercial) as the main explanatory variable. Village fixed effects are included to control for unobserved and time-invariant determinants of productivity that could otherwise affect the analysis. The results of these regressions are shown in table 5 for three different specifications: in columns 1 and 4, the main independent variable is the share of output self-consumed, while columns 2 and 5 only include a binary variable taking value 1 when the farmers consume all of their production, and columns 3 and 6 add a dummy equal to 1 when farmers sell more than two-thirds of their production. The results are consistent across all specifications and show that subsistence farmers display significantly lower marginal productivity of both capital and land.

The nature of the efficiency-enhancing input transactions is studied by comparing the actual levels of production of each farm with those in the full efficiency counterfactual. The rationale is that, whenever farmers are operating an inefficiently low amount of input, their resulting production will be lower than the one they would achieve if factors of production were efficiently distributed.

More specifically, I estimate a probit model where the dependent variable takes value 1 whenever the farmer's production is lower than that in the full efficiency counterfactual, and the main explanatory variable captures whether the farmer operates for subsistence or is a commercial one (the explanatory variables used are the same as in table 5). The average marginal effects are displayed in table 6, where the counterfactual efficient levels of production are obtained equalizing the marginal productivities among farmers located in the same districts (columns 1–3), subcounties (columns 4–6), or villages (columns 7–9).

In line with the findings on marginal productivities of land and capital, the results indicate that subsistence farmers are more likely to operate at a larger-than-efficient scale, indicating that efficiency-enhancing land and capital transactions would imply the transfer of factors of production from subsistence to commercially oriented farmers. The findings are always statistically significant and robust across specifications. In terms of economic magnitude, the coefficients imply that subsistence farmers are 25 to 30 percent more likely to operate on a larger-than-efficient scale as compared to commercial farmers.

40 As shown in fig. 2, a certain share of the autarkic farmers are actually operating on an inefficiently small scale. This is relatively unimportant for the empirical analysis as it still implies that, compared to subsistence farmers, commercial farmers are more likely to operate on a smaller-than-efficient scale.

Table 6. Actual vs Efficient Farm Scale

	Dependent variable: The farm is smaller than efficient ($y_i < y_i^e$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
% self-consumed	−0.056**	–	–	−0.078***	–	–	−0.101***	–	–
	(0.02)			(0.02)			(0.01)		
Subsistence farm	–	−0.080***	−0.078***	–	−0.089***	−0.088***	–	−0.095***	−0.090***
		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)
Commercial farm	–	–	0.006	–	–	0.009	–	–	0.023**
			(0.01)			(0.01)			(0.01)
Counterfactual level	District	District	District	Subcounty	Subcounty	Subcounty	Village	Village	Village
N	23,719	23,719	23,719	23,719	23,719	23,719	23,719	23,719	23,719
Pseudo R^2	0.003	0.006	0.006	0.002	0.006	0.006	0.002	0.006	0.006
Mean dep var	0.261	0.261	0.261	0.312	0.312	0.312	0.383	0.383	0.383

Source: Author's own computation based on the Ugandan Census of Agriculture.
Note: The table displays the average marginal effects from the probit model. Subsistence farm is a dummy taking value 1 where the all the production was self-consumed. Commercial farm is a dummy taking value 1 when at least two-thirds of the total production was sold. The dependent variable takes value 1 when the farms' production is lower than that in the full efficiency counterfactual. In columns 1–3, the counterfactual allocation is obtained redistributing inputs within districts, in columns 4–6 within subcounties, and in columns 7–9 within villages. Standard errors clustered at the district level in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

5.4. Within Groups Dispersion

Another interesting prediction that can be tested to validate the model is that misallocation in a given village/area should be more pronounced within subsistence farmers. This is because subsistence farmers include both food sellers and autarkic producers, which in equilibrium, for $\tau > 1$, present different levels of marginal productivity. In particular (see fig. 4), not only do autarkic farmers have higher marginal productivity than net buyers, but they also have differing values among each other.

In order to test whether this pattern is observed in the data, I estimate the regression:

$$e_{gz} = \beta_1 \text{Subsistence\_Farms}_{gz} + X'_{gz}\gamma + \nu_z + \epsilon_{gz},$$

where the dependent variable is the potential gains from reallocation (in logs) for each group g (subsistence or non-subsistence) in zone z (district, subcounty, or village), and β_1 is the coefficient of interest, which according to the model should take on positive values. Additionally, I include a number of other group-specific factors that might affect the magnitude of misallocation, which are summarized by the vector X. In particular, I control for the average land and capital per worker in each group, as well as their average productivity and productivity dispersion. While the village-level analysis is less likely to be affected by differences in land quality and natural shocks, the subcounty and district-level findings are based on a higher number of observations per group and as such they can be more reliable and less susceptible to outliers.

As shown in table 7, the estimates of β_1 are positive across all the specifications considered.[41] In terms of magnitude, the percentage difference (implied by the log linear specification) appears to be larger for broader geographical areas and somewhat less pronounced within villages (the productivity dispersion is 13.7 percent higher in the case of the district-level analysis and only 5 percent in the village-level analysis). This arguably reflects the fact that a considerable share of the misallocation is captured by village fixed effects in the latter specifications. However, the estimates are always statistically significant at any conventional confidence level.

41 The underlying dispersion in productivity is always included as it leads structurally to higher levels of misallocation (as long as land and capital tend to be more uniformly distributed across farmers than their skill level).

Table 7. Misallocation within Subsistence Group

	Dep var: efficiency losses (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence farms	0.135***	0.132***	0.090***	0.101***	0.065***	0.047***
	(0.03)	(0.04)	(0.01)	(0.01)	(0.01)	(0.02)
Productivity dispersion (log)	0.056**	0.254***	0.080***	0.236***	0.070***	0.125***
	(0.03)	(0.05)	(0.01)	(0.03)	(0.01)	(0.02)
Average land (log)	–	0.036	–	– 0.010	–	– 0.027
		(0.10)		(0.02)		(0.02)
Average capital (log)	–	0.094***	–	0.072***	–	0.052***
		(0.02)		(0.01)		(0.01)
Average productivity (log)	–	– 0.214***	–	– 0.186***	–	– 0.077***
		(0.06)		(0.04)		(0.02)
Counterfactual level	District	District	Subcounty	Subcounty	Village	Village
N	158	158	1,460	1,460	4,240	4,240
Adj. R^2	0.875	0.914	0.655	0.709	0.872	0.887

Source: Author's own computation based on the Ugandan Census of Agriculture.

Note: The dummy subsistence farms indicates the households in the district/subcounty/village which did not sell any of their production. Average land and capital refers to the per unit of labor figures. Standard errors clustered at the district level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.5. Transportation Costs, Subsistence Farming, and Misallocation

The most important insight of the theoretical model is that, all else being equal, lower access to food markets increases the size of the subsistence sector (fig. 3) and in turn the magnitude of factor misallocation across farmers (fig. 5).

On a theoretical ground, the concept of market access is rather complicated and depends on a number of different factors, including some that are typically difficult to capture such as price volatility and market thinness. Transportation costs are arguably the easiest component of market access to measure, and they are used as a proxy for τ in the empirical analysis. As explained in the data section, village-level transportation costs are captured by the estimated travel time to the closest urban area given the existing infrastructures.[42] The underlying intuition is that farmers that are better connected to urban areas should, ceteris paribus, be more likely to engage in commercial agriculture. It is however important to point out that this only represents one of the possible forms of transaction cost faced by farmers and that the index is only a proxy for the actual transportation costs derived using available information on road infrastructure.

In order to study relationship between the size of the subsistence sector and this measure of market access, a village-level regression is estimated[43]

$$\text{Subsistence\_index}_v = \beta_1 \text{Travel\_time\_(log)}_v + \gamma X_v' + \delta_d + \epsilon_v,$$

where the dependent variable is an index capturing the relative importance of subsistence farming in the village and the main explanatory variable is the (log) of the estimated travel time to the closest urban area. The vector X includes a number of potential confounding factors, and district-level fixed effects δ are also added to control non-parametrically for other unobserved factors, like general degree of economic development, availability of other natural resources, non-farming employment opportunities, that might otherwise affect the results. The coefficient of interest is β_1 and it can be interpreted as an elasticity since the subsistence indexes used are expressed in percentages.

42 The closest urban area is in this case the one that takes the least time to reach, rather than the closest geographically.

43 The findings presented in this section are based on the subset of villages that I was able to geolocate (about 90 percent of the original sample), as the travel time to urban area could only be estimated for the geolocated enumeration areas.

Table 8. Time to Urban Area and Subsistence Farming

	Dep var: Subsistence farming prevalence					
	(1)	(2)	(3)	(4)	(5)	(6)
Time to urban area (log)	0.071***	0.101***	0.059**	0.095***	0.022	0.039**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Average land (log)	–	– 0.057***	–	– 0.068***	–	– 0.032***
		(0.01)		(0.01)		(0.01)
Average capital (log)	–	0.023***	–	0.023***	–	0.003
		(0.01)		(0.01)		(0.01)
Average productivity (log)	–	– 0.029***	–	– 0.032***	–	– 0.004
		(0.00)		(0.00)		(0.00)
Productivity dispersion (log)	–	0.003	–	0.003	–	– 0.001
		(0.00)		(0.00)		(0.00)
N	2,584	2,584	2,584	2,584	2,584	2,584
Adj. R^2	0.225	0.373	0.223	0.390	0.221	0.315

Source: Author's own computation based on the Ugandan Census of Agriculture.
Note: District-level fixed effects are included in all the specifications. Subsistence farming is captured by the fraction of the total land used by purely subsistence farmers (columns 1–2), the percentage of farmers who did not sell any of their output (columns 3–4), and by the percentage of the total output self-consumed (columns 5–6). Standard errors clustered at the district level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The resulting estimates are displayed in table 8 and generally show that, regardless of whether controls are included or not, the subsistence sector tends to be larger in villages with lower market access. In particular, three different outcomes are considered, namely, the percentage of land operated by purely subsistence farmers, the percentage of subsistence farmers, and the fraction of the total output consumed by farmers in each village. Although they all generally describe the same phenomenon, it is worth highlighting that the first option represents the most naturally related to the model, as increasing τ has a direct impact on the share of input operated by subsistence farmers. The controls include the average farmers' productivity and land and capital per worker, as well as productivity dispersion.

According to these estimates, the elasticity of subsistence farming prevalence with respect to this measure of market access ranges between (a non-significant) 0.02 and 0.07 in the unconditional regressions, and between 0.04 and 0.1 where additional controls are included. Although the magnitude is economically significant, there are plausibly factors other than transportation costs that determine the prevalence of subsistence agriculture. Additionally, the travel times computed might suffer from some slight measurement error, which might in turn result in some attenuation bias. Reassuringly, when controls are included, the estimates are always significant at the 95 percent level at least.

Generally, the controls have the expected sign; in particular, higher average productivity and land per capita result in lower levels of subsistence farming. Somewhat unexpectedly, the average amount of capital operated has the opposite effect, as villages where farmers tend to operate more capital have less commercial farming. This might reflect the fact that more capital-intensive production techniques are used in areas with less fertile land or less suitable climates.

Ultimately, these findings indicate that there is a correlation between subsistence farming and market access at the village level and confirm a widely accepted pattern in the existing theoretical and empirical literature (Gollin and Rogerson 2014). The conclusive part of this analysis section aims to study whether and to what extent costly market access can also be a concurrent cause of factor misallocation. Specifically, the estimated regression is

$$\mathrm{e}_v = \beta_1 \text{Travel\_time\_(log)}_v + \gamma X'_v + \delta_d + \epsilon_v,$$

where the dependent variable e is the potential gain from reallocation achievable in each village v expressed as a percentage (i.e. $(Y_v^E / Y_v^A) - 1$). The main hypothesis is that to higher transportation costs (τ) should correspond higher inefficiencies in the distribution of land and capital, and in turn higher potential

Table 9. Misallocation and Subsistence Farming

	Dep var: Efficiency losses (percentage); mean = 0.40							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time to urban area (log)	0.082**	0.068**	–	–	–	–	–	–
	(0.03)	(0.02)						
Subsistence farming	–	–	0.206***	0.100***	0.191***	0.064**	0.035	0.039
			(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
Productivity dispersion	0.083***	0.145***	0.076***	0.200***	0.076***	0.200***	0.073***	0.200***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
% land market	–	– 0.008	–	– 0.011	–	– 0.012	–	– 0.010
		(0.02)		(0.02)		(0.02)		(0.02)
% credit	–	0.012	–	0.007	–	0.009	–	0.008
		(0.01)		(0.01)		(0.01)		(0.01)
Average land (log)	–	– 0.014*	–	– 0.009	–	– 0.011	–	– 0.015*
		(0.01)		(0.01)		(0.01)		(0.01)
Average capital (log)	–	0.008	–	0.009	–	0.009	–	0.010
		(0.01)		(0.01)		(0.01)		(0.01)
Average productivity (log)	–	– 0.126***	–	– 0.186***	–	– 0.186***	–	– 0.189***
		(0.01)		(0.01)		(0.01)		(0.01)
N	2,584	2,584	2,584	2,584	2,584	2,584	2,584	2,584
Adj. R^2	0.187	0.478	0.193	0.501	0.191	0.499	0.178	0.498

Source: Author's own computation based on the Ugandan Census of Agriculture.

Note: District-level fixed effects are included in all the specifications. Percentage land markets indicate the percentage of farmers operating at least one hired or bought plot of land, while percentage credit indicates the percentage of farmers that obtained credit in the previous five years. Subsistence farming is captured by the fraction of the total land used by purely subsistence farmers (columns 3–4), the percentage of farmers who did not sell any of their output (columns 5–6), and by the percentage of the total output self-consumed (columns 7–8). Standard errors clustered at the district level in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

gains from reallocation. As shown in table 8, transportation costs can only partly explain the prevalence of subsistence farming across villages; for this reason, I also estimate an alternative specification:

$$\mathrm{e}_v = \beta_1 \text{Subsistence\_index}_v + \gamma X_v' + \delta_d + \epsilon_v,$$

which captures the direct link between the size of the subsistence sector (proxied by the same indexes considered in table 8) and the magnitude of input misallocation. These specifications capture the broad relationship between factor misallocation and subsistence farming, where the latter is not necessarily driven by variations in the proxy for market access.

Table 9 shows the resulting estimates for a number of different specifications considering alternative explanatory variables and with or without including additional controls. In line with the theoretical predictions of the model, the point estimates for β_1 are always positive and statistically significant, suggesting a positive correlation between the estimated travel time to urban markets, the prevalence of subsistence farming and the magnitude of agricultural factor misallocation.

In particular, the findings indicate that a 10 percent increase in the estimate to reach the closest urban area correlates with a 0.7 (0.8 in the case of the unconditional regression) percentage point increase in the magnitude of factor misallocation, where the average village-level misallocation is 40 percentage points. Although economically and statistically significant, it is clear that market access only explains a relatively low share of input misallocation. A potential explanation for this is that the estimated transportation costs are only one (and imperfect) of the determinants of market access and in turn of subsistence farming.

In columns 3–8, I examine the direct relationship between factor misallocation and the size of the subsistence sector across villages. The estimated correlation is positive and statistically significant when considering the share of land operated by subsistence farmers (columns 3–4) and the share of farmers that did not sell any output (columns 5–6), while the (still positive) coefficients are not statistically significant when using the share of output self-consumed (columns 7–8).

In this case, the magnitude of the coefficients suggests that a percentage point increase in the size of the subsistence sector correlates with an increase of factor misallocation of 0.2 to 0.06 percentage points. In terms of standard deviations, a unit increase in the size of the subsistence sector correlates with between 0.05 and 0.15 units increase in factor misallocation. Overall, these findings suggest the existence of a significant link between the prevalence of subsistence farming and factor misallocation at the village level, although this only explains a relatively low share of the variation in factor misallocation observed across villages.[44]

Interestingly, there is no significant link between the magnitude of input misallocation and the fraction of farmers operating land accessed through market transactions and/or of the producers who had access to credit in the previous five years. Although these findings are only suggestive, they seem to indicate that in the case of Uganda, failures and frictions in the credit and land markets play no substantial role in determining inefficiencies in the distribution of agricultural factor.[45]

Overall, these results are in line with the theoretical model and, along with the findings presented in the previous sections, suggest that costly access to the food market represents a good candidate to explain the observed agricultural factor misallocation. In particular, it appears clear that subsistence farmers tend to operate a disproportionately large amount of input and that the resulting efficiency losses are more pronounced in areas with poorer market access. Conversely, there is no correlation between misallocation and the volume of activity in the land and credit markets.

6. 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 (2022), the huge disparities observed cannot be attributed to systematic differences in land endowments, but are the outcome of suboptimal production patterns. A particularly appealing explanation for them is that, as markets for agricultural inputs and output fail systematically in low-income countries, resources are not allocated efficiently among existing farmers.

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. Although there is a lively debate on the actual magnitude of these efficiency losses (Gollin and Udry 2019; Maue et al. 2020), less attention has been paid to the potential causes of misallocation (in sharp contrast with the corresponding literature on the manufacturing sector).

Specifically, all existing studies (with the exception of Shenoy (2017)) impute agricultural factor misallocation to restrictions to 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 is robust and persuasive. However, as stressed by

44 According to these findings, misallocation is crucially driven by average productivity and productivity dispersion, which are exogenous in the model. In supplementary online appendix S5, I provide an extension of the model where selection into agriculture is endogenous, showing that higher costs of accessing the markets increase the variance in and reduces the average of agricultural productivity. It is therefore possible that the estimated impact of transportation costs is a lower bound.

45 It is worth pointing out that, as explained above, I proxy the level of activity in the land market as the share of farmers who operate at least one parcel of land that was either rented in or bought, while in the case of the credit market I consider the fraction that had received credit from any source in the previous five years. Thus, these variables are possibly inaccurate since I do not observe whether farmers are renting land out nor the nature of the land transactions (e.g. whether the owner/seller is a relative) and I am unable to differentiate between farmers who did not receive credit because they did not need it or because they had no access to it.

virtually all existing studies, high levels of misallocation seem to persist also in countries where land markets are relatively 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 to address this gap by considering the concurrent role of distortions in output markets in determining resource misallocation in agriculture. In particular, it explores the possibility that, even in the presence of frictionless input markets, severe misallocation can arise when, due to transportation costs, farmers find it more convenient to grow the food they consume rather than buy 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).

This work offers 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 an inefficient distribution of resources. The hypotheses of the model are tested using data from Uganda, where land transactions are relatively more common and not subject to formal limitations.

The estimated magnitude of the productivity gains from reallocation is around 140 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 input markets are a necessary, yet not sufficient, condition to achieve an efficient distribution of agricultural inputs across farmers. When considering only misallocation within farmers operating in the same villages, the estimated productivity losses are about 40 percent.

In line with the model proposed, the empirical analysis shows 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. Consistent with this hypothesis, there is a positive correlation between the level magnitude of misallocation and the prevalence of subsistence farming, possibly driven by high transportation costs. In contrast to previous studies, frequency of land transactions and access to credit do not seem to play any role.

To my 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, I argue that complementary policies/investment aimed at reducing frictions in food markets are needed to achieve an efficient input distribution across farmers and enhance agricultural productivity.

Data Availability

The Ugandan Census of Agriculture (UCA) data belongs to the Ugandan Bureau of Statistics (UBOS), which I would like to thank for sharing the microdata with me. I do not have the permission to share it.

References

Adamopoulos, T. 2011. " Transportation Costs, Agricultural Productivity, and Cross-Country Income Differences." *International Economic Review* 52(2): 489–521.

——. 2019. "Spatial Integration, Agricultural Productivity, and Development: A Quantitative Analysis of Ethiopia's Road Expansion Program."In *2019 Meeting Papers*, Number 86. Society for Economic Dynamics.

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia. 2022. " Misallocation, Selection, and Productivity: A Quantitative Analysis with Panel Data from China." *Econometrica* 90(3): 1261–82.
Adamopoulos, T., and D. Restuccia. 2014. "The Size Distribution of Farms and International Productivity Differences." *American Economic Review* 104(6): 1667–97.
——. 2022. " Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data." *Review of Economic Studies* 89(4): 1629–53.
Ayerst, S., L. Brandt, and D. Restuccia. 2020. "Market Constraints, Misallocation, and Productivity in Vietnam Agriculture." *Food Policy* 94: 101840
Baland, J.-M., F. Gaspart, J.-P. Platteau, and F. Place. 2007. " The Distributive Impact of Land Markets in Uganda." *Economic Development and Cultural Change* 55(2): 283–311.
Britos, B., M. A. Hernandez, M. Robles, and D. R. Trupkin. 2021. "Land Market Distortions and Aggregate Agricultural Productivity: Evidence from Guatemala." *Journal of Development Economics* 155: 102787.
Chari, A., E. M. Liu, S.-Y. Wang, and Y. Wang. 2021. "Property Rights, Land Misallocation, and Agricultural Efficiency in China." *Review of Economic Studies* 88(4): 1831–62.
Chen, C., D. Restuccia, and R. Santaeulàlia-Llopis. 2017. "The Effects of Land Markets on Resource Allocation and Agricultural Productivity." NBER Working Paper (w24099).
——. 2021. "The Effects of Land Markets on Resource Allocation and Agricultural Productivity." *Review of Economic Dynamics* 45: 41–54.
Christiaensen, L., and L. Demery. 2017. "Agriculture in Africa: Telling Myths from Facts." *Food Policy* 67: 1–11.
Costinot, A., and D. Donaldson. 2016. "How Large Are the Gains from Economic Integration? Theory and Evidence from US Agriculture, 1880-1997." NBER Working Paper (w22946).
David, J. M., and V. Venkateswaran. 2019. " The Sources of Capital Misallocation." *American Economic Review* 109(7): 2531–67.
David, J. M., V. Venkateswaran, A. P. Cusolito, and T. Didier. 2021. " Capital Allocation in Developing Countries." *World Bank Economic Review* 35(4): 1102–21.
De Janvry, A., M. Fafchamps, and E. Sadoulet. 1991. " Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained." *Economic Journal* 101(409): 1400–17.
Deininger, K. W., and P. Mpuga. 2003. "Land Markets in Uganda: Incidence, Impact and Evolution over Time." Conference Paper, International Association of Agricultural Economists (IAAE) Annual Meeting. Durban, South Africa.
Dillon, B., and C. B. Barrett. 2017. "Agricultural Factor Markets in Sub-Saharan Africa: An Updated view with Formal Tests for Market Failure." *Food Policy* 67: 64–77.
Fafchamps, M. 1992. " Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World." *American Journal of Agricultural Economics* 74(1): 90–99.
FAO. 2003. "Multipurpose Landcover Database of Uganda – AFRICOVER." [Online]; [Accessed August 2019]. https://data.apps.fao.org/map/catalog/static/api/records/9a4cc73d-ffb2-4d3f-abc9-2d83669da17a.
FAO and IIASA. 2015. "Global Agro-Ecological Zones v 3.0.". [Online]; [Accessed August 2019]. https://www.gaez.iiasa.ac.at/.
Gollin, D., S. L. Parente, and R. Rogerson. 2007. " The Food Problem and the Evolution of International Income Levels." *Journal of Monetary Economics* 54(4): 1230–55.
Gollin, D., and R. Rogerson. 2010. "Agriculture, Roads, and Economic Development in Uganda." NBER Working Paper (w15863).
——. 2014. "Productivity, Transport Costs and Subsistence Agriculture." *Journal of Development Economics* 107: 38–48.
Gollin, D., and C. Udry. 2019. "Heterogeneity, Measurement Error and Misallocation: Evidence from African Agriculture." NBER Working Paper (w25440).
Haque, Z. 2022. "The Production Engel." Working Paper. New York University.
Hsieh, C.-T., and P. J. Klenow. 2009. " Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403–48.
Hunt, D. 2004. " Unintended Consequences of Land Rights Reform: The Case of the 1998 Uganda Land Act." *Development Policy Review* 22(2): 173–91.
S Jayne, T.. 1994. " Do High Food Marketing Costs Constrain Cash Crop Production? Evidence from Zimbabwe." *Economic Development and Cultural Change* 42(2): 387–402.

Lagakos, D., and M. E. Waugh. 2013. “ Selection, Agriculture, and Cross-Country Productivity Differences.” *American Economic Review* 103(2): 948–80.

Li, N. 2023. “In-Kind Transfers, Marketization Costs and Household Specialization: Evidence from Indian Farmers.” *Journal of Development Economics* 164 103130

Maue, C. C., M. Burke, and K. J. Emerick. 2020. “Productivity Dispersion and Persistence among the World’s Most Numerous Firms.” NBER Working Paper (w26924).

Nabwire, L. 2015. “Uganda’s Dilemmas in the Transition to Modern Commercial Agriculture: Implications for the Poverty Reduction Agenda.”. Research Paper, Future agriculture.

W Omamo, S.. 1998. “ Transport Costs and Smallholder Cropping Choices: An Application to Siaya District, Kenya.” *American Journal of Agricultural Economics* 80(1): 116–23.

Perego, V. M. 2019. “Crop Prices and the Demand for Titled Land: Evidence from Uganda.” *Journal of Development Economics* 137: 93–109.

Qin, Y., and X. Zhang. 2016. “The Road to Specialization in Agricultural Production: Evidence from Rural China.” *World Development* 77: 1–16.

Restuccia, D., and R. Santaeulalia-Llopis. 2017. “Land Misallocation and Productivity.” NBER Working Paper (w23128).

Restuccia, D., D. T. Yang, and X. Zhu. 2008. “ Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis.” *Journal of Monetary Economics* 55(2): 234–50.

Shamdasani, Y. 2021. “Rural Road Infrastructure & Agricultural Production: Evidence from India.” *Journal of Development Economics*: 152 102686.

Shenoy, A. 2017. “Market Failures and Misallocation.” *Journal of Development Economics* 128–80: .

Sotelo, S. 2020. “Domestic Trade Frictions and Agriculture.” *Journal of Political Economy* 128(7): 2690–738.

TAMSAT. 2017. “TAMSAT Rainfall Estimates.” [Online]; [Accessed August 2019]. http://www.tamsat.org.uk/

Tombe, T. 2015. “ The Missing Food Problem: Trade, Agriculture, and International Productivity Differences.” *American Economic Journal: Macroeconomics* 7(3): 226–58.

UBOS. 2010. “Ugandan Census of Agriculture – 2009.”https://www.ubos.org/wp-content/uploads/publications/03_2018UCAholding.pdf

UBOS. 2012. “Roads: Uganda 2008–2009.” [Online]; [Accessed August 2019]. https://data.humdata.org/dataset/uganda-roads?force_layout=desktop

UBOS. 2018. “Statistical Abstract 2018.” Ugandan Bureau of Statistics. https://www.ubos.org/wp-content/uploads/publications/05_2019STATISTICAL_ABSTRACT_2018.pdf

Ulimwengu, J. M., and R. Ramadan, 2009. “How Does Food Price Increase Affect Ugandan Households? An Augmented Multimarket Approach.” IFPRI discussion paper, International Food Policy Research Institute (IFPRI).

Valentinyi, A., and B. Herrendorf. 2008. “Measuring Factor Income Shares at the Sectoral Level.” *Review of Economic Dynamics* 11(4): 820–35.