Agronomic Endowment, Crop Choice and Agricultural Productivity*

Rishabh Sinha — The World Bank

Xican Xi — University of Fudan

February, 2018

Abstract

Using plot level data on actual and potential yields together with actual crop choices, we perform an accounting exercise to measure aggregate gains in agricultural productivity in five large countries in Sub-Saharan Africa due to the closing of the actual-potential yield gaps and changes in crop choices. We find large aggregate gains in productivity in all countries. The two channels – changes in yield and changes in crop choice account are equally relevant in accounting for the gains for all countries except Tanzania where crop selection plays a secondary role. To discipline the accounting gains, we develop a model of crop selection in presence of imperfect substitutability in preferences across crops in presence of crop-specific subsistence requirements and risks to show why optimal crop selection might deviate from an optimization of nominal output.

Keywords: Crop Choice, Agricultural Productivity, Sub-Saharan Africa, Intermediate Inputs

*Email: rishabhsinha@worldbank.org and xicanxi@fudan.edu.cn. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1 Introduction

A principal finding of the developing accounting literature is the fact that gaps in agricultural productivity across countries are much larger compared to gaps in productivity in other sectors (Gollin *et al.* (2002), Restuccia *et al.* (2008)). The closing of the agricultural productivity gap becomes even more crucial from the perspective of closing the gap in aggregate because developing countries allocate a much large share of its resources in the agricultural production. The two general set of reasons why agricultural productivity lags in developing countries can be that either these countries have a poor endowment of resources (soil, temperature, rainfall) necessary for agricultural production or that there are inefficiencies in how these resources are used. The stark differences in policy implications from the two broad determinants make it imperative to discern their relative importance in driving low productivity in developing countries.

The role of geography in agricultural productivity is not trivial. Agricultural production differs from production elsewhere because it bears a close association with the geographical features of the region in which it is practiced. Rainfall, temperature and soil quality are some of the exogenous factors that make agricultural activity productive in certain regions compared to others. Additionally, a region may be suitable for cultivation of a particular crop based on geography but may yield dismal output if used to cultivate crops that require significantly different geographical attributes. In a recent paper, Adamopoulos & Restuccia (2017) employing an accounting framework and grid data on crop-specific yields and land use from Global Agro-Ecological Zones (IIASA & FAO (2012)) show that cross-country variations in agricultural productivity are not driven by differences in the agronomic endowment. In other words, productivity in poor countries is not low because they have lower potential yields but because the actual yields lie far from potential yields. In this paper, we first perform a deeper accounting exercise in the spirit of this recent study by focussing on five large economies in Sub-Saharan Africa that are home close to half of the total population of the region.

A key feature of our accounting exercise is that we allow for two cropping seasons. Agricultural production in many Sub-Saharan countries is characterized by cropping seasons which generally coincides with the rainy seasons. These cropping seasons also interact with other weather patterns like temperature making it possible to grow certain crops in one season but not in others. Abstracting from multiple cropping seasons in the accounting exercise with counterfactual crop choices can potentially give rise to spurious findings. To understand this, consider a farm that cultivates maize during the first season and soybean during the second. Assume that maize and soybean have the highest returns among crops that can be grown during the two seasons respectively. In such a case, the present land utilization points to an optimal scenario. An exercise that considers a common crop choice across seasons will require cultivating soybean during the first season if it has higher returns than maize. Yet, agronomic conditions make it extremely difficult to grow soybean during the first season. The choice of maize and soybean during the two seasons is not hypothetical as many regions in the Northern Guinean Savannah in Nigeria (Anambra, Edo, Enugu, Ondo etc.) face exactly the same choice.

We consider a set of counterfactual exercises. In the first exercise, we ask how much of the productivity gains can be realized via the closing of the actual-potential yield gap in these countries. Another contribution of our paper is that we perform this analysis where potential yields also respond to agricultural practice – with respect the intensity of intermediate usage and water supply, in addition to agronomic endowments. A useful feature of the GAEZ data which provides potential yields for different scenarios helps us quantify the marginal effects of these variables. Like geography, these variables have varied effects on crop-specific yields for the same farm. We find that there are significant gains in aggregate productivity from closing the yield gap for three of the five countries, even when potential yield is measured for least sophisticated agricultural practice. However, our accounting shows that aggregate output decreases for Democratic Republic of Congo and Nigeria when actual yields are counterfactually changed to potential yields under the scenario of less developed practice indicating that on aggregate these countries are producing more than the worst case scenario. Consistent with Adamopoulos & Restuccia (2017), we find huge gains in aggregate output when potential yields at high levels of inputs are used. We also find a somewhat marginal increase in aggregate output by using irrigation facilities after a high-level of intermediate use has already been put to use. Subsequently, we also highlight the tremendous variation in productivity gains across regions within the five countries.

As the data available to us is on the crop level, it is also possible to quantify gains if crop choices are allowed to vary at the yields actually realized on the farms. Our results show that productivity gains from changes in crop choices surpass the gains from closing the yield gap considering even the most optimistic yield scenario for three countries. In a final exercise, we measure gains when crop choices are allowed to change under different potential yield scenarios. We find a notable role for the crop selection channel. With the exception of Tanzania in which this channel can account for 10 percent of gains alone, it accounts for a much larger share for other countries with the range of gains being accounted for ranging between 45 – 60 percent for these countries. Interestingly enough, the marginal contributions of the two channels account for less than half of combined gains in Tanzania. This points to large gains from the interaction of the two channels in the country with crop selection amplifying the gains from the other channel three-fold.

Though our accounting framework suggests that there are large gains from either channel, a pertinent question is whether the accounting gains will hold under a richer economic environment. In other words, how realistic are these potential gains? To answer this, we propose a model of crop selection featuring elements that are important for crop selection decisions under heterogenous yields across crops. Specifically, we allow for imperfect substitutability in preferences across crop types in presence of subsistence requirements in certain crops together with crop-specific risks of production. Our counterfactual exercises produce substantial changes in crop shares in aggregate output in nominal terms which motivates our analysis in a general equilibrium setting. We derive important implications from the model that qualitatively illustrates why optimal crop selection under an economic environment may be markedly different from the naive optimal under the accounting framework which simply maximizes the nominal agricultural output. We leave the quantitative analysis for the future.

2 Accounting Framework

We use a simple aggregation framework similar to Adamopoulos & Restuccia (2017) to quantify the importance of crop selection. Each country i is endowed by a fixed number of farms given by F^i. The area of a farm f is denoted by l_f which is allocated to production of crops. An important way in which we extend the framework in Adamopoulos & Restuccia (2017) is by allowing for multiple cropping seasons. As discussed previously, due to their warm climatic conditions, agriculture in many countries in the Sub-Saharan Africa are characterized by more than one cropping season. For the purpose of our analysis, we restrict our attention to two cropping seasons which is the case for most farms. Given this, instead of choosing over individual crops, each farm f has to choose among different crop combinations. The choice set of a farm f is given by C_f that contains finite number of elements. A typical element of the crop choice set C_f can either contain two crops (for example, maize which can be grown in the first season and sorghum which can be grown in the second) or a single crop (for example, cassava which has longer growing period and engages farm for both seasons). Note that there are some sub-country level regions in which there exists only one cropping season and in these cases each element of the crop choice set contains a single crop. The nominal yield of a crop choice $c \in C_f$ is given by

$$P(c) = \sum_{g \in c} p_g A_{g_f} \tag{1}$$

where g denotes the individual crops in c and price and agricultural yield of individual crop g is given by p_g and A_{g_f} respectively. Since our data on crop choice and yields from GAEZ is available at farm (grid) level, we are able to use actual crop yield A_{g_f} at the farm level. If θ_{cf} represents the share of farm area dedicated to cultivation of crop choice c, then the total agricultural output of country i is given by

$$Y^i = \sum_{f \in F^i} l_f \sum_{c \in C_f} P(c)\theta_{cf} = \sum_{f \in F^i} l_f \sum_{c \in C_f} \theta_{cf} \left(\sum_{g \in c} p_g A_{g_f} \right) \tag{2}$$

To avoid clutter in notation, we drop the country script i from now on.

The first step in our accounting exercises is to set the benchmark by estimating the value of actual agricultural output. Given actual crop yields A_g^A and share of area allocated to different crop combinations θ_{cf}^A, actual agricultural output Y^A is estimated by plugging these values in equation 2

$$Y^A = \sum_{f \in F^i} l_f \sum_{c \in C_f} \theta_{cf}^A \left(\sum_{g \in c} p_g A_{g_f}^A \right) \tag{3}$$

Throughout our accounting exercises, we will keep prices p_g and farm land endowment l_f fixed.

This implies that the two ways in which output can change counterfactually is through changes in yield (A^A_{gf}) and crop choice (θ^A_{cf}). The crop-specific gridded yield data from GAEZ under different scenarios helps us in making counterfactual changes in the former. To fix notation, suppose the yield is a function of the state of operation that ranges from $1, \ldots, S$. Under GAEZ agricultural scenarios, the state of operation depends on two factors – the intensity of intermediate input usage and whether irrigation is practiced on the farm or not.

In the first set of counterfactuals, we ask how much will the aggregate output change if actual yields are counterfactually changed to GAEZ yields keeping the crop choices unchanged. Counterfactual output Y^s can be obtained by plugging in counterfactual yield A^s_{gf} at the farm (grid) level

$$Y^s = \sum_{f \in F^i} l_f \sum_{c \in C_f} \theta^A_{cf} \left(\sum_{g \in c} p_g A^s_{gf} \right) \tag{4}$$

In the second counterfactual exercise, we consider changes in crop choice without changing the yields. An important point to note here is that since only a small subset of all possible crops are actually cultivated on a farm, our crop choice in this exercise is restricted by crops that are actually cultivated and not encompass the entire crop choice set that can be grown on a farm obtained by the crop calendar. Counterfactual output Y^θ is obtained by choosing θ^*_{cf} to optimize the returns on the farm

$$Y^\theta = \sum_{f \in F^i} l_f \sum_{c \in C_f} \theta^*_{cf} \left(\sum_{g \in c} p_g A^A_{gf} \right) \tag{5}$$

The optimization essentially reduces down to choosing an element c^* of the crop choice set C_f that has the maximum nominal yield.

In the final counterfactual exercise, we estimate the output changes when we allow for changes in crop choices when instead of actual yields we use crop-specific yields under different GAEZ scenarios. This implies that given a set of crop-specific yields A^s_{gf} for a farm f, there is an element c^*_s in crop choice set C_f that maximizes nominal yields under the scenario s. Potentially the optimal elements c^*_s can differ across the scenarios if the yield gains under GAEZ scenarios vary disproportionately across crops. For example, it is possible that a certain crop can see dramatic rise in yields with better intermediate input use which other crops may not be able to replicate. The counterfactual agricultural output $Y^{s,\theta}$ when both yields and crop choices are changed is hence given by

$$Y^{s,\theta} = \sum_{f \in F^i} l_f \sum_{c \in C_f} \theta^{s*}_{cf} \left(\sum_{g \in c} p_g A^s_{gf} \right) \tag{6}$$

where θ^{s*}_{cf} places a unitary weight on element c^*_s.

Having laid down the accounting framework, we now briefly discuss the data that we use to perform our accounting exercises.

3 Data

We carry out our accounting exercise for five large countries in the Sub-Saharan Africa – Democratic Republic of Congo, Ethiopia, Kenya, Nigeria and Tanzania. Together these countries account for just under half of the total population of the Sub-Saharan Africa. Moreover, like most developing countries agriculture is an important activity both in terms of accounting for employment and value added in all these countries.

There are two broad sources from where we obtain our data. The data on crop specific yields (A_g), crop choices (θ_c) and land endowment (l_f) is taken from the Global Agronomic Ecological Zones (GAEZ) database. The database contains geographic gridded data on actual crop yields for more than 15 different crops including cereals crops (wheat, rice etc.), sugar crops (sugarcane, sorghum etc.), oil crops (soybean, groundnut etc.) among others. The database also contains yield data for vegetables (cabbage, carrot, onion and tomato combined) but given that actual land allocated to the vegetables is not much, we drop them from our analysis.

More importantly for our analysis, the GAEZ data also provides potential crop yields at grid-level under different scenarios. These scenarios map to the agricultural practice being adopted at the farm level and reduce to the use of intermediates in production and the nature of water supply. The intermediate usage is classified on three levels – low, intermediate and high, whereas the nature of water supply depends on whether the agricultural production depends on rainfall entirely or employs some irrigation technique in addition.

The low intermediate use yields refer to agricultural practices that is largely subsistence based and is labor intensive. The production also does not use any meaningful amount of nutrients and chemicals, and also occurs in absence of any conservation practices. At the intermediate input level, the production is also performed for market participation in addition to subsistence needs. The production can also employ better varieties of seed and use hand tools, livestock and preliminary levels of mechanization. The labor intensity is lower compared to the low input case. There is also some use of fertilizers and chemicals for pest and disease control together with fallowing and conservation. The high input case basically refers to modern agricultural practice that happens in most of developed world. The production is done entirely for market purposes with complete mechanization of the agricultural process. There are no shortfalls in the use of fertilizers, chemicals etc.

An important element of our accounting exercise is the land endowment of a farm. However, the area data obtained from GAEZ corresponds to harvested area by crops and not the actual area of the farm. We find evidence of the practice of multiple cropping in many cases where the sum of harvested area on a farm is greater than the actual area. Hence, we need to first convert the harvested area obtained from GAEZ to the actual farm area. As the process is somewhat involved we discuss the details in an appendix section. The main intuition behind the exercise is to basically identify the maximum land that could have been cultivated in the two cropping seasons given the actual crop choices, and then use this as our measure of land endowment of the farm.

We need to complement the data on actual and potential yields together with actual crop choices obtained from GAEZ data with two additional data from FAO to complete our data requirement. First, in order to aggregate agricultural output from crop level, we need prices. We tap into the FAO data and recover the country-level prices of different crops and use the median price in our analysis. The second more important piece of information that we use from the FAO database is the harmonized crop calendar.

The FAO database classifies regions of a country into different AEZs and then provides a crop calendar for each AEZ. The number of AEZs vary across countries and range from 9 in the Democratic Republic of Congo to 29 in Kenya. Table 4 provides the list of AEZs for the five countries. The first step is to construct the crop choice set for each AEZ in each country. The crop calendar data from the FAO has information on cropping onset period together with growing period for each crop. Using this information, it is possible to identify the different crop combinations that are feasible for a particular AEZ. There is a huge variation in the number of crop combinations that can be practiced in an AEZ both across and within countries. For example, the elements in the crop choice set ranges within a narrow band of 4–6 in the 6 Nigerian AEZs. On the other hand, crop combinations vary anywhere from 2 in Upper Midland 4 to 124 in Lower Midland 4 in the 29 Kenyan AEZs. The second step is to assign a particular grid in GAEZ data to a particular AEZ so that the AEZ specific crop choice set can be applied at the grid level. To do this, we map districts/states at country level to different AEZs given information in the FAO crop calendar database together with temperature and rainfall information. Once we have made this mapping, we have all elements necessary to carry out the three set of counterfactual exercises mentioned in the accounting framework. In the next section, we discuss the results of these exercises.

4 Counterfactual Exercises

In the first counterfactual exercise, we ask how much would the aggregate agricultural output change if the actual yields are changed to GAEZ potentials under various scenarios leaving the crop choices unchanged. As outlined in the framework, this primarily means using different values of crop yields under different scenarios and employing equation 4. The three different GAEZ scenarios that we use throughout the quantitative analysis are – low input use under rainfed cultivation, high input use under rainfed and high input usage under irrigated cultivation. Table 1 shows the findings of the exercise where columns (1) – (3) correspond to the three scenarios listed earlier.

The first finding from the exercise is that actual output is, in fact, higher than the potential output under the least productive scenario in which low level of inputs is used under rainfed cultivation for Democratic Republic of Congo and Nigeria. This suggests that on aggregate these countries have moved above the least productive scenario. On the other hand, the remaining three countries still can gain by closing the actual–potential gap even at the least sophisticated level of cultivation. The gains are also not insignificant for any of the countries with output nearly increasing by a half for Tanzania.

The second observation is rather expected with the output gains increasing as agricultural scenarios become more and more advanced. Looking across countries, we notice that there is a notable variation in gains. Agricultural output approximately doubles under high input use for the Democratic Republic of Congo. However, these gains are much smaller compared to other countries in the sample. Output in Ethiopia, Kenya and Tanzania increase by at least 300 percentage under the high input scenarios. The table also sheds some light on the relative importance of irrigation with respect to inputs in driving higher output. The findings suggest a slightly limited role from irrigation after high inputs have been put to use. For example, moving from low inputs to high inputs keeping cultivation dependent on rainfall generates an output gain of 7 – 11 times for the three countries in which actual output lags potential output under the least productive scenario. Yet, moving further up by using irrigation delivers much lower gains. Additional gain in aggregate output ranges from a paltry 14 percent of actual output to 95 percent of actual output for Tanzania.

The aggregate gains reported in table 1 hide the variations observed at the regional level within the countries. Figure 1 shows the changes in aggregate output under different scenarios for the district/state at the 10$^{\text{th}}$ (black bar), the 50$^{\text{th}}$ (gray bar) and the 90$^{\text{th}}$ (light-gray bar) percentiles of each country. The verticle scales are made country-specific to make the graphs readable.

First, focus on the first scenario in which agriculture is practiced using low inputs and is rainfall dependent. The aggregate output was lower under this scenario for both Democratic Republic of Congo and Nigeria. However, these losses are spread differently across the regions. Though most regions report a decline in output in the Democratic Republic of Congo, the losses for the bottom 10 percentile is a half of actual output compared to a marginal decline of 3 percent for the region at the 90$^{\text{th}}$ percentile. A similar pattern is observed in Nigeria where the top 10 percentile regions report gains compared to aggregate losses.

Looking within countries for the other two scenarios, we again find similar heterogenous gains across regions. The aggregate gains are more unevenly distributed across the regions in Democratic Republic of Congo and Nigeria, and also to some extent in Ethiopia with most gains occurring for a smaller number of regions. For example, the output gain in terms of percent gain from the actual output is more than 3.5 times as large as the bottom 10 percentile region for the top 10 percentile region in the Democratic Republic of Congo. Whereas, the gain in top regions is less than 2.5 times the gains in the bottom 10 percentile region in Tanzania. Finally, we also find that a secondary role of irrigation in driving gains after high inputs have already been put to use is preserved at various percentiles across countries.

We now move on to our second accounting exercise. The goal here is to bring crop selection as another potential factor determining aggregate output and quantitatively evaluate its importance relative to a change in yields-only approach. To do this we use the data and estimate counterfactual output using equations 5 and 6. The results of the accounting exercises are presented in Table 2.

Column (1) reports the gain in output when the optimal element of crop choice set is chosen under actual yields. Remember that in this exercise the optimal selection is restricted to only those choices which are actually observed at the farm level and this restricted set is sometimes much smaller than the choice set constructed using GAEZ crop calendar information. As expected, the counterfactual output is higher for all countries. However, the gains for each country is significant suggesting the importance of crop selection in the accounting exercise. At the lower end, Democratic Republic of Congo and Tanzania see an expansion of the size of their actual output followed by Nigeria for which gains are twice as large with output growing to three-fold its original size. However, the gains are much higher for Ethiopia (5.3 times actual output) and Kenya (around 6.5 times actual output).

Columns (2) – (4) report the output gains when crop selection is optimized under potential yields under GAEZ scenarios. Columns (2) and (3) refer to low and high input use scenarios under rainfed cultivation whereas column (4) corresponds to high input use practice under irrigation. Following from before, the output gains are highly variant across countries. The gains are smaller for Democratic Republic of Congo and Nigeria and higher for Kenya and Tanzania. Optimal crop selection at lowest productive scenario leads to an output expansion of 25 percent in the Democratic Republic of Congo and 44 percent in Nigeria. On the other hand, output grows by 1.25 times in Ethiopia and more than twice in Kenya and Tanzania. Yet, the gains increase by multiples when higher inputs are put to use even with no change in water use. The smaller output gains in Democratic Republic of Congo and Nigeria in column (2) expand by 8 – 9 times in column (3). The gains are also much higher for other countries with Kenya topping the list with almost a 10-fold increase in actual output.

Finally, column (4) reports the output gains from optimal crop choice under the most productive scenario in which irrigation techniques are used in addition to high inputs. The additional gains in column (4) compared to column (3) are somewhat lower compared to what is obtained from moving to low to high inputs. The expansion is particularly muted for Democratic Republic of Congo and Nigeria. There is a further expansion of 20 and 30 percentage points in the former and the latter respectively. Though, the marginal gains from irrigation techniques are larger for Tanzania where output expands by an impressive 13 times compared to actuals.

In order to evaluate the importance of changes in crop choice relative to changes in crop yields, we report the results of the previous exercise together in table 3. Column (1) reports the output gains under optimal crop selection at actual yields while column (2) refers to output gains when highest crop yield scenario is used under actual crop choices. Column (3) describes output gains combining the two factors – optimal crop choices under the highest yield scenario. Comparing column (1) and (2), we find that in three out of the five countries, the gains in output due to changes in crop selection only outweigh the output gains when only yields are changed. Among the five countries, the actual crop selection in Tanzania is closest to what is suggested by the simple optimal rule that maximizes nominal returns. Moving to optimal crop choices under actual yields accounts for less than 10 percent of the maximum gains for Tanzania. The channel accounts for a much larger share for other four countries and ranges from 45 – 60 percent of the maximum gains. In contrast, the gains from changes in yields as a share of maximum gains vary over a smaller range with the channel accounting for one-third of potential gains in Kenya and Tanzania and one-half of potential gains for Nigeria. The interesting fact emerging from the table is that the sum of marginal contributions of the two channels accounts for less than half of maximum gains in Tanzania. This essentially points to large gains from the interaction of the two channels in the country.

In summary, we conclude that our accounting framework suggests that both changes in crop selection, as well as crop yields, can deliver large gains at the aggregate for the countries in Sub-Saharan Africa. For most countries, the two channels can be equally effective in driving the maximum potential gains though there are large interaction gains that we find for Tanzania.

4.1 Need for an Economic Model of Crop Choice

Though our accounting framework suggests that there are large gains from either channel, the more important question is whether the accounting gains hold under economic analysis. In other words, the accounting analysis abstracts from many factors that are important for crop selection. Additionally, crop selection might also endogenously respond to changes in higher yields.

To gain perspective on large gains obtained under the accounting analysis, in the next section we present a model of crop selection that captures some of the key elements of crop selection mechanism. The first element of our theory is that there is imperfect substitutability across the various crop types – people do not just want to consume cereals but also pulses, tubers, and other crops. This means that all crops, at least when measured on a broad level, are cultivated in equilibrium. If preferences across crops are such, then as a first consequence this rules out the specialization in a particular crop. Note that this specification does not necessarily rule out specialization at the household level, but at the common market aggregation. If there are no costs to participate in the market economy, a household can specialize in a particular crop and exchange a part or whole of its output for other crops in the market.

The second feature of our model is that there are subsistence requirements in certain food crops. This is based on the idea that nutritional requirements precede consumption of other nature can be traced back to the *food problem* put forward in Schultz (1953). In this context, crops that are directly used for nutrition like cereals and pulses will have subsistence requirements as opposed to crops that are grown for non-nutritional consumption. This specification also ties to the vast literature on structural transformation where preferences are assumed to be non-homothetic in the sense that the share of agricultural consumption shifts away from agriculture to manufacturing and services as an economy grows richer.[1] The third element of our theory is that agricultural production is risky and risks are crop-specific. For instance, certain crops are more resistant to the shortfall in water and hence have lower rainfall risks compared to other crops. Additionally, the correlation of risks across a pair of crops reduces to a portfolio allocation problem studies in crop choice literature (Fafchamps (1992), Rosenzweig & Binswanger (1993) etc.).

[1] See Herrendorf *et al.* (2014) for a detailed discussion on non-homothetic preferences.

Finally, we study the crop choice problem in a general equilibrium setting where prices adjust to balance demand and supply at the crop level. The price of a crop that is relatively higher compared to other crops will face downward pressure as more of it is produced in the economy. Figure 2 shows the changes in nominal crop shares for two broad crop categories – cereals and sugars combined (left panel), and roots and tubers combined (right panel). The horizontal axis represents actual crop shares while counterfactual shares are marked on the vertical axis. The red squares correspond to the first counterfactual exercise in which only the yields are changed (to high input irrigated case) and the blue dots point to changes in crop share when crop choices are changed keeping yields at actual levels. Looking at the two panels, we recognize that the share of cereals and sugars increase under the first counterfactual at the expense of a contraction in the share of roots and tubers. An opposite state is observed when crop choices are varied keeping yields constant. The change in nominal shares is also significant as the squares and dots lie far from the 45-degree line. Such large deviations suggest an active role for price adjustment. Based on the above motivation, we develop a model of crop selection that features the above elements in the next section.

5 A Model of Crop Choice

We focus our attention to a two crop case. The first crop, which we term *subsistence (s)*, is the one that has positive subsistence requirements. On the other hand, there are no subsistence requirements in the second crop which we define as the cash crop (c). Both crops can be produced on a farm and can be sold or bought in the market at the market price. The model can be extended to include multiple crops in which subsistence requirements are crop specific. However, a two-crop case suffices to draw out the three main mechanisms determining crop choices outlined before.

There are L units of land available in the economy that can be used for agricultural production. As there is no other use for land, the basic economic problem is to allocate the land towards the cultivation of two crops. Land is the only factor of production and the production technology of a crop $g \in \{s, c\}$ given by

$$Y_g = A_g(\theta_g L) \tag{7}$$

where Y_g is the crop output and A_g is the crop yield. The share of land allocated towards cultivation of g is denoted by θ_g and feasibility requires that $\theta_c = 1 - \theta_s$ with $\theta_s \in [0, 1]$. The agricultural yield of the two crops depends on the state of the economy and is joint log-normal with mean μ and variance-covariance matrix Σ.

The utility of the representative household in state s is given by

$$\frac{1}{1-\rho}(C(s))^{1-\rho} \tag{8}$$

where ρ is the parameter capturing constant relative risk aversion and $C(s)$ is the constant elasticity of substitution aggregate of individual crop consumption augmented for subsistence needs

$$C(s) = \left(\sum_{g \in \{s,c\}} \alpha_g^{1/\sigma} (c_g(s) - \bar{c}_g)^{(\sigma-1)/\sigma} \right)^{\frac{\sigma}{\sigma-1}} \tag{9}$$

The parameter α_g are the demand-shifters and $\bar{c}_g > 0$ introduce non-homotheticity in spirit of Stone-Greary preferences. As we are concerned with showing crop choices when some crops have subsistence requirements, we set $\bar{c}_c = 0$.

First, let us consider the case where there is no uncertainty and there are no subsistence requirements in s. Then, the equilibrium land allocation is obtained using the standard property of CES preferences together with the result that prices are inverse of yields under linear technology (choosing the rental rate of land as the numeraire) described earlier. The share of land allocated to subsistence crop s under no subsistence and no uncertainty is given by

$$\theta_{NS,NU} = \frac{\alpha_S A_s^{\sigma-1}}{\sum \alpha_g A_g^{\sigma-1}} \tag{10}$$

As seen in the above equation, the assumption of CES preferences immediately rules out complete specialization if the elasticity of substitution σ is not arbitrarily high. Change in share allocated to subsistence depends on whether the two crops are gross complements ($\sigma \in (0, 1)$) or are gross substitutes ($\sigma > 1$). Using a much finer classification of crops, Allen & Atkin (2016) find the elasticity to be significantly higher than unity. Though, at a high enough aggregation in which cash crops can be interpreted as non-agricultural consumption, the evidence from the structural transformation literature would suggest the broad crop categories to be gross complements. We discuss the results here assuming the elasticity to be somewhere between 0 and 1 and note that the qualitative results would be opposite if instead the elasticity is higher than 1. Going back to the above equation, an exogenous increase in yield of subsistence crop leads to a reduction in the share allocated to its cultivation and hence cultivation moves away from the crop that becomes easier to produce.

Now consider the case in which there are subsistence requirements. The share allocated to subsistence crop in such a case is given by

$$\theta_{NU} = \theta_{NS,NU} + \frac{\bar{c}}{LA_s} \cdot \frac{1}{\left(1 + \frac{\alpha_s}{\alpha_c}\left(\frac{A_s}{A_c}\right)^{\sigma-1}\right)} \tag{11}$$

There are many valuable insights that can be derived from the above result. The first observation is that as long as subsistence requirement is strictly positive ($\bar{c} > 0$), the share allocated to subsistence crop is strictly greater than under the no subsistence case. To see how land endowment matters, consider the case when the demand-shifters α_g and the elasticity σ is set to 1. The second term on the right-hand side of equation 11 reduces to the ratio of subsistence requirement $\bar{c}$ and the maximum possible output of the subsistence crop (LA_s). This implies that the share allocated to subsistence crop becomes smaller as an economy becomes more able to satiate its need for subsistence consumption, and this increase can result either through an expansion of land resources or via an increase in exogenous yield of the subsistence crop.

The third effect links to the substitutability channel. Assuming that the two crops are gross complements, a relative increase in the exogenous yield of the cash crop c makes the subsistence requirement less strict and leads to a decrease in land share allocated to the subsistence crop. An increase in productivity of cash crop increases the total amount of cash crop that can be produced using before-increase land allocation. The complimentary in preferences means that some of these gains in cash crop output need to be balanced by increases in consumption of the subsistence crop. However, any such increase in consumption of subsistence crop via increased land share allocated to its production also partially fulfills the subsistence requirement primarily making subsistence requirements less stringent. Therefore, an economy that has lower crop yields face an increased burden of subsistence and gains in aggregate agricultural productivity will be higher as the economy transitions from lower yields due to subsistence requirements becoming less binding and aligning crop selection decisions towards aggregate output optimization conditional on complementarity assumptions.

Finally, the introduction of risks in agricultural production highlights the fact that crop choices are not independent of the volatility of the returns. The crop choice decision primarily reduces to a portfolio choice in which higher returns are traded for lower volatility. Our specification of log-normal crop yields, allows us to solve for closed-form solutions of crop choice used extensively in the portfolio allocation literature (Campbell & Viceira (2002) and deliver standard results.

In this section, we developed a model of crop choice to advance from the naive accounting framework that captures some of the essential features of the decision making. The discussion above details why actual crop choices may deviate from the naive optimal of the accounting framework. In our future work, we aim to put discipline on the gains obtained from the accounting framework by taking our economic model to the data.

References

Adamopoulos, Tasso, & Restuccia, Diego. 2017. Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data. *York University and University of Toronto*.

Allen, Treb, & Atkin, David. 2016. Volatility and the Gains from Trade. *NBER Working Paper No. 22276*.

Campbell, John Y, & Viceira, Luis M. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press.

Fafchamps, Marcel. 1992. Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World. *American Journal of Agricultural Economics*, **74**(1), 90–99.

Gollin, Douglas, Parente, Stephen, & Rogerson, Richard. 2002. The Role of Agriculture in Development. *American Economic Review*, **92**(2), 160–164.

Herrendorf, Berthold, Rogerson, Richard, & Valentinyi, Akós. 2014. Growth and Structural Transformation. *Pages 855–941 of: Handbook of Economic Growth*, vol. 2. Elsevier.

IIASA, & FAO. 2012. Global Agro-Ecological Zones (GAEZ v3.0). *IIASA, Laxenburg, Austria and FAO, Rome, Italy*.

Restuccia, Diego, Yang, Dennis Tao, & Zhu, Xiaodong. 2008. Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis. *Journal of Monetary Economics*, **55**(2), 234–250.

Rosenzweig, Mark R, & Binswanger, Hans P. 1993. Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *The Economic Journal*, **103**(416), 56–78.

Schultz, Theodore William. 1953. The Economic Organization of Agriculture. *New York: McGraw-Hill*.

Table 1: Closing Actual–Potential Yield Gap

	Percent change in Yield		
Inputs Water Supply	Low Rain-fed (1)	High Rain-fed (2)	High Irrigated (3)
Congo, Dem. Rep.	-36	88	102
Ethiopia	32	367	450
Kenya	40	314	380
Nigeria	-16	174	230
Tanzania	47	347	442

Table 2: Gains with Optimal Crop Choice

	Percent change in Yield			
Inputs Water Supply	Actual Actual (1)	Low Rain-fed (2)	High Rain-fed (3)	High Irrigated (4)
Congo, Dem. Rep.	116	25	217	238
Ethiopia	532	125	628	948
Kenya	645	275	943	1110
Nigeria	197	44	421	450
Tanzania	108	217	838	1330

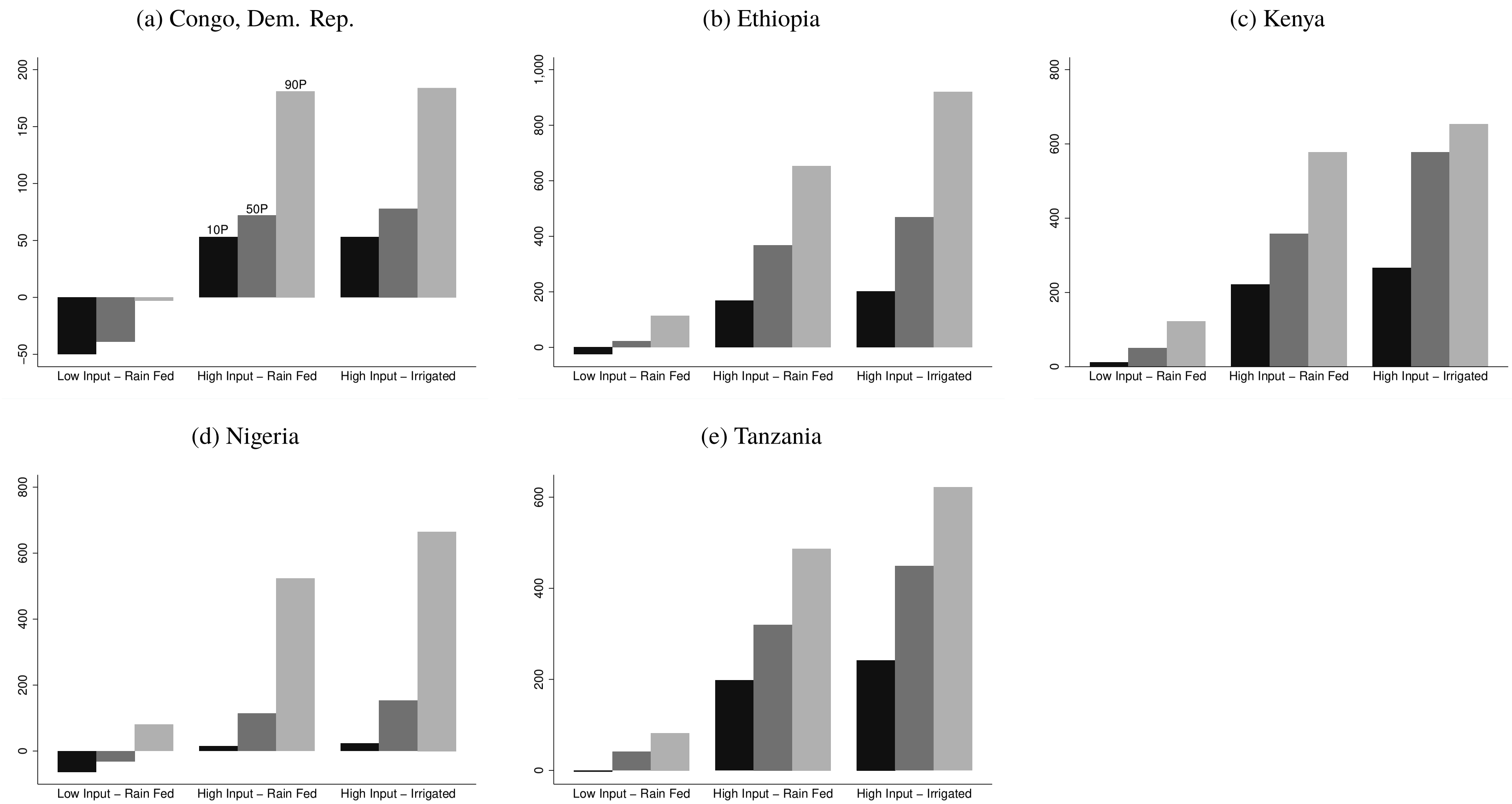
Table 3: Gains from Optimal Crop Choice vs. Actual–Potential Gap Closing

	Percent change in Yield		
Inputs	Actual	High	High
Water Supply	Actual	Irrigated	Irrigated
Crop Selection	Optimal	Actual	Optimal
	(1)	(2)	(3)
Congo, Dem. Rep.	116	102	238
Ethiopia	532	450	948
Kenya	645	380	1110
Nigeria	197	230	450
Tanzania	108	442	1330

Table 4: Agro-Ecological Zones

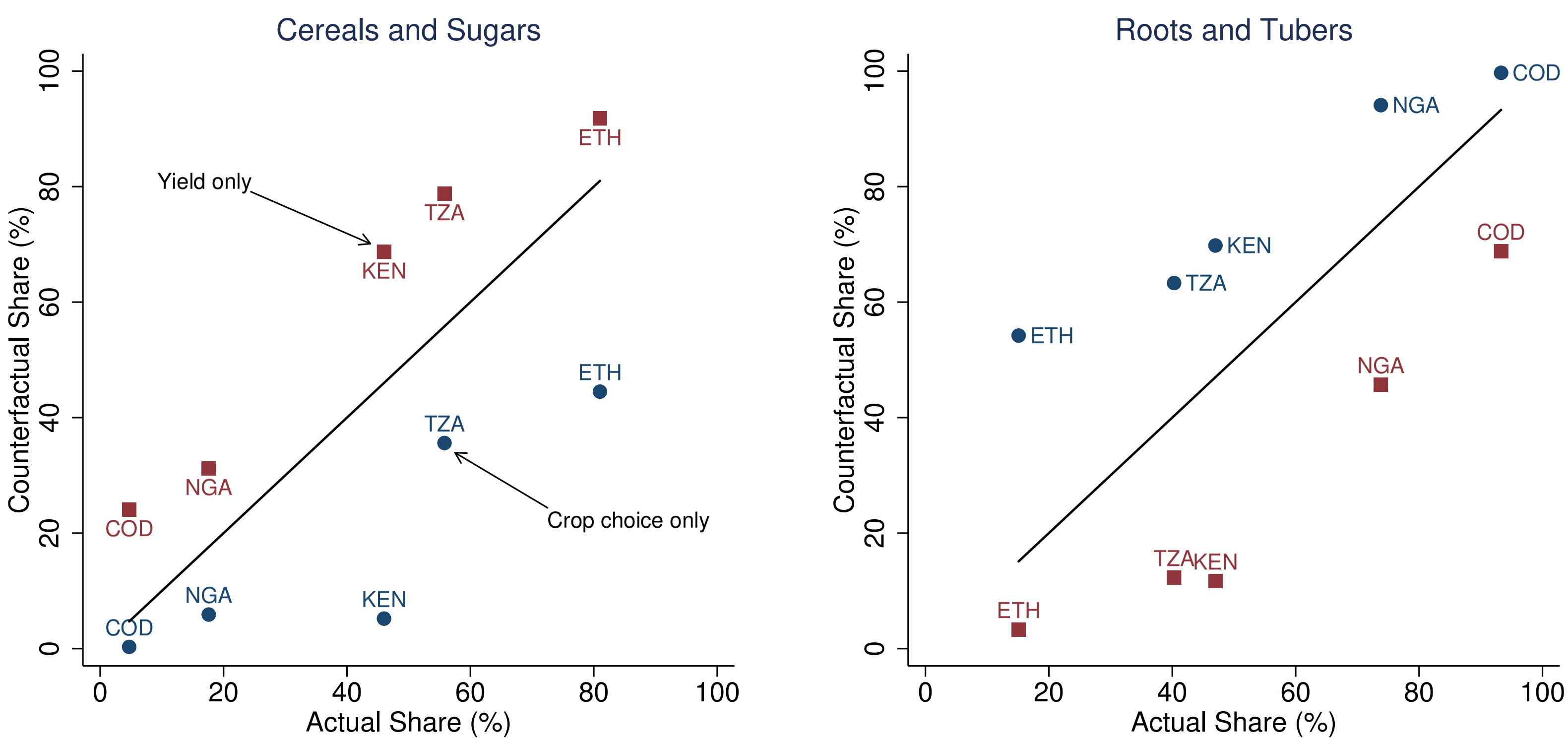
Congo, Dem. Rep	Ethiopia	Kenya	Nigeria	Tanzania
Central Basin	Arid	Lower Highland 1–5	Derived Savannah	Central Plateaux
Centre South	Humid	Lower Midland 1–6	Humid Forest	Coastal Plains
East	Moist	Lowerland 2/Inner Lowland	N. Gui. Savannah	E Plateaux & Mountain Blocks
North	Per-humid	Lowerland 3–7	S. Gui. Savannah	High Plains
Southeast	Semi-arid	Tropical Aphine 1	Sahelian Savannah	Rukwa-Ruaha
Southwest	Sub-humid	Tropical Alphine 11	Sudanese Savannah	Ufipa Plateau
	Sub-Moist	Upper Highland 1–4		Volcanoes & Rift
		Upper Midland 1–6		Western Highlands

Figure 1: Within Country Variation in Yield Gains – Percent Change in Yield



The above figures show percentage change in yields when actual yields are counterfactually changed to different GAEZ scenarios keeping the crop choices unchanged. The black, gray and light gray bars represent yield change at the 10th, 10th and 10th percentile respectively.

Figure 2: Changes in Nominal Crop Share



Appendix: Crop Choice and Agricultural Productivity

A Crop Calendar

FAO provides crop production information pertaining to sowing and harvesting periods for more than 130 crops located in close to 300 AEZs for African countries. We use the FAO crop calendar information to construct the crop choices available within a specific AEZ.

A1 Crop Choice Set Construction with an Example

To be completed.

A2 Adjustment for Multi-Season Crops

There are some crops in certain AEZs that can be grown in both the long and the short season. We are unable to attribute season specific yields for such crops. As such, we assume that the crop yields are invariant across the seasons. This assumption makes some elements of the crop choice set redundant. For example, consider a crop choice set which contains the following two elements:

1. Crop A in the long season, crop B on the short season
2. Crop B in the long season, crop A on the short season

Given that the crop yields are invariant across seasons, the total crop production is the same for both elements listed above. As such, we drop one of the elements from our constructed choice set to eliminate such redundancies.

We make one more adjustment to our crop choice. Consider the case in which the following two elements are present in the crop choice set:

1. Crop A in the long season, crop B on the short season
2. Crop A in the short season

Under our assumption of invariant crop yields, the second choice is dominated by the first as total production is higher in the former. Hence, we drop the second element from the crop calendar choice set.

B Land Endowment from Harvested Area

In this section, we outline how we estimate the total land endowment for a country from the data on harvested area by crops obtained from GAEZ. Land endowment is different from harvested area and multiple cropping of the same land over agricultural seasons will make harvested area to exceed the land endowment. To make progress, we classify the crops under three groups:

1. Annual Crops: These are crops with long growing periods such that once land is utilized for their cultivation, it is not possible to use it to grow any other crop in the same year.
2. Single Seasonal Crops: These are crops that have shorter growing periods and allows for multiple cropping of the same land over different seasons. For example, rice can be grown in combination with wheat with the cultivation of each crop spread across the two seasons. However, these crops can only be grown in one of the two seasons only.
3. Double Seasonal Crops: These crops are similar to Single Seasonal with the one exception. They can be grown in either of the two seasons. For example, rice can be grown in the long season together with growing wheat in the short season. Alternatively, it may also be possible to grow rice in the short season with maize been cropped in the long season. In this example, rice is a Double Seasonal crop that can be grown in either of the seasons.

The above classification allows to separate the crops that can be grown in a season t in three mutually exclusive and exhaustive sets. The total area of a farm is the estimated using the following expression

$$a_f = \sum_t \sum_{c \in A_t} a_c^{GAEZ} + \max_t \Big\{ \sum_{c \in S_t} a_c^{GAEZ} + \psi_t \sum_{c \in D_t} a_c^{GAEZ} \Big\} \tag{12}$$

$$\tilde{a}_{ft} = \sum_{c \in C_{S_t}} a_{fc}^{GAEZ} + \psi_t \sum_{c \in C_{D_t}} a_{fc}^{GAEZ} \tag{13}$$

where a_{fc}^{GAEZ} is the harvested area for a crop c obtained from GAEZ, and S_t and D_t are the sets of Single Seasonal and Double Seasonal Crops respectively. The parameter ψ_t represents the share of total harvested area allocated to the season t. Given that only a small fraction of total land under cultivation is used in the short season, we expect $\psi_L > \psi_S$ and use a value of $\psi_L = 0.8$ for our benchmark calculations. We also report results with different specifications of this benchmark share.

The rationale behind the above expression is rather straightforward. Total area under cultivation is the sum of the area allocated to the cultivation of crops that are annual in nature and the maximum of the area used to cultivate non-annual crops over the two seasons.