Journal of Productivity Analysis (2022) 58:221–237
https://doi.org/10.1007/s11123-022-00647-y

# Aggregate productivity and inefficient cropping patterns in Uganda

Bruno Morando[1]

Accepted: 16 September 2022 / Published online: 30 September 2022


**Abstract**
This paper measures the impact of inefficient spatial distribution of crops on aggregate agricultural productivity in Uganda. By combining village level data on land use and on crop specific land suitability, I show that agricultural TFP could be increased by one third just by reallocating crops according to the underlying structure of comparative advantage. Interestingly, a regional decomposition indicates that half of these gains can be achieved just by redistributing crop production within narrowly defined areas serving the same urban markets. The empirical analysis suggests that differences in market access are a good candidate to explain these inefficiencies: in line with the qualitative theoretical model, more isolated farmers devote systematically more land to non-perishable food crops and their production is less aligned with the agro-climatic conditions they face.

**Keywords** Aggregate productivity · Crop distribution · Comparative advantage · Market access

**JEL** Q14 · O40 · O13

## 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 share of the existing cross country income differences (Gollin et al. 2007, Restuccia et al. 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 are one of the concurring causes of lagging agricultural productivity.

In this paper, I study how costly market access affects farmers decisions of what to grow and quantify the productivity losses derived from inefficiencies in the geographical distribution of crops. While existing studies (Costinot and Donaldson 2016, Sotelo 2020) focus on the lack of integration across regional markets, I look at the cropping patterns within areas that supply the same urban markets. In particular, using village level data on land use and crop specific agroclimatic suitability, I show that, in Uganda, inefficiencies in the spatial allocation of crops within market regions are at least as important as the ones across regional markets. To preview some findings, I estimate that farming productivity could be increased by one third if crops were reallocated across villages according to the structure of agroclimatic comparative advantage. Interestingly, more than half of those gains could be achieved just by redistributing crops within areas serving the same regional markets.

The main hypothesis is that these within market area inefficiencies are driven by differences in transportation costs and market access. In order to formalise this insight, I develop a model of crop choice where consumption and production decisions are non-separable and farmers face heterogeneous transportation costs and different crop specific productivity.[1] By doing so, I show that the impact of transportation costs on crop choices is twofold. In fact, low

✉ Bruno Morando
bruno.morando@mu.ie

1 Department of Economics, Maynooth University, Maynooth, Ireland

[1] The theoretical framework borrows heavily from seminal studies on smallholder farmers' production decisions (Jayne 1994, Omamo 1998, Key et al. 2000). Crucially, these early works aimed to understand the lack of responsiveness to shocks in crops' price, but they did not include heterogeneity in land suitability for different crops.

market accessibility not only increases the share of land devoted to staple crops, but also reduces the responsiveness to crop specific comparative advantage of farmers whose production and consumption decisions are interlinked.

The empirical analysis corroborates the main predictions of the model. Namely, I find that farmers who are better connected to markets devote systematically more land to high value crops (regardless of their relative suitability) than those who are located in more isolated areas. Additionally, farmers in the most remote locations are less responsive to comparative advantage, as their production decisions are informed by their consumption needs and deviate more significantly from the revenue maximising crop choices. In particular, the impact of relative agroclimatic suitability on land use fades as the cost of reaching urban markets increases.

To my knowledge, this is the first paper that examines cropping patterns within narrowly defined market areas and estimates the relative magnitude of inefficiencies in the geographical distribution of crops occurring within regional markets. This work is obviously related to a number of studies that measure the efficiency losses resulting from the lack of integration across domestic regional markets. Costinot and Donaldson (2016) show how an increasingly pervasive transportation network in the US shaped the cropping patterns across different counties, which gradually specialised into the production of relatively more suitable crops. Adamopoulos (2019) performs 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 (2020) 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.

This paper also echoes the findings from a strand of literature that estimates the impact of market accessibility on crop choices exploiting some natural experiments: Li (2017) 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. Similarly, Qin and Zhang (2016) find that when Chinese villages are connected to the main road network, farmers start specialising in the production of the cash crops which are most suitable to their climate and soil conditions.

Furthermore, this work contributes to the broader 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 prohibitive 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 et al. 2022), 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 2021, Maue et al. 2020). Differently from the similar cross country analysis by Adamopoulos and Restuccia (2022), the estimates of the productivity losses are based on micro level data rather than projections and seem to indicate larger inefficiencies in the geographical crop distribution in Uganda.

The paper 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 estimates the magnitude of the spatial inefficiencies in cropping patterns and the relative importance of misallocation within regional markets. In Section 5, I present the empirical analysis and comment on the findings. Section 6 concludes the paper with some final remarks.

## 2 Theoretical framework

This section presents a simple model of crop choice where for sufficiently high transaction costs between villages and local urban markets, farmers production choices are non-separable from their consumption decisions. Unlike the otherwise rather similar work by Omamo (1998), it includes differences in crop specific land suitability across different villages and as such it allows to draw predictions on the impact of transportation costs on the efficiency of the spatial distribution of agricultural production.[2]

Another important departure from the seminal works on smallholder farmers production choices is that crop decisions are made at the village rather than at the household level. The implicit assumption is that farmers in the same village can freely trade crops among each other and all face the same shadow price price for each crop which in turn is a function of whether the village as a whole is a net seller, a net buyer or autarkic.[3] 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 and as such the empirical analysis is based on village level regressions.[4]

Finally, the model abstracts from other potential sources of interdependence among production and consumption decisions considered by the existing literature on farm household decision making (see De Janvry and Sadoulet (2006) for a recent review) such as self-insurance (Fafchamps 1992; Kurosaki and Fafchamps 2002) and preferences towards local crop varieties not available in the market (Arslan and Taylor 2009). Such dimension are not directly observed in the data but, along with other potentially relevant factors driving farmers crop choices, are discussed in section 2.3 and when commenting on the results.

### 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 boils down to the allocation of land to each crop. The utility is modelled as a function of the consumption of food $c_f$ and another representative good (which is also the numeraire) $c_m$ i.e. $U = U(c_f, c_m)$. Unlike food, this good cannot be produced by farmers and can be thought of as a composite good representing all the non-food items households need to source from the market. 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 $\Delta$, which depends on the location of the village as well as the quality of the transportation infrastructure.[5] Farmers face no transaction costs when buying the "other good" $m$.[6] Finally, each village has an exogenous source of income $\bar{y}$, which can be thought of as the result of some extra agricultural activities.

Formally, the resulting maximisation problem can be described as:

$$\max_{c_f, c_m, l_f, b_f, c_f} U(c_f, c_m) \tag{1}$$

[2] The implicit assumption of homogeneous land made by these seminal studies 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.

[3] The same predictions would be obtained by setting up an household level model where smallholders in the same village cannot trade crops with each other but have the same initial endowment, preferences and face the same (village specific) transportation and relative crop suitability.

[4] It is however important to point out that it is likely that preferences and transportation costs might differ also within households operating in the same villages and that as a result some inefficiencies might be driven by wedges in shadow prices across households in the same village. Since these differences are not observed in the data, they are not modelled in the theoretical section.

[5] For simplicity, transaction costs are assumed to be linear in the quantity bought/sold. Also, unlike Omamo (1998), transaction costs are not crop specific. I discuss the implications of crop specific transportation costs in the next sections.

[6] 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.

subject to:

$$c_m + p_f c_f + \Delta(b_f + s_f) \leq \bar{y} + p_f g_f(l_f) + (p_c - \Delta) g_c(l_c) \quad (2a)$$

$$c_f = g_f(l_f) + b_f - s_f \quad (2b)$$

$$l_f + l_c = \bar{l} \quad (2c)$$

$$b_f \geq 0 \quad (2d)$$

$$s_f \geq 0 \quad (2e)$$

where $b_f$ and $s_f$ indicate the quantity 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 $\bar{y}$ 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. The remaining inequality constraints state that the amount of food crop bought and sold must be non-negative.

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

$$\begin{aligned}\mathcal{L} = & U(c_f, c_m) \\ & + \lambda_1[\bar{y} + p_f g_f(l_f) + (p_c - \Delta) g_c(\bar{l} - l_f) - c_m - p_f c_f - \Delta(b_f + s_f)] \\ & + \lambda_2[g_f(l_f) + b_f - c_f - s_f] + \mu_1 b_f + \mu_2 s_f\end{aligned} \quad (3)$$

The most immediate way to study the qualitative impact of transportation costs on crop choice is to compute the shadow price of food crop $f$, which represents the value farmers attach to the crop and on which the land use decisions are based.

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

$$\frac{\partial \mathcal{L}}{\partial c_f} = 0 \Rightarrow \frac{\partial U(.)}{\partial c_f} = \lambda_1 p_f + \lambda_2 \quad (4a)$$

$$\frac{\partial \mathcal{L}}{\partial c_m} = 0 \Rightarrow \frac{\partial U(.)}{\partial c_m} = \lambda_1 \quad (4b)$$

$$\frac{\partial \mathcal{L}}{\partial l_f} = 0 \Rightarrow (\lambda_1 p_f + \lambda_2)\frac{\partial g_f(.)}{\partial l_f} = \lambda_1 (p_c - \Delta)\frac{\partial g_c(.)}{\partial l_c} \quad (4c)$$

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.[7] By plugging in the first order conditions, the expression becomes $p_f^* = \frac{\lambda_1 p_f + \lambda_2}{\lambda_1} = p_f + \frac{\lambda_2}{\lambda_1}$.[8]

The value of the multipliers can be found by solving the remaining conditions:

$$\frac{\partial \mathcal{L}}{\partial b_f} = 0 \Rightarrow \lambda_2 = \lambda_1 \Delta - \mu_1 \quad (5a)$$

$$\mu_1 \geq 0 \quad (5b)$$

$$b_f \mu_1 = 0 \quad (5c)$$

$$\frac{\partial \mathcal{L}}{\partial s_f} = 0 \Rightarrow \lambda_2 = -\lambda_1 \Delta + \mu_2 \quad (5d)$$

$$\mu_2 \geq 0 \quad (5e)$$

$$s_f \mu_2 = 0 \quad (5f)$$

where the multipliers $\mu_1$ and $\mu_2$ are set to be non-negative since the underlying problem is a maximisation.

This implies that the food crop decision price depends on the village net position in the food market. In particular, where 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, if 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 lays between the consumer and the producer prices.

It follows that unless the village is a net food exporter, the producer price represents a lower bound of the shadow price on which farm households base their production decisions. Therefore, in maximising their utility, farmers might optimally deviate from the profit maximising land allocation giving rise to interdependence between consumption and production decisions. This implies that, *ceteris paribus*, villages in more isolated areas where $\Delta$ is higher will tend to devote a higher share of their land to the food crop. 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 augment this benchmark model to account for heterogeneous land endowment which is captured by different crop specific land productivity.

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

[8] 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: $p_f^* \frac{\partial g_f(.)}{\partial l_f} = (p_c - \Delta)\frac{\partial g_c(.)}{\partial l_c}$.

## 2.2 Land heterogeneity and comparative advantage

Differences in crop specific land suitability are incorporated through the production functions $g_f$ and $g_c$. For the sake of tractability, I assume that they are linear in the quantity of land devoted to each crop:

$$g_f = Fl_f$$

$$g_c = Cl_c$$

where both F and C are village specific.

I assume that utility is increasing in food consumption only until the saturation point $\overline{c_f}$ is met. Once this threshold is reached, utility is a generic function $u$ which is increasing in $c_m$:

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

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.

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:

$$\frac{F}{C} \geq \frac{p_c - \Delta}{p_f - \Delta} \tag{6a}$$

$$\frac{F}{C} \leq \frac{p_c - \Delta}{p_f + \Delta} \tag{6b}$$

$$\frac{p_c - \Delta}{p_f + \Delta} < \frac{F}{C} < \frac{p_c - \Delta}{p_f - \Delta} \tag{6c}$$

The first case applies to villages with relatively high food crop suitability, to the extent that it guarantees higher monetary returns than the cash crop. 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. IN the former case, the excess production is sold to the market to maximise revenues while in the latter it will be produced for self consumption only. Either way, producing the cash crop is never optimal as it generates lower revenues 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); thus, the whole land endowment is optimally devoted to the cash crop. Indeed, when condition (6b) is met, 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.

In both of these scenarios, transaction costs are not pronounced enough to determine any deviation from the profit maximising land allocation and production and consumption decisions are independent. In fact, $\Delta$ only affects the feasible consumption set (as higher $\Delta$s reduce the value of the production and in turn the amount of goods that can be afforded) but it does not impact crop choices. This also implies that villages are fully specialised in the production of the crop that maximises monetary returns.

Conversely, when the transaction costs are high enough, farmers face the situation depicted in Equation (6c). In this case, although the cash crop guarantees higher monetary returns, 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 $m$). In this instance, there is interdependence between consumption and production decisions.

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

$$l_f^* = \begin{cases} \bar{l} & \text{if} \quad \frac{F}{C} \geq \frac{p_c - \Delta}{p_f - \Delta} \\ 0 & \text{if} \quad \frac{F}{C} \leq \frac{p_c - \Delta}{p_f + \Delta} \\ \frac{\overline{c_f}}{F} & \text{if} \quad \frac{p_c - \Delta}{p_f + \Delta} < \frac{F}{C} < \frac{p_c - \Delta}{p_f - \Delta} \quad \text{and} \quad \overline{c_f}F \leq \bar{l} \\ \bar{l} & \text{if} \quad \frac{p_c - \Delta}{p_f + \Delta} < \frac{F}{C} < \frac{p_c - \Delta}{p_f - \Delta} \quad \text{and} \quad \overline{c_f}F > \bar{l} \end{cases} \tag{7}$$

This shows that transportation costs not only result in deviations from profit maximising cash crops, 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 $F$ is, the lower the amount of land devoted to it in equilibrium $l_f^*$ will be.

Figure 1 plots 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. Where there are no transportation costs ($\Delta = 0$), there is full specialisation and villages grow either the cash or the food crop only, depending on which one generates higher revenue. On the other hand, when $\Delta > 0$, 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

**Table 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 percentage 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). *Source: Ugandan Census of Agriculture*.

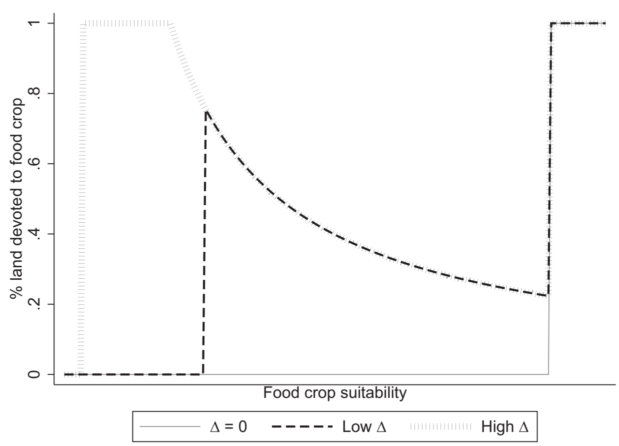


**Fig. 1** Transportation costs and optimal land use

maximising one, the share of land devoted to it is a negative function of F, and therefore crop choices move against 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 $\frac{p_c-\Delta}{p_f+\Delta}<\frac{F}{C}<\frac{p_c-\Delta}{p_f-\Delta}$ i.e. where specialisation does not occur and the crop choice deviates from the structure of comparative advantage.

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

## 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 than what is assumed in the model. The complexity of Ugandan cropping pattern is clearly depicted in Table 1, which provides statistics on the nine most important crops grown in the country by acreage. Two things in particular need to be noticed: first, there is no 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. Coffee is also 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.[9] For this reason, it fully reflects the characteristics of the textbook cash crop.

[9] 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).

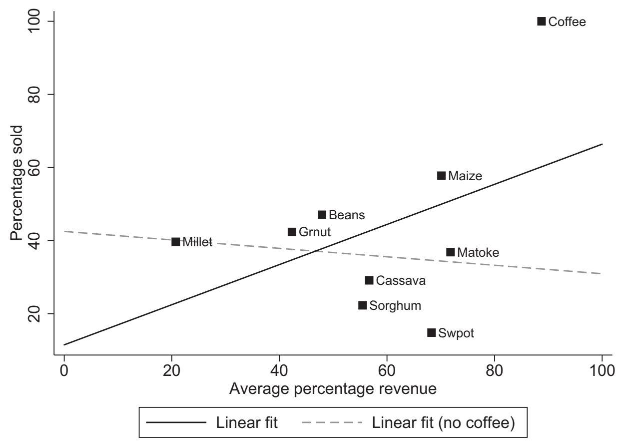


**Fig. 2** Average potential revenue and percentage sold. Percentage sold on the y axis refers to the aggregate production for each crop. Average percentage revenue is obtained as the average ratio of the expected value of the crop and the revenue maximising crop across all villages. The solid line represents the linear fit between the two variables while the dashed one plots the linear fit when excluding coffee

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 revenues and percentage sold is not as clear when all crops are taken into account. In particular, as can be seen in Fig. 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 driven only by coffee). For example, very valuable crops like sweet potatoes and matoke are less commonly sold than less or equally profitable crops like millet, beans and maize.

However, a clear pattern can be identified: Indeed, less perishable/bulky food crops such as cereals, pulses and oil seeds display systematically higher levels of marketisation given their expected monetary returns. 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 light 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 applies equally to the food and the cash crop. However, the fact that in the real 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.[10] However, this possibility is accounted for in the interpretation of the findings from the empirical analysis.

## 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, 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.

### 3.1 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 farms located in 3557 enumeration areas spread throughout the 80 districts of Uganda.[11]

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.

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

[11] 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 reform district borders.

Most of the villages sampled were also geolocated. Thus, it is possible to combine the information contained in the UCA with spatial datasets. Unfortunately, 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 (≤40 percent) and those where the crops under analysis cover a relatively small fraction of the agricultural area (≤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 2941 villages.[12]

## 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 and of terrain and soil characteristics, returns the expected yield per hectare for a number of crops.[13] 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 differences in crop specific productivity even within small areas. Crucially, GAEZ is estimated for the nine most widely grown crops in Uganda (see Table 1). Therefore, it allows to define a very accurate picture of the underlying structure of the agricultural comparative advantage.

Since the model features two crops only, the relative crop suitability was fully captured by the ratio of land productivity of food and cash crop $\frac{F}{C}$. The most obvious way to generalise this to account for more than two crops is to compare the expected yield of each crop to some average fertility measure. Formally, for each crop $c$ in village $v$, I define an index of comparative advantage for $c$ as:

$$cadv_v^c = \frac{dec_v^c}{\sum_c dec_v^c * share^c} \tag{8}$$

where $dec_v^c$ is the decile of land suitability for crop $c$ in village $v$ (obtained from GAEZ), and $share^c$ is the overall fraction of land devoted to crop $c$ in the sample.[14] 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 an 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 Fig. 3 and in Table 2.[15] For each crop, the mean is very close to 1 and the distribution is rather 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.[16]

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

[12] 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.

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

[14] 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.

[15] Plots and descriptive statistics are based on the 2,941 observations corresponding to all the villages included in the final sample.

[16] 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).

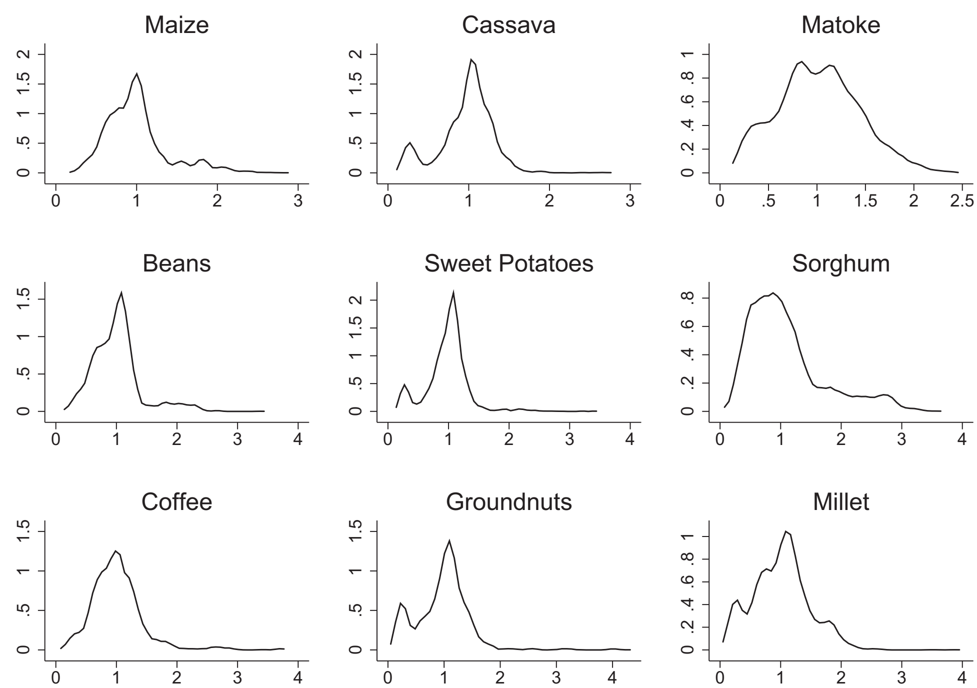


**Fig. 3** Distribution of comparative advantage (Kernel density)

**Table 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 |

interpretation of the results and on the way the index should be entered in regression models.

Finally, Table 3 displays the cross correlation matrix for the indexes of comparative advantage of the nine crops included in the analysis. Since the indexes capture the relative crop suitability, the majority of the correlation coefficients are negative. However, this is not the case for crops that require analogous soil and climatic conditions to grow like cassava and sweet potatoes or maize and sorghum.

### 3.3 Transportation costs and market accessibility

While existing studies define regional markets on the basis of administrative boundaries (i.e. US counties in Costinot and Donaldson (2016) or Peruvian provinces in Sotelo (2020)), 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. Information on the road network and its characteristics come from the map provided by Ubos (2012) and depicted in Fig. 4 which refer to the transport infrastructure available in 2010. To calculate crossing times, I assume that the quickest mean of transportation available is 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 residential 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.[17]

As the villages in the sample are geolocated, I am able to match each of them to the corresponding cell's estimated

[17] 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.

**Table 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 |

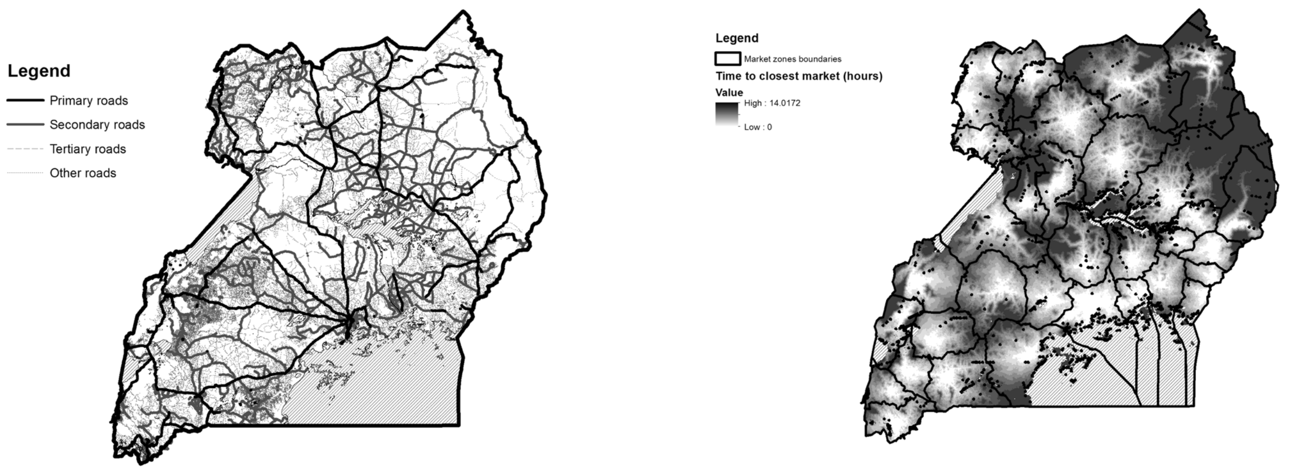


**Fig. 4** Road network and market access

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 better connected to a given major city.[18] As a result, a "catchment zone" is assigned to each of the 32 cities, each representing a regional market. Contrarily to the definitions based on administrative boundaries, this procedure relies on the existing transportation network and therefore it is more likely to capture the real subdivision into market areas. Figure 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 than trying to compute a realistic travel time, the procedure derives a more general market connectivity index based on the existing road network. For this reason, I enter them in the regression models as quintiles which are presumably less affected by the underlying assumptions on road speed. Moreover, this reduces the statistical noise caused by potential errors in the geolocation procedure as quintiles display lower variations among neighbouring cells.

### 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 useful to provide some estimates 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.

[18] 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 market are considered in the computation of the market access index.

**Table 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

The results presented in Table 4 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 (8), is driven by differences within villages located in the same market zones. The same is true when considering the variation in the quintile of the comparative advantage indexes, which will be used in some empirical specifications.

These findings highlight the importance of understanding the determinants of crop choice also within market zones, as well as the frictions that might cause the cropping patterns to deviate from the optimal ones given the underlying system of comparative advantage. In the next section, I compute 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 within the 32 local market zones identified.

## 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 (2022). Intuitively, the procedure aims to compute the difference between the actual value of aggregate output and the one potentially 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:

$$Y^E = \max_{l_{cv}} \sum_{c \in C} \sum_{v \in V} p_c \hat{z}_{cv} l_{cv} \tag{9}$$

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 $c$.[19]

Subject to:

$$\sum_{c \in C} l_{cv} \leq L_v \quad \forall\, v \in V \tag{10a}$$

$$\sum_{v \in V} l_{cv} \leq L_c \quad \forall\, c \in C \tag{10b}$$

$$l_{cv} \geq 0 \quad \forall c \in C \quad , \quad \forall\, v \in V \tag{10c}$$

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.[20]

$Y^E$ can therefore be estimated 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 one third.[21]

In order to estimate how much of the gap in productivity is accounted for by misallocation within the 32 market zones identified in the previous section, I generate 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:

$$Y^m = \max_{l_{cv}} \sum_{c \in C} \sum_{v \in V^m} p_c \hat{z}_{cv} l_{cv} \tag{11}$$

[19] 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 1.

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

[21] This estimate is higher than the one by Adamopoulos and Restuccia (2022), as their corresponding figure for Uganda is around 19%. This difference is due to the fact that I am using micro level data as opposed to regional projections.

subject to:

$$\sum_{c \in C} l_{cv} \leq L_v^m \quad \forall\, v \in V^m \tag{12a}$$

$$\sum_{v \in V^m} l_{cv} \leq L_c^m \quad \forall\, c \in C \tag{12b}$$

$$l_{cv} \geq 0 \quad \forall c \in C \quad , \quad \forall\, v \in V^m \tag{12c}$$

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

In practice, the output (value) maximising crop distribution is calculated 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 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:

$$Y^M = \sum_{m \in M} Y^m \tag{13}$$

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 more than half of the potential 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 within each market zone appear to be inefficient and to play as important role in determining the aggregate spatial 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.

## 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, have a direct impact on crop choices, leading farmers in more remote areas to grow crops with lower market value and/or easier to store and transport (like cereals and grains), and if the impact of comparative advantage is mitigated or even reversed for producers with poor access to markets.

### 5.1 Responsiveness to comparative advantage

First of all, it can be useful to estimate the relationship between the village level share of land devoted to each crop and its relative productivity, abstracting for the time being from the transportation costs. In practice, I compute the following regression:

$$\text{Land share}_{cv} = \beta_1 \text{comparative advantage}_{cv} + X'_{cv}\gamma + \mu_m \tag{14}$$

for each crop $c$. I use a left censored tobit model, in consideration of the the non trivial number of zeros in the dependent variable. Since I have identified significant variation in the measure of comparative advantage both within and across market zones, the regression is estimated both with and without the market fixed effects $\mu$ in order to obtain results that indicate the responsiveness of land use both within and across local markets. The dependent variable is the fraction of agricultural land 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 (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.[22] Controlling for general land fertility is particularly crucial as agricultural potential might be one of the determinants driving the endogenous development of transport infrastructure and as such represent a potential confounding factor for the analysis.

The estimates of $\beta_1$ are displayed in Table 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

[22] 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.

not market level fixed effects are included. Given the above mentioned issues in the interpretation of the results based on the continuous index, the 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 4, there is sizeable variation 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 (2020), there are considerable departures from the efficient crop distribution also within domestic local markets.[23] 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), there is 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.

## 5.2 Market accessibility and cropping patterns

According to the theoretical framework, the first-order impact of low market accessibility is a reduction of the area devoted to cash crops. As argued in section 3.2.3, this

**Table 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 |

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

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, robust standard errors in parentheses

prediction can be generalised to food crops with high market value but less easy to transport due to their perishability or bulkiness. While this 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 Fig. 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:

$$\begin{aligned}\text{Land share}_{cv} = {} & \beta_1 \text{comparative advantage}_v + \beta_2 \text{time to market}_{cv} \\ & + \beta_3 \text{comparative advantage}_v * \text{time to market}_{cv} + X'_{cv}\gamma + \mu_m \end{aligned} \tag{15}$$

[23] 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.

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.[24]

The coefficient of $\beta_2$ 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.[25]

Tables 6 and 7 present the estimates of coefficients $\beta_1$, $\beta_2$ (with and without interaction term) and $\beta_3$ for each of the nine crops considered, where comparative advantage are either captured by the index described in Equation (8) or by the quintile of the corresponding distribution. Due to the above mentioned issues with the interpretation of a linear change in the index, most of the discussion will be focused on the model based on the quintiles (Table 6). For most of the crops, $\beta_2$ 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 1, they are a low value crop mostly grown for consumption, similar to beans and millet across these two dimensions. On this 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).

**Table 6** Transportation crop and crop choice (quintile specification)

| | Time to market | C.adv | Interaction | $R^2$ |
|---|---|---|---|---|
| Maize | 0.43*** | −0.77*** | | 0.37 |
| [av = 16.9] | 0.82** | −0.37 | −0.13 | 0.38 |
| Cassava | 0.03 | 0.23 | | 0.45 |
| [av = 15.5] | 0.05 | 0.25 | 0.01 | 0.46 |
| Matoke | −1.27*** | 1.46*** | | 0.68 |
| [av = 16.6] | 0.09 | 2.78*** | −0.44*** | 0.68 |
| Beans | 0.48*** | 0.28* | | 0.34 |
| [av = 10.6] | 1.11*** | 0.91*** | −0.21*** | 0.34 |
| Spot | −0.57*** | 0.17* | | 0.30 |
| [av = 7.8] | −0.92*** | −0.18 | 0.12 | 0.31 |
| Sorghum | 0.91*** | 0.99*** | | 0.62 |
| [av = 5.8] | 0.73*** | 0.81 | 0.06 | 0.62 |
| Coffee | −0.94*** | 1.38*** | | 0.37 |
| [av = 6.9] | −0.68*** | 1.61*** | −0.08 | 0.38 |
| Gnut | 0.12 | 0.96*** | | 0.35 |
| [av = 5.2] | 0.60*** | 1.46*** | −0.16*** | 0.37 |
| Millet | 0.39*** | 0.71*** | | 0.32 |
| [av = 3.8] | 1.29*** | 1.60*** | −0.29*** | 0.33 |

The controls included are land fertility, average farm size, number of farms. The figures represent the estimated impact of a unit change in the comparative advantage (or transportation costs) 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. Square brackets indicate the average share of land devoted to each crop

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, robust standard errors in parentheses

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 $\beta_3$ 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, $\beta_3$ 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.

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

[25] 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 7** Transportation crop and crop choice (index specification)

| | Time to market | C.adv | Interaction | $R^2$ |
|---|---|---|---|---|
| Maize | 0.42*** | −2.74*** | | 0.32 |
| [av = 16.9] | −0.21 | −4.69*** | 0.66 | 0.32 |
| Cassava | 0.01 | 2.21*** | | 0.42 |
| [av = 15.5] | 0.39 | 0.93 | 0.43 | 0.43 |
| Matoke | −1.28*** | 3.87*** | | 0.69 |
| [av = 16.6] | −0.25 | 6.99*** | −1.04*** | 0.70 |
| Beans | 0.48*** | 0.43 | | 0.31 |
| [av = 10.6] | 0.19 | −0.48 | 0.30 | 0.32 |
| Spot | −0.58*** | 1.24*** | | 0.33 |
| [av = 7.8] | −1.01*** | −0.16 | 0.46 | 0.32 |
| Sorghum | 0.90*** | 1.40*** | | 0.64 |
| [av = 5.8] | 1.25*** | 2.33*** | −0.32* | 0.65 |
| Coffee | −0.92*** | 3.05*** | | 0.38 |
| [av = 6.9] | −1.16*** | 2.32*** | 0.23 | 0.38 |
| Gnut | 0.14 | 1.57*** | | 0.35 |
| [av = 5.2] | 0.61*** | 1.57*** | −0.51*** | 0.39 |
| Millet | 0.39*** | 2.27*** | | 0.33 |
| [av = 3.8] | 1.33*** | 5.06*** | −0.93*** | 0.36 |

The controls included are land fertility, average farm size, number of farms. The figures represent the estimated impact of a unit change in the transportation cost quintile (or comparative advantage index) 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. Square brackets indicate the average share of land devoted to each crop

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, robust standard errors in parentheses

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 maximising crop mix and have lower values of land concentration.

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

$$\mathrm{Y}_v = \beta_1 \text{time to market}_v + X'_{cv}\gamma + \mu_m \tag{16}$$

where X contains the usual controls plus nine variables indicating the village specific suitability for each crop.[26]

The estimates are shown in Table 8. 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 Fig. 1, the higher transportation costs are, the less specialised the production is. In particular, when Δ is equal to zero, farmers always fully specialise in the production of the most profitable crop, while for high Δs full specialisation occurs only when the comparative advantage for either crop is very pronounced. According to these findings however, specialisation (captured by the Herfindahl index of land concentration) is higher 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 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.

## 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

[26] 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).

**Table 8** Marketization and specialisation

| | % 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* | 2941 | 2941 | 2941 | 2941 | 2941 | 2941 |
| adj. $R^2$ | 0.097 | 0.176 | 0.116 | 0.272 | 0.528 | 0.662 |

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)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, robust standard errors in parentheses

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 maximising 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 2020) 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 specialise 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 on crop specific land suitability from 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 spatial distribution of agricultural production.

**Acknowledgements** I am grateful for comments from Carol Newman, Tara Mitchell, Doug Gollin and Salvatore di Falco. I would also like to thank two anonymous referees for their insightful input.

## Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

## References

Adamopoulos T (2011) Transportation costs, agricultural productivity, and cross-country income differences. Int Econ Rev 52(2):489–521

Adamopoulos T (2019) Spatial integration, agricultural productivity, and development: a quantitative analysis of Ethiopia's road expansion program. In: Proceedings of Meeting Papers, number 86. Society for Economic Dynamics

Adamopoulos T, Restuccia D (2022) Geography and agricultural productivity: cross-country evidence from micro plot-level data. Rev Econ Stud 89(4):1629–1653

Arslan A, Taylor JE (2009) Farmers' subjective valuation of subsistence crops: the case of traditional maize in Mexico. Am J Agric Econ 91(4):956–972

Chen C, Restuccia D, Santaeulàlia-Llopis R (2022) The effects of land markets on resource allocation and agricultural productivity. Rev Econ Dyn 45:41–54

Costinot A and Donaldson D (2016) How large are the gains from economic integration? Theory and evidence from US agriculture,

1880–1997. Technical report, National Bureau of Economic Research
De Janvry A and Sadoulet E (2006) Progress in the modeling of rural households' behavior under market failures. In *Poverty, inequality and development*, Springer, p 155–181
Dorosh P, Wang HG, You L, and Schmidt E (2010) Crop production and road connectivity in Sub-Saharan Africa: a spatial analysis
Fafchamps M (1992) Cash crop production, food price volatility, and rural market integration in the third world. Am J Agric Econ 74(1):90–99
Gollin D, Parente SL, Rogerson R (2007) The food problem and the evolution of international income levels. J Monet Econ 54(4):1230–1255
Gollin D, Rogerson R (2014) Productivity, transport costs and subsistence agriculture. J Dev Econ 107:38–48
Gollin D, Udry C (2021) Heterogeneity, measurement error, and misallocation: evidence from African agriculture. J Political Econ 129(1):1–80
Jayne TS (1994) Do high food marketing costs constrain cash crop production? Evidence from Zimbabwe. Econ Dev Cult Change 42(2):387–402
Key N, Sadoulet E, Janvry AD (2000) Transactions costs and agricultural household supply response. Am J Agric Econ 82(2):245–259
Kurosaki T, Fafchamps M (2002) Insurance market efficiency and crop choices in Pakistan. J Dev Econ 67(2):419–453
Li N (2017) Marketization costs and household specialization: evidence from Indian farmers
Maue CC, Burke M, and Emerick KJ (2020) Productivity dispersion and persistence among the world's most numerous firms. Technical report, National Bureau of Economic Research
Omamo SW (1998) Transport costs and smallholder cropping choices: an application to Siaya District, Kenya. Am J Agric Econ 80(1):116–123
Qin Y, Zhang X (2016) The road to specialization in agricultural production: evidence from rural China. World Dev 77:1–16
Restuccia D and Santaeulalia-Llopis R (2017) Land misallocation and productivity. Technical report, National Bureau of Economic Research
Restuccia D, Yang DT, Zhu X (2008) Agriculture and aggregate productivity: a quantitative cross-country analysis. J Monet Econ 55(2):234–250
Sotelo S (2020) Domestic trade frictions and agriculture. J Political Econ 128(7):2690–2738
Stifel D, Minten B (2008) Isolation and agricultural productivity. Agric Econ 39(1):1–15
Tombe T (2015) The missing food problem: trade, agriculture, and international productivity differences. Am Econ J: Macroecon 7(3):226–58
Uganda Bureau of Statistics (2012) Roads: Uganda, 2008–2010. Uganda Bureau of Statistics. The link is https://geodata.lib.berkeley.edu/catalog/stanford-rd848my5087. Accessed in August 2020