Essays in Development Economics

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE

IN APPLICATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

Bruno Morando

Supervised by

Tara Mitchell

2020

Declaration

I declare this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library condition of use and acknowledgement.

Abstract

This dissertation is a collection of three essays with a geographical and thematic common denominator. The broad objective of the thesis is to study how the Ugandan agricultural sector is affected by the countrys poor infrastructural network which prevents the smooth flow of goods and information. Overall, the dissertation can be split in two main sections which differ slightly in terms of topic and methodology. The first part is composed of the first two papers, where I use micro level farm data from the Ugandan Census of Agriculture combined with spatial data on the Ugandan road network to explore the detrimental impact of transportation costs on the agricultural sector in terms of misallocation of resources. More specifically, in the first paper I show that in more remote areas less productive farmers operate a larger share of agricultural inputs which significantly reduces the aggregate total factor productivity. I attribute it to the lack of access to output (food) markets, which discourages commercial agriculture while making subsistence farming more profitable. This prevents efficiency enhancing land transactions in spite of the lack of formal restrictions to land markets. The second paper builds on a similar intuition, but rather than studying the inefficiencies in the distribution of factors across heterogeneous producers, it examines the resulting distortions in crop choices. Using detailed spatial data on crop specific agroclimatic suitability developed by the GAEZ project, I generate the patterns of natural comparative advantages across crops within the main market areas of the country and show how costly market access brings about deviations from the optimal geographical distribution of crops. In particular, I find that not only do more isolated villages devote a systematically lower area to cash and perishable crops, but they are also less responsive to comparative advantages in their production decisions. In both papers, the empirical analysis is supported by theoretical models which illustrate the main intuition and derive testable predictions. Conversely, the third paper is purely empirical and aims to study how the extension of mobile telephony in rural Uganda impacted farmers welfare. In particular, I want to test whether the resulting easier access to price information was more beneficial to more educated farmers a priori more likely to adopt the new technology or to less schooled individuals more likely to suffer from information asymmetry when dealing with traders. The empirical analysis corroborates the latter hypothesis,

showing that farmers with a lower level of education receiving mobile phone coverage were able to negotiate higher prices when selling agricultural produce at their farm gate and displayed an increase in income and wealth levels.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

The fact that low income countries are disproportionately less productive in the agricultural sector, which nevertheless employs a large share of their labour force, is perhaps the best established empirical regularity in the literature of development economics. This phenomenon is known as the agricultural productivity gap and is considered to be the very root of cross country income differences (Gollin, Parente, and Rogerson 2007, Restuccia, Yang, and Zhu 2008, Lagakos and Waugh 2013). Why aren't productivity differences as pronounced in the manufacturing sector? Why don't workers reallocate to relatively more profitable activities? Why aren't developing countries importing food products from rich economies given their apparent comparative advantage in farming? Those are all important questions that have animated the economic debate in recent years.

Such a phenomenon is even more puzzling in consideration of the recent work by Adamopoulos and Restuccia, 2018, which shows that this productivity gap cannot be attributed to systematic differences in the climate/land quality endowments between developing and rich countries. Instead, the authors find that the gap is solely driven by differences in production efficiency and (although to a lesser extent) in crop choices. In the second and third chapter of this dissertation, I look separately at these two issues, and examine the factors that might determine these sub optimal outcomes. In particular, I will consider the rather emblematic case of Uganda, whose geographical and climatic characteristics are deemed to be particularly suitable for farming, but nevertheless displays very low levels of agricultural productivity.

Both chapters build on the same intuition: namely, that prohibitive transportation costs, which result from the lack of effective infrastructures, crucially shape the agricultural sectors of low income countries. In particular, they determine the number, size and productivity of operating farms as well as the crop choices made by farmers. Indeed, other features that characterize farming activities in poor economies are the smaller average size of the production units (Adamopoulos and Restuccia 2014) and the fact that they typically produce for self-consumption rather than marketization purposes (Li 2017). Both these facts can be seen as the result of the difficulties rural agents face in connecting to the food markets either as sellers (and this explains the low marketization of the agricultural produce) or buyers (which causes many people to produce their own food).

In the second chapter, I show that costly market access is a concurring cause of productivity losses through agricultural factor misallocation across heterogeneous farmers. In particular, transportation costs can drive a wedge between consumer and producer prices and therefore between the shadow value attached to agricultural output by subsistence farmers (who are typically net food buyers) and the commercial ones (who are net sellers by definition). This prevents efficiency and (in a frictionless world) welfare enhancing land transactions from less to more productive individuals even when (as is the case of Uganda) they are not formally restricted. The empirical analysis corroborates this hypothesis. Specifically, I find that in more isolated areas subsistence farming is more widespread and, in turn, misallocation more pronounced. This represents an important contribution to the related literature that typically focuses on the misallocation caused by legal limitations to land transactions (Restuccia and Santaeulalia-Llopis 2017, Chen, Restuccia, and Santaeulàlia-Llopis 2017, Adamopoulos et al. 2017, Chari et al. 2017).

In the third chapter, I focus instead on the impact of transportation costs on cropping patterns. Rather than looking at lack of integration across different regional markets (Costinot and Donaldson 2016, Sotelo 2019), I examine how they affect land use decisions made by farmers who serve the same urban market. In line with the theoretical model developed, I find first order and second order effects. Firstly, producers in more isolated areas tend to devote a higher share of land to food crops and/or cash crops that are easier to store and transport to the market. Secondly, since farmers with lower market access are more likely to produce for self-consumption, they fail to specialise and might ultimately use more land to grow crops which are relatively less suitable to meet their consumption needs. Such dynamics can explain why cropping patterns are inefficient not only when considering their spatial distribution across regional markets, but also within relatively narrow areas.

The fourth chapter has a slightly different focus and deals with the issue of lack of access to price information, which is deemed to be crucial for farmers' welfare, especially when they negotiate their sales at the farm gate interacting with more informed traders (Courtois and Subervie 2014). In order to study the effect of the relaxation of this constrain, I exploit the rollout of mobile phone coverage in rural Uganda between 2003 and 2009. Unlike similar studies, I relax the implicit assumption of a homogeneous effect across the population, allowing it to be contingent on the household specific level of formal schooling, acknowledging the theoretically ambiguous relation between modern ICTs' provision and human capital. On the one hand, more educated individuals are generally quicker in adopting new technologies and grasping their benefits. On the other, easier access to information might be more valuable to farmers with *a priori* worse knowledge about market dynamics. The empirical analysis corroborates the second view, showing that farmers with lower human capital manage to receive higher prices when selling their crops to traders once mobile phone coverage is made available thus increasing their return to farming.

Chapter 2

Subsistence farming and factor misallocation: Evidence from Ugandan agriculture

2.1 Introduction

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

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

Consistently with this theory, reforms aimed at relaxing long lasting limitations to land transactions have been shown to increase efficiency in resource allocation and foster agricultural productivity in Malawi (Restuccia and Santaeulalia-Llopis 2017), Ethiopia (Chen, Restuccia, and Santaeulàlia-Llopis 2017) and China (Adamopoulos et al. 2017, Chari et al. 2017). These studies have opened a debate on the actual severity of agricultural input misallocation in developing countries and its implications in terms of overall productivity losses (Restuccia and Rogerson 2017). On the one hand, Adamopoulos et al., 2017 claim that the consequences of agricultural factor misallocation are more significant than it would appear from a static analysis as it affects also the sorting of workers across sectors. On the other hand, Gollin and Udry, 2019 suggest that the existing estimates of misallocation might be inflated as they fail to disentangle the actual dispersion in farmers' productivity from other unobserved sources of output variation such as measurement errors and land quality heterogeneity.

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

More specifically, another distinctive fact regarding agriculture in developing countries is that a large share of the output is consumed by the individuals who cultivate it on small scale holdings (Adamopoulos and Restuccia 2014). This suggests that accessing the food market (as buyers or sellers) might be costly and farmers are better off producing for self consumption rather than for commercial purposes. Since the seminal paper by De Janvry, Fafchamps, and Sadoulet, 1991, a number of studies have been proposed to explain the behaviour of farm households facing implicit (Fafchamps 1992) or explicit (Omamo 1998) output marketization costs. The common denominator of these models is that the producers' output shadow price is a function of their specific consumption needs and production capacity rather than being constant and equal to the market price. Namely, net food buyers attach higher value than net food sellers to their output as a result of the above mentioned frictions that create a wedge between producer and consumer prices.

In this paper, I develop this basic intuition into a model showing that resource misallocation can occur even in the presence of virtually frictionless input markets as long as, due to costly access to the food market, subsistence farmers attribute higher value to their produce and thus are less

[1] A significant exception is Shenoy, 2017, who also considers imperfections in the credit markets.

willing to transfer their factors of production to more able market oriented farmers.[2] Crucially, since these frictions are structurally correlated with productivity (since less skilled farmers are more likely to operate for self consumption only), they are expected to have a more detrimental impact on factor distribution Restuccia and Rogerson, 2008.

The empirical analysis shows that in Uganda in spite of the large share of farmers active in the land market, misallocation is still a preponderant issue. I find that redistributing agricultural inputs efficiently among the existing farmers would result in a 142 percent increase in overall productivity. Interestingly, TFP would increase by 40 percent just by redistributing factors of production among farms operating in the same villages. In line with the model, it can be shown that these gains imply the redistribution of land from subsistence to market oriented farmers. According to the theoretical framework, reducing transaction increases the willingness to pay for land of more productive commercial farmers, symmetrically reducing the one of subsistence producers. Thus, it facilitates the transfer of inputs from less to more productive individuals. In line with this prediction, I find that the magnitude of village and district specific misallocation is a positive function of the pervasiveness of subsistence agriculture, which in turn correlates negatively with market accessibility.

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

This chapter develops as follows: section 2.2 presents a broad overview of the agricultural sector in Uganda, with a particular attention to the functioning of the land markets and the issue of lack of efficient transport infrastructures. Section 2.3 provides the theoretical framework used to measure misallocation and a model linking it to frictions in output markets and subsistence farming. Section 2.4 describes the data used for the empirical analysis. Section 2.5 is devoted to the presentation of the empirical findings corroborating the theoretical model and comments on the results. Section 2.6 concludes with some final remarks.

2.2 Agriculture in Uganda

Uganda is one of the poorest countries in the world, and systematically ranks at the bottom of the distribution of several economic, social and health indicators. In spite of relatively good

[2]The idea that factor misallocation might be the result of frictions in the output market is not new. However, it has never been explored fully nor tested empirically with the partial exception of Chen, Restuccia, and Santaeulàlia-Llopis, 2017.

economic performance since the 2000s, poverty remains a widespread issue. According to the most recent estimates (Ubos, 2018), the poverty rate is just below 20 percent, but the aggregate figure masks huge regional and urban/rural imbalances. In particular, the rural areas in the Northern and Eastern regions present remarkably higher poverty indicators.

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

Interestingly, as pointed out by Ulimwengu and Ramadan, 2009, although from an aggregate point of view Uganda is a net food exporter, most of the farmers are net food buyers. In particular, they find that 65 percent of the net food buyers in Uganda declare that subsistence agriculture is their primary 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 beign employed in farming was the necessity to obtain food for their household.

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

Unlike other developing countries 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 English domination, and land rentals and sales are common throughout the whole country: according to Deininger and Mpuga, 2002, as far back as the early 50s, 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 inexistent social ties, suggesting that these transfers are generally inspired by market logic.

Although land transactions seem to be common practice 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% of the land was titled in 2010 (due to a cumbersome and costly administrative procedure) and most farmers had no formal claim 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.

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.[3]

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

2.3 Theoretical framework

The theoretical framework used to measure misallocation is in line with the cutting edge literature (see Restuccia and Santaeulalia-Llopis 2017, Chen, Restuccia, and Santaeulàlia-Llopis 2017, Adamopoulos et al. 2017 and Chari et al. 2017). In the first subsection, I present a basic framework used to measure the magnitude of misallocation. In the second, I set up a theoretical model featuring frictionless markets for inputs and where resource misallocation is presented as the equilibrium outcome of costly access the output (food) market for buyers and sellers. A formal derivation of the results and a brief discussion on the underlying conditions are provided in the Appendix A.

[3]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.

2.3.1 Measuring Misallocation

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

$$\widetilde{Y}_i = (s_i X_i)^{1-\gamma}[(q_i L_i)^{\alpha} K_i^{1-\alpha}]^{\gamma} \tag{2.1}$$

where q_i is a land augmenting parameter included to control for the quality of the soil. The parameters α and γ define the three factors' shares.[4]

Following the literature, I focus on misallocation of land and agricultural capital, while labour is treated as a non tradable factor. It is therefore convenient to rearrange the production function in labour units terms:

$$\widetilde{y}_i = s_i^{1-\gamma}[(q_i l_i)^{\alpha} k_i^{1-\alpha}]^{\gamma} \tag{2.2}$$

in this setting, the parameter γ determines the returns to scale in the variable factors of production.[5]

Finally, heterogeneous land quality is controlled for by normalizing the output $\widetilde{y}_i / q_i^{(1-\alpha)\gamma} = y_i$. By doing so, I obtain the operative form of the production function that will be used to infer farm specific productivity:

$$y_i = s_i^{1-\gamma}(l_i^{\alpha} k_i^{1-\alpha})^{\gamma}. \tag{2.3}$$

The magnitude of resource misallocation and its impact on aggregate factor productivity can be estimated comparing the production resulting from the actual distribution of factors across heterogeneous farmers $(l_1, \ldots, l_N;\, k_1, \ldots, k_N)$ to the one characterising the counterfactual scenario where land and capital are distributed efficiently. The latter can be derived as the solution of the social planner problem, that boils down to the maximisation of the total production imposing only the resource constraint. Since the framework is meant to capture static misallocation only, the total quantity of inputs available in the economy is kept fixed.[6] The problem can formally

[4] More specifically, labour's share is $1-\gamma$, land's share is $\alpha\gamma$ and capital share is $(1-\alpha)\gamma$.

[5] As $\gamma < 1$, the model features decreasing returns to scale. This is crucial to obtain a non degenerate input distribution in the efficiency counterfactual when producers' output is homogeneous and single farms do not face negatively sloped demand curves. The value of this parameter is also crucial in determining the potential gains from reallocation. Indeed, a higher γ generates a counterfactual efficient distribution that allocates a larger share of inputs to more productive farmers and thus results in higher estimated gains from reallocation when inputs are homogeneously distributed among farmers.

[6] This crucially implies that in the model, workers do not move across sectors and there is no accumulation of agricultural capital over time.

be described as:

$$\max_{l_i,k_i} \sum_{i}^{N} y_i \text{ subject to: } \sum_{i}^{N} l_i = L \text{ and } \sum_{i}^{N} k_i = K \tag{2.4}$$

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

Intuitively, the first order conditions state that the the marginal products of both factors of production must be equalized across farmers. As the production function features decreasing returns to scale, these conditions result in a non-degenerate distribution of inputs, where each farmer operates a share of each input that is proportional to her productivity s_i. Formally, this implies:

$$l_i^e = \frac{s_i}{\sum_j^N s_j} L \quad \text{and} \quad k_i^e = \frac{s_i}{\sum_j^N s_j} K \tag{2.5}$$

where the superscript e denotes the counterfactual of full efficient factor distribution.

By plugging in these values into the production function, I obtain the farm specific efficient output level: y_i^e as $s_i^{1-\gamma}(l_i^{e\alpha} k_i^{e1-\alpha})^\gamma$. The magnitude of misallocation can therefore be obtained as the ratio between the total production in the efficient scenario to the one actually observed:

$$e = \frac{\sum_j^N y_j^e}{\sum_j^N y_j} = \frac{Y^e}{Y} \tag{2.6}$$

Conveniently, the same procedure can be applied to compute the potential gain from efficient factor reallocation at narrower geographical levels and/or within some categories of farmers. In the remainder of the paper, I will take advantage of this opportunity to test some predictions of the model.[7]

2.3.2 Misallocation and subsistence farming

Although the debate on the magnitude of the potential gain from reallocation of agricultural inputs and its potential contribution to the reduction of the agricultural productivity gap in low income countries is very active, less attention has typically been devoted to the causes of the misallocation. Typically, frictions in land rental and sales market are indicated as the major cause of the inefficient resource distribution (Chari et al. 2017, Chen, Restuccia, and Santaeulàlia-Llopis 2017).

[7]More specifically, I will estimate the village/county/district level misallocation as the ratio between the total production that would be obtained if marginal products were equalized across all the farmers in the administrative units considered and the actual level of output. I will also estimate the misallocation within farmers participating and non participating in land markets and within subsistence/commercial farmers in a given location.

In support of this hypothesis, existing findings show that more active land rental markets are correlated with lower level of factor misallocation and that policies aimed at liberalizing land transactions have an efficiency enhancing effect on factor distribution. However, the same studies highlight that resource distribution remains far from efficient after land reforms and even when input markets appear to be well functioning. It is therefore important, especially from a policy perspective, to identify some concurrent sources of misallocation. In the following, I provide a simple theoretical framework aimed at showing that input misallocation can occur when there are no barriers to input trade as long as accessing the food market (both as a producer and as a consumer) is costly.[8]

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

$$y_i = s_i^{1-\gamma} \xi_i^{\gamma}. \tag{2.7}$$

In line with a well established literature on production decisions of farm households (see, among others, the seminal work by Finkelshtain and Chalfant 1991), farmers are not modeled as producers willing to maximize their profits, but rather as consumers and producers, who evaluate their output differently according to their net position in the food market. For simplicity, I assume that farmers consume agricultural produce (food) until the satiation point $\overline{y}$ is met deriving constant marginal utility normalized to 1.[9] The production exceeding $\overline{y}$ is sold, and for each unit the farmer receives a constant utility normalized to 1-τ.[10] The resulting maximization problem is:

$$\max_{\xi_i} \begin{cases} y_i - r\xi_i & \text{if } y_i \leq \overline{y} \\ (1-\tau)y_i - r\xi_i & \text{if } y_i > \overline{y} \end{cases} \tag{2.8}$$

In general, τ can be thought of as a combination of more and less tangible factors (such as transaction/transportation costs, volatile market prices, thin markets) which incentivize farmers to grow their own food rather than buying it at the market and symmetrically make selling agricultural output less profitable. Intuitively, a larger τ implies higher incentives towards food self production and lower profitability of commercial agriculture.

[8]The assumption of frictionless input markets is obviously functional to the model and its tractability but it is not meant to describe the actual conditions of Ugandan land markets.

[9]A more realistic model would feature decreasing marginal utility of food consumption. I maintain the the assumption of constant marginal utility for the sake of manageability but it is worth mentioning that decreasing marginal utility would leave the main predictions of the model untouched.

[10]Here, I am assuming that the threshold $\overline{y}$ is the same across the different households. Note that this assumption is less problematic than it might appear as output y is expressed in per labour terms. Therefore, as long as one assumes that all the individuals in a households are employed in agriculture and labour is not tradeable (in line with the model), the assumption could be rephrased as: "each individual farmer has equal food consumption threshold".

In order to see how different values of τ affect the equilibrium allocation of resources and how the resulting distribution compares to the social planner solution, the solution to the optimization problem is presented for different values of τ, starting from the extreme scenarios of $\tau = 0$ and $\tau = 1$.

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

$$\max_{\xi_i} y_i - r\xi_i \tag{2.9}$$

where r is the price of the composite agricultural input ξ.[11]

Similarly to the case of the social maximizer problem depicted in equation 2.4, the resulting first order condition requires the marginal productivity of inputs to be equalized across farmers:

$$\gamma s_i^{1-\gamma} \xi_i^{*\gamma-1} = MP_i^* = r \tag{2.10}$$

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

$$\xi_i^* = As_i \tag{2.11}$$

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

At the opposite end of the spectrum, markets for food do not exist (or are too costly to access) as in De Janvry, Fafchamps, and Sadoulet, 1991. Following the model's notation, this implies setting $\tau = 1$. In practice, this mirrors the fact that farmers are unable to sell their produce and therefore do not derive any utility from output exceeding the consumption threshold $\overline{y}$. Therefore, farmers maximise:

$$y_i - r\xi_i \text{ subject to: } y_i \leq \overline{y} \tag{2.12}$$

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

[11] The fact that there is no difference between the cost of renting the input in and the returns to rent it out (as in Chari et al. 2017) and that there is no maximum threshold of input that can be rented in (as in Shenoy 2017) reflects the assumption that input markets are perfectly functioning.

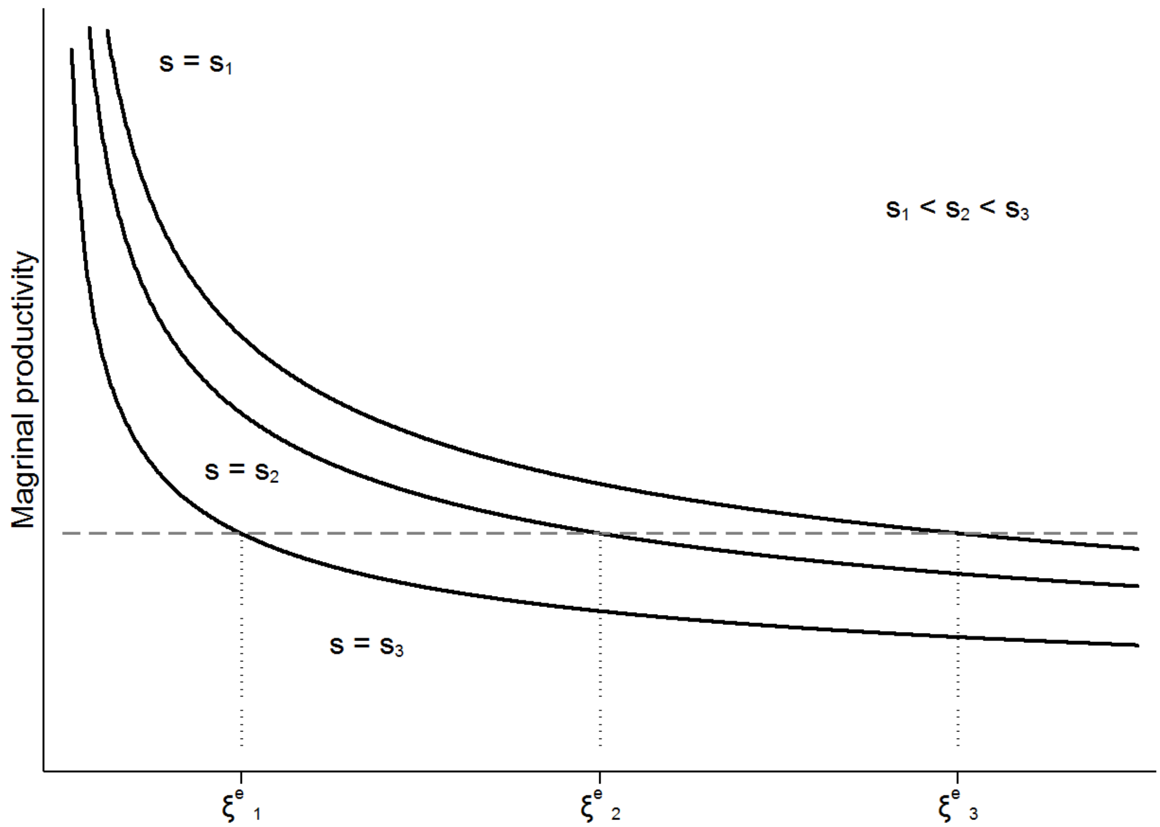


Figure 2.1: Equilibrium with no frictions in output market

The figure represents the equilibrium rental cost r of the tradable input ξ, as well as the marginal productivity for three farmers with increasing agricultural skills ($s_1 < s_2 < s_3$). In equilibrium, they operate a share of the composite input ξ that is proportional to their productivity s. The superscript $*$ indicates that this equilibrium also guarantees a fully efficient distribution of the composite input ξ across farmers.

As a result, the share of factors of production employed is proportional to productivity only for the first set of (non constrained) farmers, whereas it is a negative function of s for the second group. Formally:

$$\xi_i^* = \begin{cases} \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i & \text{if } s_i \leq \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} \\ \left(\frac{\overline{y}}{s_i^{1-\gamma}}\right)^{\frac{1}{\gamma}} & \text{if } s_i > \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} \end{cases} \tag{2.13}$$

This equilibrium outcome, depicted in Figure 2.2, features different marginal productivity across farmers. As such, the factor's distribution is inefficient and differs from the one of a hypothetical social planner.[12]

$$MP_i^* = \begin{cases} r & \text{if } s_i \leq \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} \\ Bs_i^{\frac{1-\gamma}{\gamma}} & \text{if } s_i > \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} \end{cases} \tag{2.14}$$

where $B = \frac{\gamma}{\overline{y}^{\frac{1-\gamma}{\gamma}}}$.

More specifically, misallocation can be observed along two different dimensions. Firstly, constrained farmers have higher marginal productivity than the unconstrained ones. This happens because the former have no incentive to scale up as they would not be able to sell the excess output, while the latter are producing up until the point where their marginal productivity equals the input price, and would employ more input if they could afford it. Additionally, within constrained farmers, those with better agricultural skills need less input to produce $\overline{y}$ and display higher marginal productivity when doing so.

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

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

[12]The only case in which the equilibrium would be exactly equal to the one without any friction in the food market is when the distribution of farmers' productivity is such that the constraint $y_i \leq \overline{y}$ is not binding for any of them. In the opposite scenario of the constraint being binding for all the farmers, the share of input operated by all producers will be a negative function of their skills.

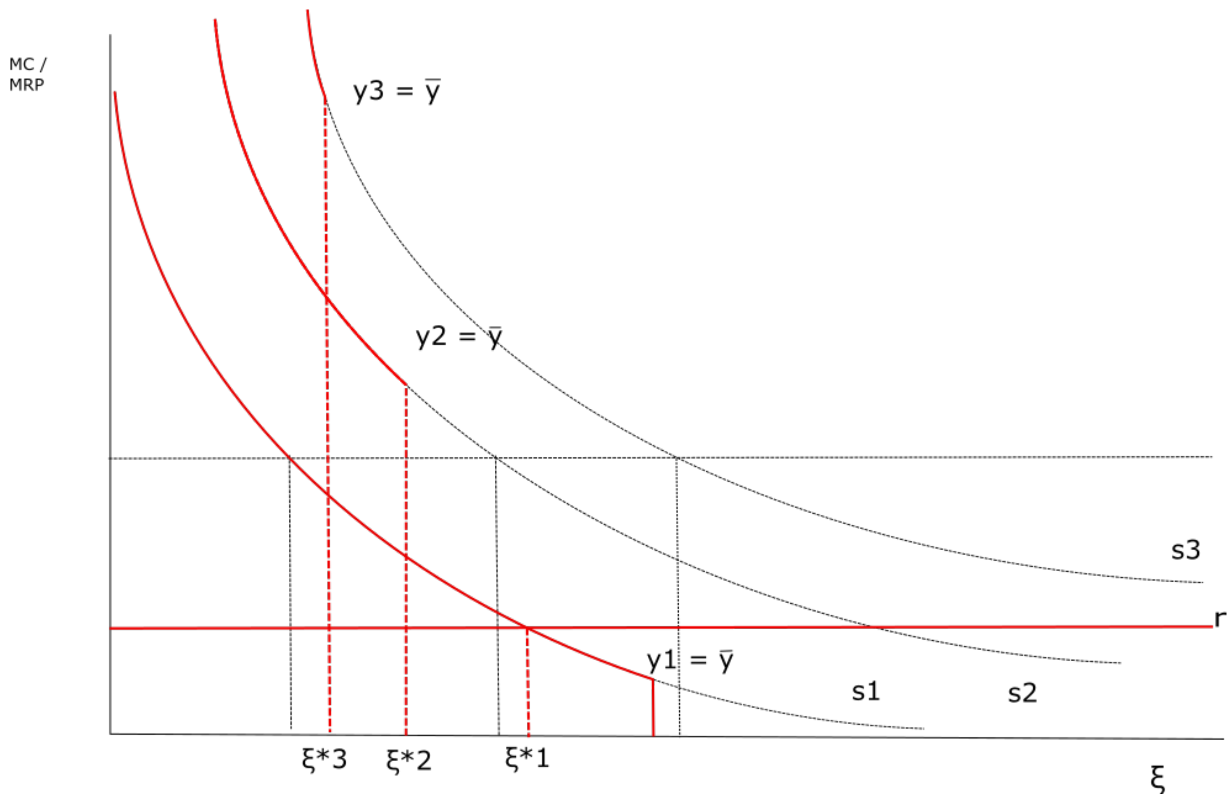


Figure 2.2: Equilibrium with no food market

The three red solid lines represent the marginal return farmers get from an additional unit of composite good ξ Since the excess production cannot be sold in this case, it drops to 0 as farmers reach the level of production $\overline{y}$. Dotted lines represent farmers' marginal productivities and are no longer equal across farmers in equilibrium. The new equilibrium quantities $\xi*$ indicate that now the amount of land and capital operated is a negative function of farmers' productivity. The new equilibrium price of the input r is plotted in red and obtained by setting $\xi*_1 + \xi*_2 + \xi*_3 = \xi_1^e + \xi_2^e + \xi_3^e = \Xi$ and can be compared to the one in the full efficiency counterfactual (the dotted horizontal line).

a) those with $s_i \leq \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}}$, who would not commercialize their produce even if there were no output market frictions as their production is lower than the consumption threshold. They operate a share of land such that the marginal product equals the cost of input r. The farmer with ability s_1 in Figure 2.3 belongs to this category;

b) those with $\overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} < s_i \leq \overline{y}\left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}}$, who only hire input ξ to meet their consumption threshold $\overline{y}$, but would produce a surplus if there were no frictions in the market for agricultural produce. The total quantity of input ξ_i used is equal to $\left(\frac{\overline{y}}{s_i^{1-\gamma}}\right)^{\frac{1}{\gamma}}$ and therefore those with relatively higher level of productivity operate a lower quantity of input. Accordingly, marginal productivity will be an increasing function of s_i: $Bs_i^{\frac{1-\gamma}{\gamma}}$. The farmer with productivity s_2 in figure 2.3 belongs to this category;

c) those with $s_i > \overline{y}\left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}}$, who engage in commercial agriculture regardless of the frictions in output market and will hire input until the marginal productivity equals $\frac{r}{1-\tau}$. Similarly to case **a**, the input distribution will reflect farmers' productivity within the group. Nevertheless, their marginal productivity will be higher than the one in **a** and **b**. The farmer with productivity s_3 in Figure 2.3 belongs to this category.

Formally, in equilibrium, the quantity of input used is:

$$\xi_i^* = \begin{cases} \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i & \text{if } s_i \leq \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} \\ \left(\frac{\overline{y}}{s^{1-\gamma}}\right)^{\frac{1}{\gamma}} & \text{if } \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} < s_i \leq \overline{y}\left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}} \\ \left[\frac{\gamma(1-\tau)}{r}\right]^{\frac{1}{1-\gamma}} s_i & \text{if } s_i > \left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}} \end{cases} \tag{2.15}$$

With a resulting marginal productivity of:

$$MP_i^* = \begin{cases} r & \text{if } s_i \leq \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} \\ Bs_i^{\frac{1-\gamma}{\gamma}} & \text{if } \overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} < s_i \leq \overline{y}\left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}} \\ \frac{r}{1-\tau} & \text{if } s_i > \left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}} \end{cases} \tag{2.16}$$

A graphical representation of this equilibrium is provided in Figure 2.3. Similarly to the case with $\tau = 1$, farmers feature different levels of marginal productivity across the three groups. Additionally, resources are inefficiently distributed within farmers belonging to group **b** whose members operate a share of land that is a negative function of their agricultural skills. Figure 2.4 illustrates the resulting input distribution in the three scenarios considered.

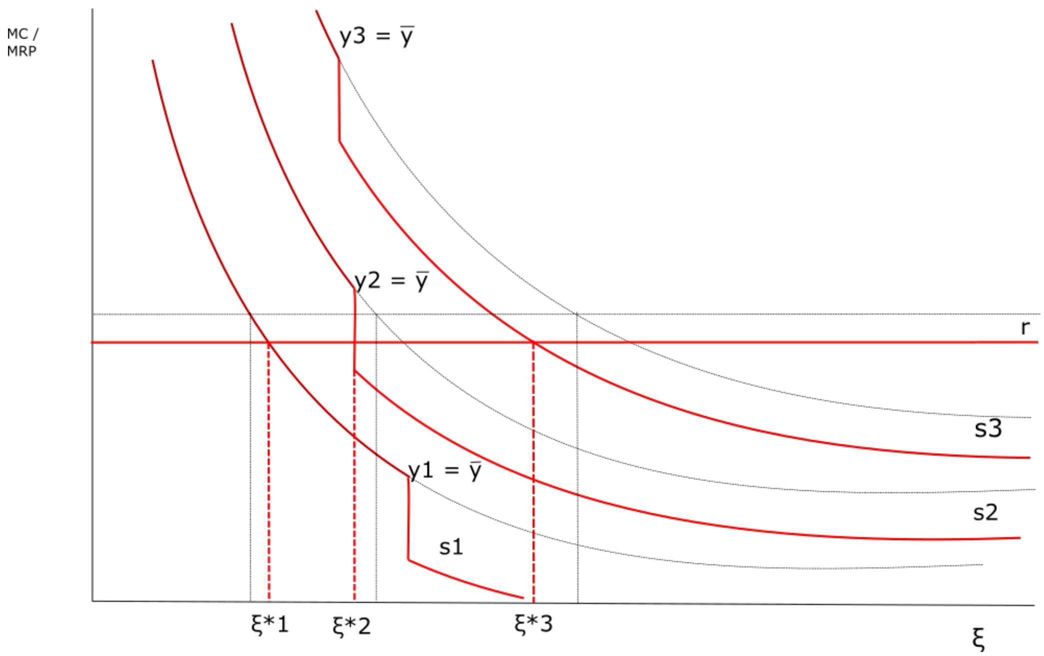


Figure 2.3: Equilibrium with frictions in the food market

The three red solid lines represent the marginal return farmers get from an additional unit of composite good ξ Since the excess production cannot be sold in this case, it drops to 0 as farmers reach the level of production $\overline{y}$. Dotted lines represent farmers' marginal productivities and are no longer equal across farmers in equilibrium. The new equilibrium quantities $\xi*$ indicate that now the amount of land and capital operated is a negative function of farmers' productivity. The new equilibrium price of the input r is plotted in red and obtained by setting $\xi*_1 + \xi*_2 + \xi*_3 = \xi_1^e + \xi_2^e + \xi_3^e = \Xi$ and can be compared to the one in the full efficiency counterfactual (the dotted horizontal line).

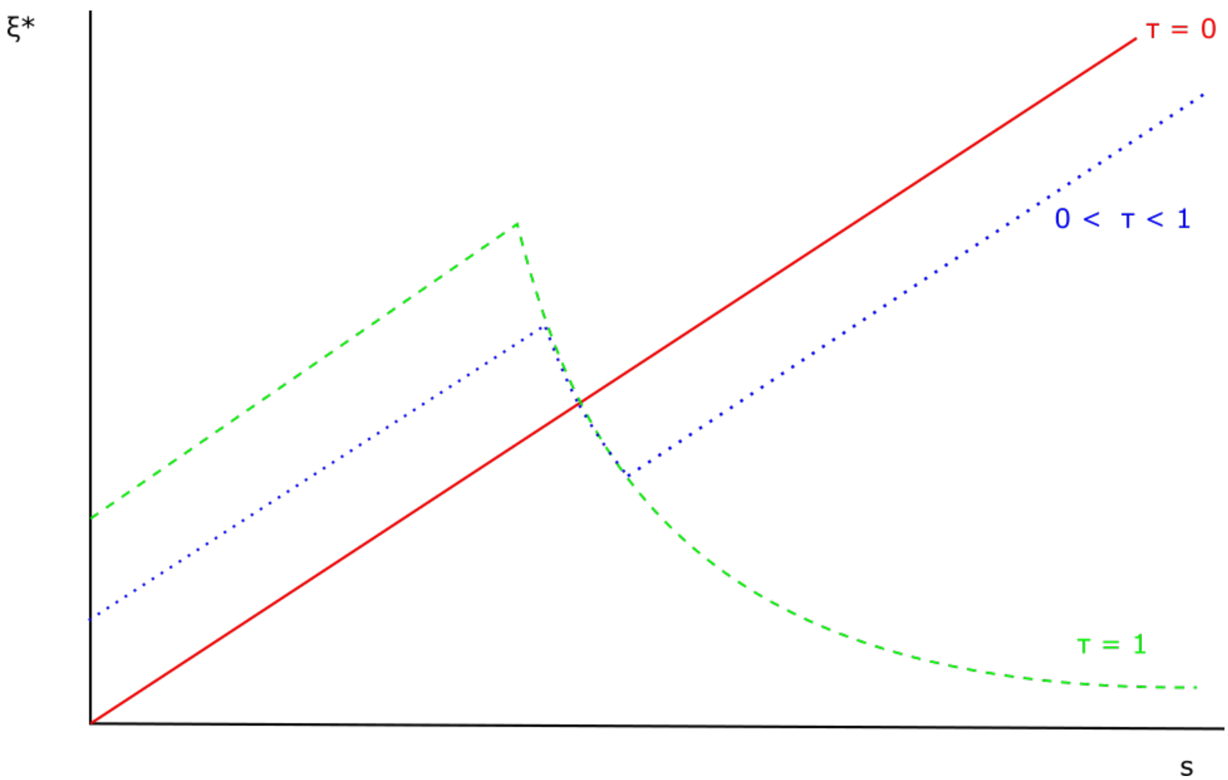


Figure 2.4: Input operated in equilibrium and agricultural skills

The red solid line represents the input operated in equilibrium per for each level of skills when there are no frictions in the food market. The blue pointed line represents the intermediate case with $0 < \tau < 1$ while the green dotted line the scenario where food markets do not exist and $\tau = 1$. The lines depict the "normal" situation in which the total input Ξ is lower than $\overline{y}^{\frac{1}{\gamma}} \sum_i \left(\frac{1}{s_i}\right)^{\frac{1-\gamma}{\gamma}}$, $\gamma = 0.597$ and there is at least one farmer i such that $s_i > \overline{y} \left(\frac{\sum_i s_i}{\Xi}\right)^{\gamma}$. See the appendix for an explanation of these conditions.

In practice, the model returns four main predictions on the magnitude and patterns of factor misallocation in a context where food markets are thin and costly to access.[13] The first one is that since output frictions affect the scale of the production units as opposed to the input mix, the resulting factor distribution should deviate from the most efficient one with respect to the size of the farms, while the input mix operated by farmers should be (close to) optimal. Secondly, in order to achieve gains from reallocation, land should be transferred from subsistence to market oriented farmers. Indeed, farmers who sell a positive share of their production are shown to have a higher marginal productivity in equilibrium than farmers who only farm for self consumption purposes.

Thirdly, according to the model, the misallocation in any given village will be higher among subsistence farmers than among marked oriented farmers. Indeed, while in equilibrium commercial farmers have all the same marginal productivity, subsistence farmers are a heterogeneous group composed by those who do not meet their consumption threshold in equilibrium (category **a**) who have all the same marginal productivity, and those who produce for self consumption only but would exceed this quota if the marketization costs were not in place (category **b**). Individuals belonging to the latter group will not only have higher marginal productivity than the ones belonging to the former, but also have different marginal productivities one another. Lastly, I expect that farmers facing more prohibitive marketization costs τ will be characterized by higher levels of subsistence agriculture and in turn display more severe factor misallocation. Although τ can in principle capture any factor (tangible or not) causing a wedge between consumer and producer price, I will focus on transportation costs as they are easier to compute.

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

2.4 Data

In performing the empirical analysis, I use data from a number of different sources, as briefly summarised in Table 2.1. The most relevant is the Uganda Census of Agriculture (UCA) by the

[13] These predictions are based on the most realistic scenario of $0 < \tau < 1$, where the market for food exists but presents some imperfections in the form of transaction costs, thinness and high price uncertainty.

Ugandan Bureau of Statistics, which collects detailed holding level information on more than 25,000 Ugandan farms and is meant to be nationally representative of the population of medium and small farms. More specifically, it is a cross section which refers to the 2008/09 agricultural year.[14] Variables on agricultural input and output, as well as on participation on land markets and disposition of crops produced are derived from this dataset. Consequently, the farm will be our main unit of analysis as output and inputs are aggregated at the holding level.[15]

Cultivated land is defined as the total area of the plots farmed in the two agricultural seasons under analysis. As every parcel was measured through GPS techniques, the resulting variables are arguably less subject to measurement errors than figures based on farmers' own estimations. The questionnaire also asks about the mode of acquisition of each of the plots operated so that it can be assessed whether each farm included at least some land that was acquired through market transactions as opposed to inheritances or donations or other forms of informal appropriation.[16]

As for the labour input, farmers were asked about the exact amount of working days provided by family members and hired labourers as well as their gender and whether they were children or adults. Unfortunately, these data were missing in 56 percent of the observations, and the only information available was whether each worker was employed full time or only temporarily for a given season. From the available observations, I infer that on average, temporary workers provided about half the working days than their full time counterparts. Therefore, in measuring the total quantity of labour input, all part time workers are considered as half a unit. Additionally, I introduce different weights depending on the gender and on the age of the workers on the basis of the relative median agricultural wage of each group as measured in the 2009/10 LSMS survey.[17] As a result, labour input is defined in terms of full time adult male equivalents rather than total working hours. Although this is not ideal and it would be better to have more detailed measures

[14] Namely, the dataset collects information regarding the second agricultural season 2008 and the first agricultural season 2009.

[15] Recent studies on agricultural misallocation in Uganda (Gollin and Udry 2019; Maue, Burke, and Emerick 2020) have used the World Bank's LSMS as main source of data on agricultural production. The main advantages of the LSMS with respect to the UCA are that it is a panel dataset and that it contains information on households' welfare and non-consumption activities. However, the UCA is much more suitable for this analysis as 1) it collects information on many more villages (the sample is roughly ten times larger), which allows to run more credible village level regressions 2) 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, 3) the coordinates of the enumeration areas are not altered 4) farms' output is expressed in homogeneous crop units that can be easily and consistently aggregated.

[16] 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, I have 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.

[17] More specifically, the median hourly agricultural wage of a woman was 0.8 times the one of a man, while the one of a child, regardless of the gender, was 0.55 times the one of a man.

Table 2.1: Datasets used

	Name	**Source**	**Level**
Main analysis			
Input level, 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
*Transportation costs**			
Cells crossing time	Ugandan Road Network	Ubos	1 km^2 resolution
Urban conglomerates	Africover	FAO	1 km^2 resolution

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

of labour input, this measure still captures a lot of variation across different holdings based on realistic assumptions and reliable entries of the survey.

Finally, capital is defined as the sum of the value of the agricultural asset used in the farming activity.[18] Differently from previous studies (Restuccia and Santaeulalia-Llopis 2017, Chen, Restuccia, and Santaeulàlia-Llopis 2017), I do not include the value of means of transportation and livestocks as there is no way to establish whether they were actually employed in agricultural production.[19]

In terms of final output, farmers were asked about their total production for the 18 major crops that overall account for more than 97 percent of the total farmed area.[20] Conveniently, the total production for each crop is reported in kilos of dry final produce by applying different conversion factors depending on the self reported state of the harvest and units. Similarly to the case of capital items, prices of the different crops were not available. Therefore, in order to aggregate the total production, I use a set of median prices derived from the 2009/10 LSMS survey.

Furthermore, the survey collects information on the destination of the final output. In particular, farmers were asked, for each crop, the share of the total production that was self consumed, used

[18]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.

[19]As the price/value of each capital item are not available in the UCA, I derived them 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.

[20]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.

Table 2.2: Descriptive Statistics

	Mean	Median	SD	N
Household level				
Labour (full time adult male equivalent)	2.92	2.30	2.08	24,431
Land (hectares, 2 seasons)	2.00	1.30	2.40	24,431
Capital (1000 Shillings)	219.722	21.00	1590.4	24,431
Value added per labout unit (1000 Shillings)	4,675.36	255.23	663.72	24,431
Participation in land markets (1 = yes)	0.70	1.00	0.46	24,431
% produce marketed	0.31	0.25	0.30	24,431
Subsistence farmer*	0.28	0	0.34	24,431
Market oriented farmer†	0.15	0	0.45	24,431
Village level				
Number of farmers	7.68	8	2.31	3,181
% of farmers participating in land markets	0.70	0.8	0.32	3,181
% produce marketed	0.40	0.33	0.23	3,181

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

for other purposes by the households, sold or stored. I define the percentage of self consumption as the fraction of the non stored output that was not sold.[21]

The descriptive statistics for the most relevant variables are shown in Table 2.2. In line with the defining characteristics of the Ugandan agricultural sector listed above, I find that farmers tend to operate on a very low scale and using a rather limited amount of capital. In particular, the median farmer was cultivating 0.65 hectares and with capital equivalent to 10 US dollars.[22] The distribution of both land and capital is highly skewed to the right, with median values much lower than average values, suggesting that a high share of the holdings is rather small and farmers tend to operate on a close to subsistence level. Accordingly, it seems that a large share of the agricultural input is not commercialised but self consumed. Only less than 15 percent of the farmers in our sample were selling at least two thirds of their output and nearly 30 percent of them were producing for self consumption only. The average and median farmers were selling 31 and 25 percent of their output respectively.

As expected, land markets are shown to be very active, especially when compared to the cases of Malawi and Ethiopia (Restuccia and Santaeulalia-Llopis 2017; Chen, Restuccia, and San-

[21]The variable is generated at the crop level and then aggregated using crop prices. Therefore, the underlying assumption is that, for each crop, the amount stored will be sold and self consumed in the same proportion as the output sold or self consumed before the survey. Farmers reported to store the whole production only in less 0.1 percent of the cases which were dropped from the subsequent analysis.

[22]The figure on land refers to the average land operated per season.

taeulàlia-Llopis 2017): indeed more than 70 percent of the farmers in the sample are operating at least one plot of land that was either acquired in the past or rented in for the current agricultural season. This figure does not come as a surprise considering the Ugandan context, characterized by the lack of formal barriers to land transfers. However, given the large prevalence of customary rights, a large fraction of these transfers might be of informal nature and thus not necessarily efficiency enhancing (Chen, Restuccia, and Santaeulàlia-Llopis 2017).

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

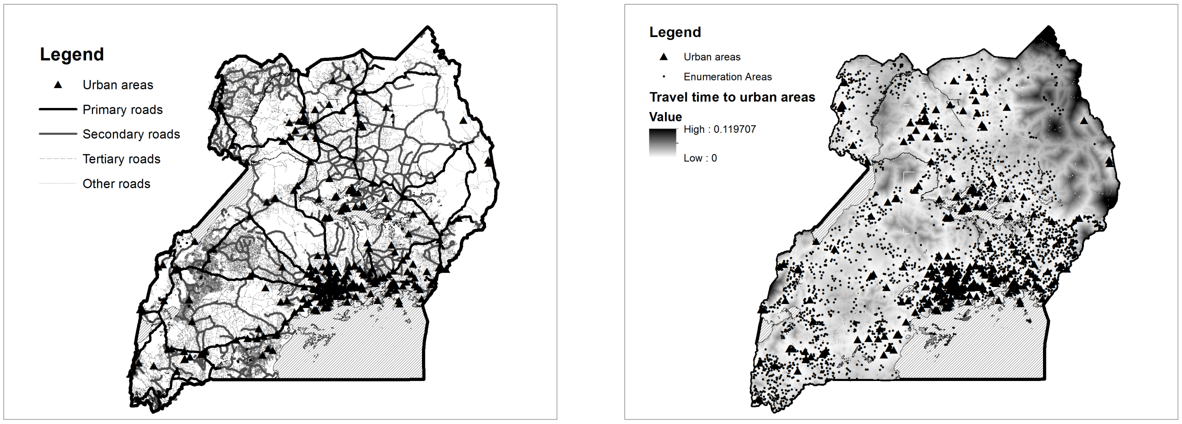
Finally, in order to measure village and district specific transportation costs I use two additional spatial datasets. More specifically, I define market accessibility as the travel time to the closest urban area. In order to identify Ugandan urban areas, I use the Africover dataset by FAO, which maps the most important urban conglomerates on the basis of satellite data.[23] To identify the travel time to these urban areas, I employ a dataset provided by Ubos describing the Uganan road network. In practice, following the procedure of Adamopoulos, 2019, I divide the country in 1×1 kilometer cells and assign to each a crossing time on the basis of the terrain roughness and whether the cell is crossed by a road.[24] 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. In the case of the village, the reference travel time is the one of the cell the village is located in. In the case of the district, the proxy considered is a simple average across their inhabited cells (excluding forests, mountains, national parks and water bodies). The outcome of this procedure is shown in Figure 2.5.

In the empirical analysis, I will explore to which extent these transportation costs are linked to subsistence agriculture and in turn misallocation.

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

[24] 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 tertiaty roads and 10 for rural roads. However, the results are not sensitive to sensible changes in these travel timee.

Figure 2.5: Road network and urban areas accessibility



2.5 Empirical specifications and results

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

2.5.1 The magnitude of misallocation

The first step required to provide some figures on the magnitude of misallocation consists in estimating the farm specific productivity s_i that can be obtained as, rearranging equation 2.3:

$$s_i = \left(\frac{y_i}{l_i^{\alpha} k_i^{1-\alpha}} \right)^{\frac{1}{1-\gamma}} \tag{2.17}$$

As data on farm level input use and production are readily available, I only need to estimate the parameters α and γ, which define the elasticities of the three inputs entering the production function: labour X, land L and capital K. Following the procedure suggested by Chari et al., 2017, I obtain them by estimating a production function as follows:[25]

$$\log \widetilde{Y_{ivt}} = \beta_1 \log X_{ivt} + \beta_2 \log q_{iv} L_{ivt} + \beta_3 \log K_{ivt} + W'_{vt}\delta + \phi_i + \tau_t + \epsilon_{ivt} \tag{2.18}$$

[25] Unlike Chari et al., 2017, 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., 2017, using crop specific or general coefficients does not affect the results of the estimation significantly.

As the UCA data are cross sectional, the production function is estimated using the World Bank's LSMS data for Uganda referring to the 4 agricultural seasons between 2010 and 2011.[26] In practice, the logarithm of the value added for each farmer i and season t is regressed on the logarithm of the quantities of the three inputs as well as a vector W of time and village (indexed by the subscript v) specific shocks plus farmers' and time fixed effects.[27] The land input is augmented by the parameter q_{iv} that is meant to capture heterogeneous land quality across farms.[28]

According to these estimates, the shares of labour, land and capital are 0.403, 0.475 and 0.122 respectively.[29] This implies $\alpha = 0.796$ and $\gamma = 0.597$. Reassuringly, as shown in Table 2.3, these values are quite close to the ones estimated by similar studies involving developing countries. Consistently with the literature, the capital share is significantly lower than the one in more industrialized economies such as the US.

These parametres can be plugged into equation 2.17 to obtain an estimate of the farm specific productivity s_i for each of the N observations in the sample.[30] In order to address some potential concerns related to measurement errors, the top and bottom 2.5% of the observations in the productivity distribution are trimmed. However, all the results are robust to more and less conservative trimming strategies.

[26]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.

[27]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.

[28]Unlike the analysis, 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 dataset. The procedure used to compute the land quality index follows quite closely the one in Restuccia and Santaeulalia-Llopis, 2017 and Chen, Restuccia, and Santaeulàlia-Llopis, 2017 and is based on the 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 labour days of the farmer i in village v on the left hand side and a vector X of farm specific and Z of village specific variables that capture the soil and terrain characteristics, as well as the ratio between capital and land to control for different capital intensities across fields. The farm specific vector X includes self reported soil quality, occurrence of erosion problems, presence of irrigation facilities and prevalent soil type and topography. As the value reported were parcel rather than farm specific, they are aggregated at the farm level by considering the weighed average of the continuous variable (using the size of the cultivated parcels as weights) or the weighted mode for the categorical variables. The index will be simply defined as $X'_{iv}\beta_1 + Z'_v\beta_2$.

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

[30]In estimating s_i, I control for land quality creating an index based on the 10 dimensions of soil suitability available in the GAEZ dataset (soil workability, toxicity, terrain slope, rooting conditions, oxygen availability, nutrient retention and availability, altitude, excess salinity and dominant soil type) and for weather shocks using rainfall data from TAMSAT. More specifically, I follow the procedure outlined by Chen, Restuccia, and Santaeulàlia-Llopis, 2017 and control for different rainfall patterns by regressing the output on a set of dummies, each representing a decile of total precipitations during the two rainy seasons (Sept-Nov 2008 and Mar-May 2009) and using the residuals as a measure of total production.

Table 2.3: Shares of agricultural inputs in selected countries

	Uganda	Malawi	China	Uganda*	Ghana	US
Labour	0.403	0.42	0.243	0.389	0.212	0.46
Land	0.475	0.39	0.417	0.473	0.536	0.18
Capital	0.122	0.19	0.088	-	0.181	0.36

*The alternative figures from Uganda come from Gollin and Udry, 2019, who do not include capital input in the production function. The figures presented come from Restuccia and Santaeulalia-Llopis, 2017 for Malawi, Chari et al., 2017 for China, Gollin and Udry, 2019 for Uganda (with no capital input) and Ghana and Valentinyi and Herrendorf, 2008 for the US. In the case of China, the total estimates (without allowing from crop specific shares) are presented.

On the basis of the distribution of the farm level productivity, it is possible to estimate the total production in the full efficiency scenario, namely, when inputs are redistributed proportionally to each individual's agricultural skills, as shown in equation 2.5. I find that, if factors of production were efficiently distributed among the existing farmers, the agricultural factor productivity would increase by 142 percent. Table 2.4 provides a comparison between my figures and the ones estimated by similar studies.

Interestingly, misallocation seems to be less severe than in contexts where land markets had just been liberalised at the time of the study such as Malawi and Ethiopia, suggesting that facilitating agricultural input transactions has a positive impact on the efficiency of factor distribution.[31] Also, unlike the case of the two above mentioned studies, there seems not to be any significant difference in the magnitude of within group misallocation between farmers who operate land acquired through the market and farmers who do not. Overall, these findings suggest that input distribution (even within farmers who bought and/or rent land in) can be far from efficient also in the absence of formal restrictions to market based land transactions.

Intuitively, misallocation reflects the fact that more productive individuals do not necessarily operate larger quantities of inputs, and that consequently the marginal productivity of factors of production differs greatly across farmers and is a positive function of their productivity. Figures 2.6 and 2.7 confirm that the distribution of both land and capital seems to be virtually independent of farm specific productivity.

Furthermore, I show that a relevant productivity gain could be achieved if resources were allocated efficiently within the villages in our sample. Namely, redistributing land and capital at the

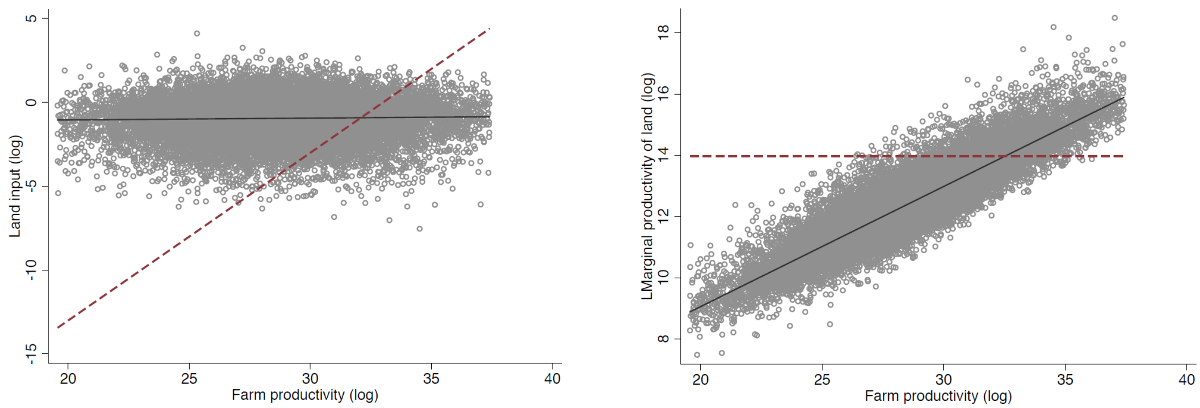
[31] It is worth mentioning that the estimates of the gains are very sensitive to the assumption of the curvature of the production function (in this case, γ), as the same joint input and productivity distribution can lead to sizeably different estimates of the potential gains depending on the return to scale/demand structure imposed. For this reason, comparison across studies can be misleading. However, in the cases of Restuccia and Santaeulalia-Llopis 2017 and Chen, Restuccia, and Santaeulàlia-Llopis 2017, the curvature imposed is very similar ($\gamma \approx 0.58$).

Table 2.4: Comparison to existing studies

	Uganda	Malawi	Ethiopia	China	Vietnam	Thailand
Nationwide analysis						
Total gain	2.42	3.59	3.07	1.57	1.78	-
Operating marketed land	2.43	3.00	2.61	-	-	-
Non operating marketed land	2.38	4.60	3.18	-	-	-
% operating market land	70.1	16.6	24.3	-		-
Village level analysis						
Total gain*	1.39	1.65	-	1.27	-	1.19
% Contribution	27	39	-	53	-	-

* In this case, total gain refers to the aggregate TFP gain that would be obtained by reallocating resources efficiently within villages. The estimates are derived using different methodologies and as such they are not directly comparable. The figures from Malawi come from Restuccia and Santaeulalia-Llopis, 2017, for Ethiopia from Chen, Restuccia, and Santaeulàlia-Llopis, 2017, for China from Adamopoulos et al., 2017, for Vietnam from Ayerst, Brandt, and Restuccia, 2018 and from Shenoy, 2017 for Thailand.

Figure 2.6: Efficient vs actual land input distribution

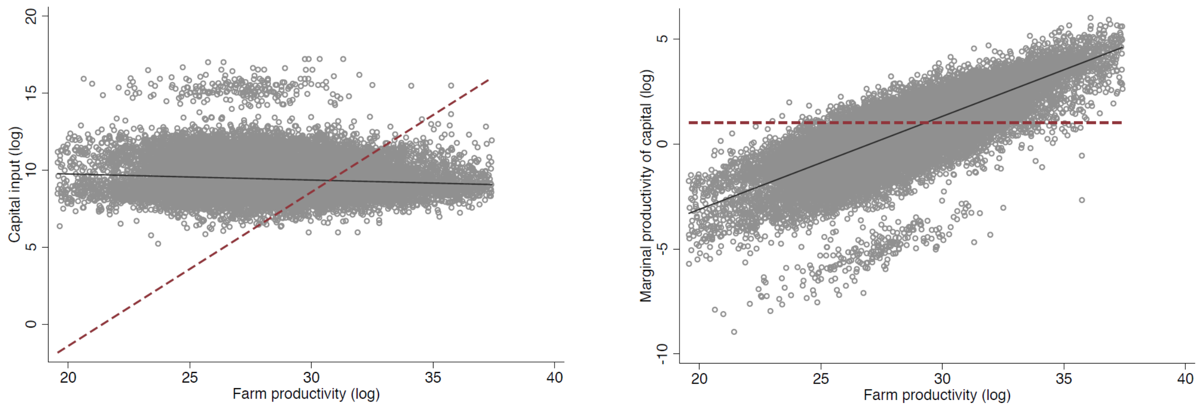


Note: In both graphs, the black line represents the linear fit of the actual observations while the red dashed line represents the efficient distribution of input (proportional to the farmer specific productivity) and the resulting marginal productivity (equalized across agents). The land input is defined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

village level in proportion to the farmers' productivity would increase the value of the agricultural output by nearly 40 percent. As land quality, cropping patterns as well as weather shocks experienced are arguably homogeneous within villages, this finding provides some additional evidence that factors of production are not efficiently allocated.

In the following section, I will illustrate the decomposition of the observed misallocation in two components, capturing the deviations from the efficient distribution attributable to the scale of

Figure 2.7: Efficient vs actual capital input distribution



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

the farms and the input mix operated by each farmer respectively. According to the model, the former should be more relevant since, as long as there are no restrictions in input transactions, farmers should be able to operate the right mix of land and capital, and the only source of misallocation should be the quantity of inputs used, which is directly affected by distortions in the food (output) market.

2.5.2 Mix vs scale misallocation

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

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

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

$$\begin{aligned} &\max \sum_i^N y_i \text{ subject to } pl_i + ck_i = p\bar{l}_i + c\bar{k}_i, \\ &\qquad\qquad \sum_i^N l_i = \sum_i^N \bar{l}_i = L \text{ and } \sum_i^N k_i = \sum_i^N \bar{k}_i = K \end{aligned} \tag{2.19}$$

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

Differently from the case of the social planner maximization problem depicted in equation 2.4, the "scale" constraint $pl_i + ck_i = p\bar{l}_i + c\bar{k}_i$ is added to the resource constraint. Intuitively, it imposes that the value of the optimal l and k must be the same as the one of the inputs actually utilized in the agricultural production at the market clearing prices p and c so that the resulting allocation leaves each farm scale unaffected. The first order conditions imply:

$$\frac{l_i^*}{L} = \frac{k^*}{K} = \left[\alpha\frac{\bar{l}_i}{L} + (1-\alpha)\frac{\bar{k}_i}{K}\right] \tag{2.20}$$

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

$$\frac{\sum_i^N y_i^*}{\sum_i^N y_i} = \frac{Y^*}{Y} \tag{2.21}$$

The estimate obtained is 1.17, indicating that by reallocating inputs optimally across farmers maintaining their operational scale unaffected, the overall productivity would increase by only 17 percent.

The total gain from reallocation $G = \frac{Y^e - Y}{Y}$ can be decomposed as the sum between the gain from redistributing factors to guarantee an optimal mix of land and capital G^{MIX} plus the gains from redistributing the inputs (in their optimal combination) proportionally to the farmers' agricultural skills G^{SCALE}, where $G^{MIX} = \frac{Y^* - Y}{Y}$ and $G^{SCALE} = \frac{Y^e - Y^*}{Y^*}$.

This implies that the total misallocation estimated in Section 2.3.1 is predominantly caused by inefficiencies in the distribution of farms' scale, as postulated by the model. In particular, in our sample the gains from an efficient redistribution of factor mix within farms would reduce the total misallocation only by 11 percent. The predominance of scale misallocation is similarly very

Table 2.5: Decomposition of misallocation

	No controls	Weather	Weather & Land quality
Overall	2.55	2.49	2.42
	(0.89)	(0.89)	(0.88)
District	2.07	2.07	2.04
	(0.88)	(0.88)	(0.87)
County	1.97	1.98	1.96
	(0.87)	(0.87)	(0.87)
Village	1.39	1.39	1.39
	(0.85)	(0.85)	(0.85)

Percentage scale contribution in parentheses.

clear when considering narrower administrative units, and controlling or not for weather shocks and land quality, as shown in Table 2.5.[32]

It is important to mention that this methodology presents some relevant flaws. Most remarkably, as explained in Shenoy, 2017, the results should be interpreted as a lower bound of the relative importance of factor market misallocation as imperfections in land and capital markets might affect the scale and not only the mix of the input used. Intuitively, perfectly functioning input markets should at least guarantee a perfect combination of the tradable factor at the farm level, and this procedure only allows us to assess to which extent this is true. Whether the remaining amount of misallocation is actually due to other factors affecting the scale of the farms or to frictions in the input markets is a question that cannot be answered by this decomposition. Nevertheless, the fact that the input combination at the farm level seems to be almost fully optimal represents a valuable argument in support of the assumption of well functioning factor markets.[33]

2.5.3 Testing the model's predictions

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

[32]Note that the estimates presented in the last row are exactly the same as the ones displayed in Table 2.4.

[33]In the only comparable study, Shenoy, 2017 finds a considerably more relevant contribution of misallocation in input combination, ranging from 36% in 1996 to 75% in 2008 in the case of Thailand.

Table 2.6: Marginal factor productivities

Dependent Variable	Marginal productivity of L (log)			Marginal productivity of K (log)		
	(1)	(2)	(3)	(4)	(5)	(6)
% output consumed	-0.266***			-0.486***		
	(0.04)			(0.06)		
Substistence farm		-0.236***	-0.224***		-0.481***	-0.472***
		(0.02)	(0.02)		.(0.03)	(0.03)
Market oriented farm			0.067**			0.051
			(0.03)			(0.04)
N	24,364	24,364	24,364	24,364	24,364	24,364
adj. R^2	0.562	0.565	0.565	0.466	0.473	0.473

Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Subsistence farm is a binary variable taking value 1 when all the farm's produce is self consumed; 28 percent of the farms are purely subsistence. Market oriented farm takes value 1 when at least two thirds of the farm's output is sold, this is the case for 15 percent of the whole sample.

I will start by assessing whether this is the case by comparing farmers who operate in the same village in order to control for village specific land characteristics and/or cropping patterns that might affect agricultural productivity in a way that is not captured by the land quality index q. Formally, the regression estimated is:

$$\log(MP_{jiz}) = \beta_1 \text{ subsistence index}_{iz} + \nu_z \tag{2.22}$$

where the marginal productivity of input j for farmer i in the village z is regressed on some variable capturing whether i is a commercial or a subsistence farmer as well as a set of village level fixed effects. Practically, depending on the specification, I will either adopt a continuous index of subsistence indicating the percentage of the output value self consumed by the household (columns 1 and 4) or a dummy taking value one when the farmer produces exclusively for self consumption (columns 2 and 5) or two dummies identifying purely subsistence or commercial farmers (defined as these selling at least two thirds of their produce).

As expected, subsistence farmers display significantly lower marginal productivity of both land and capital. As shown in Table 2.6, the magnitude of the gap is rather substantial. When considering the percentage of agricultural output self consumed, I estimate a 30 and 63 percent difference in marginal productivity of land and capital between a farmer consuming all of her output and one selling the whole produce to the market. As the specification is log-linear, this translates into an estimated 0.3 and 0.6 percent decrease of marginal productivity of labour and capital for each percentage point increase in the share of output consumed. Interestingly, virtually all this gap is captured by the subsistence dummy included in the models estimated in columns 2 and 4, which seems to corroborate the view that there is a discontinuity in the shadow price of output between purely subsistence farmers and the others.[34] The coefficients of the "market oriented farm" dummy are positive for both inputs, but their magnitude is much smaller and not significantly different from zero in the case of capital.

These findings indicate that an efficient reallocation of factors of production would involve the transfers of land and capital from subsistence to market oriented farmers.[35] This can be shown and tested directly by comparing the actual land distribution (l_i) to the one in the full efficiency counterfactual: (l_i^{ez}).[36] More specifically, I compare the actual distribution to the efficient one calculated across different administrative areas, namely districts, counties and villages.[37]

In the first three regressions, the dependent variable is the probability of operating a quantity of land that is larger than it would be if factors were distributed proportionally to the agricultural productivity within the zone under analysis. As it is bounded between 0 and 1, I estimate a simple probit model as follows:

$$Pr(l_i^{ez} > l_i) = \Phi(\beta_1 \text{subsistence farm}_{iz} + \beta_2 \text{Market oriented farm}_{iz} + \epsilon_{iz}) \tag{2.23}$$

where Φ is the cumulative normal distribution and z refers to the three different zones the counterfactual is based on. The two farm specific dummies included are the same as Table 2.6.

I also estimate a similar set of regressions where the dependent variable is the difference between the land actually operated and the efficient level. The functional form accordingly changed from probit to OLS. The resulting expression is:

$$l_i^{ez} - l_i = \beta_1 \text{subsistence farm}_{iz} + \beta_2 \text{Market oriented farm}_{iz} + \epsilon_{iz} \tag{2.24}$$

where the coefficients β_1 and β_2 capture the difference between the residual group and the subsistence and the market oriented farmers respectively.

[34]According to the model, the marginal productivity should be equalized across all the farmers who sell a non zero share of their output and so there should only be a gap between pure subsistence farmers and the others. However, it is realistic to assume that farmers selling a high share of their produce and semi subsistence farmers who only sell a very small amount of output (but are still potentially net buyers in the market for food) attribute different shadow value to their production.

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

[36]I present and discuss the findings for land as it represents by far the most important treadable factor of production, but it can be shown that the results are very similar for capital input.

[37]In this and the following regressions, moving from broader to more narrowly defined areas reduces the group size, making the results more susceptible to measurement errors and outliers (especially when computing the magnitude of misallocation). However, I also gain in credibility as farmers who are closer are more likely to exchange factors of production and land in particular. Reassuringly, the results are generally similar across specifications.

Table 2.7: Actual vs efficient land scale

Dependent Variable	$\Pr(l_i > l_i^{ez})$ Probit (AME)			$\log(l_i) - \log(l_i^{ez})$ OLS coefficients		
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence farm	0.071***	0.068***	0.119***	0.661***	0.645***	0.632***
	(0.01)	(0.01)	(0.01)	(0.08)	(0.07)	(0.05)
Market oriented farm	0.017	0.020	-0.059**	0.082	0.026	-0.162**
	(0.01)	(0.01)	(0.01)	(0.09)	(0.08)	(0.07)
Level of analysis (z)	District	County	Village	District	County	Village
N	24,364	24,364	24,364	24,364	24,364	24,364
Mean y	0.78	0.77	0.64	1.91	1.83	0.83
adj. R^2	-	-	-	0.075	0.095	0.064

Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Subsistence farm is a binary variable taking value 1 when all the farm's produce is self consumed; 28 percent of the farms are purely subsistence. Market oriented farm takes value 1 when at least two thirds of the farm's output is sold, this is the case for 15 percent of the whole sample. The dependent variable in the first three columns is the probability of each farm to be operating a larger then efficient quantity of land. In columns 4 to 6 it is the difference between the (log) of land operated and the (log) of efficient farm scale. The level of analysis refers to the counterfactual scenario considered (efficient reallocation at the district/county/village level).

Table 2.7 presents the estimates of these regressions. Consistently with the hypothesis, subsistence farmers are operating on a larger than efficient scale both on the intensive and on the extensive margin. Indeed, the findings are remarkably similar across the different specification, and show that at the district and county level subsistence farmers are 9 and 8 percent more likely to be operating too much land (as compared to the unconditional probability). At the village level, this figure increases to 19 percent.[38]

The results of the OLS regressions are rather comparable, confirming that the prediction of the model holds also when comparing the actual scale of the actual and counterfactual holdings. Finally, it is worth mentioning that, as far as the village level specifications are concerned, also the market oriented farm dummy is statistically significant, but negative.

Estimates presented in Tables 2.6 and 2.7 show that the input distribution within villages is consistent with the patterns presented in the model, where subsistence farmers operate a higher than efficient shares of input and consequently have lower marginal productivity. Importantly, it has to be remarked that rather than being the result of a more or less arbitrary allocation procedure by the village leaders (potentially motivated by personal preferences or equality concerns), the observed factor distribution is better described as the outcome of voluntary transactions, as

[38]The fact that the magnitude is higher in this last case might reflect that factor transactions typically occur at the village level, and so the dynamics predicted by the model are more evident.

in most of the villages (see Table 2.2) a significant share of the farmers are operating plots that have been acquired, either permanently or temporally, through the land market. Thus, it might be argued that misallocation is not necessarily due to frictions in land markets, but rather the consequence of the different shadow values individuals attach to their produce.[39].

Another prediction of the model is that in each zone, misallocation should be larger within subsistence farmers. This mirrors the fact that in equilibrium subsistence farmers, unlike commercial ones, display different marginal productivities (see Equation 2.16).[40]

Formally, the estimated regression takes the form:

$$e_{gz} = \beta_1 \text{Subsistence Farms}_{gz} + X'_{gz}\gamma + \nu_z + \epsilon_{gz} \tag{2.25}$$

where the dependent variable represents the potential gains from reallocation (in logs) for each group g (subsistence or non-subsistence) in zone z (district, county and village). 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, that are summarized by 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.

As shown in Table 2.8, the estimates of β_1 are positive across all the administrative boundaries considered and whether controls are added or not.[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 village. 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 statistically significant at any conventional confidence level.

As expected, higher variance in the underlying group specific distribution in productivity dispersion correlates positively with the misallocation estimates. As far as the other controls are concerned, the model does not really provide any particular indication. The estimates show that larger amounts of capital per farm are correlated to more pronounced misallocation while higher average skill levels have the opposite impact.

[39]This does not necessarily imply that land markets actually operate in a frictionless way and that all the transactions are motivated by purely economic reasons

[40]In the model, the misallocation arises within constrained farmers (belonging to group **b**) as well as across groups (as the more skilled constrained farmers have higher marginal productivity). By allowing for decreasing marginal utility in food consumption, there would be misallocation also within non constrained farmers. As long as the frictions in output markets have also an individual specific component, one would expect some (yet arguably less significant) misallocation within commercial farmers as well.

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

Table 2.8: Misallocation within subsistence and commercial farmers

Dependent Variable	Group specific misallocation (log)					
	Districts		Counties		Villages	
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence Farms	0.152***	0.142***	0.148***	0.139***	0.062***	0.047***
	(0.03)	(0.04)	(0.04)	(0.05)	(0.01)	(0.02)
Productivity dispersion (log)	0.091***	0.352***	0.121***	0.283***	0.066***	0.120***
	(0.02)	(0.07)	(0.03)	(0.09)	(0.01)	(0.02)
Average land per farm (log)		0.060		0.030		-0.020
		(0.12)		(0.08)		(0.02)
Average capital per farm (log)		0.073***		0.085***		0.052***
		(0.02)		(0.02)		(0.01)
Average farm productivity (log)		-0.301***		-0.222**		-0.075***
		(0.09)		(0.10)		(0.02)
N	158	158	307	307	4,431	4,431
adj. R^2	0.647	0.779	0.636	0.733	0.367	0.438

Standard errors clustered at the district level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable represents the poterntial gain from efficient reallocation of land and capital within subsistence or commercial farmers in a district (first two columns) county (columns 3 and 4) and village (coulmns 5 and 6). Subsistence farm is a binary variable taking value 1 when considering the subset of farmers who consumed 100% of their produce. Fixed effects at the zone level (district for columns 1 and 2, county for columns 3 and 4 and village for columns 5 and 6) are always included.

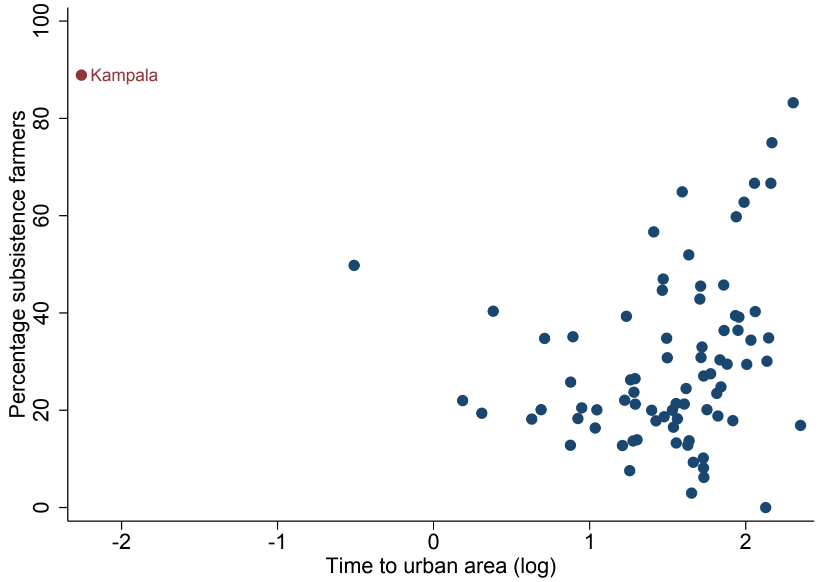
To conclude, I want to test whether there is actually a link between the prominence of subsistence agriculture and factor misallocation and to which extent this is driven by transportation costs as captured by our indicator of market accessibility. As explained in the data section, I will do so both at the village and at the district level. If on the one hand the village level specification is more credible as transportation costs are realistically village specific and input transaction occurs principally between farmers who reside in the same area, the district level specification, in light of the larger group size and lower sensitivity to errors in the geolocation process, provides some considerable advantages.

In the district level analysis, only 79 out of the 80 districts are considered as the capital district of Kampala is dropped from the analysis. This is due to the fact that in spite of the high market access score, most of the farmers are operating for subsistence purposes only, possibly mirroring the fact that in this highly urbanized environment agriculture represents a side activity only. This would explain why, as clear from a visual examination of Figure 2.8 Kampala is an obvious outlier.

The first step consists in assessing whether facing higher costs to access the market, as proxied by longer travel times to the most easily reachable urban areas, actually correlates with more pervasive subsistence agriculture. I will do so by estimating the following regression:

$$\text{Subsistence index}_z = \beta_1 \text{Travel time}_v + \gamma X'_z + \epsilon_z \tag{2.26}$$

Figure 2.8: District level travel time and subsistence farming



where a number of different subsistence indexes computed at the village or the district level are used as dependent variables. In particular, I compute four different ones, namely: the percentage of land operated by purely subsistence farmers, the percentage of farmers who are in pure subsistence (not selling any of the output), the fraction of output consumed by the median farmer and finally the fraction of the total output consumed by farmers in the zone.[42] Although they all generally capture 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 land operated by subsistence farmers (whether this happens by farmers changing group or a partial redistribution of land from commercial to subsistence farmers) and not on the number of subsistence farmers or any of the other variables considered. Results are shown for the village and district level specification and in Table 2.9 and 2.10 respectively.

It is apparent that subsistence agriculture is more widespread in villages located farther away from urban areas and in district with lesser transport infrastructure and connectivity. This finding seems to be robust across all the different dimensions of subsistence, with the sole exception of the percentage of output consumed by the median farmer in the village level regression. The point estimates in the tables represent the elasticities multiplied by 100. Thus, in terms of magnitude, a 1 percent increase in travel time for village results in an increase in 0.09 percent in the land farmed by purely subsistence farmers, which translates in a beta coefficient of nearly 0.1. Notably, the magnitude of the impact is more pronounced at the district level, as

[42]The pairwise correlation of the indexes ranges from 0.59 between percentage of total output sold and percentage of subsistence farmers at the district level and 0.90 between percentage of subsistence farmers and percentage of land operated by subsistence farmers at the village level.

Table 2.9: Travel time to urban areas and subsistence agriculture (Village Level)

	% land subst.	% subst	% consumed (median)	% consumed (overall)
Time to urban area	9.37***	8.92***	1.26	3.00*
	(2.48)	(2.43)	(1.85)	(1.74)
	[0.09]	[0.09]	[0.01]	[0.04]
Average land per person	-5.90***	-7.09***	-4.14***	-3.23***
	(0.76)	(0.74)	(0.66)	(0.62)
	[-0.17]	[-0.20]	[-0.13]	[-0.11]
Average capital per person	1.96***	2.00***	0.25	0.24
	(0.62)	(0.63)	(0.60)	(0.53)
	[0.07]	[0.07]	[0.01]	[0.01]
Average productivity	-4.01***	-4.48***	-1.63***	-0.48
	(0.44)	(0.43)	(0.42)	(0.40)
	[-0.28]	[-0.31]	[-0.13]	[-0.04]
Productivity dispersion	0.23	0.43	0.00	-0.54
	(0.43)	(0.42)	(0.40)	(0.38)
	[0.02]	[0.03]	[0.00]	[-0.05]
N	2,957	2,957	2,957	2,957
R^2	0.111	0.128	0.121	0.108

Standard errors clustered at the district level in round brackets: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standardized beta coefficients in square brackets. All variables are in logs. The dependent variables are: % of land cultivated by subsistence farmers (who consume all of their produce) in column 1, the percentage of farmers who are in pure subsistence in column 2, the percentage of output (value) consumed by the median farmer in column 3 and the percentage of the output (value) consumed in the village in column 4. They are all expressed in values from 0 to 100.

Table 2.10: Travel time to urban areas and subsistence agriculture (District Level)

	% land subst.	% subst	% consumed (median)	% consumed (overall)
Time to urban area	9.75***	6.92***	6.17***	9.92***
	(4.46)	(4.68)	(4.26)	(2.95)
	[0.37]	[0.25]	[0.24]	[0.50]
Average land per person	-3.02	-3.15	-2.79	1.40
	(3.85)	(3.97)	(3.75)	(2.84)
	[-0.10]	[-0.09]	[-0.09]	[-0.06]
Average capital per person	-0.73	-0.39	-0.63	-0.51
	(1.32)	(1.43)	(1.46)	(1.13)
	[-0.05]	[-0.03]	[-0.05]	[-0.05]
Average productivity	-11.12***	-12.84***	-9.34**	-6.34*
	(5.89)	(5.35)	(4.56)	(3.95)
	[-0.60]	[-0.66]	[-0.52]	[-0.46]
Productivity dispersion	8.52	10.34*	8.21*	5.04
	(5.31)	(5.43)	(4.32)	(3.96)
	[0.35]	[0.41]	[0.36]	[0.28]
N	79	79	79	79
R^2	0.375	0.381	0.342	0.366

Standard errors clustered at the district level in round brackets: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standardized beta coefficients in square brackets. All variables are in logs. The dependent variables are: % of land cultivated by subsistence farmers (who consume all of their produce) in column 1, the percentage of farmers who are in pure subsistence in column 2, the percentage of output (value) consumed by the median farmer in column 3 and the percentage of the output (value) consumed in the village in column 4. They are all expressed in values from 0 to 100.

the corresponding coefficient takes value 0.37. This suggests that village level estimates might be undermined by an imprecise geolocation and/or that there are other unobserved factors that explain the predominance of subsistence agriculture which are not captured by the estimated travel time to urban conglomerate. Reassuringly, all the other controls have the expected signs, as zones with higher availability of land per capita and more able farmers correlate with lower values of the subsistence indexes.[43]

Once established that the estimated travel time is indeed one of the determinants of subsistence agriculture, I proceed by computing its impact on agricultural factor misallocation. However, since the findings seem to indicate that it can only explain a relatively small share of the variation in the importance of subsistence farming (especially across villages), I will also study the direct impact of the indexes of subsistence agriculture we calculated on misallocation, in consideration of the fact that there might be other factors, such as price volatility (see Fafchamps 1992), affecting the gap between commercial and subsistence farmer's shadow prices.

Formally the impact of transportation costs on misallocation can be estimated as:

$$e_v = \beta_1 \text{travel time}_v + \beta_2 \text{land market participation}_v + X_v' \gamma + \epsilon_v$$

Alternatively, the direct relationship between subsistence and potential gains from trade is computed:

$$e_v = \beta_1 \text{subsistence index}_v + \beta_2 \text{land market participation}_v + X_v' \gamma + \epsilon_v$$

In both cases, I add to the standard controls a variable describing the village or district specific land market participation (captured by the percentage of farmers who are operating at least one plot that was either bought or rented in) in order to check whether, in line with the findings of Restuccia and Santaeulalia-Llopis, 2017 and Chen, Restuccia, and Santaeulàlia-Llopis, 2017, geographical variability in land transactions correlates with the level of factor misallocation.

The village level estimates are displayed in Table 2.11, and show that although the coefficients attached to all the subsistence indexes are significant at the conventional levels, this is not the case for the estimated time to reach an urban conglomerate, which is positive but not statistically different from 0. This is not necessarily surprising given its weak explanatory power highlighted in Table 2.9.

One possibility is that there are other, more relevant factors that affect the pervasiveness of subsistence farming, that might or not be related to other frictions in output market not captured

[43] The fact that capital per person has the opposite impact on some specification is admittedly quite puzzling, and might reflect agricultural intensification in areas more land constrained.

Table 2.11: Misallocation, travel time to urban areas and subsistence (Village level)

	Dep variable: Village level misallocation (log)				
	(1)	(2)	(3)	(4)	(5)
Time to urban area	0.91	-	-	-	-
	(1.72)	(-)	(-)	(-)	(-)
	[0.01]	[-]	[-]	[-]	[-]
Subsistence index†	-	6.26***	4.36***	3.49**	3.24*
	(-)	(1.59)	(1.46)	(1.45)	(1.73)
	[-]	[0.08]	[0.05]	[0.04]	[0.03]
Pct participation land market	0.54	1.01	1.06	0.96	0.91
	(1.41)	(1.33)	(1.34)	(1.34)	(1.34)
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Average land per person	-1.35**	-1.45**	-1.55***	-1.74***	-1.79***
	(0.59)	(0.61)	(0.60)	(0.60)	(0.60)
	[-0.05]	[-0.05]	[-0.05]	[-0.06]	[-0.06]
Average capital per person	0.59	0.62*	0.62*	0.61*	0.60*
	(0.39)	(0.35)	(0.35)	(0.35)	(0.35)
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]
Average productivity	-10.71***	-10.44***	-10.46***	-10.55***	-10.59***
	(0.39)	(0.39)	(0.39)	(0.39)	(0.39)
	[-0.90]	[-0.88]	[-0.88]	[-0.89]	[-0.89]
Productivity dispersion	11.46***	11.42***	11.41***	11.41***	11.43***
	(0.38)	(0.37)	(0.37)	(0.37)	(0.37)
	[1.03]	[1.03]	[1.02]	[1.02]	[1.03]
N	2,957	2,957	2,957	2,957	2,957
R^2	0.102	0.122	0.125	0.127	0.119

Standard errors clustered at the district level in round brackets: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standardized beta coefficients in square brackets. † the subsistence indexes used vary from model to model, they are: % of land cultivated by subsistence farmers (who consume all of their produce) in column 2, the percentage of farmers who are in pure subsistence in column 3, the percentage of output (value) consumed by the median farmer in column 4 and the percentage of the output (value) consumed in the village in column 5. All (non percentage) variables are in logs. Coefficients (elasticities) and standard errors are multiplied by 100 for readability.

by transportation costs. On the other hand, it is also possible that the estimated travel time is too noisy a measure considering the potentially imperfect geolocation procedure or it captures poorly the actual access to rural markets.

For this reason, the same analysis is carried out also at the district level, where the estimated travel time is more likely to capture the general quality of the transportation infrastructure as opposed to an inevitably more volatile cell specific estimate. As shown in Table 2.12, the results in this case are in line my hypothesis. In particular, a standard deviation increase in the estimated time to reach an urban conglomerate is correlated to a 0.31 standard deviations higher level of misallocation. The direct impacts of the district level subsistence indexes are rather similar, excluding the percentage of output consumed whose positive coefficient is yet not statistically significant.

Although these results might be driven by other factors that correlate with a low quality transport

Table 2.12: Misallocation, travel time to urban areas and subsistence (District level)

	Dep variable: District level misallocation (log)				
	(1)	(2)	(3)	(4)	(5)
Time to urban area	18.79***	-	-	-	-
	(8.93)	(-)	(-)	(-)	(-)
	[0.31]	[-]	[-]	[-]	[-]
Subsistence index†	-	47.90*	60.37***	61.30**	20.85
	(-)	(26.73)	(23.24)	(29.71)	(37.16)
	[-]	[0.20]	[0.27]	[0.25]	[0.07]
Pct participation land market	11.89	1.21	1.59	-2.17	9.81
	(21.70)	(17.36)	(15.94)	(17.46)	(18.74)
	[0.39]	[0.01]	[0.01]	[-0.01]	[0.06]
Average land per person	-0.94	1.09	1.56	1.64	-0.22
	(3.64)	(3.95)	(3.87)	(3.95)	(3.97)
	[-0.03]	[0.03]	[0.05]	[0.05]	[-0.01]
Average capital per person	0.42	0.50	0.66	-0.11	0.22
	(2.37)	(2.50)	(2.50)	(2.54)	(2.65)
	[0.02]	[0.02]	[0.03]	[-0.00]	[0.01]
Average productivity	-5.43***	-4.07***	-3.80***	-4.45***	-4.76***
	(1.82)	(1.80)	(1.78)	(1.82)	(1.86)
	[-0.39]	[-0.29]	[-0.27]	[-0.32]	[-0.34]
Productivity dispersion	5.28	6.40	6.09	4.66	5.44
	(3.76)	(4.90)	(5.10)	(4.82)	(4.25)
	[0.09]	[0.11]	[0.10]	[0.08]	[0.09]
N	79	79	79	79	79
R^2	0.289	0.291	0.287	0.290	0.285

Standard errors clustered at the district level in round brackets: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standardized beta coefficients in square brackets. † the subsistence indexes used vary from model to model, they are: % of land cultivated by subsistence farmers (who consume all of their produce) in column 2, the percentage of farmers who are in pure subsistence in column 3, the percentage of output (value) consumed by the median farmer in column 4 and the percentage of the output (value) consumed in the village in column 5. All (non percentage) variables are in logs. Coefficients (elasticities) and standard errors are multiplied by 100 for readability.

infrastructure such as poor access to other facilities and markets, the link between travel time, subsistence farming and factor misallocation is rather robust and prevalent in the context under analysis, and seems to be corroborated both by within and across zones patterns.

Suggestively, land market participation seems not to play a role in reducing misallocation in the context under analysis. This is in clear contrasts with the findings of Chen, Restuccia, and Santaeulàlia-Llopis, 2017, who look at the impact of land market liberalization in the aftermath of a land reform. Such apparently puzzling result is plausibly due fact that land markets have been active for a relative long time in Uganda and have progressively realised a great share of the total gains from trade achievable given the other existing frictions. However, it cannot be excluded that this is due to the fact that the observed land transactions are not based on market logic and as such, they are not necessarily efficiency enhancing.

2.6 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, 2018, the huge disparities observed cannot be attributed to systematic differences in land endowments but rather must be the outcome of sub optimal production patterns. A particularly appealing explanation for them is that, as markets for agricultural inputs and output fail systematically in low income countries (see Dillon and Barrett 2017 for a recent study on Sub Saharan Africa), resources are not allocated efficiently among existing farmers. In a similar fashion to Hsieh and Klenow, 2009, the main hypothesis is that consistent productivity gains could be achieved by redistributing the existent factors of production from less to more productive individuals.

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

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

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

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

The estimated magnitude of the productivity gain 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 inputs markets are a necessary, yet not sufficient condition to achieve an efficient distribution of agricultural inputs across farmers.

In line with the model proposed, I find that subsistence farmers operate systematically on a larger than efficient scale. This suggests that they attribute a higher value to their output and therefore are reluctant to separate from their land and agricultural capital. This prevents more productive commercial farmers from increasing their holdings' size. Misallocation therefore seems to be caused by frictions in food markets that induce a wedge between subsistence and commercial farmers' shadow prices rather than explicit barriers to input transactions. Consistently with my hypothesis, I show a positive correlation between the village and district 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 does 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, we argue that complementary policies/investment aimed at reducing frictions in food markets are needed to achieve a efficient input distribution across farmers and enhance agricultural productivity.

Chapter 3

Transportation costs and cropping patterns in Uganda

3.1 Introduction

In Sub Saharan Africa, a large share of the population relies on farming as the main source of income and livelihood. This fact, paired with low agricultural productivity, explains a large part of the existing cross country income differences (Gollin, Parente, and Rogerson 2007, Restuccia, Yang, and Zhu 2008). Since the bulk of farming activities takes place in rural and often remote areas with poor access to markets and infrastructures, it is natural to suspect that transportation costs might be one of the concurring factors of lagging agricultural productivity. Specifically, prohibitive transportation costs are believed to lead to resource misallocation within the agricultural sector. In particular, lack of integration across markets (caused by poor transportation networks) can have a distortionary impact on crop choices, which can significantly deviate from the optimal patterns implied by heterogeneous agro-climatic conditions.

Costinot and Donaldson, 2016 show how an increasingly pervasive transportation network in the US shaped the cropping patterns across different counties, which gradually specialized into the production of relatively more suitable crops. Adamopoulos, 2019 perform a similar exercise to study the impact of a state intervention in Ethiopia aimed at improving the road network, showing that it accounted for a relevant share of the agricultural productivity gains experienced by the country between 1996 and 2014, partly through a more efficient allocation of land between food and cash crops. Sotelo, 2019 estimates a model that predicts that both welfare and productivity in Peru would benefit from an abatement of the trade frictions across different region of the country, "by unlocking the forces of comparative advantages" Sotelo (2019, p.2).

Rather than looking at the issue of lack of integration across regional markets, this paper focuses on the transportation frictions faced by rural farmers who serve the same main urban market. This is justified by some estimates which suggest that, in the context under analysis, inefficiencies in cropping patterns within market regions are at least as important in magnitude as the ones across regional markets. Thus, the theoretical framework is largely in line with De Janvry, Fafchamps, and Sadoulet, 1991 and the following works on farm households' land use decisions (see De Janvry and Sadoulet, 2006 for a recent overview on the literature).[1] On the theoretical ground, the main contribution is that, unlike Jayne, 1994 and Omamo, 1998, the model incorporates differences in crop specific land suitability.[2] By doing so, I show that the impact of transportation costs on crop choices is twofold. Indeed, low market accessibility not only increases the share of land devoted to staple crops (and conversely reduces the acreage of high value cash crop), but also reduces or even reverses the responsiveness to crop specific comparative advantage of farmers whose production and consumption decisions are interlinked. Intuitively, this happens as individuals who rely on self production to meet their consumption target for a given food crop will need to devote a share of land to its production which is inversely proportional to the crop specific suitability.

The empirical analysis corroborates the main predictions of the model. Namely, I find that substantial gains from crop reallocation could be achieved also within areas realistically producing for the same markets, and that the misallocation observed is driven by differences in market access across heterogeneous producers who face different agro-climatic conditions. When comparing farmers with different transportation costs, results show that those who are better connected to the markets devote systematically more land to high value crops (regardless of their relative suitability) than those who are located in more isolated areas, and vice versa. Additionally, I find that farmers in the most remote locations are less responsive to comparative advantage of both food and cash crops, as they produce mostly for self consumption purposes.

This paper relates closely to the work by Li, 2017, who shows that Indian farmers exposed to a food programme relaxing their subsistence constrain reallocated land from less suitable/ profitable crops produced for self consumption to more valuable ones, enhancing the overall allocative efficiency of cropping patterns across and within regional market areas. It also echoes the findings of Qin and Zhang, 2016 who show that when Chinese villages are linked to the

[1] These early works were originally aimed at understanding the lack of responsiveness to shocks in crops' price, but their framework can be generalised to explain apparently sub optimal crop choices that are not in line with the agronomic comparative advantage.

[2] The implicit assumption of land being equally suitable for each crop can be quite problematic and unrealistic. For example, Stifel and Minten, 2008 found that in the case of Madagascar the conventional negative relationship between isolation and cash crop production was reversed and that was entirely driven by the fact that the areas with more favourable conditions for the main export crops (vanilla, cloves and coffee) also happened to be the least well connected to urban areas.

main road network, farmers start specializing in the production of the cash crops which are most suitable to their climate and soil conditions.

This work also contributes to the more general literature on the impact of poor transport infrastructures on agricultural outcomes in the developing world. A number of dual sector macro models have incorporated transportation costs to study their effect on the distribution of resources across sectors. Adamopoulos, 2011 shows that they magnify the impact of the subsistence constraint on the overall economic performance and widen the productivity wedge between farming and the manufacturing sector. Similarly, Gollin and Rogerson, 2014 provide a model where costly trade between rural agricultural and urban areas results in a larger and less productive agricultural sector and hinders aggregate productivity. In a cross country framework, Tombe, 2015 shows that transportation costs prevent low income countries from importing food from more developed economies with huge comparative advantage in agricultural production, thus aggravating the agricultural productivity gap and its negative impact on poor economies.

Poor road access has also been found to directly dampen agricultural production. Gollin and Rogerson, 2014 show that on a theoretical ground, transaction costs between urban and rural areas do not only increase total food production needed to sustain cities, but also reduce the adoption of modern intermediate inputs by farmers and in turn their yields. A similar intuition underlies the empirical works by Dorosh et al., 2010 and Stifel and Minten, 2008, who find a strong link between land productivity and road connectivity and attribute it to easier access to advanced technologies and inputs. Interestingly, this relationship does not seem to be driven exclusively by differences in land quality that might correlate with transport infrastructure placement.

Finally, this paper provides further evidence and estimates of misallocation of agricultural factors of production in Sub-Saharan countries. Unlike the existing studies (Restuccia and Santaeulalia-Llopis 2017, Chen, Restuccia, and Santaeulàlia-Llopis 2017), rather than looking at input distribution across farmers who differ in their inherent productivity, I focus on the distribution of crops across areas with different agro-climatic conditions, whose impact on crop specific productivity is well understood and less subject to measurement issues that might undermine the analysis (Gollin and Udry 2019; Maue, Burke, and Emerick 2020). Differently from the similar cross country analysis by Adamopoulos and Restuccia, 2018, the empirical results are based on micro level data rather than projections and seem to indicate larger inefficiencies in the geographical crop distribution in Uganda. To my knowledge, this is the first paper using GAEZ and farm level data to estimate the magnitude of misallocation in geographical crops' distribution. In particular, the analysis indicates that the aggregate agricultural productivity could be increased by one third just by reallocating crops production efficiently.

The essay is organised as follows: Section 2 illustrates the theoretical framework and links the main predictions of the model to the data. Section 3 presents the data and the derivation of the main variables, with a particular focus on the definition of comparative advantage and transportation costs. Section 4 is devoted to the description of some motivating facts on crop misallocation in Uganda that justify my interest in distortions within market areas. In Section 5, I present the empirical analysis and comment on the findings. Section 6 concludes the paper with some final remarks.

3.2 Theoretical Framework

This section presents a very simple model of crop choice. Unlike the otherwise rather similar work by Omamo, 1998, the decision is made at the village rather than at the household level. This reflects the fact that the two most important explanatory variables, namely transportation costs and crop specific suitability, are defined at the village level in the data. This implies that every village will be presented as a representative agent that makes production and consumption decisions to maximise a unique utility function.

For the sake of tractability, I will abstract from a number of very relevant dimensions such as differences in factor shares across crops and the choice between farm and off-farm labour. Furthermore, the model ignores potential complementarities between different crops (that might explain some crop combination due to efficient intercropping patterns) and does not take into account the (potentially relevant) role played by the livestock sector in smallholder Ugandan agriculture. The potential issues arising from these simplifications are discussed in greater detail in the following sections.

3.2.1 The Model

Each village has a fixed endowment of land $\bar{l}$ that can be used to produce either a cash c or a food f crop according to two different production functions: $g_c(l_c)$ and $g_f(l_f)$ respectively. Land is the only input in agricultural production and it is inelastically supplied to the farming sector. Thus, the decision problem involves only the share to be devoted to each crop. The utility is a function of the consumption of food c_f and another representative good (which is also the numeraire) c_m i.e. $U = U(c_f, c_m)$. This good cannot be produced by the farmers and can be thought as a composite good representing all the non-food items households necessarily need to source from the market (similarly to Omamo, 1998). The cash crop does not enter the utility function, but it can be sold at the regional market at the exogenous price p_c. Similarly, the food crop can be either bought or sold at the same market at the exogenous price p_f. The fact that each village faces the same price reflects that they are all located in the same region.

When selling or buying crops, farmers face a per unit transport cost Δ, which depends on the location of the village as well as the quality of the transportation infrastructure.[3] Unlike when buying/selling crops, farmers face no transaction costs when buying the "other good" m.[4]

Each village has an exogenous source of income $\overline{y}$, which can be thought of as the result of some extra agricultural activities.

Formally, the maximisation problem can be described as:

$$\max_{c_f, c_m, l_f, b_f, c_f} U(c_f, c_m) \tag{3.1}$$

subject to:

$$c_m + p_f c_f + \Delta(b_f + s_f) \leq \overline{y} + p_f g_f(l_f) + (p_c - \Delta) g_c(l_c) \tag{3.2a}$$

$$c_f = g_f(l_f) + b_f - s_f \tag{3.2b}$$

$$l_f + l_c = \bar{l} \tag{3.2c}$$

$$b_f \geq 0 \tag{3.2d}$$

$$s_f \geq 0 \tag{3.2e}$$

where b_f and s_f stand for units of food bought and sold respectively.

The first constraint imposes that the total amount spent on consumption goods and transportation cannot exceed the total income (obtained as the sum between $\overline{y}$ and the value obtained for the cash crop produced and sold). The second is the food consumption constraint, simply stating that the amount of food consumed must be equal to the difference between the total production (plus the amount bought) and the quantity sold. The third one imposes that the total land available is used for food or cash crop production only.[5] The remaining inequality constraints state that the amount of food crop bought and sold must be non-negative.

[3]For simplicity, transaction costs are assumed to be linear in the quantity bought/sold. Also, unlike Omamo, 1998, transaction costs are not crop specific in this version of the model, although I discuss the potential implication of the relaxation of this assumption in the following subsections.

[4]Potentially, this could alter households' optimal consumption bundle by making the representative market good relatively more costly for more isolated households. However, given the simple structure imposed on the preferences (with utility increasing only in food consumption up to the satiation point), relaxing the assumption on frictionless good does not make any difference in terms of production decisions. Also, a more flexible preference structure combined with non zero transaction costs when buying the market good m would not alter (and actually reinforce) the main predictions of the model as more isolated farmers would also consume relatively more (self produced) food.

[5]These three constraints are expressed as equalities as the model is static and does not allow to save (or store food crops) for future periods, and we assume that the only use for land is agricultural production and therefore it can be only used to grow cash or food crop.

The resulting Lagrangian (once the the land use constraint is substituted for) takes the form:

$$
\begin{aligned}
\mathcal{L} =& U(c_f, c_m) \\
& + \lambda_1[\overline{y} + p_f g_f(l_f) + (p_c - \Delta) g_c(\overline{l} - l_f) - c_m - p_f c_f - \Delta(b_f + s_f)] \\
& + \lambda_2[g_f(l_f) + b_f - c_f - s_f] + \mu_1 b_f + \mu_2 s_f
\end{aligned}
\tag{3.3}
$$

The most immediate way to qualitatively study the impact of transportation costs on crop choice is to find the shadow price of food crop f, which represents the value farmers attach to the crop (which might differ from regional market price p_f) and on which the land use decisions are based.

The first order conditions for c_m, c_f and l_f are:

$$\frac{\partial \mathcal{L}}{\partial c_f} = 0 \Rightarrow \frac{\partial U(.)}{\partial c_f} = \lambda_1 p_f + \lambda_2 \tag{3.4a}$$

$$\frac{\partial \mathcal{L}}{\partial c_m} = 0 \Rightarrow \frac{\partial U(.)}{\partial c_m} = \lambda_1 \tag{3.4b}$$

$$\frac{\partial \mathcal{L}}{\partial l_f} = 0 \Rightarrow (\lambda_1 p_f + \lambda_2) \frac{\partial g_f(.)}{\partial l_f} = \lambda_1 (p_c - \Delta) \frac{\partial g_c(.)}{\partial l_c} \tag{3.4c}$$

Food shadow price p_f^* can be obtained as:

$$p_f^* = \frac{\frac{\partial U(.)}{\partial c_f}}{\frac{\partial U(.)}{\partial c_m}}$$

which is the value that sets the ratio of the marginal utilities equal to the ratio of prices.[6] By plugging in the first order conditions, the expression becomes $p_f^* = \frac{\lambda_1 p_f + \lambda_2}{\lambda_1} = p_f + \frac{\lambda_2}{\lambda_1}$.[7]

The remaining first order conditions as well as the corresponding mutually slackness constraints

[6] In this case, m is the numeraire good, therefore p_f^* can be obtained as the ratio between marginal utilities.

[7] The same result could be obtained by looking at the supply side by imposing the equality of the marginal value of food and cash crop production, formally: $p_f^* \frac{\partial g_f(.)}{\partial l_f} = (p_c - \Delta) \frac{\partial g_c(.)}{\partial l_c}$.

are as follows:

$$\frac{\partial \mathcal{L}}{\partial b_f} = 0 \Rightarrow \lambda_2 = \lambda_1 \Delta - \mu_1 \tag{3.5a}$$

$$\mu_1 \geq 0 \tag{3.5b}$$

$$b_f \mu_1 = 0 \tag{3.5c}$$

$$\frac{\partial \mathcal{L}}{\partial s_f} = 0 \Rightarrow \lambda_2 = -\lambda_1 \Delta + \mu_2 \tag{3.5d}$$

$$\mu_2 \geq 0 \tag{3.5e}$$

$$s_f \mu_2 = 0 \tag{3.5f}$$

where the multipliers μ_1 and μ_2 are set to be non-negative since the underlying problem is a maximisation.

This implies that the food decision price depends on the village balance of trade in the food market. In particular, where if the village is a net importer ($b_f > 0$) $\lambda_2 = \lambda_1 \Delta \Rightarrow p_f^* = p_f + \Delta$ thus the food decision price equals the consumer price (which is the sum of the regional market price and the transportation cost). Conversely, where the village is a net seller, $\lambda_2 = -\lambda_1 \Delta \Rightarrow p_f^* = p_f - \Delta$ and therefore the shadow price equals the producer price. Finally, where the village is autarchic, the decision price equals $p_f - \Delta + \mu_2$, with $\mu_2 > 0$ (or $p_f + \Delta - \mu_1$ with $\mu_1 > 0$) and therefore it lies between the consumer and the producer prices.

It follows that unless the village is a net food exporter, the producer price (which is the price that would be received when selling food to the market net of transportation costs) does not represent the decision price. More specifically the, food producer price represents a lower bound of the shadow price on which farm households base their production decisions. Therefore, in maximizing their utility, farmers might optimally deviate from the profit maximizing land allocation. This implies that *ceteris paribus* villages in more isolated areas where Δ is higher will tend to devote a higher share of their land to the food crop (or generally speaking to any produce entering their utility function). Up to this point, the theoretical framework is mirroring the one developed by, among others Fafchamps, 1992 and Omamo, 1998 and the predictions obtained are in line with the existing literature.

In the following section, I will augment this benchmark model to account for heterogeneous land endowment which is captured by different crop specific land productivity. This allows me to study the distortions in crop choices brought about by transportation costs beyond deviations from the profit maximising land allocation.

3.2.2 Land heterogeneity and comparative advantage

Variations in agro-climatic conditions across villages result in different levels of crop specific productivity across decision makers. These differences are included in the model through the production functions. For the sake of tractability, I assume that they are linear in the amount of land devoted to each crop; in formal terms:

$$g_f = Fl_f$$

$$g_c = Cl_c$$

where both F and C are village specific.

Following Gollin, Parente, and Rogerson, 2007, I assume that utility is increasing in food consumption only until a certain threshold $\overline{c_f}$ is met. After the satiation point is reached, utility is an increasing function of c_m. Formally:

$$U(c_f, c_m) = \begin{cases} c_f, & \text{if } c_f \leq \overline{c_f} \\ u(c_m), & \text{if } c_f > \overline{c_f} \end{cases}$$

This implies that, for each village, the priority is to meet the food consumption need (either by producing or by buying the food crop) and good m is bought and consumed only once food subsistence is guaranteed. Since $\overline{c_f}$ is a saturation point, food consumption will never exceed it in equilibrium.

The production decisions are a function of the relative magnitude of the transportation costs and the crop specific suitability. In particular, three different scenarios can be identified:

$$\frac{F}{C} \geq \frac{p_c - \Delta}{p_f - \Delta} \tag{3.6a}$$

$$\frac{F}{C} \leq \frac{p_c - \Delta}{p_f + \Delta} \tag{3.6b}$$

$$\frac{p_c - \Delta}{p_f + \Delta} < \frac{F}{C} < \frac{p_c - \Delta}{p_f - \Delta} \tag{3.6c}$$

The first case applies to villages with relatively high food crop suitability, to the extent that it guarantees higher returns than the cash crop when sold at the market. For this reason, regardless of whether the resulting production exceeds or falls short of the subsistence threshold, all the land available is devoted to the food crop. Intuitively, in the former case the excess production is sold to the market to maximise revenues (and c_m) while in the latter it will be produced for self consumption. Either way, producing the cash crop is never optimal as it generates lower revenues than the food crop and it does not enter the utility function.

The second scenario depicts the opposite situation where land suitability is much higher for cash crops (relative to the magnitude of the transportation costs), which means that the whole land endowment is optimally devoted to cash crop production. This is a consequence of the fact that producing and selling C units of cash crops generates enough revenue to buy at least F units of food, even once transportation costs are accounted for. Thus, growing the cash crop is the best option regardless of the consumption choices.

In both of these scenarios, transaction costs are not pronounced enough to determine any deviation from the profit maximising land allocation. In fact, Δ only affects the feasible consumption set (as higher Δs reduce the value of the production and in turn the amount of goods that can be afforded) but it does not alter the crop choices made. More precisely, production and consumption decisions are completely independent and farmers act as pure profit maximisers.[8] This also implies that villages are fully specialised in the production of either crop, which is the one that maximises revenues.

Conversely, when the transaction costs are high enough, farmers face the situation depicted in Equation 3.6c. In this case, although the cash crop guarantees higher revenues per hectare, due to the high cost of accessing agricultural markets, farmers are better off producing food for self consumption purposes and use the residual land (if any) to grow cash crop. Intuitively, this happens because producing and selling C units of cash crops allows them to buy less than F units of food crop. Doing so is therefore only convenient once the village's food consumption needs are satisfied (in order to maximise the consumption of the other good m). This in turn implies that the quantity of land devoted to the food crop, *ceteris paribus* is a negative function of its land suitability. As a result, for sufficiently high transportation costs, crop choices can be opposite to the efficient one, given the underlying structure of comparative advantage.

Formally, the quantity of land devoted to the food crop is:

$$l_f^* = \begin{cases} \bar{l} & \text{if} \quad \frac{F}{C} \geq \frac{p_c-\Delta}{p_f-\Delta} \\ 0 & \text{if} \quad \frac{F}{C} \leq \frac{p_c-\Delta}{p_f+\Delta} \\ \frac{\overline{c_f}}{F} & \text{if} \quad \frac{p_c-\Delta}{p_f+\Delta} < \frac{F}{C} < \frac{p_c-\Delta}{p_f-\Delta} \quad \text{and} \quad \overline{c_f}F \leq \bar{l} \\ \bar{l} & \text{if} \quad \frac{p_c-\Delta}{p_f+\Delta} < \frac{F}{C} < \frac{p_c-\Delta}{p_f-\Delta} \quad \text{and} \quad \overline{c_f}F > \bar{l} \end{cases} \tag{3.7}$$

This shows explicitly that transportation costs can not only result in deviations from profit maximizing crops (when the profit maximising crop is not the one also consumed by farmers),

[8]Note that this does not necessarily imply that farmers grow cash crop only, as the so-called food crop can be the one that guarantees the higher returns given the expected yields and prices.

but also reduce (and potentially reverse) the impact of agronomic comparative advantage on land use patterns. Indeed, in the case of interdependence between consumption and production decisions, the higher the suitability to food crop F, the lower the amount of land devoted to it in equilibrium l_f^*.[9]

Figure 3.1 summarises these findings, plotting the equilibrium share of land devoted to the food crop as a function of the food crop suitability F and maintaining everything else (and crucially cash crop suitability C) fixed. When there are no transportation costs, there is full specialization and villages grow either the cash or the food crop only, depending on which one generates higher revenue. Once transportation costs Δ are included, there are deviations from this outcome. In particular, for some values of F, there is interdependence between production and consumption decisions, which results in both crops being grown. More specifically, when food suitability is not high enough for the food crop to be the profit maximising one, the share of land devoted to it is a negative function of F, and therefore crop choices reverse the underlying structure of comparative advantage. This pattern is particularly marked for high values of Δ, which widen the set of values F such that $\frac{p_c-\Delta}{p_f+\Delta} < \frac{F}{C} < \frac{p_c-\Delta}{p_f-\Delta}$ and so specialization does not occur and crop choice deviates from the structure of comparative advantage. In the case of "High Δ" depicted in Figure 3.1, this is the case for nearly each possible value of F.

To sum up, the model shows that transportation costs not only generate deviations from the profit maximizing crop choice, but also reduce the responsiveness of farmers to comparative advantage. Importantly, although the base intuition is perfectly in line with the trade literature on integration across different regional markets, this framework describes the distorsive impact of transportation costs on production within producers operating in the same local area and highlights the issue of market accessibility of rural farmers rather than of connectivity across regional markets.

3.2.3 From theory to practice

Before moving to the presentation of the data and the econometric analysis, it is useful to point out some differences between the model and the necessarily more complicated structure of the data and to generalise its predictions for the real world scenario.

The most obvious departure from the theoretical framework is that farmers can choose between more than two crops, and that the distinction between food and cash crop is often less clear-cut

[9]Whether or not this results in a reversal of the comparative advantage depends on the correlation between the food and the cash crop suitability and the extent to which there is actual variation in that dimension. This will be explored in the data section.

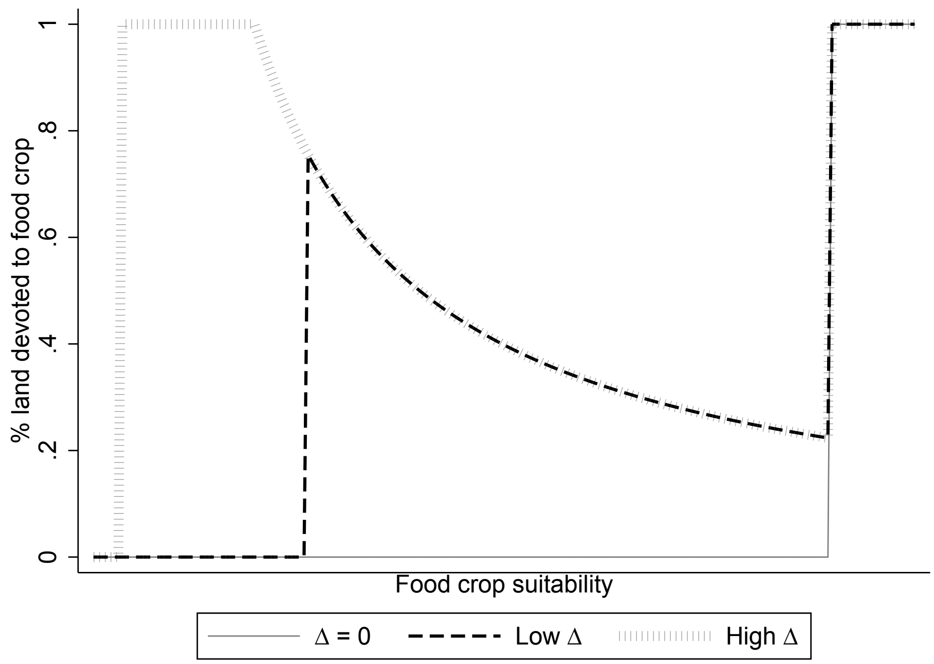


Figure 3.1: Transportation costs and optimal land use

than what is assumed in the model. The complexity of Ugandan cropping pattern is clearly depicted in Table 3.1, which provides statistics on the nine most important crops grown in the country (by acreage).[10] Two things in particular need to be noticed: first, there is no such thing as a clearly dominant crop, as the most widely grown (maize) only accounts for less than 20 percent of the total agricultural land, closely followed by cassava and matoke. Even put together, these crops account for less than half of the total acreage.

In addition, with the sole exception of coffee, there is no crop that is grown for selling purposes only. Indeed, the percentage of the total production sold ranges from 13 percent for sweet potatoes to 58 percent for maize. However, these crops differ fundamentally in terms of potential revenue per acre which can be used as a proxy for the market value of the crops.[11] Unsurprisingly, coffee seems to be by far the crop with the highest market value. Indeed, it is the produce that maximises expected revenue in half of the villages in the sample and on average guarantees a return equal to 88 percent of the revenue maximising crop.

A viable way to identify other cash crops would be to choose the ones that present similar values in terms of monetary returns and marketisation. However, the relationship between potential

[10]The figures refer to the Ugandan Census of Agriculture, which is one of the datasets used in the empirical analysis. I will elaborate on the structure of the data in the next section.

[11]Potential revenue per acre is computed as the market price (which is the median obtained by pooling all the transactions for each of the crops in the World Bank's 2009 LSMS for Uganda) and the expected yield estimated by the GAEZ agronomic model (see next section for more information on it).

Table 3.1: Crops characteristics

	Percentage land		Potential revenue		Percentage sold	
	Overall	Village (mean)	% Rev max (village)	Mean % (village)	Mean (village)	Overall
Maize	18.48	16.88	14.17	70.12	37.75	57.76
Cassava	16.26	15.47	0.03	56.69	20.27	29.14
Matoke	14.67	16.61	16.47	71.81	20.60	36.85
Beans	10.54	10.56	0.03	47.92	29.88	47.07
Swpot	7.94	7.76	13.53	68.28	11.20	14.82
Sorghum	6.82	5.79	5.04	55.46	20.57	22.31
Coffee	6.35	6.90	49.78	88.73	100	100
Grnut	6.19	5.25	0.95	42.33	30.00	42.37
Millet	4.77	3.76	0.00	20.73	24.70	39.68

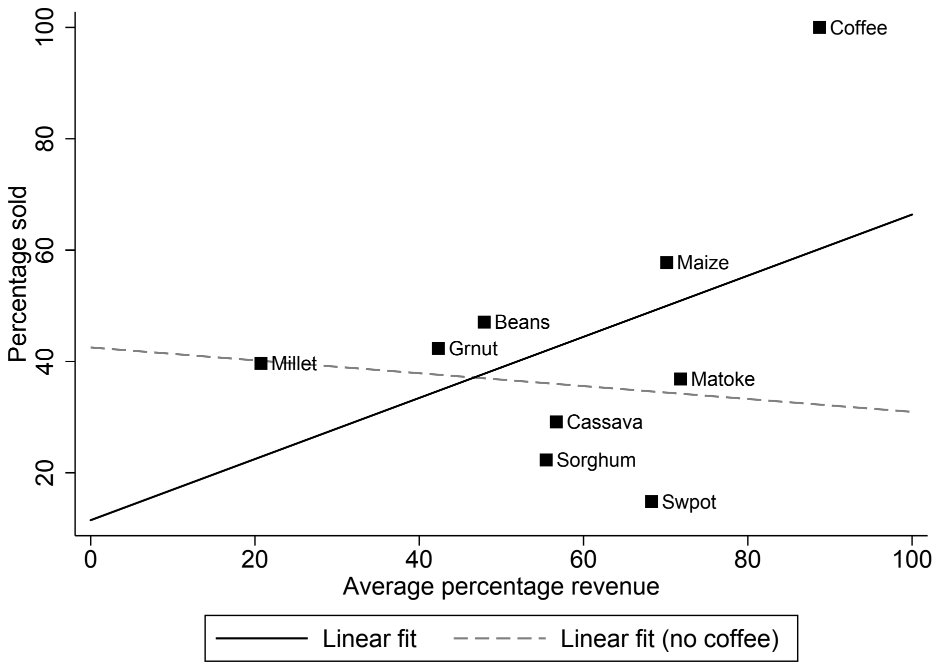
The first column reports the percentage of the total land devoted to each crop, while the second the average land per village. Third and fourth columns refer to the potential revenue per hectare. Column 3 presents the pervcentage of villages where the crop is the one that maximises revenue, while column 4 indicates the average ratio of the potential revenue of the crop and of the revenue maximising one. The share of the total production that is sold for each crop is presented both as the mean across villages with non zero production (column 5) and as a fraction of the aggregate output (column 6).

revenues and percentage sold is not as clear when all crops are taken into account. In particular, as can be seen in Figure 3.2, it is not necessarily the case that crops with higher potential revenue are systematically more likely to be sold (the positive slope of the linear fit is indeed solely driven by coffee). In particular, very valuable crops like sweet potatoes and matoke are less commonly sold than less or equally profitable crops like millet, beans and maize. This seems to question the dichotomy between cash and food crop stated in the model and introduces some additional complexity in farmers' cropping decisions.

However, a clear pattern can be identified: Indeed, less perishable/bulky food crops such as cereals, pulses and oil seeds (with the only exception of sorghum) display higher levels of marketisation given their market value. This is possibly driven by the fact that these crops are easier to transport and/or store and therefore can be transported for longer distances without losing their market value whereas tubers like sweet potatoes and fruit like matoke are both heavier and more perishable and so less convenient to move to far markets. Therefore, these crops are realistically very profitable to be sold where markets are easy to access and only valuable as food where they are not.

In consideration of that, the predictions of the model can be generalised to a context where crop choice involves a mix of pure cash crops (coffee) and food crops which differ in their sensitivity to transportation costs. In particular, lower access to markets will plausibly reduce the amount of land devoted to cash crops (as indicated by the model) and food crops with high market value, but very susceptible to transportation costs (like matoke and sweet potatoes) and increase the share of land used to grow less perishable food crops (like maize, millet and sorghum).

In terms of responsiveness to comparative advantage, the predictions of the model are left unaffected, as it equally applies to the food and the cash crop. However, the fact that in the real



Percentage sold on the y axis refers to the aggregate production for each crop. Average percentage revenue is obtained as the average ratio of the expected value of the crop and the revenue maximizing crop across all villages. The solid line represents the linear fit between the two variables while the dashed one plots the linear fit when excluding coffee.

Figure 3.2: Average potential revenue and percentage sold

world farmers have different food crops they can pick from might result in a lower distortionary impact of transportation costs as they might adjust their consumption decisions depending on the agronomic condition they face (i.e. consuming predominantly crops that are easy to grow in their area). The extent to which they are willing and able to do so depends crucially on the degree of substitutability of their preferences and modelling it is beyond the scope of this paper.[12] However, this possibility is accounted for in the interpretation of the findings from the empirical analysis.

Additionally, crops differ also along dimensions that are more difficult to capture empirically and are not included in the model. One obvious example is the labour/land intensity, which might play an important role in the decision making process but it is abstracted from in the theoretical framework.[13] Additionally, some crops like cassava and sorghum are typically grown by virtue of their resistance to droughts, which can make them relatively more valuable than the otherwise similar maize and sweet potatoes. Since it does not feature any form of uncertainty, the model is unable to capture this fact. Again, this will be accounted for in the interpretation of the results rather than explicitly incorporated in the theoretical framework.

Finally, some cropping choices might be motivated, rather than by the mere suitability for the produces themselves, by the existence of sizeable complementarities across crops and between some specific crops and livestock related activities. Such complementarities can be either of technical nature (e.g. to contain pest damage, as shown in Sekamatte, Ogenga-Latigo, and Russell-Smith, 2003), or reflect *ex ante* risk hedging strategies farmers might optimally undertake (Bezabih and Di Falco 2012). These factors might ultimately explain part of the deviations from the optimal cropping patterns dictated solely by crop specific comparative advantages. Thus, it would be misleading to interpret such deviations as inefficiencies as they rather reflect choices that are optimal in some dimensions that are not accounted for in the theoretical framework. However, it is reasonable to state that the magnitude of the measured inefficiencies is still informative on the occurrence of distortions due to transportation costs and lack of market access.

3.3 Data and descriptive statistics

The empirical analysis is based on three main datasets, namely the Ugandan Census of Agriculture (UCA), the Global Agro-Ecological Zones (GAEZ) and the Road Network of Uganda,

[12]Unfortunately, I do not have data on households' consumption and therefore I cannot directly test this hypothesis.

[13]e.g. perennial crops like coffee, bananas and cassava tend to require lower quantities of labour per acre and so might be relatively more convenient for labour scarce households or simply relatively more profitable than what captured just looking at the yields per hectare.

which will be used respectively to extract information on cropping choices across the country, patterns of crop specific comparative advantage and market accessibility. In the following, I provide a brief description of each of the data sources and explain how the main variables are derived.

3.3.1 The Ugandan census of agriculture

The Ugandan census of Agriculture (UCA) is a micro level dataset providing accurate information on the economic activities of a large and nationally representative sample of agricultural holdings. The survey refers to the 2008/09 agricultural year and originally includes 35,407 holdings located in 3,557 enumeration areas spread throughout the 80 districts of Uganda.[14]

The most relevant variable derived from this dataset is land use, which is available at the village level. In particular, for each enumeration areas, the survey provides information on the crop choices made by each respondent for both the second agricultural season of 2008 and the first agricultural season of 2009. This allows me to compute the village specific fraction of land devoted to each of the nine major crops included in the analysis, which overall covers more than 90 percent of the agricultural land. Since all the plots are measured through GPS, the resulting measure is very accurate.

Most of the villages sampled were also geolocated. Thus, it is possible to combine the information contained in the UCA with spatial datasets. Coordinates were not available for around 10 percent of the enumeration areas in the original sample. These villages are therefore not included in the econometric analysis since no reliable information on their transportation costs and land specific crop specific land suitability could be obtained. Additionally, I remove the enumeration areas with low response rate ($\leq$ 40 percent) and those where the crops under analysis cover a relatively small fraction of the agricultural area ($\leq$ 60 percent) as the resulting figures on land use might not be informative on the actual crop choices made in the area. This reduces the sample to 2,941 villages.[15]

[14]Enumeration areas are survey specific localities which are roughly the same size as a village (for rural areas) or a parish (for urban areas). In the following, I will use the terms enumeration area and village interchangeably. The sampling follows the pre-2006 district borders.

[15]This represents 83 percent of the original villages sampled and 87 percent of the holdings. The results are robust if all the geolocated EAs are considered and/or when attributing to non geolocated EAs the median values of crop suitability and market access of other observations in the same district.

3.3.2 Crop specific land suitability and comparative advantage

Information on crop specific land suitability is taken from the GAEZ dataset, created by FAO. It consists of a sophisticated agronomic model, which on the basis of information on climate patterns (averaged over 30 years between 1960 and 1990) and of terrain and soil characteristics, returns an estimate of the expected yield per hectare for a number of crops.[16] Since the dataset is defined on a very granular scale (5 Arc-minutes, i.e. 10 square kilometres cells at the Equator), it is able to capture variation in crop specific productivity even within small areas. Crucially, GAEZ is estimated for the nine most widely grown crops in Uganda (see Table 3.1). Therefore, it allows to define a rather comprehensive description of the underlying structure of comparative advantage in agriculture across villages.

Since the model features two crops only, the structure of comparative advantage was fully captured by the ratio of land productivity for food and cash crop. The most obvious way to generalise this to account for more than two crops is to compare the land suitability for each crop to some average across the others. Formally, for each crop c in village v, I define an index of comparative advantage for c as:

$$cadv_v^c = \frac{dec_v^c}{\sum_c dec_v^c * share^c} \tag{3.8}$$

where dec_v^c is the decile of land suitability for crop c in village v, and $share^c$ is the overall fraction of land devoted to crop c in the sample.[17] In line with the concept of comparative advantage, this index is constructed in such a way that it can take up a lower value in areas that experience absolute advantage in the production of a crop. For example, a village that is located in an area which is very favourable for the production of all crops (e.g. lies in the highest decile of the suitability distribution of each crop) will have a index equal to 1 for all crops, while a village that has an average productivity for c and lower than average for the others, will have an index ≥ 1 for crop c.

The distributions and the descriptive statistics of the resulting indexes are shown in Figure 3.3 and in Table 3.2.[18] For each crop, the mean is very close to 1 and the distributions are concentrated around it. This suggests that there are a number of agro-climatic factors that are either favourable or detrimental to agricultural production in general and therefore crops' expected yields are generally positively correlated.[19]

[16] The estimates are available for a number of different combinations of input used in the agricultural production (low, intermediate or high) and the irrigation method. I consider a low input level and rain fed irrigation, which represent by far the most common scenario in Sub Saharan Africa.

[17] I weight the contribution of each crop by its acreage in the full sample, so that having relatively high/low productivity for a marginal crop has a lower impact on the measure of comparative advantage. By doing so, the findings are more robust to the number of crops considered as removing or adding a relatively less commonly grown crops would have a lower impact on the resulting index.

[18] Plots and descriptive statistics are based on the 2,941 observations corresponding to all the villages included

Table 3.2: Distribution of comparative advantage indexes

	Mean	Median	St. Dev	Min	Max
Maize	0.95	0.91	0.36	0.22	2.70
Cassava	0.93	0.98	0.32	0.15	2.59
Matoke	0.97	0.96	0.40	0.20	2.27
Beans	0.96	0.95	0.38	0.19	3.23
Swpot	0.94	0.97	0.33	0.17	3.25
Sorghum	1.04	0.89	0.61	0.16	3.39
Coffee	0.96	0.94	0.41	0.14	3.54
Grnut	0.93	0.99	0.45	0.11	4.05
Millet	0.96	0.99	0.44	0.12	3.70

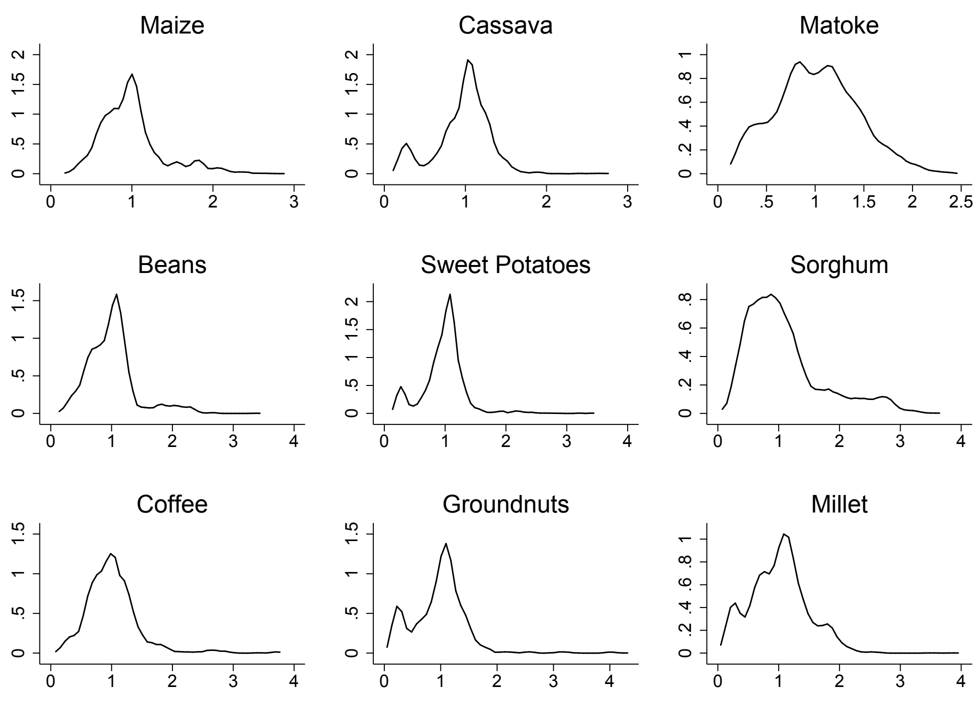


Figure 3.3: Distribution of comparative advantage (Kernel density)

Table 3.3: Cross correlation of indexes of comparative advantage

	Maize	Cassava	Matoke	Beans	Spot	Sorghum	Coffee	Gnut	Millet
Maize	1.000								
Cassava	-0.797	1.000							
Matoke	-0.548	0.397	1.000						
Beans	0.704	-0.752	-0.613	1.000					
Spot	-0.738	0.741	0.221	-0.661	1.000				
Sorghum	0.660	-0.757	-0.606	0.671	-0.513	1.000			
Coffee	-0.264	0.271	0.256	-0.363	0.047	-0.466	1.000		
Gnut	-0.463	0.262	-0.145	-0.275	0.342	-0.306	0.026	1.000	
Millet	-0.496	0.301	-0.132	-0.257	0.456	-0.136	-0.306	0.650	1.000

Interestingly, the distributions of these indexes tend to have longer right tails. This is not surprising as changes in the index where the values are below 1 capture larger differences in relative crop suitability. As an example, while an index of 0.5 for crop c implies that on average land is twice as suitable for the other crops (with higher weights attached to more common crops) than for c, while an index of 1.5 indicates that crop c is only 50 percent more suitable than the others. Additionally, the values taken up by the index are mechanically affected by the relative share of the crop the measure is computed for. This reflects the fact that when computing the index for a crop c, the denominator is a weighted average of all the crops, including c itself, and if c is given a higher weight due to its high acreage, the difference between numerator and denominator will be reduced. Accordingly, the range of values taken by the index is narrower for more widely grown crops like maize and matoke than for more marginal crops like groundnut and millet.

These features of the measure imply that, depending on the value taken by the index and on the crop considered, a unit change can capture different changes in the crop's relative productivity. This serves as a warning for the interpretation of the results and on the way the index should be entered in regression models.

Finally, Table 3.3 displays the cross correlation coefficients across the indexes of comparative advantage for each crop. As they capture the relative crop suitability, the majority of the coefficients have a negative sign, indicating that villages facing comparatively better agronomic conditions for a crop tend to have symmetrically relatively lower suitability for the others. Unsurprisingly, this is not the case for crops that are very similar and as such require analogous soil and climatic conditions to grow like cassava and sweet potatoes or maize and sorghum. However, even in these instances the coefficients never exceed 0.75.

in the final sample.

[19] In fact, when computing the correlation coefficients across crops, they are virtually all positive. The highest value is 0.84 (beans and maize), while the the lowest is -0.10 (sorghum and matoke).

3.3.3 Transportation costs and market accessibility

While existing studies define regional markets on the basis of certain administrative boundaries (i.e. US counties in Costinot and Donaldson, 2016 or Peruvian provinces in Sotelo, 2019), I adopt a different approach which allows me to determine the extension of local markets and at the same time define a village specific market access based on the estimated transportation costs. The latter will be defined as a function of the existing road network.

In order to do so, I divide the whole country in cells of one square kilometre, and for each of them compute the crossing time, as a function of whether there is a road, which type of road it is, the terrain roughness and land cover. Roads placement and characteristics are based on the road network map provided by Ubos (2015) and depicted in Figure 3.4. When calculating each cell's crossing time, I assume that the quickest means of transportation available are used i.e. cars for motorways and that the speed is a function of the type of road. Specifically, I assume 70 km/h for primary roads, 50 for secondary, 20 for tertiary and 10 for the residual ones. For cells without roads, the speed is a function of the terrain roughness (from 4 km/h for flat to 2 for very steep) and of the land cover (correction coefficient of 1.5 for forests and no transit for inland water bodies and natural reserves). Then, for each cell, I run an algorithm returning the time to reach a main city using the fastest route, where I consider as main city any town with more than 30,000 inhabitants.[20]

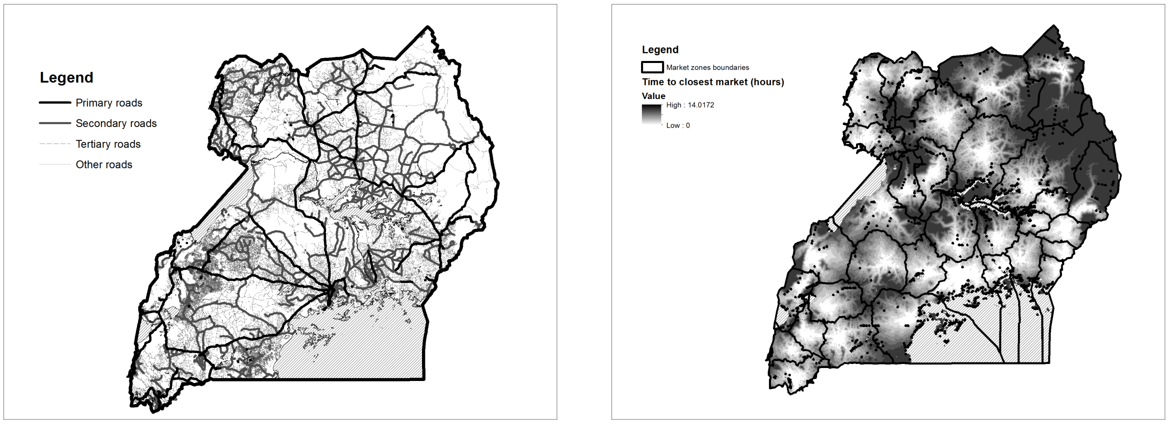
As the villages in the sample are geolocated, I am able to match each of them to the corresponding cell's estimated travel time, that represents my proxy for market accessibility. Additionally, this procedure allows me to define regional markets as the set of cells that given the existing transportation network are best connected to a given major city.[21] As a result, I assign a "catchment zone" to each of the 32 cities , each representing a regional market. Contrarily to the definitions based on administrative boundaries, this procedure is based on the existing transportation network and therefore it is more likely to capture the real subdivision into market areas. Figure 3.4 provides a graphical representation of this partition and of the market access index.

A potential drawback of this methodology is that the estimated travel times are rather sensitive to the assumptions made in terms of cells' crossing time. It is worth pointing out that rather

[20] I add to this set of city Kaabong in the North Eastern region of the country as it nearly meets the threshold of 30,000 and there is a cluster of enumeration areas in its proximity that are otherwise quite far from any other city. Changing the threshold to similar values do not affect the analysis in significant way.

[21] As there are a number of cities with more than 30,000 inhabitants in some areas, such as the surroundings of Jinja and the capital Kampala, in these instances I do not define a market for each of those cities but I rather consider only the most important city among them when defining the boundaries of a regional market. The implicit assumption made is that these markets are perfectly integrated due to their proximity. However, also the cities that do not define a regional markets are considered in the computation of the market access index.

Figure 3.4: Road network and market access



than trying to compute a realistic travel time, the procedure aims at calculating a more general market connectivity index based on the existing road network. For this reason, I do not expect the computed travel time value to be necessarily informative and thus I will enter them in the regression models as quintiles which are presumably less affected by the underlying assumptions on road speed. Additionally, this could reduce the potential statistical noise caused by slight errors in the geolocation procedure as quintiles display lower variations among neighbouring cells.

3.3.4 Heterogeneity of crop suitability and comparative advantage within market areas

The main contribution of this paper is the identification of deviations from the patterns of comparative advantage observed at the market area level, as opposed to the ones caused by lack of integration across regional markets. For future reference it is therefore useful to provide some estimate of the variation of land suitability and the comparative advantage index within the market zones defined in the previous subsection. Indeed, if variation in land suitability were solely the reflection of differences in agro-climatic conditions across markets, there would be no reason to be concerned about the spatial distribution of crops within market areas.

The results presented in Table 3.4 clearly indicate that this is not the case. Indeed, a large part of the variation in crop specific land suitability, as well the indexes of comparative advantage defined in Equation 3.8, seems to be driven by differences within villages located in the same market zone. As far as land suitability is concerned (the raw index of expected yield per hectare estimated by the GAEZ model), for all the crops under analysis, around 50 percent of the overall variation is observed within the 32 market areas considered. As for the index of comparative

Table 3.4: Within market area variability in land suitability and comparative advantage

	Suitability	C.adv Index	C.adv Quintiles
Maize	67.71	71.58	78.50
Cassava	49.75	65.50	61.58
Matoke	42.99	61.36	60.93
Beans	63.40	62.50	72.51
Swpot	46.08	76.16	68.46
Sorghum	56.91	53.81	54.08
Coffee	47.66	71.42	64.90
Grnuts	48.34	70.06	68.71
Millet	43.87	61.10	59.30

Figures are obtained as 1-R^2 computed by regressing land suitability and comparative advantage (index and quintiles) on a full set of market fixed effects.

advantage, the figures are even higher, exceeding 60 percent for each crop excluding sorghum, and above 70 percent for maize, sweet potatoes, coffee and groundnuts. As indicated in the third column, the share of variation captured by differences within market areas is similarly very high also when using the quintiles in the crop specific comparative advantage distribution as opposed to the continuous index.

These findings highlight the importance of understanding the determinants of crop choice also within market zones, as well as the potential frictions that might cause the cropping patterns to deviate from the ones implied by the underlying system of comparative advantage. The next section aims at computing the efficiency gains in terms of increased value of production that could be reached by reallocating crops on the basis of their relative suitability, both across the whole sample and within market areas. This exercise will provide an idea of the magnitude of the existing misallocation in cropping patterns and the extent to which it is brought about by inefficiencies at the local market zone level.

3.4 Efficiency gains from crops reallocation

In order to compute the potential gains from an efficient spatial reallocation of crops, I follow the steps suggested by Adamopoulos and Restuccia, 2018.[22] Intuitively, the procedure aims to compute the difference between the actual value of aggregate output and the one potentially

[22] More specifically, I will estimate the "second counterfactual", to follow the terminology of Adamopoulos and Restuccia (2018, p.21).

achievable just by reallocating crops to the relatively most productive locations without changing the aggregate land shares nor the amount of agricultural land in each location. Formally, the total value of production obtained from efficiently reallocating crops $c \in C$ across all the villages $v \in V$ is computed as:

$$Y^E = \max_{l_{cv}} \sum_{c \in C} \sum_{v \in V} p_v \hat{z}_{cv} l_{cv} \tag{3.9}$$

Where $\hat{z_{cv}}$ is the expected yield estimated by the GAEZ model for crop c in village v, l_{cv} is the quantity of land used for crop c in village v and p_c is the price of crop c.[23]

Subject to:

$$\sum_{c \in C} l_{cv} \leq L_v \ \ \forall \, v \in V \tag{3.10a}$$

$$\sum_{v \in V} l_{cv} \leq L_c \ \ \forall \, c \in C \tag{3.10b}$$

$$l_{cv} \geq 0 \ \ \forall c \in C \ \ , \ \ \forall \, v \in V \tag{3.10c}$$

Where the first constraint imposes the aggregate share of land devoted to each crop remains fixed, the second one imposes that the amount of farm land for each village does not change, while the last one is a non negativity constraint.[24]

Y^E can therefore be estimated simply by solving a linear programming problem. The comparison between the resulting value and the actual value of the production Y^A indicates the level of misallocation in land use. I find that the ratio $\frac{Y^E}{Y^A}$ takes value 1.33, which suggests that by reallocating crops according to their relative productivity across all the villages in the sample would increase the total value of the production by around one third.[25]

In order to compute how much of the gap in productivity is accounted for by misallocation within the 32 market zones identified in the previous section, I compute another counterfactual Y^M, which represents the total value of the production that would be achieved if crops were reallocated efficiently only across villages located in the same market zones. Formally, for each market $m \in M$, I solve the maximisation problem:

[23]I use $\hat{z}$ as opposed to the actual yields since I am only interested in the efficiency of crop distribution and I do not focus on production. The prices are obtained as the median market value of each crop according to the LSMS and they are the same used to generate the statistics in Table 3.1.

[24]The constraints guarantee that in the efficient counterfactual, only crop distribution is altered, whereas the aggregate share of land devoted to each crop and the amount of land used in the agricultural process in every location are the same. This implies that every difference between the actual output value and the counterfactual is attributable to the inefficient geographical distribution of the crops.

[25]Note that this estimate is higher than the one by Adamopoulos and Restuccia, 2018, as their corresponding figure for Uganda is around 19%. This difference is realistically due to the fact that I am using micro level data as opposed to projections. However, the order of magnitude of the estimates is reassuringly similar.

$$Y^m = \max_{l_{cv}} \sum_{c \in C} \sum_{v \in V^m} p_v \hat{z}_{cv} l_{cv} \tag{3.11}$$

subject to:

$$\sum_{c \in C} l_{cv} \leq L_v^m \ \ \forall \, v \in V^m \tag{3.12a}$$

$$\sum_{v \in V^m} l_{cv} \leq L_c^m \ \ \forall \, c \in C \tag{3.12b}$$

$$l_{cv} \geq 0 \ \ \forall c \in C \ \ , \ \ \forall \, v \in V^m \tag{3.12c}$$

Where V^m is the set of villages v located in the market area m.

In practice, the output (value) maximising crop distribution is computed separately for each of the 32 market zones, holding constant the market specific land shares of each crop. In order to compute these gains, I assume that the relative crop prices are not affected by the geographical reallocation of crops and the resulting differences in quantities produced. This mirrors the structure of the model where regional market prices are exogenously fixed and not affected by crop choice and production. The achievable increase in the value of output varies significantly across the 32 markets, ranging from 5.01 percent for the lowest to 61.05 for the highest, with an average of 21.28 percent.

I can then use the solutions to these maximisation problems to estimate:

$$Y^M = \sum_{m \in M} Y^m \tag{3.13}$$

the ratio $\frac{Y^M}{Y^A}$ is informative on the relative importance of within market crop misallocation. In particular, a value close to 1 would indicate that all the misallocation observed is due to inefficiencies across markets, while a value close to $\frac{Y^E}{Y^A}$ would suggest that most of the potential gains could be achieved just by reallocating crops within villages in the same market zone. In this case, the ratio takes value 1.17. This implies that 53 percent of the gains from reallocation could be achieved just by changing the cropping patterns within the 32 market areas identified in the previous section.

These descriptive findings suggest that although regional markets seem to be far from integrated, also the cropping patterns of farmers within each zone appear to be inefficient and to play as important a role in determining the aggregate misallocation. In the next section, I will examine the possibility that these inefficiencies are driven by differences in market accessibility across production units operating in the same market.

3.5 Empirical Results

This section presents the results of the main empirical analysis. I combine the information on land use, comparative advantage and market access obtained as explained in Section 3.3 to study to which extent the misallocation in land use patterns is due to transportation costs. In particular, I will test the two hypotheses derived from the model, namely whether transportation costs, controlling for comparative advantage, affect crop choices, incentivising farmers better linked to markets to grow crops with higher market value and lower crop specific transportation costs (coffee and/or tubers) and whether the impact of comparative advantage is mitigated or even reversed among production units that are more isolated.

3.5.1 Responsiveness to comparative advantage

First of all, it can be useful to have a rough idea of the aggregate response of farmers by computing for which extent, regardless of the transportation costs, the share of land devoted by each village to a crop changes in the crop's relative productivity. In practice, this will be done by estimating the regression:

$$\text{Land share}_{cv} = \beta_1 \text{comparative advantage}_{cv} + X'_{cv}\gamma + \mu_m \tag{3.14}$$

using a left censored tobit model, in consideration of the the non trivial number of zeros in the dependent variables. Since I have identified significant variation in the measure of comparative advantage both within and across market zones, the regression will be estimated both with and without the market fixed effects μ in order to obtain results that indicate the responsiveness of land use both within and across local markets. The dependent variables are the enumeration level land shares devoted to crop c in village v expressed as a number between 0 and 100. The explanatory variable of interest is the comparative advantage expressed either as an index (see Equation 3.8) or as the quintiles in the relative index distribution. X includes a number of village specific controls; namely the overall land fertility for generic agricultural purposes, the number of farms in the enumeration area and the average holding size.[26]

The estimates are displayed in Table 3.5 and show that, with the sole yet notable exception of maize, the share of land devoted to each crop correlates positively to the corresponding relative productivity. This is true whether or not market level fixed effects are included. Given the above mentioned issues in the interpretation of the results based on the continuous index, the

[26] Fertility is computed as $\sum_c dec_v^c * share^c$, which is also the denominator of the comparative advantage formula and is a weighted average of the decile in the distribution of land suitability for each crop, where the weights are the land shares of the crop in the sample. It is worth pointing out that the results are not affected by the exclusion of these controls or by removing any subset of them.

Table 3.5: Crop specific responsiveness to comparative advantage

	Index		Quintiles	
Maize	-4.09***	-3.08***	-1.14***	-0.83***
	(0.83)	(0.76)	(0.21)	(0.19)
Cassava	10.64***	2.22***	1.42***	0.24
	(0.91)	(0.84)	(0.21)	(0.20)
Matoke	18.27***	3.34***	5.30***	1.33***
	(1.34)	(1.09)	(0.37)	(0.32)
Beans	1.41**	0.39	0.30*	-0.26
	(0.55)	(0.66)	(0.15)	(0.17)
Swpot	2.31***	1.00**	0.42***	0.13
	(0.48)	(0.46)	(0.12)	(0.11)
Sorghum	5.77***	1.23***	4.12***	0.92***
	(0.48)	(0.48)	(0.23)	(0.19)
Coffee	9.12***	3.08***	2.85***	1.38***
	(0.87)	(0.80)	(0.27)	(0.31)
Grnut	5.73***	1.68***	2.28***	1.03***
	(0.45)	(0.43)	(0.11)	(0.17)
Millet	7.27***	2.61***	2.41***	0.83***
	(0.43)	(0.48)	(0.13)	(0.16)
Market FE	No	Yes	No	Yes

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robust standard errors in parentheses. The figures represent the estimated impact of a unit change in the comparative advantage index (or quintile) on the percentage of land devoted to the same crop (from 0 to 100). The estimates are obtained using a tobit model left censored at 0.

magnitude of these effects is best appreciated by looking at the results based on the quintile distribution. According to these findings, when market fixed effects are not included, moving one quintile up in the distribution of crop suitability increases significantly the share of land devoted to each crop. The impact ranges from 5.30 percentage point for matoke to only 0.42 for sweet potatoes. The results are qualitatively similar when using the index. In the case of maize, the converse is true, as moving up a quintile reduces the share of land covered by it goes down by around 1 percentage point both with and without market zone fixed effects.

The differences in magnitudes between the models with and without market level fixed effects are otherwise rather pronounced, and in some instances the point estimates of the former are not statistically significant. This is particularly notable for some crops like matoke, cassava and millet, where the differences are five-fold to threefold. As shown in Table 3.4, there are significant differences in the relative productivity of crops even within local markets. Thus, the relatively low (and sometimes insignificant) coefficients of the models with fixed effects cannot be entirely attributed to low variation in the measures of comparative advantage. Rather, they indicate that, unlike what is assumed in Costinot and Donaldson, 2016 and Sotelo, 2019, there are considerable departures from the efficient crop distribution also within domestic local markets.[27] This conclusion corroborates the findings of the previous section, proving the existence of sizeable productivity gains from geographical redistribution of land use also within the 32 market areas.

Thus, the aggregate analysis shows that generally farmers are not particularly responsive to the underlying system of comparative advantage, especially when considering the geographical distribution of crops within each market zone. Interestingly, in the case of maize (the most important crop by acreage), these findings show a negative relationship between land shares and relative crop productivity, which is one of the key prediction of the model for high levels of transportation costs. In the following section, I include my measure of transportation cost to examine their impact on land use decisions.

3.5.2 Market accessibility and cropping patterns

According to the theoretical framework, the first order impact of low market accessibility (captured here by the estimated travel time to main markets) is a reduction of the area devoted to cash crops. As argued in section 3.2.3, this prediction can be generalised to food crops with high market value but less easy to transport due to their perishability or bulkiness. While this

[27] It is however worth stressing that the local markets considered by Costinot and Donaldson, 2016 are much smaller (US counties) than the ones considered here, which makes the assumption of efficient crop distribution more credible and justifies the prevalent interest in the integration across rather than within them.

result is widely acknowledged by the literature, existing studies have typically overlooked the fact that land is heterogeneous in terms of crop specific suitability (Stifel and Minten 2008). In the following, this is done by including the index of comparative advantage.

Additionally, transportation costs can reduce farmers' responsiveness to comparative advantage by incentivising subsistence farming. More specifically, farmers operating in remote areas might optimally use a larger share of land to grow relatively less suitable crops just to meet their consumption need and avoid the high costs they would sustain if they were to buy the same crops at the markets. As shown in Figure 3.1, the higher the transportation costs are, the lower is the responsiveness of production choices to the structure of comparative advantage of each crop.

In order to identify both these effects, the following village level regression is estimated:

$$\begin{aligned}\text{Land share}_{cv} = {} & \beta_1 \text{comparative advantage}_v + \beta_2 \text{time to market}_{cv} \\ & + \beta_3 \text{comparative advantage}_v * \text{time to market}_{cv} + X'_{cv}\gamma + \mu_m\end{aligned} \tag{3.15}$$

Where X represents the same set of controls as in the previous regressions and I capture market accessibility as the quintile in the distribution of the estimated time to reach the major market city.[28]

The coefficient of β_2 captures the direct impact of transportation costs, which I expect to be crop specific. In particular, it should take on positive values for food crops which are easy to transport, while it should be negative for cash crops as well as food crops with high market value but which are hard to store and move to the markets in light of their perishability and bulkiness.

The mitigation (and potential reversal) of the supply response to comparative advantage is instead identified through the interaction term between crop specific comparative advantage and transportation costs. In this case, I expect the coefficient to be negative as the effect does not depend on crops' characteristics.[29]

Table 3.6 presents the estimates of coefficients β_1, β_2 (with and without interaction term) and β_3 for each of the nine crops considered, where comparative advantages are either captured by the index described in Equation 3.8 or by the quintile in the corresponding distribution. Due to the above mentioned issues with the interpretation of the coefficients attached to the index, most of the discussion will be focused on the model based on the quintiles. For most of the

[28] As in the previous case, I use a left censored left probit model in light of the non trivial share of 0 values.

[29] As it is straightforward in the two crops example, a reduction in the farmers' responsiveness to food crop's comparative advantage mechanically implies that the responsiveness to cash crop is symmetrically reduced.

Table 3.6: Transportation costs and crop choice

	Time to market quintile	C. adv index	Interaction	Time to market quintile	C. adv quintile	Interaction
Maize	0.42***	-2.74***		0.43***	-0.77***	
	(0.18)	(0.79)		(0.18)	(0.20)	
	-0.21	-4.69***	0.66	0.82**	-0.37	-0.13
	(0.47)	(1.60)	(0.47)	(0.42)	(0.42)	(0.13)
Cassava	0.01	2.21***		0.03	0.23	
	(0.17)	(0.91)		(0.17)	(0.20)	
	-0.39	0.93	0.43	0.05	0.25	-0.01
	(0.51)	(1.80)	(0.51)	(0.40)	(0.39)	(0.12)
Matoke	-1.28***	3.87***		-1.27***	1.46***	
	(0.23)	(0.93)		(0.23)	(0.28)	
	-0.25	6.99***	-1.03*	0.09	2.78***	-0.44***
	(0.62)	(2.03)	(0.56)	(0.55)	(0.57)	(0.16)
Beans	0.48***	0.43		0.48***	0.28*	
	(0.13)	(0.59)		(0.13)	(0.15)	
	0.19	-0.48	0.30	1.11***	0.91***	-0.21***
	(0.33)	(1.16)	(0.33)	(0.29)	(0.30)	(0.09)
Swpot	-0.58***	1.24***		-0.57***	0.17*	
	(0.11)	(0.50)		(0.11)	(0.11)	
	-1.01***	-0.16	0.46	-0.92***	-0.18	0.12
	(0.30)	(0.96)	(0.29)	(0.25)	(0.26)	(0.07)
Sorghum	0.90***	1.40***		0.91***	0.99***	
	(0.16)	(0.46)		(0.16)	(0.21)	
	1.25***	2.33***	-0.32*	0.73***	0.81***	0.06
	(0.28)	(0.83)	(0.19)	(0.35)	(0.40)	(0.12)
Coffee	-0.92***	3.05***		-0.94***	1.38***	
	(0.25)	(0.87)		(0.25)	(0.30)	
	-1.16***	2.32***	0.23	-0.68***	1.61***	-0.08
	(0.56)	(1.77)	(0.47)	(0.57)	(0.58)	(0.16)
Grnut	0.14	1.57***		0.12	0.96***	
	(0.11)	(0.38)		(0.11)	(0.14)	
	0.61***	3.13***	-0.51**	0.60***	1.46***	-0.16***
	(0.23)	(0.79)	(0.23)	(0.22)	(0.25)	(0.07)
Millet	0.39***	2.27***		0.39***	0.71***	
	(0.12)	(0.46)		(0.12)	(0.15)	
	1.33***	5.06***	-0.93***	1.29***	1.60***	-0.29***
	(0.27)	(0.86)	(0.25)	(0.27)	(0.29)	(0.08)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robust standard errors in parentheses. The table displays the coefficient estimates when regressing the percentage of the land devoted to each crop on the measure of comparative advantage, the transportation costs (measured by the time to reach the closest market centre) and their interaction. In model 1, the comparative advantage is entered as the index, whereas in model 2 quintiles are considered. The time to reach the market is always entered as quintiles.

crops, β_2 has the expected signs. Low value staple crops like beans, sorghum and millet are shown to be more commonly grown in remote areas, while the only "pure" cash crop: coffee, along with high value perishable food crops like sweet potatoes and matoke are less so. The coefficients are insignificant only for cassava and groundnuts. In the first case, this is hardly a surprise, as cassava is typically grown for reasons that are not captured by the model and do not necessarily correlate with market access, such as its resistance to droughts and low labour intensive production methods. As for groundnuts, according to Table 3.1, they are a low value crop mostly grown for consumption, similar to beans and millet across these two dimensions. On these basis, it would be plausible to expect a positive sign. The point estimate is indeed positive but not statistically different from zero. In terms of magnitude, on first inspection the coefficients appear to be quite small ranging from 0.39 for millet to -1.27 for matoke. They are however economically significant if considering the relatively low mean variable of the dependent variable (namely, the percentage of land devoted to each crop).

Adding the interaction term allows me to test whether transportation costs reduce farmers' responsiveness to comparative advantage, as suggested by the model. I find that β_3 is negative in seven out of nine cases, and statistically significant for matoke, beans, groundnut and millet when considering the quintiles. When I use the indexes of comparative advantage as main explanatory variable, the sign of the point estimates is negative only for four out of nine crops, but they are always statistically significant, while reassuringly they never are when they take on a positive value. Overall, β_3 is negative and significant in at least one of the two specifications for five crops and in both for three and it is never positive and statistically significant.

The fact that these findings are not as clear cut as the model suggests is not surprising. Indeed, as long as farmers are flexible in their consumption decisions and more food crops are available, they might opt to produce and consume crops their land is relatively more suitable for and choose not to consume crops that cannot be easily grown given the agro-climatic conditions faced. However, the fact that the estimates are never positive and significant and that are mostly negative in the quintile specification indicates that at least to some extent, the mechanism identified by the model is at play.

In order to provide some additional empirical support to the theoretical framework, I also examine the impact of higher transportation costs on a number of related outcomes such as percentage of output self consumed, land concentration and the estimated revenue of the crop choice as a fraction of the revenue maximising land use. According to the stylised two-crop model, farmers in more remote areas should be more likely to consume (or less likely to sell) the crop they grow, depart more markedly from the revenue maximizing crop mix and have lower values of land concentration.

Table 3.7: Marketization and specialization

	% self consumption		Land concentration		% potential revenue	
	(1)	(2)	(3)	(4)	(5)	(6)
Time to market	1.185***	1.044***	0.445***	0.161	-0.522***	-0.365**
	(0.30)	(0.31)	(0.16)	(0.16)	(0.16)	(0.15)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Market FE	No	Yes	No	Yes	No	Yes
Mean Y	61.29	61.29	30.60	30.60	63.21	63.21
N	2,941	2,941	2,941	2,941	2,941	2,941
adj. R^2	0.097	0.176	0.116	0.272	0.528	0.662

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robust standard errors in parentheses. The controls included are land fertility, average farm size, number of farms and a the (log) of each crop's suitability expressed in expected yields (tonnes per hectare) according to the GAEZ model. All dependent variables are expressed on a scale from 0 to 100. % self consumption refers to the share of the output (value) produced that is self consumed, land concentration is an index computed as the sum of the square of the shares of the land devoted to each crop (Herfindahl index) and % potential revenue represents the ratio between the value of the expected yield and the maximum value obtainable given the soil and climate characteristics (obtained combining LSMS price data to GAEZ expected yields).

I test whether this is the case by estimating a set of simple village level regressions taking this form:

$$\mathrm{Y}_v = \beta_1 \text{time to market}_v + X'_{cv}\gamma + \mu_m \tag{3.16}$$

where X contains the usual controls plus nine variables indicating the village specific suitability for each crop.[30]

The estimates are shown in Table 3.7. As expected, farmers located in more isolated areas are more likely to produce for self-consumption. Although the percentage of the output (where different crops are aggregated using their output value) is generally quite large (more than sixty percent on average), the estimates suggest that the average differential between enumeration areas in the best and in the worst connected quintile is rather sizeable, being close to 5 percentage points. In line with this, the findings also show (columns 5 and 6) that farmers with lower market accessibility tend to depart more from the profit maximising crop mix, suggesting that their production decisions are driven also by other considerations like the perishability of the crops or their value as consumption goods. Both these results are robust to the inclusion of market zone fixed effects.

On the other hand, the results depart from the model's prediction with regard to the land concentration index. Indeed, as shown in Figure 3.1, the higher transportation costs are, the less

[30]This new set of control is especially fundamental in estimating the regression having as dependent variable the percentage of potential revenue as the denominator (maximum potential revenue) is largely dependent on the suitability of crops with high market value (and typically of coffee).

likely farmers are to specialize in one single crop. In particular, when Δ is equal to zero, farmers always fully specialize in the production of the most profitable crop, while for high Δs full specialization occurs only when the comparative advantage for either crop is very pronounced. According to these findings however, specialization (captured by the Herfindahl index of land concentration) increases in more remote villages, although the relation is only statistically significant in the specification without market fixed effects. A possible explanation for this finding is that, as suggested by the high average levels of self consumption, farmers do not fully specialise in cash crops even where the transportation costs are low. Thus, while better connected farmers grow a mix of cash and food crops, those who face higher barrier in accessing markets might only specialise in production of food crops which are most suitable given the agro-climatic conditions they face. This is also in line with the results presented in Table 3.6 which indicates that although transportation costs reduce farmers' responsiveness to comparative advantage, the impact is less pronounced than what the two-crop model suggests.

3.6 Conclusion

Transportation costs are known to have a negative impact on agricultural productivity, especially in developing countries where road networks and infrastructure are lacking. In fact, they are considered to be one of the concurring causes of the disproportionately large scale of the farming sector in poor economies (Adamopoulos 2011 Tombe 2015) and its backwardness in terms of technology and input use (Gollin and Rogerson 2014).

Another way through which transportation costs shape agricultural activities is by affecting farmers' crop choices. Although the related literature is rather varied, two main strands/perspectives can be identified. On the one hand, a number of studies on farm households' decision making (Jayne 1994, Omamo 1998) have looked at the impact of transportation costs in terms of reduction in farmers' market access. These papers typically show that when agricultural markets are missing or are costly to access, farmers tend to deviate from profit maximizing land use and produce potentially lower valued crops for self consumption purposes. On the other hand, some recent works (Costinot and Donaldson 2016, Adamopoulos 2019, Sotelo 2019) have examined how transportation costs lead to scarce integration across local domestic markets, causing inefficiencies in the spatial distribution of crops. In particular, they show that by reducing frictions across regional markets, producers could specialize in growing those crops which are most suited to the agro-climatic conditions they face, increasing the aggregate output.

This paper shows that inefficiencies in cropping patterns do not only occur at the national level (as typically assumed by the literature on domestic market frictions), but also within more narrowly defined market areas. Combining data on farm level land use from the Ugandan Census of Agriculture and the GAEZ agronomic model, I find that aggregate output value could be increased by 33% just by reallocating crops following the underlying structure of comparative advantage, and that more than half of these gains could be achieved just by redistributing crops within 32 domestic markets. In order to explain this, I develop a model where, unlike in Fafchamps, 1992, Jayne, 1994 and Omamo, 1998, farmers are heterogenous not only in their market access, but also in the relative crop productivity. Combining these two dimensions shows that transportation costs also cause deviations from the optimal crop spatial distribution across produces operating within the same market area.

The empirical analysis corroborates this hypothesis. In particular, I show that farmers located in areas with low market access not only devote systematically more land to the production of low value food crops, but also that their cropping choices are less responsive to the agro-climatic conditions they face. These findings suggest that investing in rural infrastructures is as important as reducing frictions across regional markets to improve the efficiency in the geographical distribution of cropping patterns.

Chapter 4

Complements or Substitutes? Mobile phone coverage and education as means of farmers' empowerment

4.1 Introduction

Information (and notably price information) plays a crucial role in guaranteeing a correct functioning of markets (Stigler 1961), which in turn is a key determinant of economic development and growth. The historical lack of widespread networks of Information and Communication Technologies (ICTs) in the developing world has thus been considered a possible explanation for the persistent gap between low and high income countries. This phenomenon is often referred to as digital divide. Its most immediate consequence is scarce market integration, which not only compromises efficiency, reducing the total surplus, but also leaves room for rent extraction by more advantaged agents who can exploit information asymmetries, with obvious detrimental implications in terms of welfare distribution. The most typical and widely studied example is middlemen taking advantage of rural smallholders with no or weak knowledge of realized market prices (Courtois and Subervie 2014).

The recent spread of mobile phone technology is often regarded as a solution to this long lasting problem (The Economist 2009). Indeed, the last two decades have witnessed a rapid expansion of mobile phone coverage and possession throughout the developing world. Such innovation is believed to be particularly valuable in rural and isolated areas, where low population densities and natural obstacles have systematically prevented the creation of landline networks and the long distances as well as the lack of complementary infrastructures that may reduce search costs (e.g. paved roads) make the issue of lack of reliable and timely sources of information more crucial.

Although some macro (Waverman, Meschi, and Fuss 2005) and micro (Klonner and Nolen 2010) level evidence on the impact of mobile telephony services extension in rural areas of developing countries has been provided, the existing studies are generally inconclusive, especially when it comes to the evaluation of the benefits for the poorest of the poor: smallholder farmers (Bhavnani et al. 2008). In particular, a common drawback of the existing literature is that it disregards that the digital divide is a more complex issue than often assumed (Fuchs and Horak 2008), since the physical lack of technology and infrastructure represents only one of the relevant dimensions to be accounted for and therefore the sole provision of modern ICTs might fail to address it.

More specifically, Van Dijk and Hacker, 2003 identify three more components determining the digital divide. Namely: the existence of usage opportunities (contingent on the economic framework) as well as skill and mental barriers to access to modern ICTs (strictly related to the individual specific level of education). According to this view, ICTs are economically unprofitable as long as there is no concrete use for the information (e.g. because of the occurrence of other market failures) or the potential users are unable/unwilling to adopt the newly available technology to search, select, process and apply information. Following this logic, only individuals with a sufficient level of human capital would actually be able to exploit the new opportunities made available by modern ICTs. As stated by Huffman, 2001: "rapid advances and fall prices of communication and information technologies, farm people of the future will need strong basic schooling to adopt and use these technologies as to participate successfully in the new global information system of the 21^{st} century." (abstract). As a result, their provision might actually widen the already existing social and economic gaps within population in rural communities, with ambiguous consequences on poverty reduction.

On the other hand, it might also be the case that those with a lower level of formal human capital are the ones benefitting the most from these new technologies. Indeed, in scenarios dominated by agricultural and informal activities, human capital is believed to affect only the efficiency in decision making process, which can be improved through the provision of information so that the easier availability of the latter might eventually turn out to be more valuable to those with lower schooling; as Welch 1970, p. 51 argued: "Activity of disseminating information can short-cut the gains to education".[1] Ultimately, whether education and mobile telephony are complements or substitutes for farm households is an empirical matter, given that the related theoretical insights are conflicting.

[1]Note that availability of mobile phone coverage does not imply any form of direct provision of information, but only a reduction in its search cost. Moreover, more educated individuals might still have some advantage in implementing these new methods of information search activity.

This paper contributes to the literature by looking at whether the expansion of mobile phone coverage in rural areas had a positive impact on farmers and if so, whether and how the effect depends on the level of schooling of the individuals. In particular, I will focus on the most direct and obvious channel through which modern ICTs can affect rural farmers' welfare in the short term, namely: as a mean to gather information on market prices. In the context under analysis, producers typically interact with traders at the farm gate and have little knowledge about realised market prices. Therefore, accessing price information can improve their bargaining power and in turn obtain higher prices and increase their return to farming. As stated above, whether this benefits to a lerger degree more educated farmers (who are more likely to adopt new technology in a timely manner) is ambiguous as they might *a priori* have better access to information or more bargaining power.

The empirical analysis relies on a panel household survey carried out among rural Ugandan communities composed of three waves (2003, 2005 and 2009). Notably, during the time frame covered by the dataset the country has witnessed a rapid expansion of mobile phone services, which were only available in less than half of the sample in 2003 and were progressively extended to virtually all the villages by 2009.[2].

To preview some of the results, I find empirical support for the "substitutability" hypothesis. While an aggregate analysis suggests a non significant impact of mobile phone rollout on prices obtained by farmers, specifications allowing for heterogeneous effects show that less educated farm households were able to sell their produce at a higher price with a positive impact on agricultural income and welfare measures. This does not seem to be the result of a change in marketing practices as they keep on selling to traders at the farm gate (even more so after mobile communications are made available) but rather the consequence of an improved bargaining position when dealing with traders. Although there is no way to test it empirically given the data available, this suggests that there is enough competition among traders for an easier access to information to be valuable in the negotiation process (Mitchell 2017). The implications of the findings are less clear cut for what concerns the more educated farmers. In fact, the results seem to suggest that some of them might end up receiving lower prices as a result of the rollout of mobile phone coverage. This might suggest that they face more competition from previously less informed sellers when market prices are high.

This chapter is organized as follows: section 4.2 presents review of the existing literature on the impact of mobile phones on development and outline the theoretical framework. Data and descriptive statistics are introduced in section 4.3, while the empirical specification and the

[2]Thanks to the proactive governmental action to foster competition in the telecommunication industry and encourage a widespread extension of mobile phone services, by the end of 2007 coverage was available in 79.3% of the territory and to most of its population. The corresponding figure for the whole Sub-Saharan Africa was of 16.7% only (Buys et al. 2009).

results are shown and commented on in section 4.4. Section 4.5 presents some robustness checks and section 6 concludes with some final remarks.

4.2 Conceptual framework

4.2.1 Mobile phones and the digital divide

The pioneering study by Jensen, 2007 on the impact of mobile phone towers on the efficiency of sardines' markets in the Indian state of Kerala contributed to the spread of enthusiasm and optimism about the poverty reduction and development enhancing potential of this technology. The paper argues that mobile phone coverage, by allowing fishermen to search for price information and perform spatial arbitrage, increased substantially the degree of market efficiency, with positive impact on welfare of both buyers and sellers.[3]

A number of subsequent studies involving grain markets partly corroborated these encouraging findings. Indeed, they found that mobile telephony brought about a significant and long lasting reduction in consumers' (Aker 2008) and producers' (Aker and Fafchamps 2014) price volatility, as a result of an improved market integration. However, in these cases the welfare implications are mixed. Indeed, the results suggest (Aker 2010) that traders, as early adopters of mobile phones, were the only direct beneficiaries of the resulting increased surplus.[4] Indeed, the market structures of agricultural produces, unlike the one studied by Jensen, are typically complex and involve a number of middlemen and traders (Fafchamps and Hill 2008), so that increased efficiency does not necessarily result in welfare improvements for all the participants, and most notably, for producers.[5]

Acknowledging that, some scholars have moved the attention to the direct measurement of the welfare impact of this technology. Klonner and Nolen, 2010 estimated a positive effect of mobile phone coverage rollout on rural South African households' income resulting in a significant reduction in poverty levels. Analogously, Beuermann, McKelvey, and Sotelo, 2012 found that availability of mobile telephony services gradually phased-in in rural Peru increased farm households' expenditures and asset value by 7.5 and 13.5 percent respectively. On the contrary, no

[3]Since fishermen were selling their catch directly in the coastal markets, there were no additional agents (intermediaries) involved.

[4]It has to be noted that risk adverse farmers might still have experienced some welfare improvements due to the reduction in the price variability.

[5]In the case of the studies mentioned, mobile phone coverage was not available to most Nigerien farmers in the time frame considered (2001-2006), so that (at least at that earlier stage) it possibly made the issue of asymmetric information between farmers and middlemen even more severe than before mobile telephony.

impact on households' livelihood was associated with the extension of village phones in Rwanda by Futch and McIntosh, 2009. However, the channels leading to the welfare impact (if any) are not evident in these studies, as well as the extent to which they are actually related to the reduction of search costs.[6] In particular, it is not clear if farmers are actually adopting the newly available technology to search for price information so to increase their bargaining power vis-á-vis traders and/or engage in spatial arbitrage.[7]

A related strand of literature tries to shed some light on that point by assessing whether and in which circumstances the direct provision of price information (often delivered through SMS) is actually beneficial to farmers.[8] These studies typically rely on randomized control trials consisting in informing a randomized subsample of the population and comparing its outcomes with the ones of the virtually identical control groups. Even in this case, the conclusions are rather mixed.

On the one hand, Courtois and Subervie, 2014 and Hildebrandt et al., 2015 found a positive impact of the Esoko program, a mobile phone based Market Information Service (MIS) run in Ghana, on the producers' price for maize, groundnuts and yam, suggesting that price information represents a valuable resource to farmers.[9] A similar experiment was carried out by Nakasone, 2013, who estimated a positive effect of 13 to 14 percentage points on the price received by rural Peruvian farmers provided with timely information on realized market prices. Analogous results were found by Svensson and Yanagizawa, 2009 and Goyal, 2010 for MIS programme delivered through different means. The former evaluated the impact of the radio broadcasting of prevalent market prices in Uganda, estimating a 15 percent increase in maize producer price; while the latter found that the provision of internet kiosks in the Indian state of Madhya Pradesh, which allowed the market prices to be checked, increased by 3 percent the soybeans' price farmers obtained in local markets.

However, similar programmes were found to be ineffective: Camacho and Conover, 2010 showed that providing Columbian farmers with price and weather information did not result in any welfare improvement. Furthermore, Fafchamps and Minten, 2012 found that the MIS named

[6] While Beuermann, McKelvey, and Sotelo, 2012 only considers the impact on welfare, Futch and McIntosh, 2009 identify some changes in farmers' marketing behaviours, that nevertheless did not result in any welfare benefit. Klonner and Nolen, 2010 explain the results through a higher participation of women to labour market, without discussing in depth the role of easier access to information as a possible determinant of that.

[7] Although there are potentially other applications of mobile phones for rural farmers (Aker and Mbiti 2010), the search for price information represents the most obvious business application of mobile telephony in these rural contexts.

[8] In this case, the focus is on the impact of information on farmers' behaviours and welfare, whilst the ability of farmers to use their phones to obtain the information is not assessed as they are only used as a mean to convey the price information.

[9] Curiously, Hildebrandt et al., 2015 estimated inignificant coefficients for groundnuts and maize, unlike Courtois and Subervie, 2014.

Reuters Market Lights, offered to Indian farmers in the state of Maharashtra did not increase the price they received for a range of agricultural products. Consistently, Mitra et al., 2018 showed that the provision of price information to potato farmers in West Bengal (India) did not affect their marketing outcomes, leaving both the quantity traded and the producers' price unaffected.

These contradictory results have led some scholars to question the actual relevance of price information to farmers: Burrell and Oreglia, 2015 ventured to claim that it is nothing but "a myth frequently promulgated by mass media outlets" (abstract). More realistically, the value of (price) information, and in turn of all the technologies guaranteeing a cheaper access to it, is contingent on the economic background rather than absolute. As pointed out by Nakasone, Torero, and Minten, 2014 and Aker and Ksoll, 2016, the occurrence of other market failures can prevent the spread of information and ICTs from achieving the desired goals in terms of efficiency and welfare distribution. Anecdotal evidence show that credit constrained farmers relying on traders for financial support cannot exploit new information flows to improve their bargaining position, given that they are forced to maintain their usual trade partners regardless of the conditions offered (Molony 2008). Similarly, the relevance of price information can be contingent on the context specific market structure: Mitra et al., 2018 demonstrated that collusive behaviour among village traders was a plausible explanation for the lack of impact of the diffusion of price information among farmers. That is consistent with the theoretical model proposed by Mitchell, 2017, who showed that reduction of asymmetric information is beneficial to farmers only for intermediate levels of competition among middlemen.

Thus, the lack of usage opportunities for modern ICTs has been acknowledged as a component of the digital divide by the most up to date economic literature, consistently with Van Dijk and Hacker, 2003. However, most of the studies mentioned are still based on the implicit assumption of homogeneous impact across individuals, overlooking the possibility that information and new telecommunication technologies may have different effects within a given population, in spite of the theoretical insights provided by Jensen, 2007's model in this sense. In particular, the presence (or lack) of impact is typically attributed to the underlying market structure rather than by differences in individual specific characteristics and/or endowments. Therefore, the existing aggregate analyses might mask a great degree of heterogeneity, being only partially informative on the actual impact of information and ICTs on the livelihood of rural smallholders as a result. In the next subsection, I will motivate our interest in the relation between education and mobile phone coverage and explain why it needs to be taken into account when evaluating the impact of mobile telephony services on rural households' welfare.

4.2.2 Education and mobile phone coverage

The main contribution of this paper is to test empirically whether the impact of mobile phone coverage on farmers' welfare depends on the individual specific level of education. More specifically, the channel considered is the potential increase the price obtained when selling agricultural produce. First of all, it is necessary to justify the choice of the extension of mobile telephony as main explanatory variable. As already mentioned in the previous section, this has not always been the case in previous studies, that often revolved around the impact evaluation of market information services (MIS) conceiving mobile phones only as a recipient for information directly provided by some external sources rather than an autonomous tool allowing for an abatement of information search costs. Although these works are of the utmost importance to identify which are the conditions for price information to be valuable, there is a lot of heterogeneity in the availability as well as in the quality of MISs throughout the developing world. Additionally, they often fail to target the category we are more interested in: poor rural households (Cecchini and Scott 2003). Thus, no general conclusions on the true impact on welfare of the "ICTs revolution" in developing countries can be drawn from their analyses.

Mobile phone possession is another alternative variable that could in principle be considered. The two most frequently cited studies adopting such approach are Muto and Yamano, 2009 that, using my same dataset, found that farm households possessing a mobile phone received higher prices for perishable crops and adjusted their production choices accordingly, and Labonne and Chase, 2009 who estimated a significant increase in per capita consumption of rural Filipino households owning a handset. However, recent works have questioned the adoption of mobile phone ownership as the principal regressor both from a technical and a conceptual point of view. Importantly, it has been argued that it is incredibly challenging to disentangle the impact of mobile phossession from other household characteristics.[10] Moreover, in many developing countries, handset possession is neither a necessary nor a sufficient condition for a profitable use of the new technology.[11] Shimamoto, Yamada, and Gummert, 2015 provided evidence that only Cambodian farmers actually using mobile phones (regardless of whether they owned one) or other modern ICTs to gather price information were able to obtain higher prices in their sales. Similarly, Lee and Bellemare, 2013 showed that household level possession of mobile phones by itself was not associated with an increased price for the rice sold by Filipino farmers, since the only effects were found when the handset was owned by the farmer or his spouse, suggesting once again that usage patterns are more relevant than ownership. Interestingly, Tadesse and Bahiigwa, 2015 found no impact whatsoever of mobile phone ownership on Ethiopian farmers'

[10] See Aker and Mbiti, 2010 for a critique to Muto and Yamano, 2009's identifi
cation strategy.

[11] Mobile telephony services are available also to non possessors through village phones or phone sharing, two widespread phenomena in most developing countries, see Burrell and Oreglia, 2015 for the case of Uganda.

marketing habits and outcomes, but a significant increase in the probability of selling cash crops associated to the village specific level of mobile telephony penetration. For this reason, I will consider the extension in the availability of mobile telephony as the treatment of interest.

As argued before, there are no reasons to believe that its impact is homogeneous across the population. In particular, I will allow the welfare impact to depend on households' level of education. There are a number of reasons supporting this choice: first of all, as highlighted by the most recent literature on the digital divide (Fuchs and Horak 2008; Van Dijk and Hacker 2003), the availability of modern ICTs in previously uncovered areas represents only a necessary but not sufficient condition to observe any impact on potential users' livelihood, since skill/mental barriers might prevent less educated sections of the population from using them in a profitable way. This view is in line with the literature on the advantage in technology adoption of educated workers in agricultural scenarios, stating that formal human capital becomes more relevant as new technologies become available Patrick and Kehrberg 1973; Bartel and Lichtenberg 1987 This claim typically refers to innovative production techniques or improved inputs, but can be easily generalized to modern ICTs. According to this hypothesis, only more educated individuals able to adopt and make a profitable use of the technology should experience some economic gain. If this were the case, the often mentioned poverty reducing potential of mobile phone telephony would be questioned, since it could end up widening the pre-existing economic gap, leaving the poorest of the poor further behind rather than making them more connected to the markets. Some studies on adoption and usage of mobile phones in developing countries appear to corroborate this theory: Blumenstock and Eagle, 2012 found that the level of formal human capital was a crucial determinant for both ownership and usage patterns of mobile phones in Rwanda. Specifically, more educated people were not only more likely to possess a handset, but also to use it for business purposes and typically interacted with a wider network. Similarly, Mittal and Mehar, 2016 reported that, in spite of the fact that virtually all farmers possessed a mobile phone in the 120 villages of five Indo-Gangetic states surveyed, only a very small minority of individuals (less than 20 percent) with higher schooling levels, were using them to collect information for agricultural purposes.

These studies suggest a complementarity between formal education and mobile phone coverage. However, these insights are hardly conclusive. First of all, it is worth pointing out that the main scope for this ICTs (when used for strictly economic/business purposes) is to collect information to make more efficient production and/or selling decisions. In a mainly agricultural scenario, that is exactly the same role played by human capital, given the lack of any "worker'seffect", borrowing from the terminology of Welch, 1970.[12] Therefore, the information accessible through

[12] According to this theory, human capital in agriculture would only improve allocative efficiency (the capacity to allocate resources in the most pro
table way) given that there is no way formal capital can make a farmer more efficient in executing a task.

mobile phones might ultimately be more valuable for less educated individuals. Moreover, as argued by Aker and Mbiti, 2010 the skill/mental barriers typically delaying technology adoption by farmers with lower human capital might not represent a big constraint given that using a mobile phone does not necessarily require advanced skills. Finally, it could be the case that less educated people were indirectly advantaged by mobile phone coverage regardless of their adoption/usage patterns thanks to some sort of general equilibrium or spillover effects, even without any explicit information sharing process, such as the ones in Hildebrandt et al. (2015) and Tadesse and Bahiigwa (2015).

Thus, it is also plausible that less educated farmers have benefitted more from mobile phone coverage. Some evidence in this sense was provided by Welch, 1970, who showed how extension services aimed at spreading agricultural information reduced the productivity gap between more and less educated farmers. Additionally, Klonner and Nolen, 2010 and Beuermann, McKelvey, and Sotelo, 2012 found a pro-poor impact on welfare of mobile coverage rollout in rural South Africa and Peru respectively. This suggests that, at least in some instances, the socially and economically disadvantaged were positively affected rather than left further behind by the introduction of this technology.

Ultimately, there is no solid theoretical ground to back either the complementarity or the substitutability hypothesis and the issue is open for empirical investigation. In the following section, I will describe the dataset used in the analysis and the empirical strategy adopted.

4.3 Data and descriptive statistics

The empirical analysis is based on the data from a survey carried out in Uganda between 2003 and 2009 as part of the RePEAT (Research on Poverty, Environment and Agricultural Technology) project, carried out by Makerere University, FASID (Foundation for Advanced Studies on International Development) and GRIPS (National Graduate Institute for Policy Studies). The sample is composed by 10 randomly selected households in each of the 93 rural villages distributed across the Western (23), Central (29) and Eastern (43) region.[13] The villages were chosen to be representative of the rural population of the three regions.[14] The households were interviewed in 2003, 2005 and 2009. Therefore, the initial sample was composed of 930 households, 889 of whom were re-interviewed in 2005 and 816 in 2009. 550 households were added to the original sample in the third wave. These new observations were unevenly distributed across

[13]The original sample comprehended 94 villages, but one was dropped from the analysis due to the high attrition rate (no households were interviewed in the third wave).

[14]The survey was originally intended to cover the whole country, but the Northern region was eventually excluded.

the villages sampled. As a result, in the last wave there are 8 to 21 household per village, with a mean of 15. The survey collected information on the economic activities of the households as well as on the characteristics of the villages.

Crucially, the dataset indicates, for each village, the year when mobile phone coverage was first made available.[15] As shown in Table 4.1, at the time of the baseline survey only less than 50 percent of the villages were covered, the proportion increased to 80 percent by 2005 and in 2009 virtually all the location were linked to the mobile telephony network.[16]

Table 4.1: Mobile phone coverage extension by region

Region	**2003**	**2005**	**2009**
Western (23)	11	17	22
Central (29)	15	26	29
Eastern (41)	16	32	41
Total (93)	**42**	**75**	**92**

As anticipated in the introduction, this paper aims at estimating the impact of the availability of mobile phone coverage on the prices farmers are able to negotiate with the traders and consequently on farmers' welfare. Table 4.2 presents some descriptive statistics of these variables by year and availability of coverage. Unsurprisingly, both income and wealth per capita are significantly higher (on average) in villages where coverage is available.[17] This is more likely to reflect the fact that coverage expansion was market driven and realistically targeted at an earlier stage more developed/wealthier areas. The difference is particularly pronounced when considering wealth (computed as the sum of the value of the assets possessed by each household), while it is less evident and only statistically significant in 2005 when looking at the households' income (measured as the difference of monetary revenues and costs of farm and off farm activities).[18] On the contrary, prices obtained by farmers when selling their crops seem not to differ systematically between villages with and without mobile phones when considering the most frequently traded

[15] For the purpose of the analysis, I will consider mobile telephony services to be available when they have been available for at least one year in the village (e.g., if mobile phone coverage was available in the village since 2002 for the first wave of the survey (2003). The rationale behind this choice is that most of the dependent variables refer to the year previous the survey and it would be unrealistic to expect any impact of the coverage in the aftermath of its rollout (Beuermann, McKelvey, and Sotelo 2012).

[16] When talking about mobile phone coverage, I am referring to the Global System for Mobile Communications (GSM) which is the second generation digital cellular network, allowing to send text messages and make phone calls.

[17] In 2009 the statistics are only presented at the aggregate level as coverage was available in all but one village (see Table 4.1).

[18] Admittedly, these welfare measures might fail to capture perfectly the households' welfare. A particular concerns derive from the fact that agricultural activities are often meant as sources of goods for internal consumption and therefore would not be captured by considering monetary revenues only. Nevertheless, as the channel through which mobile coverage is expected to affect income is purely monetary (higher prices received for the crops sold), any alternative approach based on median prices to aggregate self consumed and sold output would be problematic.

Table 4.2: Descriptive statistics by year and coverage (dependent variables)

	2003			**2005**			**2009**
	Coverage	No Coverage	Diff	Coverage	No Coverage	Diff	
Welfare measures							
Income	11.51	11.30	0.21*	11.62	11.47	0.15	12.44
	(1.68)	(1.60)	(0.11)	(1.51)	(1.65)	(0.14)	(0.84)
Wealth	10.32	10.07	0.25***	10.32	10.13	0.19**	10.60
	(1.06)	(1.12)	(0.07)	(1.17)	(0.97)	(0.10)	(1.18)
Prices							
Maize	180.72	180.17	0.54	168.87	176.81	-7.94	464.98
	(3.51)	(3.97)	(5.29)	(2.22)	(5.60)	(5.07)	(119.19)
Coffee	461.27	572.22	-110.96***	481.77	513.80	-32.03	1220.68
	(25.56)	(33.14)	(41.23)	(23.20)	(50.57)	(58.16)	(850.38)
Beans	320.96	312.43	8.54	373.72	361.84	11.87	904.42
	(7.44)	(8.11)	(10.99)	(6.22)	(9.21)	(11.59)	(279.46)
Matoke	104.91	111.69	-6.78	121.32	108.37	12.96	410.99
	(4.43)	(7.23)	(9.10)	(4.39)	(6.66)	(7.89)	(213.95)
Cassava	112.65	114.31	-1.65	176.73	157.41	19.32	282.92
	(10.35)	(8.57)	(13.41)	(7.27)	(15.38)	(16.45)	(161.60)

The figures refer to the year and coverage specific averages. The welfare measure presented are the logarithm of income and wealth per capita, where figures are expressed in 2003 Ugandad Shillings using the deflactors provided by the Ugandan Bureau of Statistics. Standard deviations are in parentheses. The stars represent the significance level (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$) of the difference between the sample with and without coverage the against the null hypothesis of zero difference. The differences are not presented for 2009 since coverage was available in 92 out of the 93 villages.

produce. The only instance in which the difference is significant (coffee in 2003) appears to be mostly incidental.[19]

As discussed more in depth in the next section, in order to control for other confounding factors (and to identify potential heterogeneous effects in the case of education), a number of other variables defined in the dataset, both at the village and at the household level) are considered. Their means are presented by year and coverage availability in Table 4.3. In this case, the household level variables do not differ significantly between community with and without mobile telephony. At the baseline, the wide majority of households were headed by a male, with on average 6 years of education and 45 years old. The average household was composed by 8 members, half of whom in working age (between 15 and 65 years old). The main source of monetary income was agriculture.[20] These characteristics remain rather costant throughout the three waves, excluding a slight reduction in the relative importance of agricultural income over time.

[19] The figures in the table are obtained as a weighted average of the price received by each households for the crops sold throughout the each year.

[20] As these figures do not include self-consumed output, they are likely to underestimate the actual contribution of farm activities.

Table 4.3: Descriptive statistics by year and coverage (independent variables)

	2003			2005			2009
	Coverage	No Coverage	Diff	Coverage	No Coverage	Diff	
Household level							
Education	5.77	5.56	0.21	5.62	5.71	-0.09	6.12
	(3.86)	(3.84)	(0.25)	(3.76)	(4.16)	(0.33)	(3.99)
Age	45.76	43.85	1.91*	47.30	46.21	1.09	47.71
	(14.75)	(14.98)	(0.98)	(14.79)	(16.19)	(1.28)	(14.62)
Fem. head	0.13	0.10	0.03	0.12	0.12	0.00	0.13
	(0.33)	(0.30)	(0.02)	(0.33)	(0.33)	(0.03)	(0.34)
Hh size	7.97	7.59	0.38	8.14	8.44	-0.29	8.09
	(4.27)	(4.35)	(0.28)	(4.04)	(5.36)	(0.37)	(4.00)
% wrk. age	0.51	0.50	0.01	0.49	0.46	0.03*	0.52
	(0.19)	(0.20)	(0.01)	(0.19)	(0.18)	(0.02)	(0.20)
% agric. income	0.69	0.69	0.00	0.62	0.66	-0.04	0.62
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)
Village level							
Time to road (dry)	39.3	63.3	-24.0**	38.2	84.7	-46.6***	38.9
	(37.71)	(61.21)	(10.95)	(25.51)	(112.87)	(14.23)	(33.38)
Time to road (rainy)	57.6	82.3	-24.7**	58.7	93.8	-35.0**	60.49
	(50.48)	(61.73)	(12.09)	(46.16)	(82.31)	(14.99)	(60.52)
Electricity (1 = yes)	0.24	0.04	0.20***	0.20	0.00	0.20**	0.25
	(0.43)	(0.20)	(0.07)	(0.40)	(0.00)	(0.10)	(0.39)

The figures refer to the year and coverage specific averages. The stars represent the significance level (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$) of the difference between the sample with and without coverage the against the null hypothesis of zero difference. The differences are not presented for 2009 since coverage was available in 92 out of the 93 villages. The "Time to road" variables refer to the time in minutes it takes to reach an accessible road in the rainy or in the dry season.

On the other hand, the comparison of the village level variables seems to corroborate the insight that the extension of mobile telephony services did not follow a random path. In particular, it is clear that better accessible villages were targeted first (as reflected by the statistically significant difference in access to roads). This might be the result of either supply (easier to provide the service) or demand (more economic activity) factors, or more realistically a mix of both. The fact that villages covered were more likely to have access to electricity both in 2003 and 2005 corroborates this view further. A more in depth analysis of the determinants of mobile phone coverage is provided in the Appendix B.

4.4 Empirical specification and results

As argued in the previous sections, the most immediate channel through which mobile coverage availability could affect rural farmers' welfare is by guaranteeing access to agricultural prices. This can be particularly crucial in a context like the one under analysis, where farmers usually sell their produce at their farm gates to more informed traders who might exploit the asymmetric information to their advantage.[21] Indeed, Jaleta and Gardebroek, 2007 and Woldie and Nuppenau, 2008 showed that in similar scenarios, farmers who are more aware of the market prices generally manage to negotiate better conditions when bargaining with traders. The role played by education in this bargaining process less clear cut in these works. In fact, the coefficients of the formal schooling variables are positive yet statistically insignificant. However, it is worth to point out that these regressions included self reported levels of knowledge of market prices (that are positive and significant), therefore as long as the main impact of education is through a better awareness of prices and market dynamics, it is hardly surprising that the relative coefficients are not significant.

In the specification I adopt, education enters both linearly and interacted with the dummy variable indicating whether mobile phone coverage is available. This allows to assess whether higher human capital allows to negotiate higher prices both when mobile telephony is not available (linear term) and when it is (linear term plus interaction). In particular, the sign and significance of the interaction term will signal whether the availability of mobile phone coverage acts as a substitute (negative sign) or complement (positive sign) of formal education.

The way the price variable is obtained and computed deserves some attention.[22] In particular, in this first regression, prices are defined at the household level only, whereas farmers realistically

[21] As explained more extensively in the following, information on where sales took place is only available for 2005 and 2009. In these two years, more than 85 percent of the transactions happened at the farm gate and the buyers were professional traders who were not contacted or known in advance by the farmers.

[22] I will only estimate the regression for the price of crops that were sold by at least 100 households in each of the three waves. This excludes all crops but maize, coffee, beans, matoke (cooking bananas) and cassava. Having too few observations would jeopardise the credibility of the estimates and generally reduce the power of the regression. The results for other crops are however never significant and are available upon request.

engage in more transactions throughout each year, potentially obtaining different prices that might in turn be affected by the time of sale. Unfortunately, due to the structure of the survey, it is not possible to obtain information on each of these transactions. In fact, variables in the section on crop production and sales are defined at the plot level only. Therefore, only the average price obtained for all the transactions of a crop grown in a given plot are reported. This implies that there can be more than one price value for a crop only when it was grown by an household in two or more separate plots (and if it the produce from more than one plot was also sold) and they do not necessarily refer to a single transaction. Additionally, information on the mode of the sales are only available for the last two waves (2005 and 2009), where for each crop plot pair the identity of the buyer (trader, direct customer or friend/relative), the way he/she was contacted and the season and the location where the most important transaction took place are specified. In order to be able to include all the three years available, in the first specification I only consider the weighted average (where the weights considered are the quantity of crop sold for each plot) prices obtained for each of the five major crops as a dependent variable.[23]

More formally, the estimated regression takes the form:

$$\begin{aligned}\log(\text{Price})_{ivt} = \beta_1 \text{Coverage}_{vt} + \beta_2 \text{Coverage}_{vt} * \text{Education}_{ivt} + \beta_3 \text{Education}_{ivt} \\ + X'_{ivt}\gamma + W'_{vt}\delta + \nu_v + \tau_t\end{aligned}$$

where the dependent variable is regressed on a (village specific) dummy taking value 1 when mobile phone coverage is available and its interaction with the household level of education (captured by the years of schooling completed by the head). It has already been pointed out that the expansion of coverage has not followed a random pattern. Therefore, a simple comparison between covered and not covered areas would realistically capture also the structural differences between villages receiving earlier and later coverage. In order to control for that, village level fixed effects ν are included. As a result, the effect will be only identified through variation within villages over time.[24] However, fixed effects are not enough to address all concerns related to identification. Indeed, some unobserved shocks might have occurred during the considered time span, affecting both infrastructural development and prices (and/or welfare measures, which will be looked at in the following) and possibly biasing our estimates. In order to take this possibility into account, I include also time varying village level controls (labelled as W in the following equations) aimed at describing local communities' development level (transport network and electrification) to control for infrastructural enhancements.[25] Finally, a vector of household

[23] Since prices are self reported and are expressed in different units (such as bags, heap etc.) there is some room for measurement errors. In order to eliminate the most obvious outliers, I dropped from the sample prices whose z-score were higher than 2. This eliminates around 3 percent of the sample for each of the crops. However, the results are robust when adopting more conservative trimming strategies or when I keep all the observations.

[24] Conveniently, as shown in Table 4.1, coverage has been extended to several villages that were previously not supplied in the sample.

[25] In particular, the variables included are those presented in Table 4.3, namely the time to reach the closest

Table 4.4: Prices received for selected crops

	Maize		Coffee		Beans		Matoke		Cassava	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Coverage	-0.018	0.058	0.019	0.099	-0.042	0.128**	0.085	0.289***	0.013	0.420*
	(0.03)	(0.06)	(0.08)	(0.18)	(0.04)	(0.06)	(0.07)	(0.10)	(0.13)	(0.24)
Coverage * Education		-0.009		-0.009		-0.019***		-0.033***		-0.047**
		(0.01)		(0.02)		(0.01)		(0.01)		(0.02)
Education	0.005*	0.012**	0.003	0.011	0.005	0.019***	0.012**	0.026**	0.005	0.043**
	(0.00)	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,258	1,258	1,131	1,131	1,052	1,052	1,011	1,011	384	384
Clusters	92	92	74	74	86	86	88	88	80	80
adj. R^2	0.741	0.741	0.498	0.497	0.717	0.719	0.676	0.675	0.391	0.397
p-value $\beta_2 = -\beta_3$		0.28		0.71		0.97		0.29		0.74

Standard errors clustered at the village level in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The household level controls included are: age and gender of the household head, percentage of income derived from agriculture, number of components of the household and dependency ratio. The village level controls included are the time in minutes to a practicable road in the dry and in the rainy season and whether the households have access to electricity. Prices are defined as the weighted average of the prices received throughout the year by an household (where weights are the quantity of output sold). Only crops with at least 100 household/year observations are included.

level covariates X is included to capture the effect of other confounding factors such as the level of dependency on agricultural income (fraction of income from agricultural activities) and households' demographics.

Table 4.4 shows the results of this regression for the five most widely sold produce in the sample. The first interesting result reported is the apparent lack of direct impact of mobile phone coverage rollout on the prices obtained by farmers for their agricultural output when no interaction effects are considered (columns 1,3,5,7 and 9). In fact, the coefficient β_1 is never statistically significant in these instances, and it takes o negative values for maize and beans. Also, the impact of education seems to be rather negligible, since higher schooling appears to be correlated with higher prices only in the case of maize (where the coefficient is significant only at the 10 percent level) and matoke. However, the relative coefficients are always estimated to be positive, suggesting that generally farmers with better endoements of human capital tend to be able to negotiate more favourable conditions when selling their output. Unfortunately, given the data available it is not possible to determine exactly the channel through which this effect takes place. Given the prevalence of farm gate transactions (according to the data available for the last two waves only), this is likely to be related to the bargaining power of the farmers which is realistically a

road in the dry and in the rainy season and whether or not electrification was available in the village. This does not only allow to disentangle the effect of mobile phone coverage extension from these improvements in the infrastructural endowment, but also to at least partially control for some other general idiosyncratic shocks that might drive changes in the dependent variables considered (prices and welfare measures) and in the provision of infrastructures.

positive function of their level of schooling. This might underline a better awareness of market dynamics and prices or just the ability of these households (due to lower credit constraints) to sell when prices are higher.

The introduction of the interaction effect has a relevant impact on the estimates. First of all, the related coefficient β_2 always takes on negative values. Accordingly, the estimates of the impact of the coverage and of formal education on prices (β_1 and β_3, respectively) increase and turn positive and significant in the case of beans, matoke and cassava, which are also the three crops for which the interaction term is significant. These findings indicate that (at least when considering perishable and semi perishable crops), mobile phone coverage allows less educated individuals to negotiate higher prices for their produce. More specifically, when mobile telephony is not available, there seems to be a significant and sizeable "education premium" for farmers in the negotiation process, with estimates ranging from a price increase of 1.9 for beans to 4.3 percent per completed year of schooling for cassava. According to these results, the introduction of mobile phone coverage eliminates this advantage. Formally, for every crop, I fail to reject the hypothesis that the interaction term has the same magnitude (with opposite sign) of the impact of education. Therefore, the introduction of this technology was only beneficial to a subsample of the population, namely, those with less than 7 to 9 years of schooling, for whom the positive impact of the coverage was not offset by the cancellation of the education premium.[26]

This might reflect the fact that easier access to information was mostly beneficial to those who were previously less aware of market prices and/or having lower bargaining power when dealing with traders. However, this had potentially a negative impact on farmers who already had the means to inform themselves about market prices/dynamics and had some advantage in negotiating sales deal with traders by virtue of their higher human capital endowments. This can be the result of the fact that they are now facing the competition of other farmers when selling crops when prices are high.

In order to corroborate these findings further, I estimate a similar regression at the sale level, pooling together all the crops c and including among the controls a number of transaction specific controls η. Unfortunately, transaction specific variables are only available for the last two waves of the survey and do actually refer to the most important sale of each crop/plot pair rather than to actual single sales. The estimated regression takes the form:

$$\begin{aligned}\log(\text{Price})_{ivt} &= \beta_1 \text{Coverage}_{vt} + \beta_2 \text{Coverage}_{vt} * \text{Education}_{ivt} + \beta_3 \text{Education}_{ivt} \\ &\quad + Z'_{sivt}\eta + X'_{ivt}\gamma + W'_{vt}\delta + \nu_v + \tau_t + \kappa_c\end{aligned}$$

While the variables included in vectors X and W are the same as in the previous specification, η includes a number of covariates that are defined at the transaction level. Namely, they indicate

[26]The exact cut off can be estimated as $-\frac{\beta_1}{\beta_2}$ and ranges from 7 for beans to 9 for matoke and cassava.

Table 4.5: Transaction level analysis

	Price received all transactions		Price received farm gate transactions		Selling at farm gate	
	(1)	(2)	(3)	(4)	(5)	(6)
Coverage	0.040	0.113**	0.042	0.110**	-0.011	0.135***
	(0.03)	(0.06)	(0.03)	(0.04)	(0.02)	(0.03)
Coverage * Education		-0.008*		-0.007		-0.016***
		(0.01)		(0.01)		(0.00)
Education	0.008***	0.015***	0.007***	0.013**	0.002**	0.016***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Crop FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	9,235	9,235	8,021	8,021	9,235	9,235
Clusters	93	93	93	93		
adj. R^2	0.689	0.689	0.685	0.685	0.194	0.196
p-value $\beta_2 = -\beta_3$		0.00		0.00		0.92

Standard errors clustered at the village level in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The household level controls included are: age and gender of the household head, percentage of income derived from agriculture, number of components of the household and dependency ratio. The village level controls included are the time in minutes to a practicable road in the dry and in the rainy season and whether the households have access to electricity. The transaction level controls are the quantity of crop sold (log), whether it happened in the rainy or in the dry season, the identity of the buyer, and whether it was a farm gate transaction (columns 1 and 2 only). Prices are defined as the price per kilo net of transportation cost obtained in each transaction.

the (log) quantity of crop traded, the season when the sale happened (dry or rainy), where the transaction took place (farm gate versus local market) and the identity of the buyer (whether he or she was known in advance or an unknown trader).[27] Unlike the previous case, all crops will be included in this regression. In order to control for systematic differences in price among them, crop specific fixed effects κ are added.

As shown in Table 4.5, this specification returns very similar findings, both for the whole sample and when considering only the wide majority of the transactions that happened at the farm gate (columns 1-2 and 3-4 respectively). In particular, when the interaction effect is not included, coverage is estimated to have a positive yet non significant impact on prices received by farmers, while the education premium is quantifiable as a 0.8 (0.7 in the case of farm gate sales only)

[27] Unfortunately, the data on the quantity refers to the total produce of the crop from a given plot's production rather than the most important transaction (as the other variables do).

percent increase in price per year of schooling. This show that by virtue of their higher human capital, more educated farmers manage to obtain better conditions when selling their crops. Although it is not possible to identify a specific channel through which education impacts prices, the fact that in this case transaction specific controls are included, it can be ruled out that this is due to different amount sold or timing of the sale (at least as far as differences in season are concerned).[28] Realistically, this is driven by a higher bargaining power when dealing with trader, arguably a result of better knowledge of market prices. Indeed, the estimated impact is very similar when only looking at the farm gate transactions.

Introducing the interaction term (columns 2 and 4) once again makes the difference. In fact, in line with the previous results presented, it is estimated to be negative (although significant only when looking at the full sample). This in turn increases the estimates of β_1 which turns statistically significant in both specifications (columns 2 for the whole sample and column 4 for farm gate transactions only), suggesting a positive impact of coverage on farmers' prices, although decreasing in the level of education. Specifically, the estimates suggest a price increase of 12 percent for farmers with no formal education whatsoever.

At the same time, the impact of education is shown to differ depending on whether mobile telephony services are available or not. More specifically, when they are not, the estimated premium to an additional year of education is of a price increase of around 1.5 percent, that is reduced to 0.7 percent when coverage is made available. Interestingly, unlike the results presented in Table 4.4, in this case the positive impact of education on prices is only reduced rather than completely eliminated when mobile coverage is extended to a village, as formally shown by the rejection of the null hypothesis of $\beta_2 = -\beta_3$.

This also implies that the positive impact of mobile phone on price is only offset by the reduction of the returns to education for farmers with over 14 years of schooling, that represent a mere 5 percent of the sample under analysis. This finding suggests that an easier access to information benefits lesser educated individuals (who would otherwise be poorly informed about market prices) without necessarily being harmful to more schooled farmers, who just experience a lower price increase or, for very high levels of education, no benefit at all since they could already negotiate good prices to begin with by means of either a higher bargaining ability or better knowledge of market prices.[29]

[28] It cannot be excluded that they are able to sell when prices are higher during a specific season, but this is more likely to reflect better price awareness than the ability to sell when prices are systematically higher throughout the year being less credit constrained.

[29] This is partially in contrast with the results presented in Table 4.4, which suggest a negative impact for a number of educated individuals. While this can be interpreted as an increase in competition from other farmers when prices are high, the possibility of a negative impact might just be driven by imprecise estimates of the coefficients. This will be further discussed in the robustness checks section.

The equivalence of the estimates obtained for all the observations (columns 1 and 2) and the ones referring to the subsample of the farm gate transactions (columns 3 and 4) suggests that the findings are driven by an enhancement in the bargaining position of more informed farmers when dealing with traders at the farm gate.

In order to see whether and how mobile phone technology has affected farmers' marketing behaviours, I estimate a linear probability model where the dependent variable takes value 1 when the transaction happens at the farm gate.[30] The resulting estimates are shown in columns 5 and 6 of Table 4.5. In line with the other regression, I allow for mobile phone coverage to have different impact for farmers with different levels of human capital.

The findings corroborate the thesis that having easier access to information allows less educated farmers to have more bargaining power in negotiations with traders at the farm gate. In particular, when a village is disconnected to the mobile network, more educated farmers are shown to be more likely to sell their crops at the farm gate (1.6 percentage point more likely for each additional year of schooling), arguably because they suffer less from asymmetric information when dealing with traders, whereas less informed farmers might prefer to reach a market centre to sell their output avoiding the interaction with intermediaries. The introduction of mobile phone coverage eliminates this difference.

Finally, re-estimating the regression separately for a number of sub set of crops sharing similar characteristics does not lead to any new insight. The results are shown in Table 4.6 and generally mirror the ones presented in Table 4.5. In particular, with the sole exception of fruit and vegetables, the point estimate of the interaction term is negative, and significant at least at the 90 percent level for cereals, tubers and coffee, indicating substitutability of mobile phone coverage and human capital. Correspondingly, education is shown to have an impact on prices just before mobile phone telephony was introduced (i.e. $-\beta_3 = \beta_2$) in the case of tubers and coffee, whereas there is evidence that the education premium is only reduced in the case of cereals.

The findings presented so far suggest that mobile phone coverage has allowed farmers to obtain price information more easily, improving their bargaining position when dealing with traders and the conditions negotiated when selling their crops. However, not all farmers benefitted equally from this technological change. In particular, only less educated farmers are shown to obtain higher prices as a result, possibly because those with better human capital had other ways to obtain timely price information or enjoyed some advantage in bargaining with traders at the farm gate.

[30] There are potentially other outcomes of interest such as the identity of the buyer, but there is very little variation in that dimension within transaction at the farm gate and at the market.

Table 4.6: Transaction level analysis (by crop group)

	Pulses		Cereals		Tubers		Coffee		Fruit & Veg	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Coverage	-0.038	0.049	-0.013	0.095*	0.131	0.355**	-0.028	0.345*	0.029	-0.029
	(0.04)	(0.07)	(0.03)	(0.06)	(0.12)	(0.18)	(0.09)	(0.20)	(0.06)	(0.11)
Coverage * Education		-0.010		-0.013**		-0.027*		-0.038*		0.006
		(0.01)		(0.01)		(0.02)		(0.02)		(0.01)
Education	0.003	0.011*	0.007***	0.018***	0.009	0.033**	0.007	0.043**	0.005	0.000
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)	(0.00)	(0.02)	(0.00)	(0.01)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crop FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,850	1,850	2,031	2,031	773	773	1,896	1,896	2,687	2,687
Clusters	92	92	93	93	89	89	76	76	90	90
adj. R^2	0.723	0.723	0.757	0.757	0.398	0.399	0.577	0.577	0.565	0.565
p-value $\beta_2 = -\beta_3$		0.47		0.02		0.36		0.33		0.17

Standard errors clustered at the village level in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The household level controls included are: age and gender of the household head, percentage of income derived from agriculture, number of components of the household and dependency ratio. The village level controls included are the time in minutes to a practicable road in the dry and in the rainy season and whether the households have access to electricity. The transaction level controls are the quantity of crop sold (log), whether it happened in the rainy or in the dry season, the identity of the buyer, and whether it was a farm gate transaction (columns 1 and 2 only). Prices are defined as the price per kilo net of transportation cost obtained in each transaction.

As it has already been pointed out, these findings do not provide any strong evidence on the specific dimension which generates the observed education premium on crops' prices negotiated at the farm gate where mobile phone coverage is not available. In particular, formal education can correlate with several factors affecting the final price negotiated by farmers (Jaleta and Gardebroek 2007; Woldie and Nuppenau 2008) like a better knowledge of market prices thanks to a more widespread personal network, better information processing skills and knowledge of market dynamics and more refined bargaining skills. Unfortunately, the available data does not allow to disentangle these components, so I am unable to identify which is the precise component of formal education leading to the price premium. However, it is clear that the premium is greatly reduced (if not completely eliminated) by the introduction of mobile phone services. This suggests that modern ICTs, by making (price) information more readily available, disproportionally benefitted less educated individuals who were previously more penalised in the farm gate negotiations with traders. This seems to be corroborated by the findings in presented in Table 4.6 (columns 5 and 6), showing that less educated farmers were more likely to sell their produce at the farm gate once mobile phones are made available in their villages.

The obvious question that arises is whether this led to an improvement of farmers' welfare. Intuitively, as households depend heavily on agriculture (see Table 4.2), an increase in the price received for their crops might result in a substantial enhancement in their living conditions.

Unfortunately, given the structure of the data, it is very difficult to obtain consistent welfare measures throughout the three waves.[31] However, it is possible to compute two proxies per household: namely monetary income and wealth. In particular, monetary income is computed as the difference between monetary inflow and outflow in the economic activities performed by the households.[32] Wealth is instead defined as the sum of the self reported value of a number of assets listed in the survey.[33]

In order to study the impact on these welfare measures, I estimate the following regression, where Y are the logs of the income and wealth per capita respectively:

$$\begin{aligned} Y_{ivt} = {} & \beta_1 \text{Coverage}_{vt} + \beta_2 \text{Coverage}_{vt} * \text{Education}_{ivt} + \beta_3 \text{Education}_{ivt} \\ & + X'_{ivt}\gamma + W'_{vt}\delta + \nu_v + \tau_t \end{aligned}$$

where the controls included in X are the same as in the previous household level regressions.

The resulting findings are shown in Table 4.7 and are rather consistent with the results obtained so far. In particular, when the interaction term is not introduced, the estimated impact on these welfare measures is not statistically significant (columns 1 and 3). The addition of the interaction term to the model shows that less educated households experienced a statistically and economically significant increase in their monetary income and wealth. Unsurprisingly, households with higher levels of formal schooling are shown to perform better even after the introduction of mobile technology, although the schooling premium is reduced from 11 to 7 percent per year for income and 14 to 11 percent for wealth. Although I am not able to rule out other channels, given the central role played by agricultural production, the results are likely to be driven by the increase in the prices received for crops.

The magnitude of the impact, especially for lesser educated farmers is quite large. Namely, I estimate a 24 percent increase in income for households whose head has no formal schooling, and

[31] In particular, the items in the expenditure section change from year to year, and information on self consumption is only available in the third and in the fourth wave. Also, the recalling time varies substantially across years. Therefore, measures on consumption and expenditures would be inherently flawed.

[32] Three main activities are considered: agricultural production, milk production and other non-agricultural business and wage activities. I consider only monetary inflow and outflow in order to capture the impact of systematic changes in prices received for output (or paid for inputs). Ideally, I would like to compute also a measure of agricultural income. However, in order for it to be meaningful, I would need information on the input used in crop production, crucially: hours worked and land. Unfortunately, data on hours of labour is only available in the second wave. Therefore, there would be no way to disentangle the effect of reduced input usage, perhaps due to a switching to more profitable, non-agricultural activities. For this reason, I focus on total income only and control for household components and dependency ratio. However, as shown in Table 4.2, most of the income comes from crop production.

[33] In this context, it is possible to expect an impact on wealth as most of the items considered are non-durables with low unit value (e.g. spades, bicycles, watering cans, wheelbarrows) and are likely to be purchased in greater measure as a result of a positive income shock. Indeed, Diga, 2008 and Beuermann, McKelvey, and Sotelo, 2012 found a positive impact of mobile phone coverage on similarly measured wealth indexes in Uganda and Chile respectively.

Table 4.7: Impact on welfare

	Income		Wealth	
	(1)	(2)	(3)	(4)
Coverage	0.027	0.322**	0.012	0.245**
	(0.11)	(0.17)	(0.07)	(0.12)
Coverage * Education		-0.040**		-0.028**
		(0.02)		(0.01)
Education	0.075***	0.107***	0.109***	0.131***
	(0.01)	(0.02)	(0.01)	(0.01)
N	2,856	2,856	3,183	3,183
Clusters	93	93	93	93
adj. R^2	0.170	0.171	0.263	0.264
p-value $\beta_2 = -\beta_3$		0.00		0.00

Standard errors clustered at the village level in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The household level controls included are: age and gender of the household head, percentage of income derived from agriculture, number of components of the household and dependency ratio. The village level controls included are the time in minutes to a practicable road in the dry and in the rainy season and whether the households have access to electricity.

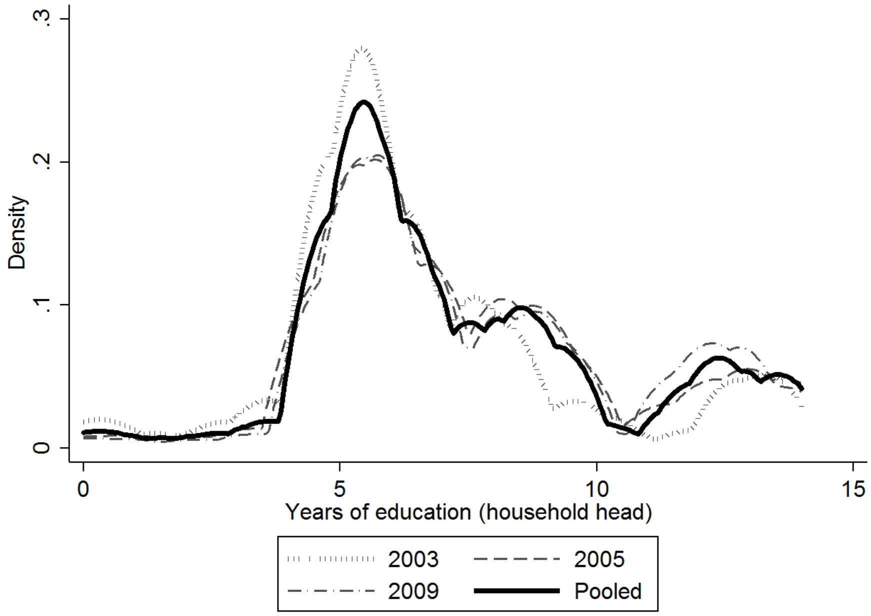
a 18 percent increase in wealth. They are however in line with the qualitative and quantitative findings of Diga, 2008.

4.5 Robustness checks

In the results presented so far, education was entered linearly in the regression models, which imposes the assumption of constant (percentage) impact of each year of education on the variables of interest.[34] Although the resulting findings are informative, it might be useful estimate a more flexible model with less restrictive assumptions on the impact of education and its interaction with mobile phone coverage availability.

[34] This is because the dependent variables are entered in the log form in the regression models.

Figure 4.1: Distribution of years of education



More specifically, I will estimate a number of regressions in the following form:

$$\begin{aligned}\log Y_{ivt} = \beta_1 \text{Coverage}_{vt} + \sum_{j=2}^{4} \beta_j \text{Coverage}_{vt} * \text{Education quartile}_{ivt} \\ + \sum_{k=5}^{8} \beta_k \text{Education quartile}_{ivt} + X'_{ivt}\gamma + W'_{vt}\delta + \nu_v + \tau_t\end{aligned} \tag{4.1}$$

where the parametres β_1 to β_8 will capture the different impact of belonging to any quartile in the distribution of household head's education on the variables of interest both before and after the introduction of mobile phone services. This specification allows for more flexibility as both the effect of education and the interaction terms are estimated separately for each quartile rather than assumed to be constant.

As shown in Figure 4.1, the levels of education have been increasing between 2003 and 2009 but the overall distribution has not changed dramatically, therefore, I will compute quartiles on the basis of the pooled distribution. The thresholds I obtain by doing so are of 5, 6 and 9 years of education. This reflects the fact that most observations are concentrated around the mean.[35]

I estimate such regressions using both prices and welfare measures as dependent variables. The

[35] The resulting categories are not all equal size as 32% of the sample belongs to the first quartile, 21% to the second, 27% to the third and 20% to the fourth.

Table 4.8: Impact on prices (quartiles)

	Maize		Coffee		Beans		Matoke		Cassava	
Coverage	-0.017	0.063	0.022	0.103	-0.003	0.133**	0.091	0.302***	0.018	0.392
	(0.03)	(0.06)	(0.07)	(0.19)	(0.02)	(0.05)	(0.07)	(0.09)	(0.02)	(0.26)
Interaction terms										
2^{nd} quartile		-0.021		-0.031		-0.026		0.02		-0.082
		(0.02)		(0.02)		(0.02)		(0.03)		(0.07)
3^{rd} quartile		-0.023		-0.043		-0.034*		-0.016		-0.076
		(0.02)		(0.03)		(0.02)		(0.02)		(0.08)
4^{th} quartile		-0.085*		-0.091		-0.171***		-0.461***		-0.527*
		(0.05)		(0.06)		(0.07)		(0.12)		(0.33)
Schooling quartiles										
2^{nd} quartile	0.032	0.041	0.030	0.036	-0.002	0.013	0.031	0.031	0.058	0.095
	(0.02)	(0.02)	(0.04)	(0.04)	(0.01)	(0.01)	(0.03)	(0.03)	(0.05)	(0.06)
3^{rd} quartile	0.026	0.045*	0.021	0.034	0.034	0.045*	0.064**	0.072***	0.084	0.097
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.07)	(0.07)
4^{th} quartile	0.048*	0.072**	0.043	0.071	0.048*	0.075***	0.116**	0.145***	0.175*	0.236**
	(0.03)	(0.04)	(0.03)	(0.06)	(0.03)	(0.02)	(0.04)	(0.04)	(0.08)	(0.11)
N	1,258	1,258	1,131	1,131	1,052	1,052	1,011	1,011	384	384
Clusters	92	92	74	74	86	86	88	88	80	80
adj. R^2	0.761	0.763	0.532	0.541	0.703	0.707	0.682	0.685	0.452	0.456

Standard errors clustered at the village level in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The household level controls included are: age and gender of the household head, percentage of income derived from agriculture, number of components of the household and dependency ratio. The village level controls included are the time in minutes to a practicable road in the dry and in the rainy season and whether the households have access to electricity.

resulting estimates are displayed in Tables 4.8 and 4.9 respectively. Reassuringly, the results do not differ qualitatively to the findings obtained when employing a linear model. With regard to the prices, it can be noticed that the impact is always estimated to be positive (albeit not always statistically significant) for all farmers except, in some cases, the ones in the highest education quartile.

Additionally, in line with the previous estimations, the absolute value of both the education dummies and the interaction terms are shown to increase monotonically when moving across quartiles, corroborating the results and the interpretation provided in the previous section. In particular, this implies that educated farmers experience a premium when the village has no mobile phone coverage, which is eliminated once mobile telephony is made available.

The analysis of the impact of the welfare measures is in line with the estimated impact on prices. Indeed, the estimates show that farmers in the three lowest quartiles experienced income and wealth gains as a result of mobile phone coverage rollout, while the most educated ones saw their welfare significantly diminished. As pointed out in the previous section, this does not imply that welfare differences across farmers with different level of schooling were completely eliminated, but only reduced by the introduction of this new technology.

Table 4.9: Impact on welfare (quartiles)

	Income		**Wealth**	
Coverage	0.090	0.212**	0.024	0.169*
	(0.11)	(0.11)	(0.07)	(0.09)
Interaction terms				
2^{nd} quartile		-0.039		-0.026
		(0.19)		(0.12)
3^{rd} quartile		-0.043		-0.197*
		(0.19)		(0.11)
4^{th} quartile		-0.319***		-0.231**
		(0.22)		(0.11)
Schooling quartiles				
2^{nd} quartile	0.192**	0.496***	0.364***	0.447***
	(0.08)	(0.17)	(0.05)	(0.10)
3^{rd} quartile	0.383***	0.421**	0.536***	0.689***
	(0.08)	(0.17)	(0.05)	(0.10)
4^{th} quartile	0.858***	1.420***	0.965***	1.229***
	(0.10)	(0.21)	(0.06)	(0.13)
N	2,873	2,873	3,183	3,183
Clusters	93	93	93	93
adj. R^2	0.128	0.131	0.247	0.248

Standard errors clustered at the village level in parentheses. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The household level controls included are: age and gender of the household head, percentage of income derived from agriculture, number of components of the household and dependency ratio. The village level controls included are the time in minutes to a practicable road in the dry and in the rainy season and whether the households have access to electricity.

4.6 Conclusion

Mobile phones had a transformative impact on the economies in Sub Saharan Africa. By reducing the otherwise high information searching costs (as stressed by Aker and Mbiti 2010, mobile phones were typically the first form of digital communication available in rural areas), this technology has allowed economic agents to connect their action and improve significantly the efficiency of markets (Jensen 2007, Aker 2010).

By addressing information asymmetry, mobile telephony is also believed to have a positive impact on the welfare of most vulnerable. In particular, rural farmers (typically the poorest sector of the population) are potentially among the most direct beneficiaries of this new technology as they could use it to obtain price information and improve their bargaining power when negotiating with traders at their farm gate (Courtois and Subervie 2014, Svensson and Yanagizawa 2009). However, evidence in this sense are rather mixed. On the one hand Goyal, 2010, Nakasone, 2013 and Svensson and Yanagizawa, 2009 found that farmers provided with price information were able to obtain higher prices for their produce with positive consequences in terms of welfare. On the other, Camacho and Conover, 2010, Fafchamps and Minten, 2012 and Mitra et al., 2018 found that similar programmes had no impact whatsoever on farmers' prices and welfare.

As pointed out by Hildebrandt et al., 2015, Mitchell, 2017 and Mitra et al., 2018, these mixed findings are not completely unexpected, as the value of information is not intrinsic but contingent on the economic background the agents operate in and particularly on the occurrence of other market failures and the underlying market structure. In this paper, I argue that the impact of ICT depends also on individual characteristics which determine the ability to use the technology to obtain the desired information as well as the utility of the information itself.

In particular, in this paper I look at whether education matters in determining the impact of mobile telephony services on farmers' outcomes in terms of prices received and welfare measures. The effect is theoretically ambiguous: if on the one hand farmers with higher human capital are more likely to adopt the new technology and use it profitably, they might also be less in need of the price information as they are *a priori* more aware of the existing market conditions and/or advantaged in the bargaining process with traders. The results show that, in the context of rural Uganda, less educated farmers were the ones benefitting the most from mobile phone coverage extension. Indeed, I show that when mobile telephony is made available into a village, the individuals with lower level of formal schooling were able to negotiate higher prices, and experienced significant improvements in their welfare.

The aggregate estimates are generally positive but not statistically significant. This is due to the fact that the most educated farmers experienced a slightly negative impact, potentially due to the higher competition faced by the lower educated ones in periods of high market prices. These findings suggest that mobile phones can be profitably used also by individuals with low level of schooling, who are also the ones who evaluate price information the most as it acts as a substitute for human capital in the bargaining process with traders.

Chapter 5

General Conclusion

This dissertation examines how the constraints faced by farmers in Uganda (but generally by farmers in the developing world) shape their choices leading to sub-optimal aggregate outcomes. The first two essays explicitly model the process of agents decision making in terms of the quantity of inputs used and the crops to be cultivated. They show that the rational decisions of individual constrained in their market accessibility are likely to result in an inefficient distribution of factors of production across individuals with inherently different skills and cropping patterns that are not in line with the underlying system of comparative advantage. The empirical analysis mirrors the theoretical framework as it employs micro data to generate aggregate estimates of the magnitude of the efficiency gains that could be achieved and to identify the causes of the observed inefficiencies.

Both chapters make an important contribution to the general literature on factor misallocation, as they shed light on two topics that have been so far overlooked. In particular, the first essay shows that abating barriers to land transactions is not a sufficient condition to remove misallocation as other factors might prevent the flow of inputs from more to less productive individuals. This is crucial in the case of the agricultural sector in developing countries where a large share of the workforce produces for self consumption purposes and therefore might face incentives that are not entirely captured by the standard models on profit maximization.

Also, the paper contributes to the discussion on the negative relationship between aggregate productivity in agriculture and the share of employment in farming. The general consensus is that low productivity implies that a higher share of the labour force must be employed to meet the country's food consumption need and therefore suggests that lagging productivity causes a large farming sector. However, if the food problem is "moved" to the household level, the opposite conclusion can be reached. Indeed, it might be the case that individuals are driven to (or reluctant to leave) farming by the need to provide for their own consumption, preventing the formation of a more modern, commercial oriented and productive farming sector through consolidation.

The findings of the second essay question the implicit assumptions of many related trade models that inefficiencies in the cropping distribution are caused by lack of integration across regional markets. Indeed, an empirical decomposition of the spatial misallocation of cropping patterns shows that more than half the total efficiency gains could be achieved just by reallocating the production within relatively narrow market areas. This shows that in the case of agriculture linking rural producers to markets might be as beneficial as connecting urban markets to unleash the force of comparative advantages.

Finally, the last essay looks at the conditions under which mobile telephony (and more generally the provision of information) can be beneficial to rural farmers and improve their bargaining power when negotiating prices with traders. The existing literature points out that the effectiveness of information technologies is contingent on the underlying market structure. In particular, the occurrence of other market failures such as lack of financial services or the existence of particular market structures (perfect competition or perfect collusion among traders) might make the provision of information useless. However, existing studies have mostly overlooked the possibility that the impact might be heterogeneous across individuals. In the essay, I study whether the availability of mobile phone coverage has had a different effects on farmers depending on their level of education. Since the theoretical effect is ambiguous, the approach is mostly empirical.

I find that less educated farmers, possibly because they are less aware of market prices or inherently less able in dealing with traders, were the ones benefitting the most from this technology. Indeed, the results suggest that the extension of mobile phone coverage helped less educated farmers to obtain better prices at the farm gate, eliminating any difference with respect to the ones with higher formal schooling.

Bibliography

[1] Adamopoulos, Tasso. "Spatial Integration, Agricultural Productivity, and Development: A Quantitative Analysis of Ethiopias Road Expansion Program". *2019 Meeting Papers*. 86. Society for Economic Dynamics. 2019.

[2] Adamopoulos, Tasso. "Transportation Costs, Agricultural Productivity, And Cross-Country Income Differences". *International Economic Review* 52.2 (2011), pp. 489–521.

[3] Adamopoulos, Tasso and Restuccia, Diego. *Geography and agricultural productivity: Cross-country evidence from micro plot-level data*. Tech. rep. National Bureau of Economic Research, 2018.

[4] Adamopoulos, Tasso and Restuccia, Diego. "The size distribution of farms and international productivity differences". *American Economic Review* 104.6 (2014), pp. 1667–97.

[5] Adamopoulos, Tasso, Brandt, Loren, Leight, Jessica, and Restuccia, Diego. *Misallocation, selection and productivity: A quantitative analysis with panel data from China*. Tech. rep. National Bureau of Economic Research, 2017.

[6] Aker, Jenny C. "Does digital divide or provide? The impact of cell phones on grain markets in Niger". *Center for Global Development working paper* 154 (2008).

[7] Aker, Jenny C. "Information from markets near and far: Mobile phones and agricultural markets in Niger". *American Economic Journal: Applied Economics* 2.3 (2010), pp. 46–59.

[8] Aker, Jenny C and Fafchamps, Marcel. "Mobile phone coverage and producer markets: Evidence from West Africa". *The World Bank Economic Review* 29.2 (2014), pp. 262–292.

[9] Aker, Jenny C and Ksoll, Christopher. "Can mobile phones improve agricultural outcomes? Evidence from a randomized experiment in Niger". *Food Policy* 60 (2016), pp. 44–51.

[10] Aker, Jenny C and Mbiti, Isaac M. "Mobile phones and economic development in Africa". *Journal of economic Perspectives* 24.3 (2010), pp. 207–32.

[11] Ayerst, Stephen, Brandt, Loren, Restuccia, Diego, et al. *Market Constraints, Misallocation, and Productivity in Vietnam Agriculture*. Tech. rep. 2018.

[12] Baland, Jean-Marie, Gaspart, Frédéric, Platteau, Jean-Philippe, and Place, Frank. "The distributive impact of land markets in Uganda". *Economic Development and Cultural Change* 55.2 (2007), pp. 283–311.

[13] Bartel, Ann P and Lichtenberg, Frank R. "The comparative advantage of educated workers in implementing new technology". *The Review of Economics and statistics* (1987), pp. 1–11.

[14] Beuermann, Diether, McKelvey, Christopher, Sotelo, Carlos, et al. *The effects of mobile phone infrastructure: Evidence from rural Peru.* Tech. rep. 2012.

[15] Bezabih, Mintewab and Di Falco, Salvatore. "Rainfall variability and food crop portfolio choice: evidence from Ethiopia". *Food Security* 4.4 (2012), pp. 557–567.

[16] Bhavnani, Asheeta, Chiu, Rowena Won-Wai, Janakiram, Subramaniam, Silarszky, Peter, and Bhatia, Deepak. "The role of mobile phones in sustainable rural poverty reduction". *retrieved November* 22 (2008).

[17] Blumenstock, Joshua Evan and Eagle, Nathan. "Divided we call: Disparities in access and use of mobile phones in Rwanda". *Information Technologies & International Development* 8.2 (2012), pp–1.

[18] Burrell, Jenna and Oreglia, Elisa. "The myth of market price information: Mobile phones and the application of economic knowledge in ICTD". *Economy and Society* 44.2 (2015), pp. 271–292.

[19] Buys, Piet, Dasgupta, Susmita, Thomas, Timothy S, and Wheeler, David. "Determinants of a digital divide in Sub-Saharan Africa: A spatial econometric analysis of cell phone coverage". *World Development* 37.9 (2009), pp. 1494–1505.

[20] Camacho, Adriana and Conover, Emily. "The impact of receiving price and climate information in the agricultural sector". *Documento CEDE* 2010-40 (2010).

[21] Cecchini, Simone and Scott, Christopher. "Can information and communications technology applications contribute to poverty reduction? Lessons from rural India". *Information Technology for development* 10.2 (2003), pp. 73–84.

[22] Chari, Anusha, Liu, Elaine M, Wang, Shing-Yi, and Wang, Yongxiang. *Property Rights, Land Misallocation and Agricultural Efficiency in China.* Tech. rep. National Bureau of Economic Research, 2017.

[23] Chen, Chaoran, Restuccia, Diego, and Santaeulàlia-Llopis, Raül. *The Effects of Land Markets on Resource Allocation and Agricultural Productivity.* Tech. rep. National Bureau of Economic Research, 2017.

[24] Costinot, Arnaud and Donaldson, Dave. *How large are the gains from economic integration? Theory and evidence from US agriculture, 1880-1997.* Tech. rep. National Bureau of Economic Research, 2016.

[25] Courtois, Pierre and Subervie, Julie. "Farmer bargaining power and market information services". *American Journal of Agricultural Economics* 97.3 (2014), pp. 953–977.

[26] De Janvry, Alain, Fafchamps, Marcel, and Sadoulet, Elisabeth. "Peasant household behaviour with missing markets: some paradoxes explained". *The Economic Journal* 101.409 (1991), pp. 1400–1417.

[27] De Janvry, Alain and Sadoulet, Elisabeth. "Progress in the modeling of rural households behavior under market failures". *Poverty, inequality and development.* Springer, 2006, pp. 155–181.

[28] Deininger, Klaus and Mpuga, Paul. "Land Markets in Uganda: Incidence, Impact, and Evolution over Time". *World Bank Discussion Paper. Washington DC: World Bank* (2002).

[29] Diga, Kathleen. "Mobile cell phones and poverty reduction: Technology spending patterns and poverty level change among households in Uganda" (2008).

[30] Dillon, Brian and Barrett, Christopher B. "Agricultural factor markets in Sub-Saharan Africa: an updated view with formal tests for market failure". *Food policy* 67 (2017), pp. 64–77.

[31] Dorosh, Paul, Wang, Hyoung Gun, You, Liang, and Schmidt, Emily. "Crop production and road connectivity in Sub-Saharan Africa: A spatial analysis" (2010).

[32] Fafchamps, Marcel. "Cash crop production, food price volatility, and rural market integration in the third world". *American journal of agricultural economics* 74.1 (1992), pp. 90–99.

[33] Fafchamps, Marcel and Hill, Ruth Vargas. "Price transmission and trader entry in domestic commodity markets". *Economic Development and cultural change* 56.4 (2008), pp. 729–766.

[34] Fafchamps, Marcel and Minten, Bart. "Impact of SMS-based agricultural information on Indian farmers". *The World Bank Economic Review* 26.3 (2012), pp. 383–414.

[35] Finkelshtain, Israel and Chalfant, James A. "Marketed surplus under risk: Do peasants agree with Sandmo?" *American Journal of Agricultural Economics* 73.3 (1991), pp. 557–567.

[36] Fuchs, Christian and Horak, Eva. "Africa and the digital divide". *Telematics and informatics* 25.2 (2008), pp. 99–116.

[37] Futch, Michael Douglas and McIntosh, Craig Thomas. "Tracking the introduction of the village phone product in Rwanda". *Information Technologies & International Development* 5.3 (2009), pp–54.

[38] Gollin, Douglas, Parente, Stephen L, and Rogerson, Richard. “The food problem and the evolution of international income levels”. *Journal of Monetary Economics* 54.4 (2007), pp. 1230–1255.

[39] Gollin, Douglas and Rogerson, Richard. *Agriculture, roads, and economic development in Uganda*. Tech. rep. National Bureau of Economic Research, 2010.

[40] Gollin, Douglas and Rogerson, Richard. “Productivity, transport costs and subsistence agriculture”. *Journal of Development Economics* 107 (2014), pp. 38–48.

[41] Gollin, Douglas and Udry, Christopher. “Heterogeneity, measurement error and misallocation: Evidence from African agriculture”. *NBER Working Paper* w25440 (2019).

[42] Goyal, Aparajita. “Information, direct access to farmers, and rural market performance in central India”. *American Economic Journal: Applied Economics* 2.3 (2010), pp. 22–45.

[43] Hildebrandt, Nicole, Nyarko, Yaw, Romagnoli, Giorgia, and Soldani, Emilia. “Price Information, Inter-Village Networks, and Bargaining Spillovers: Experimental Evidence from Ghana”. *Preliminary draft). Retrieved from http://sites. bu. edu/neudc/files/2014/10/paper_345. pdf* (2015).

[44] Hsieh, Chang-Tai and Klenow, Peter J. “Misallocation and manufacturing TFP in China and India”. *The Quarterly journal of economics* 124.4 (2009), pp. 1403–1448.

[45] Huffman, Wallace E. “Human capital: Education and agriculture”. *Handbook of agricultural economics* 1 (2001), pp. 333–381.

[46] Hunt, Diana. “Unintended consequences of land rights reform: the case of the 1998 Uganda Land Act”. *Development policy review* 22.2 (2004), pp. 173–191.

[47] Jaleta, Moti and Gardebroek, Cornelis. “Farm-gate tomato price negotiations under asymmetric information”. *Agricultural Economics* 36.2 (2007), pp. 245–251.

[48] Jayne, Thomas S. “Do high food marketing costs constrain cash crop production? Evidence from Zimbabwe”. *Economic development and cultural change* 42.2 (1994), pp. 387–402.

[49] Jensen, Robert. “The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector”. *The quarterly journal of economics* 122.3 (2007), pp. 879–924.

[50] Klonner, Stefan and Nolen, Patrick J. “Cell phones and rural labor markets: Evidence from South Africa” (2010).

[51] Labonne, Julien and Chase, Robert S. *The power of information: The impact of mobile phones on farmers' welfare in the Philippines*. The World Bank, 2009.

[52] Lagakos, David and Waugh, Michael E. “Selection, agriculture, and cross-country productivity differences”. *American Economic Review* 103.2 (2013), pp. 948–80.

[53] Lee, Kyeong Ho and Bellemare, Marc F. "Look who's talking: The impacts of the intrahousehold allocation of mobile phones on agricultural prices". *The Journal of Development Studies* 49.5 (2013), pp. 624–640.

[54] Li, Nicholas. "Marketization Costs and Household Specialization: Evidence from Indian Farmers" (2017).

[55] Maue, Casey C, Burke, Marshall, and Emerick, Kyle J. *Productivity Dispersion and Persistence Among the World's Most Numerous Firms*. Tech. rep. National Bureau of Economic Research, 2020.

[56] Mitchell, Tara. "Is Knowledge Power? Information and Switching Costs in Agricultural Markets". *American Journal of Agricultural Economics* 99.5 (2017), pp. 1307–1326.

[57] Mitra, Sandip, Mookherjee, Dilip, Torero, Maximo, and Visaria, Sujata. "Asymmetric information and middleman margins: An experiment with Indian potato farmers". *Review of Economics and Statistics* 100.1 (2018), pp. 1–13.

[58] Mittal, Surabhi and Mehar, Mamta. "Socio-economic factors affecting adoption of modern information and communication technology by farmers in India: Analysis using multivariate probit model". *The Journal of Agricultural Education and Extension* 22.2 (2016), pp. 199–212.

[59] Molony, Thomas. "Running out of credit: The limitations of mobile telephony in a Tanzanian agricultural marketing system". *The Journal of Modern African Studies* 46.4 (2008), pp. 637–658.

[60] Muto, Megumi and Yamano, Takashi. "The impact of mobile phone coverage expansion on market participation: Panel data evidence from Uganda". *World development* 37.12 (2009), pp. 1887–1896.

[61] Nabwire, Leocardia. *Ugandas Dilemmas in the Transition to Modern Commercial Agriculture: Implications for the Poverty Reduction Agenda*. Tech. rep. Future agriculture, 2015.

[62] Nakasone, Eduardo. "The role of price information in agricultural markets: Experimental evidence from rural Peru" (2013).

[63] Nakasone, Eduardo, Torero, Maximo, and Minten, Bart. "The power of information: The ICT revolution in agricultural development". *Annu. Rev. Resour. Econ.* 6.1 (2014), pp. 533–550.

[64] Omamo, Steven Were. "Transport costs and smallholder cropping choices: An application to Siaya District, Kenya". *American Journal of Agricultural Economics* 80.1 (1998), pp. 116–123.

[65] Patrick, George F and Kehrberg, Earl W. "Costs and returns of education in five agricultural areas of eastern Brazil". *American Journal of Agricultural Economics* 55.2 (1973), pp. 145–153.

[66] Perego, Viviana ME. "Crop prices and the demand for titled land: Evidence from Uganda". *Journal of Development Economics* 137 (2019), pp. 93–109.

[67] Qin, Yu and Zhang, Xiaobo. "The road to specialization in agricultural production: Evidence from rural China". *World Development* 77 (2016), pp. 1–16.

[68] Restuccia, Diego and Rogerson, Richard. "Policy distortions and aggregate productivity with heterogeneous establishments". *Review of Economic dynamics* 11.4 (2008), pp. 707–720.

[69] Restuccia, Diego and Rogerson, Richard. "The causes and costs of misallocation". *Journal of Economic Perspectives* 31.3 (2017), pp. 151–74.

[70] Restuccia, Diego and Santaeulalia-Llopis, Raul. *Land misallocation and productivity*. Tech. rep. National Bureau of Economic Research, 2017.

[71] Restuccia, Diego, Yang, Dennis Tao, and Zhu, Xiaodong. "Agriculture and aggregate productivity: A quantitative cross-country analysis". *Journal of monetary economics* 55.2 (2008), pp. 234–250.

[72] Sekamatte, BM, Ogenga-Latigo, M, and Russell-Smith, A. "Effects of maize–legume intercrops on termite damage to maize, activity of predatory ants and maize yields in Uganda". *Crop Protection* 22.1 (2003), pp. 87–93.

[73] Shenoy, Ajay. "Market failures and misallocation". *Journal of Development Economics* 128 (2017), pp. 65–80.

[74] Shimamoto, Daichi, Yamada, Hiroyuki, and Gummert, Martin. "Mobile phones and market information: Evidence from rural Cambodia". *Food Policy* 57 (2015), pp. 135–141.

[75] Sotelo, Sebastian. "Domestic trade frictions and agriculture". *University of Michigan Research Seminar in International Economics Discussion Paper*. Vol. 641. 2019.

[76] Stifel, David and Minten, Bart. "Isolation and agricultural productivity". *Agricultural Economics* 39.1 (2008), pp. 1–15.

[77] Stigler, George J. "The economics of information". *Journal of political economy* 69.3 (1961), pp. 213–225.

[78] Svensson, Jakob and Yanagizawa, David. "Getting prices right: The impact of the market information service in Uganda". *Journal of the European Economic Association* 7.2-3 (2009), pp. 435–445.

[79] Tadesse, Getaw and Bahiigwa, Godfrey. "Mobile phones and farmers marketing decisions in Ethiopia". *World development* 68 (2015), pp. 296–307.

[80] Tombe, Trevor. "The missing food problem: Trade, agriculture, and international productivity differences". *American Economic Journal: Macroeconomics* 7.3 (2015), pp. 226–58.

[81] Ulimwengu, John M, Ramadan, Racha, et al. *How does food price increase affect Ugandan households?: An augmented multimarket approach*. Tech. rep. International Food Policy Research Institute (IFPRI), 2009.

[82] Valentinyi, Akos and Herrendorf, Berthold. "Measuring factor income shares at the sectoral level". *Review of Economic Dynamics* 11.4 (2008), pp. 820–835.

[83] Van Dijk, Jan and Hacker, Kenneth. "The digital divide as a complex and dynamic phenomenon". *The information society* 19.4 (2003), pp. 315–326.

[84] Waverman, Leonard, Meschi, Meloria, and Fuss, Melvyn. "The impact of telecoms on economic growth in developing countries". *The Vodafone policy paper series* 2.03 (2005), pp. 10–24.

[85] Welch, Finis. "Education in production". *Journal of political economy* 78.1 (1970), pp. 35–59.

[86] Woldie, Getachew Abebe and Nuppenau, Ernst-August. "The influence of marketing cooperatives on the bargaining position of producers in farm-gate banana prices: Evidences from smallholders in Southern Ethiopia". *IV International Symposium on Banana: International Conference on Banana and Plantain in Africa: Harnessing International 879.* 2008, pp. 161–168.

Chapter 6

Appendix A

Social planner problem

The solution for the social planner problem can be obtained by plugging in the production function (equation 2.3) in equation 2.4, and obtaining the Lagrangian:

$$\mathcal{L} = \sum_i s_i^{1-\gamma}(l_i^{\alpha} k_i^{1-\alpha})^{\gamma} - \lambda_1 \left(\sum_i^N l_i - L \right) - \lambda_2 \left(\sum_i^N k_i - K \right) \tag{A.1}$$

with first order conditions:

$$\alpha\gamma s_i^{1-\gamma} l_i^{\alpha\gamma-1} k_i^{(1-\alpha)\gamma} = \lambda_1 \tag{A.2a}$$

$$(1-\alpha)\gamma s_i^{1-\gamma} l_i^{\alpha\gamma} k_i^{(1-\alpha)\gamma-1} = \lambda_2 \tag{A.2b}$$

$$\sum_i^N l_i = L \tag{A.2c}$$

$$\sum_i^N k_i = K \tag{A.2d}$$

By taking the ratio of the first two conditions for a generic farmer i, I obtain that:

$$k_i = \frac{1-\alpha}{\alpha} \frac{\lambda_1}{\lambda_2} l_i \tag{A.3}$$

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

By plugging this into equation A.2a and solving for $\frac{l_i}{l_j}$ for any two farmers i and j, it can be found that:

$$\frac{l_i}{l_j} = \frac{s_i}{s_j} \tag{A.4}$$

this implies that land (and consequently capital, given equation A.3) input is distributed across farmers proportionally to their agricultural skill s. In turn, given the resource constraints (equations A.2c and A.2d), this results in:

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

that is the final result presented in Section 3.

Equilibrium with $\tau = 0$

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

$$\max s_i^{1-\gamma} \xi_i^{\gamma} - r\xi_i \tag{A.6}$$

with first order condition:

$$\gamma s_i^{1-\gamma} \xi_i^{\gamma-1} = r \tag{A.7}$$

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

$$\xi_i = \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i \tag{A.8}$$

The cost of input r can be obtained as the market clearing price given the exogenous total input available Ξ and plugging A.8 into the resource constraint:

$$\sum_i^N \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i = \Xi \tag{A.9}$$

Solving for r leads to:

$$r = \gamma \left(\frac{\sum_i s_i}{\Xi}\right)^{1-\gamma} \tag{A.10}$$

This expression represents a useful benchmark to assess how changes in the output frictions τ affects r and in turn the distribution of the input across the farmers. Moreover, this can be plugged into equation A.8 obtaining:

$$\xi_i = \left(\frac{s_i}{\sum_i s_i}\right) \Xi \tag{A.11}$$

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

Equilibrium with $\tau = 1$

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

$$\max s_i^{1-\gamma}\xi_i^{\gamma} - r\xi_i \text{ subject to } s_i^{1-\gamma}\xi_i^{\gamma} \leq \overline{y} \tag{A.12}$$

and the corresponding Lagrangian is:

$$\mathcal{L} = s_i^{1-\gamma}\xi_i^{\gamma} - r\xi_i - \mu(s_i^{1-\gamma}\xi_i^{\gamma} - \overline{y}) \tag{A.13}$$

whose first order conditions are:

$$(1-\mu)\gamma s_i^{1-\gamma}\xi_i^{\gamma-1} = r \tag{A.14a}$$
$$s_i^{1-\gamma}\xi_i^{\gamma} \leq \overline{y} \tag{A.14b}$$
$$\mu \geq 0 \tag{A.14c}$$
$$\mu(s_i^{1-\gamma}\xi_i^{\gamma} - \overline{y}) = 0 \tag{A.14d}$$

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

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

$$\xi_i = \left(\frac{\overline{y}}{s_i^{1-\gamma}}\right)^{\frac{1}{\gamma}} \tag{A.15}$$

On the other hand, unconstrained farmers will operate a level of input such that the marginal productivity equals the marginal cost r, and so:

$$\xi_i = \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i \tag{A.16}$$

although this condition is exactly the same as the one for each farmer in the case of $\tau = 0$ depicted in equation A.8, it is worth to point out that it underlies a different quantity and share of inputs since the frictions in the output market may affect the equilibrium price r as it will be made explicit in the following.

Farmers belong to the first or the second group depending on their innate productivity s. More specifically, the threshold above which the constraint can be derived by plugging equation A.15 in the marginal productivity formula and setting it $= r$. By doing so, I will find the productivity level at which producing exactly $\overline{y}$ results in a marginal productivity that is equal to the cost of input r. Formally:

$$\gamma s_i^{1-\gamma} \left(\frac{\overline{y}}{s_i^{1-\gamma}} \right)^{\frac{\gamma-1}{\gamma}} = r \tag{A.17}$$

Solving for s_i leads to:

$$s_i = \overline{y} \left(\frac{r}{\gamma} \right)^{\frac{\gamma}{1-\gamma}} \tag{A.18}$$

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

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

$$\sum_{i^-} \xi_i + \sum_{i^+} \xi_i = \sum_{i^-} \left(\frac{\gamma}{r} \right)^{\frac{1}{1-\gamma}} s_i + \sum_{i^+} \left(\frac{\overline{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} = \Xi \tag{A.19}$$

where the first equality is obtained by plugging in the equilibrium value of input used by non constrained (equation A.16) and constrained (equation A.15) farmers.

Solving for r leads to:

$$r = \gamma \left(\frac{\sum_{i^-} s_i}{\Xi - \overline{y}^{1/\gamma} \sum_{i^+} \frac{1}{s_i^{(1-\gamma)/\gamma}}} \right)^{1-\gamma} \tag{A.20}$$

where the numerator of the ratio between brackets represents the sum of the ability of the unconstrained farmers, while the denominator is the total land used by unconstrained farmers (that is to say the difference between the total input Ξ and the the fraction used by constrained farmers to produce their consumption thresholds $\overline{y}$). The difference with respect to the formula for r when $\tau = 0$ is immediate and can be appreciated comparing the last finding to equation A.10. Depending on the quantity of total input available Ξ and the farmers' ability distribution, different scenarios might occur.

In the extreme case where:

$$\Xi \geq \sum_i \left(\frac{\overline{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} \tag{A.21}$$

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

When instead $\Xi < \sum_i \left(\frac{\overline{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}}$, it can be shown that the resulting equilibrium price r will be lower than the one in the case with $\tau = 0$ as long as some very reasonable conditions hold, Namely, from equations A.10 and A.20:

$$\gamma \left(\frac{\sum_{i^-} s_i}{\Xi - \overline{y}^{1/\gamma} \sum_{i^+} \frac{1}{s_i^{(1-\gamma)/\gamma}}} \right)^{1-\gamma} < \gamma \left(\frac{\sum_i s_i}{\Xi} \right)^{1-\gamma} \tag{A.22}$$

whenever, for at least one farmer i:

$$\frac{s_i}{\sum_i s_i} \Xi < \left(\frac{\overline{y}}{s_i^{1-\gamma}} \right)^{\frac{1}{\gamma}} \tag{A.23}$$

or, solving for s_i:

$$s_i > \overline{y} \left(\frac{\sum_i s_i}{\Xi} \right)^{\gamma} \tag{A.24}$$

this condition implies that for at least one farmer, the ability s is high enough for the total input operated in the case of $\tau = 0$ to be higher than the one necessary to produce $\overline{y}$. Intuitively, this brings about a reduction of the input price because the factors of production would be transferred from more productive (constrained) farmers, to less productive ones who are willing to pay less. The resulting allocation of input is the one depicted by the dashed green line in Figure 2.4.

In particular, in equilibrium all the farmers with $s_i \leq \overline{y} \left(\frac{\sum_{i^-} s_i}{\Xi - \overline{y}^{1/\gamma} \sum_{i^+} \frac{1}{s_i^{(1-\gamma)/\gamma}}} \right)^{\gamma}$ will not be constrained and operate a share of land proportional to their ability, while the others will operate just enough input to meet their subsistence threshold $\overline{y}$. As a result, all farmers with ability higher (lower) than $\overline{y} \left(\frac{\sum_i s_i}{\Xi} \right)^{\gamma}$ will be operating a lower (higher) than efficient quantity of inputs.

Note that as $s_i \leq \overline{y}\left(\frac{\sum_{i^-} s_i}{\Xi - \overline{y}^{1/\gamma}\sum_{i^+}\frac{1}{s_i^{(1-\gamma)/\gamma}}}\right)^{\gamma} < \overline{y}\left(\frac{\sum_i s_i}{\Xi}\right)^{\gamma}$, there will be some constrained farmers (who only operate a amount of input sufficient to produce $\overline{y}$) who are nevertheless operating more input than in the $\tau = 0$ counterfactual, this is a result of the reduction in the input price r.

If on the other hand, the condition A.24 does not hold for any of the farmers, none of them would be constrained as they would all be producing less than what their consumption threshold $\overline{y}$ and, as a result, $i^+ = \emptyset$. Therefore, the cost of input would be the same as in the case of $\tau = 0$. Indeed, when none of the farmers is constrained, $\sum_i^- s_i = \sum_i s_i$ and $\sum_{i^+} \frac{1}{s_i^{(1-\gamma)/\gamma}}^{1-\gamma} = 0$, equation A.20 becomes identical to equation A.10. Thus, the resulting input distribution would be exactly the same and fully efficient.

Equilibrium with $0 < \tau < 1$

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

$$\max_{\xi_i} \begin{cases} s_i^{1-\gamma}\xi_i^{\gamma} - r\xi_i & \text{if } s_i^{1-\gamma}\xi_i^{\gamma} \leq \overline{y} \\ (1-\tau)(s_i^{1-\gamma}\xi_i^{\gamma}) - r\xi_i & \text{if } s_i^{1-\gamma}\xi_i^{\gamma} > \overline{y} \end{cases} \tag{A.25}$$

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

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

$$\overline{y}\left(\frac{r}{\gamma}\right)^{\frac{\gamma}{1-\gamma}} < s_i < \overline{y}\left[\frac{r}{\gamma(1-\tau)}\right]^{\frac{\gamma}{1-\gamma}} \tag{A.26}$$

In line with the terminology used in the main section, I will refer to farmers with lower productivity as belonging to group **a**, to farmers with productivity in the interval as in equation A.26 as belonging to group **b** and to farmers with higher productivity as group **c**.

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

$$\text{Group } \mathbf{a}\text{: } \xi_i = \left(\frac{\gamma}{r}\right)^{\frac{1}{1-\gamma}} s_i \tag{A.27a}$$

$$\text{Group } \mathbf{b}\text{: } \xi_i = \left(\frac{\overline{y}}{s_i^{1-\gamma}}\right)^{\frac{1}{\gamma}} \tag{A.27b}$$

$$\text{Group } \mathbf{c}\text{: } \xi_i = \left[\frac{\gamma(1-\tau)}{r}\right]^{\frac{1}{1-\gamma}} s_i \tag{A.27c}$$

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

$$\sum_{i^a} \xi_i + \sum_{i^b} \xi_i + \sum_{i^c} \xi_i = \Xi \tag{A.28}$$

By plugging in the values in equations A.27a to A.27c, and solving for r:

$$r = \gamma \left[\frac{\sum_{i^a} s_i + (1-\tau)^{1/(1-\gamma)} \sum_{i^c} s_i}{\Xi - \sum_{i^b} \left(\frac{\overline{y}}{s_i^{1-\gamma}}\right)^{1/\gamma}} \right]^{1-\gamma} \tag{A.29}$$

It is easy to show that when $\tau = 0$ the formula collapses to equation A.10 as all the farmers would belong to either group **a** or **c** and they would be equally weighted at the numerator. If instead $\tau = 1$ the formula would mirror to equation A.20 as group c would not be populated and group **a** (**b**) would be composed by the same elements of group i^- (i^+). Although the impact of an change in τ depends on the productivity distribution, from a qualitative point of view a positive τ will always result in a decrease in r as long as there is at least one farmer with productivity higher than $\overline{y}\left(\frac{\sum_i s_i}{\Xi}\right)^{\gamma}$. More specifically, for sufficiently low τ, group **b** will not be populated and the lower r would be the result of a lower weighting of the more productive farmers in category **c** in the numerator. Higher values of τ will also affect the groups' composition, as more productive farmers in group **a** and less productive farmers in group **c** will move to group **b**. For sufficiently high τ, r as well as the resulting input distribution would mirror the situation described when $\tau = 1$.

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

Chapter 7

Appendix B

The determinants of mobile phone coverage

The proposed identification strategy relies heavily on village fixed effects. Such specification allows me to control non parametrically for all the time invariant factors that might have affected the timing of the coverage extension. As a result, this strategy is credible as long as most of the determinants are actually fixed over time. As argued above, the extension of the services has been driven by market forces so that we expect areas with a higher degree of urbanization /better development perspectives or with convenient physical features to be targeted before. The insights derived from the descriptive statistics provided so far support this intuition.

In this section, we will try to determine to which extent observable time invariant determinants can account for the timing of he rollout. The higher the goodness of fit and the explanatory power of a model predicting coverage timing by using fixed (or pre-coverage) regressors, the less relevant will be our initial concerns related to possible unobservable time varying factors (which might potentially induce spurious correlation between welfare measures and the). It is worth pointing out that this exercise is only meant to be suggestive and complementary to the inclusion of village level time variant regressors, allowing me to control at least for some dimensions of infrastructural development at the community level.

In the following analysis, the village level year of coverage is regressed on a number of fixed or pre-coverage characteristics that, according to the existing literature (Buys et al. 2009; Klonner and Nolen 2010), are likely to determine the timing of the rollout. Most of the variables employed consist of GIS data; the merging process was rather straightforward as the surveyed communities are accurately geolocated.

First of all, physical features have been shown to be an important determinant of mobile phone coverage extension: altitude and terrain slope affect the placement cost of phone towers as well as their range. The data for elevation comes from the STRM 30 (Shuttle Radar Topography Mission) Digital Elevation Model: consisting in a 30 arc-seconds resolution raster indicating the average altitude at each cell. Terrain slope is derived from the same dataset as the maximum change in altitude among adjacent cells.

However, terrain features are just one of the factors influencing the timing of the rollout. Realistically, development and urbanization levels and outlook play a crucial role in determining mobile telephony providers' profitability. The issue with these variables is that, unlike physical features, they are time variant. Thus, in the following analysis, I will considered values referring to the pre-coverage situation will be considered.

Firstly being close to important cities, in particular to regional or district capitals, arguably guarantees an earlier extension of the services. A similar reasoning can be applied for the distance from main transport network. In the empirical analysis, the distance to the closest primary road or railroad as described by the Digital Chart of the World, dating back to 1992, will be considered.

Furthermore, some newly available spatial datasets allow for an even more detailed analysis of the preexisting development and urbanization levels in rural Uganda. More specifically I will exploit the GLC 2000 (Global Land Coverage 2000) and the Global Grid of Probabilities of Urban Expansion to 2030 data. The first dataset describes the prevalent land coverage (out of 22 classes) at a very precise level (30 arc-seconds), while the second one provides the probability of each non urban grid cell (at a 2.5 arc-minute resolution) to become urban by 2030.[1]

Two village level dummies were derived from these datasets: Intensive Farming, taking value 1 when the village cell presents some clear signs of farming and/or similar human activities and Urban Expansion, whose value is 1 when the predicted probability of the area being urbanized by 2030 is higher than 75%. The descriptive statistics and the sources for all these variables are displayed in Table B.1.

I will estimate three different models to predict the timing of mobile phone coverage rollout across the villages in the sample. The first one is a simple OLS having as dependent variable the year coverage was first available. For a robustness check, I also estimate an ordered probit and an ordered logit model having as dependent variables whether the coverage was available in

Table B.1: Descriptive statistics and sources

	Mean	**SD**	**Source**
Distance (in km) from:			
District capital	17.05	10.52	RePEAT
Regional Capital	90.37	48.73	RePEAT
Transportation network	26.81	32.87	Digital Chart of the World
Urban expansion	0.25	0.43	Global Grid of Probabilities of Urban Expansion
Intensive farming	0.53	0.50	Global Land Cover 2000
Altitude (metres)	1315.23	272.69	STRM 30
Terrain slope (degrees)	1.76	2.05	STRM 30

All variables are defined at the village level on the basis of the coordinates given by RePEAT. The distance to the district capital is based on the pre 2005 reform administrative map of Uganda.

2003, 3005 or 2009, which are the three years the interviews were carried out. The results are displayed in Table B.2.[2]

Firstly, all the regressors are significant and their coefficients exhibit the expected signs. Unsurprisingly, villages closer to major cities/towns or primary roads and railroads received coverage at an earlier stage. Similarly, LC1 situated in the most developed Central region were targeted, on average, earlier. Consistently, areas exhibiting higher levels of human activity and higher urbanization potential (as of 2000) have more chances to get earlier coverage, as shown by the negative and statistically significant coefficients attached to Urban Expansion and Intensive Farming variables.

As for physical features, they are shown to be relevant determinants even when controlling for urbanization and development level. In line with the findings of Klonner and Nolen, 2010, slope has a non linear impact on mobile phone tower placement: indeed, high levels of terrain roughness imply major costs and so more delayed coverage (as shown by the positive and statistically significant coefficient estimated for villages in the 5th quintile), while more moderate levels seem not to be an obstacle. The rationale behind that

is that moderate levels of terrain steepness are associated with potentially higher set up costs but also wider phone towers' range. Even after controlling for terrain slope, elevation plays a

[1] The second dataset is based on estimates developed in 2000 and referring to variables that were available by that time. Therefore, it does not capture any potential shock occurring in the time frame under analysis (2003-2009). Reassuringly, none of the cells the villages in the sample are located is labelled as urban.

[2] Since I am only interest in the predicting power and the consistencies across the three models, I only show the coefficient estimates of the non-linear models, which are informative on the sign and significance of the impact of each variable.

Table B.2: Determinants of mobile phone rollout

	Linear	**Ordered Probit**	**Ordered Logit**
Log distance (km) from:			
District capital	0.530*	0.272*	0.387
	(0.31)	(0.16)	(0.30)
Regional Capital	0.763*	0.385*	0.655*
	(0.39)	(0.21)	(0.37)
Transportation network	0.625***	0.354***	0.674***
	(0.23)	(0.12)	(0.23)
Region (reference = Central)			
Western	0.793	0.413	0.601
	(0.69)	(0.36)	(0.63)
Eastern	1.347**	0.767**	1.394**
	(0.61)	(0.32)	(0.57)
Urban Expansion	-2.006***	-1.148***	-2.179***
	(0.57)	(0.31)	(0.54)
Intentive Farming	-1.267***	-0.692***	-1.123***
	(0.47)	(0.25)	(0.45)
Physical features:			
Altitude	-0.023**	-0.015**	-0.027**
	(0.01)	(0.01)	(0.01)
Altitude^2	0.00**	0.00***	0.00***
	(0.00)	(0.00)	(0.00)
Slope quintiles (reference = 1^{st})			
2^{nd}	0.423	0.213	0.486
	(0.68)	(0.36)	(0.63)
3^{rd}	1.472**	0.795**	1.658***
	(0.66)	(0.35)	(0.62)
4^{th}	0.541	0.356	0.828
	(0.74)	(0.39)	(0.69)
5^{th}	1.984**	1.014*	1.923**
	(0.99)	(0.53)	(0.91)
N	93	93	93
R^2	0.634		
pseudo-R^2		0.187	0.186

Robust standard errors in parentheses. * p ¡ 0.10, ** p ¡ 0.05, *** p ¡ 0.01. The point estimates refer to the beta coefficients in the case of the ordered probit and logit model. The dependent variable in the linear model is the year of coverage, while in the ordered models is whether the coverage was already available in 2003 (1), if it was made available between 2003 or 2005 (2), or if it was made available between 2005 and 2009 (3). The signs of the coefficients have therefore the same interpretation across the three models.

significant role. The negative coefficient associated to the linear term and the positive one to the squared, as well as the high cut off point (1560 metres, with only 15% of the villages in the sample presenting higher values) indicates that more elevated locations tended to receive coverage earlier unless the altitude was exceptionally high.

Reassuringly, the signs as well as the significance of all the coefficients computed through the linear specification are consistent with the ones of the more sophisticated ordered outcome models. As can be seen from the relatively high values of the R2 and pseudo R2, the time invariant (or pre-coverage) variables included in the regression explain to a good extent the timing of the coverage. This implies that the concerns related to unobservable shocks, that the fixed effects fail to control for (and might lead to spurious correlation between timing of the coverage and welfare measures), are possibly not a crucial issue given that the service extension followed a regular and highly predictable pattern.