Economies of scale and ecology of scope

Frederik Noack[1] and Martin Quaas *[2]

[1]University of British Columbia

[2]German Centre for Integrative Biodiversity Research & University of Leipzig

January 3, 2021

Preliminary - do not cite

Abstract

Farmers in Uganda grow substantially more crops per unit of land than their counterparts in developing countries. While specialization is often related to economic efficiency, a large ecological literature establishes the positive relationship between species diversity and productivity. Here, we model and estimate the private benefits of crop diversity in order to understand why Ugandan farmers are more diversified than their counterparts in developed countries. Based on empirical facts derived from 65,851 fields, we develop a model that combines the efficiency increasing mechanisms of specialization from the economics literature with economies of scope resulting from ecological interactions and other input complementarities. The model predicts that crop diversity enhances productivity in low input agriculture but also that these private benefits of diversity decline with increasing control over the production environment. We then test these predictions with plot level data from the low input agriculture of Uganda. We find that a 10 percent increase of farm size and a 10 percent increase of labor increases crop diversity by 1.6 and 1.8 percent respectively consistent with the presence of both economies of scale and economies of scope. Holding labor and land fixed, we also find that a 10 percent increase of crop diversity increases revenues by 3 percent, in line with economies of scope. Comparing harvest quantities of crops grown individually and in combination with other crops we show that the economies of scope are likely to result from ecological interactions. Planting a crop in combination with one other crop increases harvest quantities per unit of input by 14 % compared to monocultures. These results suggest that the high levels of crop diversity in

*For their helpful comments we thank Joséphine Gantois and Eric Strobl as well as participants of the seminars at the Development Economics Seminar at the Vancouver School of Economics, the Environmental Economics Seminar at UBC's Land and Food Systems, the Conference of the Western Economic Association, the Annual Conference of the Canadian Resource and Environmental Economics Association and the Annual Meeting of the American Economic Association.

Uganda are determined by the interaction between economies of scope and economies of scale. They also suggest that crop diversity is an important productivity enhancing input for agriculture of Uganda and not the result of uninsured risk.

Keywords: Crop diversity; biodiversity; regulating ecosystem services; agriculture; economic development

1 Introduction

Households in developing countries rarely specialize and farmers in Uganda are no exception. They grow on average four different crops per hectare, substantially more than their counterparts in developed countries. The literature on development economics and economic growth often interprets this lack of specialization as indicative for inefficiency (Romer 1987; Banerjee and Duflo 2007). The literature suggests that uninsured risk and malfunctioning or absent markets for credit, inputs and outputs prevent households from specializing in the most productive activity (e.g. Karlan et al. (2014)). The general conclusion of this stream of literature is that specialization increases efficiency due to increasing returns to scale and through utilizing comparative advantages. Applying this reasoning to agriculture in Uganda implies that the private benefits of crop diversity are small compared to the gains from specialization and that specialization increases agricultural production.

In contrast, a large literature in ecology studying the role of diversity for ecosystem functioning generally finds that diversity increases the level and stability of ecosystem productivity, due to ecological complementarities between species (Hooper et al. 2005). This literature suggests that species diversity enhances efficiency because different species utilizes different niches i.e. different combinations of resources or the same combination of resources but at different points in time. Biodiversity thus helps to use the resources more efficiently. In agriculture, this may not only apply to ecological resources such as nutrients, space or light, so-called regulating ecosystem services (Millennium Ecosystem Assessment 2005) or regulating nature's contributions to people (Díaz et al. 2018), but also to other inputs such as labor and capital. Different crops require labor and capital inputs at different points in time therefore increasing the effective labor and capital endowments of a rural household. This ecological concept is closely related to economies of scope in the economics literature (Panzar and Willig 1981).

In this paper we develop a microeconomic model based on empirical facts of agriculture in Uganda that combines the economic theory of specialization with the benefits of diversification suggested by the ecology literature. In the second part of this paper we test the predictions of the model using 65,851 field level observations from Uganda.

To establish the empirical basis for our model assumptions we first explore the relationship between crop diversity in Uganda and variables that are commonly associated with diversification or specialization. These variables include 1) risk exposure, 2) market access, 3) increasing returns to scale, 4) comparative advantages and 5) and input complementarities or ecological niches from the ecology literature. Farmers in Uganda perceive droughts as the most important reason for income shortfalls. However, we find no evidence that crop diversity in Uganda is correlated with rainfall risk. In contrast to rainfall risk, market participation is associated with specialization in Ugandan agriculture. However, the low revenue differences across crops also suggest little gains from specialization due to comparative advantages. Lastly, we do find evidence for the presence of crop specific fixed costs which suggests economies of scale as well as evidence for economies of scope resulting from complementarities between crops.

Based on these observations, we develop a model that includes complementarities between crops, increasing returns to scale as well as returns to specialization due to comparative advantages. In line with the ecological literature (Harpole and Tilman 2007) we make the importance of these ecological complementarities or niches dependent on the ability of the farmer to control the environmental factors of production. In other words, the ability to homogenize the environment e.g. through irrigation to regulate the water supply, application of fertilizer to regulate nutrients or greenhouses to regulate temperature, reduces the number of ecological niches and thus the benefits of crop diversity.

Our theory predicts that crop diversity increases with farm size and labor. This positive relationship between farm size, labor and crop diversity results from the interplay between the economies of scope and the economies of scale. While the complementarities between crops scale with farm size and labor, the mechanisms that drive the economies of scale remain constant. However, this relationship between farm size and crop diversity weakens with the increasing ability of the farmer to control the environment which reduces the ecological niches and thus the benefits of diversification. Our theory predicts further that crop diversity increases productivity as a consequence of the ecological and economic complementarities that create the economies of scope. However, this productivity enhancing impact of crop diversity declines with increasing revenue differences across crops and with the increasing control over the environmental or ecological resources for crop growth.

We then test these predictions using the plot level data from the living standard measurement survey of Uganda. The main advantage of these data is that they contain observations of most crops grown in monocultures and mixed with other crops within the same farm or even the same plot but in different years. These data allow us to distinguish between different mechanisms behind the economies of scale and the economies of scope. Because farmers in Uganda have limited control over environmental factors with almost absent fertilizer inputs, irrigation systems, pesticide uses or other possibilities to homogenize the environment, we focus on the model predictions for agricultural systems with low revenue differences and low abilities to control environmental factors.

In an ideal experiment we would randomize the labor and land endowments of a household to test our prediction about the impact of farm size and labor on crop diversity. Land and labor markets are underdeveloped in Uganda and more then 90 % of the labor is family labor while farm size changes are largely driven by the frequent inheritance of land. We therefore use changes in the household composition as well as land inheritance as instruments for labor and land respectively to mimic the ideal experiment described above. Our second model prediction concern the impact of crop diversity on productivity. To test these predictions we would like to vary crop diversity randomly and observe the corresponding revenues while holding land and labor constant. This experiment is difficult for two reasons: First, there are few exogenous changes in crop diversity. One example is the crop diversity decline caused by the cassava mosaic virus. However, these changes affect the household also directly through e.g. losses from infected crops. Second, changes in crop portfolio composition affect the agricultural outcomes not only through changes in diversity but also directly through the identity of crops. Revenues differ across crops and removing a crop type from the portfolio will also alter the expected revenues of the portfolio. Lastly, inputs such as labor respond to crop diversity. Even if we could randomize crop diversity, we would also have to convince the farmer to not adjust the input levels. We therefore use a differences-in-differences approach comparing the changes in crop diversity and changes in revenues across farms or across plots within one farm while controlling for labor. This approach has also another advantage: It allows us to distinguish between ecological and economic complementarities. While labor and capital complementarities are more pronounced at the farm level, ecological complementarities are mostly relevant when crops are grown on the same plot. The results of the estimation on plot and on farm level therefore inform us about the relative contributions of ecological and economic complementarities to the economies of scope and the economies of scale.

Measurement errors in agricultural data has recently attracted substantial attention (Gollin and Udry 2019; Desiere and Jolliffe 2018; Abay et al. 2019; Lobell et al. 2020). One advantage of our data is that they also report GPS measurements of field sizes. We use these GPS measurement to address potential biases from land measurement error. Another problem is that small harvest quantities are often over-reported which could bias our results. For example, we may underestimate or overestimate the contribution of crop diversity to production, if crop diversity is correlated with the area allocated to individual crop stands and therefore to individual harvest quantities. However, the detailed crop and plot specific data allow us to compare crop harvest quantities for the exact same crop stand area of a crop that is planted individually and in combination with other crops.

Our first set of results show that both farm size and labor increase crop diversity. A 10 percent increase in farm size increases crop diversity by 1.6 percent in our preferred regression specification. Farm labor has a similarly positive effect. Increasing farm labor by 10 percent increases crop diversity by 1.7 percent in our preferred regression specification. These results remain unchanged when using household composition and land inheritance as instruments for labor and land respectively as well as using the GPS land measurements to correct for land measurement error.

Our second set of results concern the impact of crop diversity on farm productivity. Our main finding is that crop diversity increases productivity. Every thing else equal, a 10 percent increase in crop diversity at the farm level increases revenues at the farm level by 3.4 percent. However, this result could be biased if crop diversity is correlated with crop patch size and harvest from smaller patches is systematically overreported. In a second approach we therefore estimate the impact of crop diversity on crop harvest at the crop patch level. In this approach we compare harvests from the same crops in pure and mixed stands everything else constant. These results suggest that growing a crop in combination with one other crop increases the harvest per unit of input by 14 percent. In contrast to the results at the farm level, the results on crop patch level capture exclusively the ecological interactions. However, this specification also allow us to control for diversity at the farm level to capture the complementary in other inputs such as labor and capital. Our estimates for crop diversity at the farm level suggest that crop diversity at the farm level has not additional impact on crop patch level outputs. We interpret this result as evidence for the importance of ecological interaction for agricultural productivity in the low input agriculture of Uganda.

The main contribution of our paper is to combine the proposed drivers for specialization and diversification from the economic and the ecology literature in one unified theoretical framework based on empirical facts from agriculture in Uganda. The empirical tests of our predictions support the underlying assumptions. Our study is therefore related to the literature that studies the economic consequences of crop diversity. An important contribution is the paper by Weitzman (2000) which studies the trade-off between the benefit from agricultural specialization and the risk of catastrophic infection in a theoretical framework. The main focus of his paper is the trade-off between individual gains from specialization through increased revenues and the contribution of individual specialization to the probability of aggregate agricultural collapse. Similar to the Weitzman paper, Brock and Xepapadeas (2003) assume that the benefit of crop diversity results from the impact of crop diversity on crop failure but their main focus is on developing a theoretical framework to value this contribution.[1] Bellora and Bourgeon (2019) add the externalities from pesticide use as well as the gains from specialization due to comparative advantages to these trade-offs. In the theoretical framework of our paper, we add the input complementarities or imperfect competition which is central to the ecological literature (e.g. Vandermeer (1992)) as well as fixed costs and increasing returns to

[1] Gollin et al. (2000) study the value of gene banks for developing pest are disease resistant crop varieties.

scale from the economics growth literature to this framework.

In contrast to these purely theoretical papers we motive our model by empirical observations and test the model predictions using plot level data from Uganda. Empirically, our paper builds on the work by Smale et al. (1998), Di Falco and Chavas (2009) and Bellora et al. (2017) who study the impact of crop diversity on the level and stability of crop production in Pakistan, Ethiopia and South Africa respectively.[2] In contrast to their papers, we do not take crop diversity levels as given but focus on the mechanism explaining diversification in agriculture. This mechanistic understanding of the drivers and implications of crop diversity allows us to make counterfactual predictions about the impact of economic development and market integration on the value of crop diversity.

The rest of the paper is organized as follows. In section 2, we present the data for farms in Uganda and in section 3 we derive stylized facts as the basis for our model. Our model is developed in section 4. The statistical methods to estimate model predictions are explained in section 5, empirical results in section 6. Sections 7 and 8 quantify the value of crop diversity and discuss counterfactual simulations of the impact of control over the environment, increasing price differences and labor markets on crop diversity. The final section contains the discussion and conclusion.

2 Data

We use plot level data from the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) from Uganda to motivate our modeling decisions and for testing the predictions of the model empirically. The small scale agriculture in Uganda is an ideal setting for our study because the low levels of inputs making the ecological interactions more pronounced and because of the high levels of crop diversity. Crop production in Uganda is not only diverse within farms but also within fields as intercropping or mixed cropping (planting a mix of crops on one plot) is common practice. Observing the same crop planted as monoculture and in combination with other crops allows us to isolate ecological interactions from economic mechanisms that operate at the farm level. Self-reported data have their limitations. However, for highly diverse small scale agriculture with rampant intercropping they may still be best source of information. We discuss how we address potential biases of self reported data below.

[2] Auffhammer and Carleton (2018) address a similar research question using district level data from India. Michler and Josephson (2017) and Tesfaye and Tirivayi (2020) take a very different approach estimating the direct impact of crop diversity on poverty and consumption respectively. On global scale, Renard and Tilman (2019) estimate the impact of crop diversity on output variance.

2.1 Household data

We use the data of the LSMS-ISA from Uganda. The survey is representative at the national level, the urban and rural level and the regional level (North, East, West and Central regions). The households were tracked over time but a share of the sample was replaced in 2012 by a new random sample of households based on the updated sample frames developed for the Population and Housing Census. The LSMS-ISA data are plot level information on seasonal crop production including data on crop type, inputs and other production practices, and yields. The survey is composed of structured interviews on 3,200 households chosen based on the stratified random sample of the Uganda National Household Survey in 2005/06. Interviews were conducted in five rounds between 2009/10 and 2015/16 with visits in each of the two growing seasons per year. However, we include the initial Uganda National Household Survey of 2005/06 as well but exclude the urban stratum to focus on the rural production, which results in a sample of 66,193 complete crop-plot and 17,837 complete household level observations. The survey design is described in detail on the LSMS-ISA web page. In the following we discuss the measurement of our key variables.

Land: Farmers in Uganda may not know the exact size of their plot.[3] In fact, farmers in Uganda systematically over-report the size of small plots and small farms while they under-report the size of large plots or farms (Carletto et al. 2013; Dillon et al. 2019). However, 60 % of the plots were also measured using GPS. To test for potential bias in self-reported plot sizes we estimate

$$\log(\text{land gps}_{it}) = \theta \log(\text{land reported}_{it}) + \varepsilon_{it}$$

where land gps_{it} is the plot size measured by GPS signal and land reported_{it} is the plot size reported by the farmer. We then predict plot size based on this estimated relationship for all self-reported values. In a robustness test, we use this predicted land size measure bootstrapping standard errors across both levels of estimation.

Land is measured in there levels of aggregation. The lowest level of aggregation is the *crop-plot level* or the *crop stand*. This is the share of a plot which is occupied by one crop. While the size of the crop stand equals the plot size for monocultures it diverges when multiple crops are planted in combination on the same plot. We use the crop cover percentages reported by the farmer to allocate the plot area to individual crops. The next level of aggregation is the *plot level*. A plot or field is a piece of land that is managed as one unit. It could be managed as monocultures with one crop type or in multicropping with several crop types. Lastly, *farm size* is the sum of cultivated plots. The measure excludes farm land under fallow, rangelands or native vegetation.

Previous studies show that measurement error is larger on small plots Burke and Lobell

[3] A plot is defined as a contiguous piece of land on which a specific crop or a specific crop mixture is grown.

(2017). We therefore exclude all crop stands under 100 m^2 in our main specification and under 1000 m^2 in a robustness test. Similarly, we exclude all farms below 0.1 ha in all specifications. We also exclude outliers of crop stands above 10 hectare, crop stands below 100 m^2 and farms above 100 hectares.

Labor: Labor was recorded on plot level in person days. In the following we combine family and hired labor. We allocate plot level labor to individual crops according to the crop shares discussed above.

Revenues and yields: Production is recorded in local physical units. We convert these units to kilograms (kg) using the median district level conversion factors from the survey. To aggregate outputs across different crops we use crop prices. Only a small fraction of the crops are sold (see Table 1). We therefore use median crop prices on year, season and district level from the survey to convert physical quantities to revenues. Although crop prices probably play a minor role in the planting decisions of farmers they may still approximate marginal utility from consumption and therefore contain welfare relevant information. In the crop specific regressions we use harvest quantities in kilograms without conversion to revenues. We further exclude all observations with less than 1 kg harvest or more than 10,000 kg harvest per plot to address misreporting and influential outliers but we show the results using the full range of data in the appendix.

Crop diversity: There are many different measures of diversity in ecology. Most of these measures combine the number of species and their relative abundance. The most commonly used diversity measures (Simpson index, Shannon index and species richness) are variations of the effective number of species defined by

$$D = \left(\sum_{i=1}^{n} p_i^{\alpha} \right)^{\frac{1}{1-\alpha}}$$

where p_i is the relative abundance of species i (Hill 1973). The main difference between the specific indices is the weight placed on abundance compared to mere presence-absence information. For $\alpha = 0$ the index equals species richness i.e. the number of species or crops that are present in a sample, for $\alpha \to 1$ the index equals the Shannon index, while for $\alpha = 2$ the index equals the inverse Simpson index. We use crop richness in our theoretical approach and main empirical specification compare our results to results based on the Simpson index as the other extreme of these diversity indices. Another advantage of the Simpson index is that it is equivalent to the Herfindahl–Hirschman Index of market concentration widely used in economics.

To measure relative abundance, we use the area planted with a crop relative to the total cropping area. For plots with intercroopping, we allocate the total plot area to individual crops according the the shares reported in the survey. The advantaged of using area planted compared to measures based on outcomes in the harvesting season (e.g. revenues or area harvested) is that the area planted is independent of weather events during the growing cycle as well the impact of crop diversity on the outcome itself.

2.2 Weather data

We use weather data as controls in our regression specifications and to measure production risk. Farmers reported drought as the single most important reason for income shortfalls followed by floods as the second most important reason for income shortfalls (see Figure 5). Both are directly related to weather events. We use two gridded data sets to measure positive and negative precipitation deviations that may cause droughts and floods. The first precipitation data set is from the Tropical Rainfall Measuring Mission (TRMM) which was produced by a joint space mission between NASA and the Japan Aerospace Exploration Agency to measure tropical rainfall. The data product combines measurements from precipitation radar, microwave imager, visible infrared scanner, clouds & earths radiant energy system and lightning imaging sensor. The data set starts in 1998 and has a 0.25° spatial resolution. We use this data set to measure rainfall.

The second data set is the Standardized Precipitation Evapotranspiration Index (SPEI) based on CRU TS 3.24 weather data and the FAO-56 Penman-Monteith estimation of potential evapotranspiration. The index combines temperature and precipitation data to measure the difference between the temperature dependent evapotranspiration and the measured precipitation level. The spatial resolution is 0.5°.

We calculate district level average monthly weather and merge these data to the household data using the district name as identifier.

2.3 Summary statistics

The following table provides the summary statistics of our household data. Overall, gross crop revenues per farm are low and farms are small. The mean revenues per year are 610 in 2010 USD ($\approx$ 1,776.5 in 2018 PPP dollars) while the mean farm size is 1.1 hectare (ha). In contrast to the small farm size, labor is abundant. The mean farm labor per season is 141 days combining family labor and hired labor. However, the share of hire labor is low. The mean share of hired labor is about 1.4 %. This low level of market participation is also apparent with the share of the harvest that is sold measured in percent. The mean share of the harvest that is sold is around 21 %. Crop diversity in contrast is high, both within and across plots. The median share of the farm land that is allocated to intercropping (defined as growing several crops on one plot at the same time) is 44 % and farms grow on average more than four crops on their land.

Table 1: Summary Statistics

Statistic	Mean	St. Dev.	Median	Q25	Q75	Min	Max
Revenues [2010 USD per year]	601.1	2,879.5	255.4	108.0	561.9	2.3	189,317.0
Share sold [%]	20.0	20.8	16.7	0	33.3	0	100
Farm size [ha]	1.1	2.1	0.7	0.4	1.2	0.1	82.6
Share inherited [%]	64.4	44.6	100.0	0.0	100.0	0.0	100.0
Labor [days per year]	276.9	253.9	214	120	352	2	6,074
Share hired [%]	1.4	7.6	0	0	0	0	100
Number of crops	4.3	1.9	4	3	5	1	16
Share intercropping [%]	44.1	37.9	50.0	0.0	75.0	0.0	100.0

The table reports the mean (Mean), the standard deviation (St. Dev.), the median (Median), the lowest (Q25) and highest (Q75) quartile as well as the minimum value (Min) and the maximum value (Max) of our 17,837 household level observations. Revenues are expressed in PPP international dollars per season and household. 'Share sold' is the share of the harvest that is sold. The remaining part is consumed by the household. 'Farm size' is the sum of cultivated plots. 'Share inherited' is the share of the land that was inherited by the household. 'Labor' is the sum of labor across all plots cultivated by the household. 'Share hired' is the share of this labor that was hired. 'Number of crops' is the number of crops grown by a household per season while 'share intercropping' is the share of the household's plots with intercropping.

3 Facts about crop diversity in Uganda

In this section we relate crop diversity to variables that the literature commonly associates with diversification. Specifically we focus on risk and market functioning from the economic literature as well as resource complementarities from the ecological literature. Here, we do not aim at establishing causal relationships that explain high levels of diversity but rather to identify patterns that form the basis of our theoretical framework. Before we relate crop diversity to common explanations of diversification we describe general cropping pattern.

3.1 Cropping pattern

Uganda is located at the equator with most of Uganda experiencing uniform warm and humid climate throughout the year. There are two cropping season in Uganda. The first cropping season in Uganda starts in January and ends in June. The second cropping season starts in July and ends in December. The tropical climate not only allows for almost continuous cultivation but it also supports a wide variety of crops. Figure 1 shows typical agricultural landscape and intercropping of beans, cassava, bananas and taro in Western Uganda. Figure 2 shows the frequency and the co-occurrence of the 10 most common crops in our data set. The shading of the antidiagonal elements represent the number of plot level observations of these crops in our

Figure 1: Typical farm landscape and intercropping of banana, beans, cassava and taro in Western Uganda (photo: Frederik Noack)

data set. The figure suggests that crop production in Uganda is diverse and not dominated by a few crops. The most common crop (beans) constitute about 20 % of all crop observations. The fields on both sides of the anitdiagonal (the off-antidiagonal fields) depict the co-occurrence of crops. They show how often a pair of crops occur on the same farm. For example, beans, cassava, maize, and cooking bananas (matoke) is a common combination while sorghum and coffee rarely co-occur on the same farm. Although some pattern become apparent, there is generally a smooth transition of shading suggesting that most crop combinations are commonly observed.

The focus of Figure 2 is on-farm crop diversity. In contrast, Figure 3 illustrates the planting system by crop type. Monoculture refers to cultivation of only one crop type per plot at the same time while intercropping refers to the mix of at least two crops per plot. The bars represent the total area planted with the 10 most common crop types in monocultures and intercropping. For intercropping we use the share of the plot that is allocated to the specific crop (see data section). The figure shows that all common crop types are planted in both monocultures and intercropping systems making direct comparison of yields between both systems possible.

The number of crops per farm is relatively evenly distributed within Uganda. The average number of crops per farm is higher than average in Central and Western Uganda and lower in Northern Uganda (Figure 4). Since the survey is not representative at the district level and new districts were created during the time of the survey, extreme values of individual districts may not be informative.

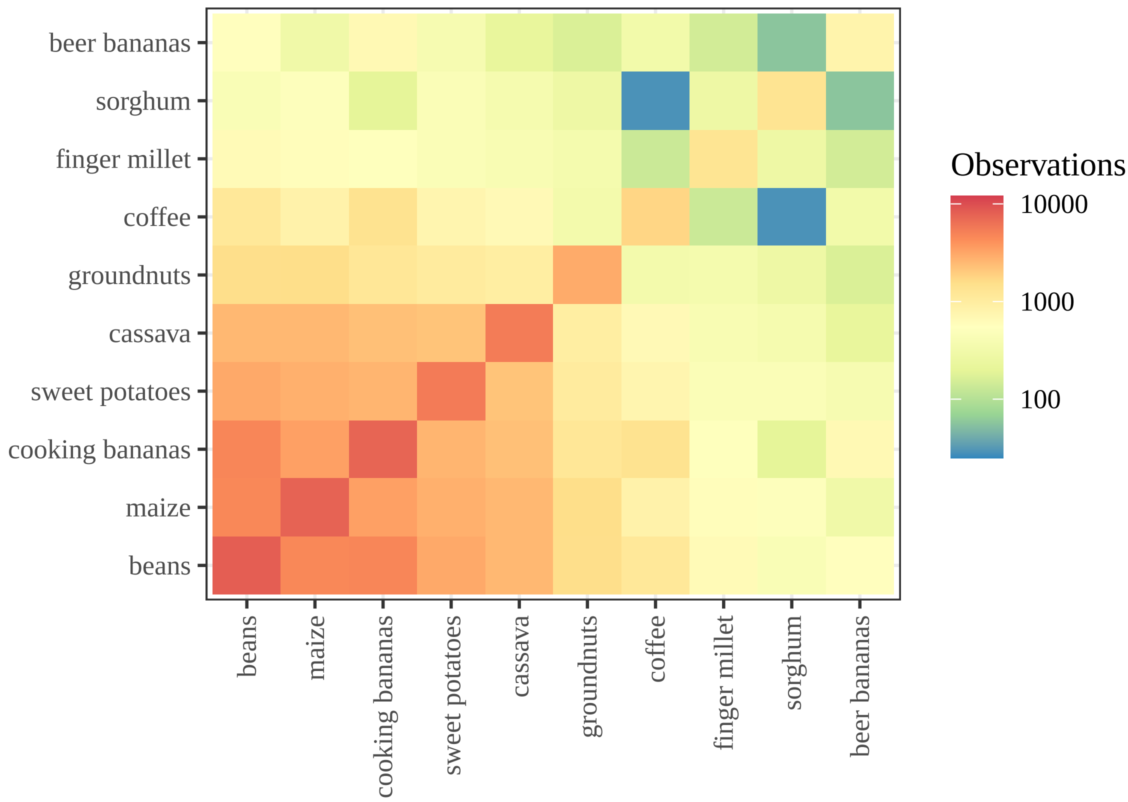

Figure 2: Co-occurrence of the main crop types

However, some spatial pattern of crop number per farm become apparent. A clear relationship between rainfall risk and crop numbers seems unlikely because the north of Uganda has low levels of crop diversity but also relatively dry weather with frequent droughts while the southern part of Uganda enjoys a relatively stable and humid climate. However, the southern parts of Uganda are also more densely populated making a relationship between market access and crop diversity possible. We explore the relationship between crop diversity and risk, markets, prices and costs in the following.

3.2 Crop diversity and risk

Diversification is often associated with risk (e.g. Eeckhoudt et al. 2005) and crop diversity is no exception (Di Falco and Chavas 2009; Auffhammer and Carleton 2018; Chuang 2019). Farmers in Uganda are facing high levels of risk of income shortfall with little access to financial markets to mitigate these shocks. By far the most common reason for income shortfalls in rural Uganda are droughts followed by floods (Figure 5). Both reasons are directly related to rainfall pattern and the resulting water balances.

Risk of droughts and floods therefore seem to be a plausible explanation for the high levels of crop diversity in Uganda. To explore the relation between risks and crop diversity, we plot

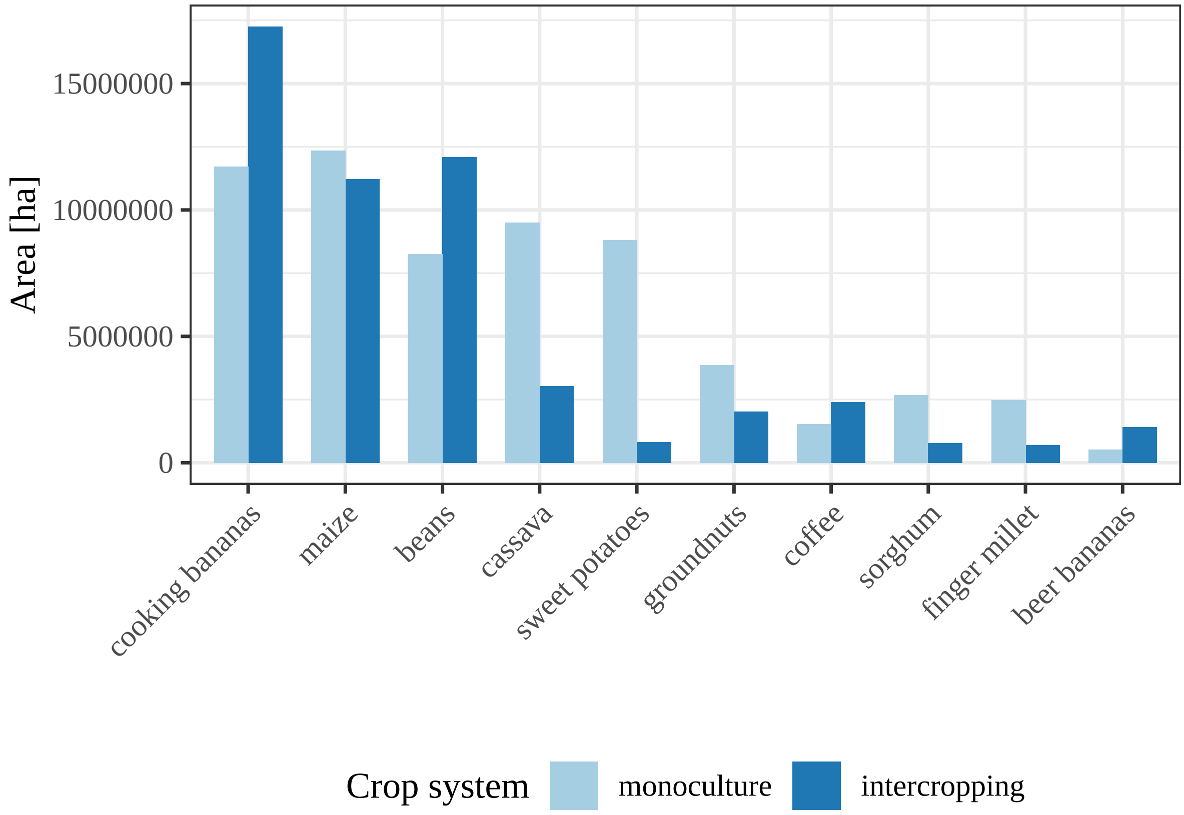

Figure 3: Production area of the most common crop types by cropping system in Uganda.

the number of crops against deciles of the coefficient of variation of annual district level rainfall between 1998 and 2016 based on the TRMM data. We use the coefficient of variation instead of the standard deviation or variance because it corrects for mean precipitation levels which could affect crop diversity levels also directly.[4] Uganda is located directly at the equator with a 12 months growing period. We therefore use annual precipitation levels.

Figure 6 suggests that there is no obvious relation between crop diversity and rainfall risk. We explore the relationship between rainfall shocks and crop diversity further in the Appendix D. Although we cannot reject a possible impact of rainfall risk on crop diversity, we conclude that it is not the main driving force behind the high levels of crop diversity in Uganda. Although this finding seems surprising at first it can be explained by the high correlation of crop responses to droughts and floods. Diversification is important to mitigate the risk of uncorrelated returns such as gambles (Gollier 2001). However, reducing the expected response to a highly correlated shock such as droughts may imply choosing the crop with the lowest expected response. In other words we expect a change in the crop identity or crop selection but

[4] Appendix A shows the correlation of the TRMM data with the precipitation data from Willmott and Matsuura of the University of Delaware as a reference. It also shows the relation between mean precipitation levels and different measures of precipitation risk.

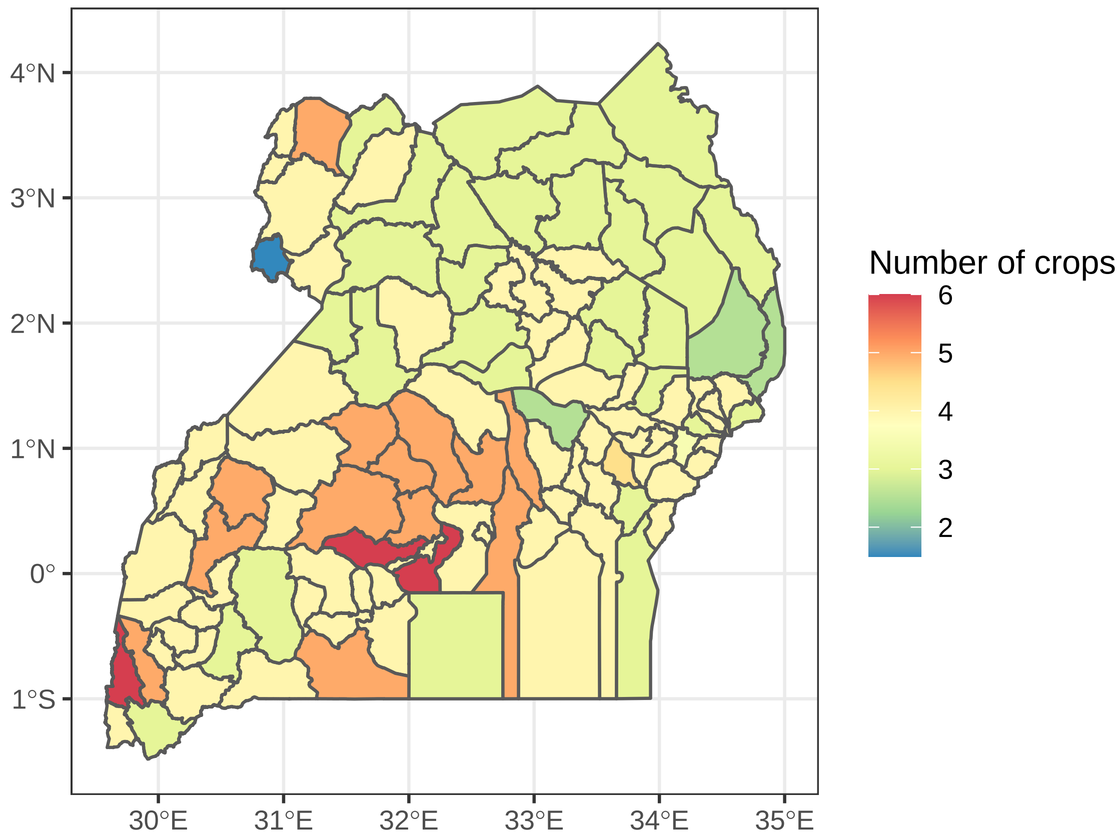

Figure 4: Mean number of crops per season and farm

not necessarily a response in the number of crops.

3.3 Crop diversity and market participation

Crop diversity may be driven by preferences for variety in consumption (e.g. Dixit and Stiglitz 1977, Quaas and Requate 2013) and missing product markets. Most farms in Uganda are subsistence farms, consuming the largest share of their production (Table 1). In consequence, if farmers have a preference for variety they may have to produce a variety of crops. It seems therefore plausible that households grow a large number of crops because of their preference for diversity in consumption. Based on this assumption we would expect that crop diversity declines with market participation because households become less dependent on their own production for consumption. Instead, they may specialize according to their comparative advantage and use their income to purchase additional crop varieties. Figure 7 shows the level of crop diversity in relation to market participation bins. Farms in the lowest market participation bin consume their complete harvest and sell nothing whereas farms in the highest market participation bin sell 90 to 100 % of their harvest. While the survey contains observations for each market participation bin, observations exceeding 50 % market participation are relatively rare (numbers below boxplots). However, the figure suggests that crop diversity may be declining

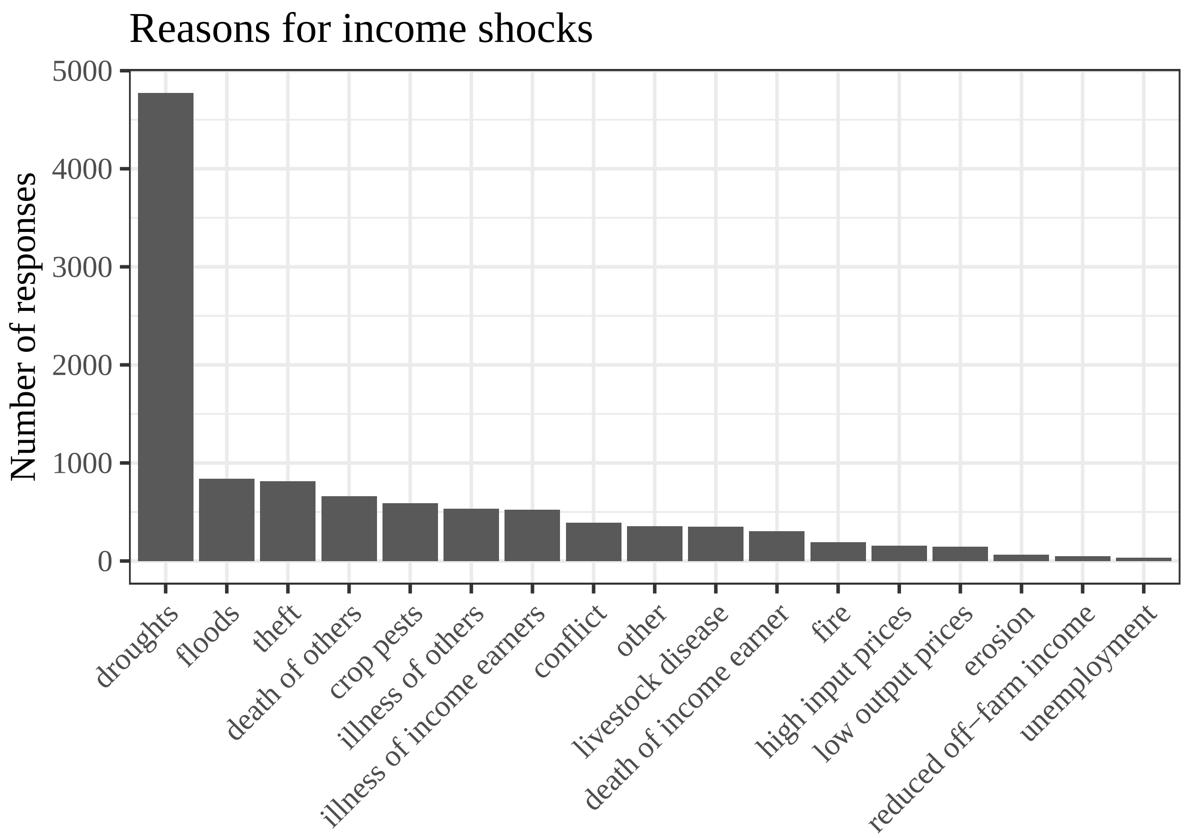

Figure 5: Reasons for income shortfalls. The size of the bar indicates the number of times the reason was named by respondents.

with market participation.

3.4 Crop diversity, economies of scope and comparative advantage

The main mechanism suggested by ecologists to explain the positive impact of biodiversity on productivity in natural ecosystems is the complementarity of species' resource use (Hooper et al. 2005). In other words, different "... species use different resources, or the same resources but at different times or different points in space..." (Hooper et al. 2005). In agriculture this is likely to occur for resources such as nutrients and light but also for other inputs such as labor. For example, if crops require labor inputs at different points in time (e.g. they have different planing and harvesting cycles) then dividing labor among more crops would increase the effective labor units of a household. Intercropping i.e. growing several crops on one plot is a common practice in Uganda. If complementaries between crops were important in crop production we would expect to see higher output per unit of resource in these intercropping systems. Figure 8 shows the output per unit of land for the ten most common crops in our data sets. The figure shows that median yields are higher for all crop types planted in combination with other crops (mixed) compared to monocultures (single).

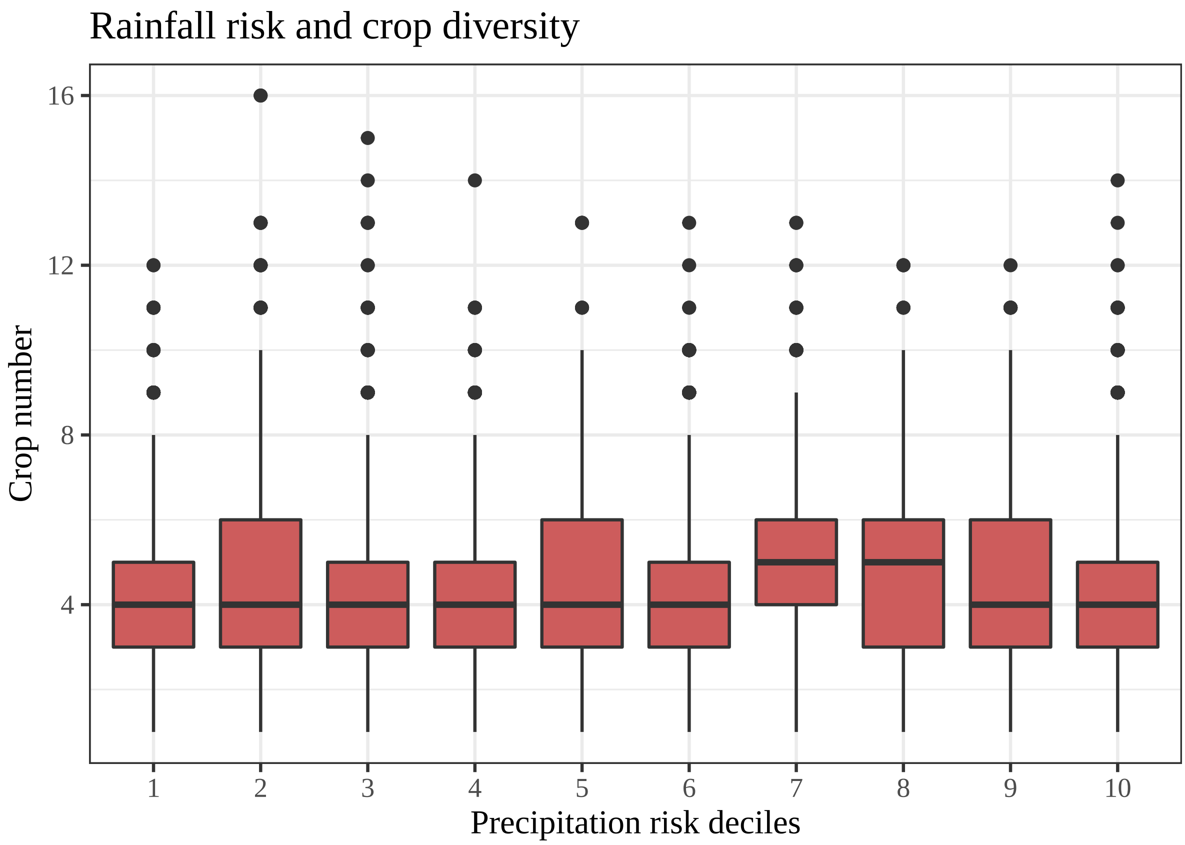

Figure 6: Boxblots of the number of crops per farm for each rainfall risk decile. Rainfall risk is measured by the coefficient of variation of annual district level rainfall using the TRMM data.

However, the complementarities between crops only increase productivity if the gains form crop complementarities in production outweigh the losses from allocating resources away from the most beneficial crops to a less beneficial crop. Figure 9 compares the revenues per unit of land across different crops grown in monocultures and intercropping. Using revenues as the welfare relevant measures of output, intercropping increases productivity if the revenue differences between crops grown in intercropping or mixed stands are smaller than the revenue difference between the same crop grown in monoculture and intercropping. Figure 9 suggests that this is the case for several crops.

These observation suggests that there are direct benefits of crop diversity from complementarities.

3.5 Crop diversity and increasing returns

In addition to revenue differences between crops, there may be crop specific fixed costs reducing the benefits from crop diversity. For example, specific crop types may need specific inputs such as machinery but also knowledge or just the additional preparation time and effort to carry out a crop specific task such as planting, weeding or harvesting. These costs create in-

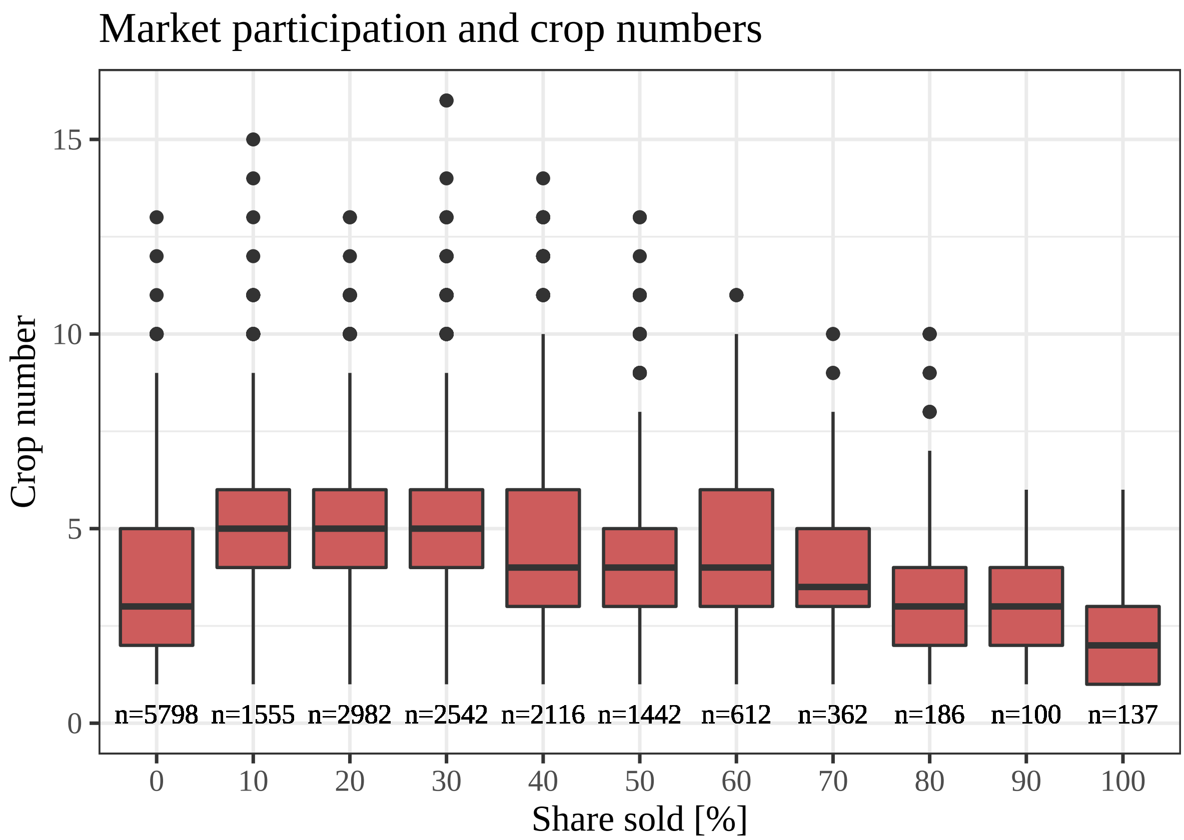

Figure 7: Boxplot of market participation bins and crop diversity. Market participation is measured in the share of the total harvest that is sold. Market participation bins are defined as 0: 0 %, 10: 1 % to 10 %, 20: 11 to 20 %,.. 100: 91 % to 100 %. The number of observations within each bin is shown under each boxplot.

creasing returns to scale in profits and are often seen as a driver for specialization (e.g.Romer (1987)).[5] These fixed costs are difficult to observe directly but we can indirectly infer about their presence. Optimal land allocations across crops may include very small values such as several square centimeters or meters. However, we would not observe these small allocations in the presence of fix costs. Figure 10 shows the land that is allocated to individual crops per farm. The median crop area per farm is around 1,400m^2 (black line) and the mean crop area per farm is around 2,600 m^2 (red line). Although these values are small according to developed country standards it is relatively large compared to the median farm size of 0.8 ha in Uganda. However, only in 1 % of the cases are crops grown on areas of less than 80 m^2 (dotted line). We interpret this finding as indicative for the presence of fixed costs.

[5] Other reasons for increasing returns to scale such as increasing returns to scale in production seem unlikely based on the farm size literature (e.g. Noack and Larsen (2019)).

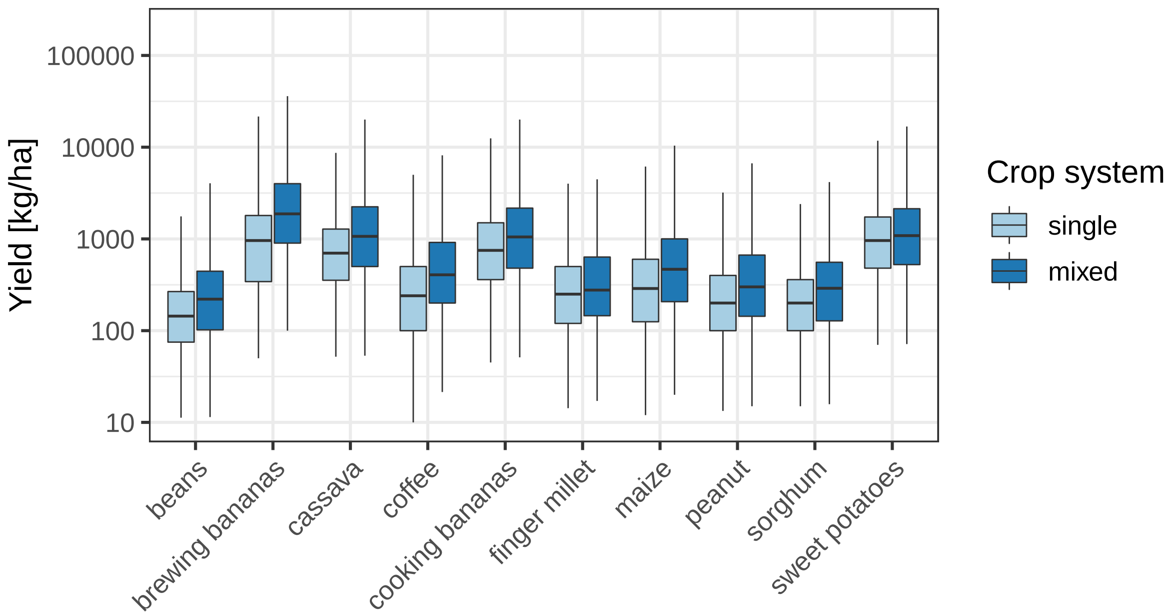

Figure 8: Yields and cropping systems. Mixed cropping systems imply that several crops are grown on one plot at the same time while only one crop is grown at the same time in monoculture cropping systems or single crop stands.

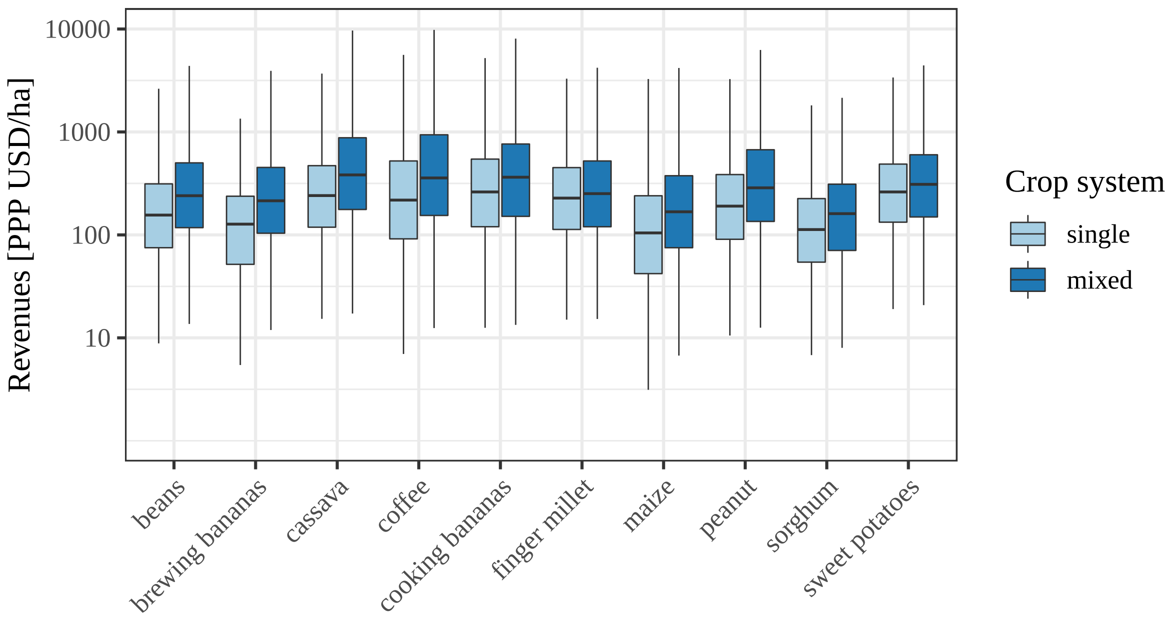

Figure 9: Yields and cropping systems. Multi cropping systems imply that several crops are grown on one plot at the same time while only one crop is grown at the same time in mono cropping systems.

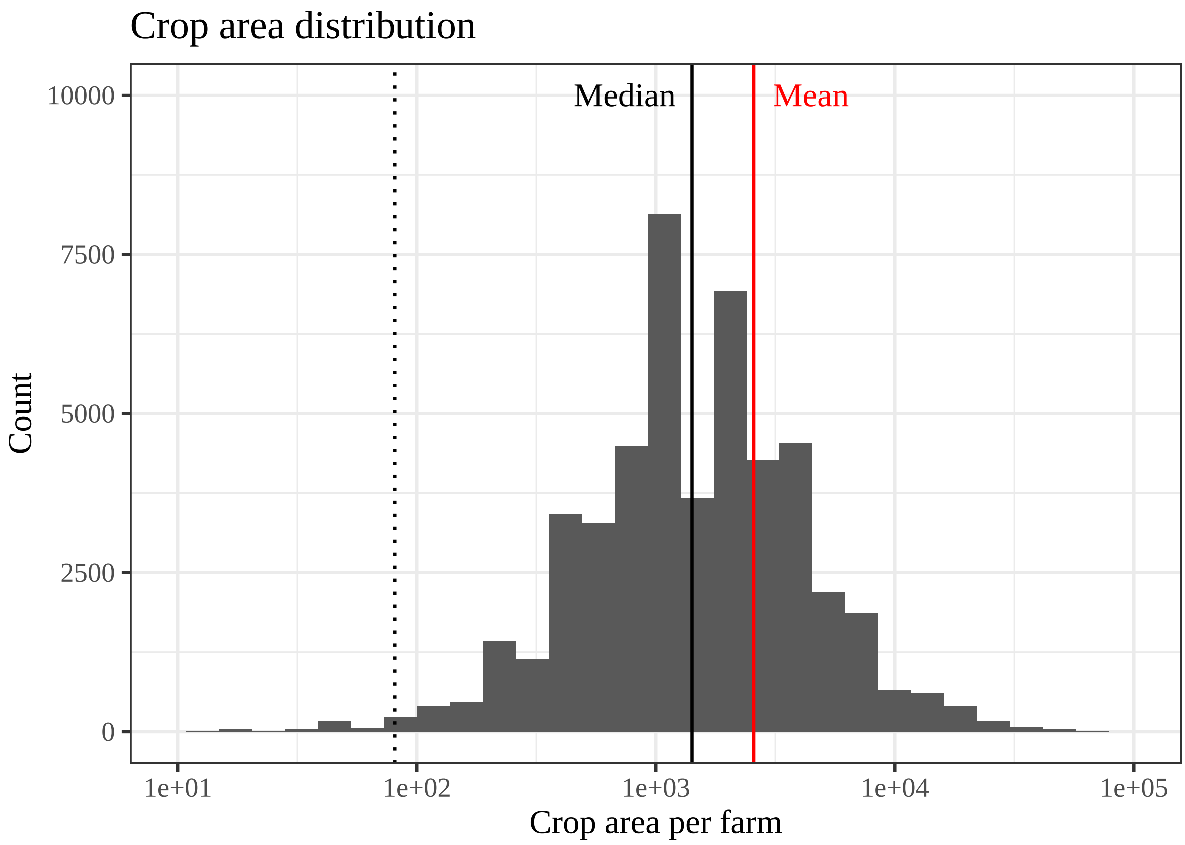

Figure 10: Farm area allocated to individual crops.

4 Model

As Foster and Rosenzweig (2017) we think of agricultural production as described by a constant returns to scale production function g with the inputs land (a) and the amount of nutrients (e) effectively available to the plants. Foster and Rosenzweig (2017) think of plant nutrient availability as produced by labor and machinery, for example used to remove weeds that compete with crops for nutrients. We also include these processes in our model by considering labor as an input that enhances nutrient availability for plants. Additionally, the overall amount of nutrients effectively available to the plants may depend on numerous other factors as well. Plants need multiple nutrients to grow in composition specific to the particular crops. Soil composition, soil structure, water availability, microbe communities, and shading are some of these factors, or regulating ecosystem services, that influence the composition and amount of nutrients available to the plant. These factors can be determined by natural processes and thus largely given to the farmer, or they can be modified artificially by application of fertilizers, tillage, irrigation, fungicides, etc. Moreover, all these factors are complementary to each other, at least to some degree. A parsimonious and tractable specification that captures these aspects of agricultural production is the Cobb-Douglas production function for crop k:

$$y_k = a_k^{\alpha}\, e_k^{1-\alpha} = a_k^{\alpha}\, \underbrace{l_k^{\beta}\, \exp\left(\int_0^{\mu} \gamma\, \ln(x_{ik})\, di + \int_{\mu}^{1} \gamma\, \ln(e_{ik})\, di\right)}_{=e_k^{1-\alpha}}, \tag{1}$$

where a_k is land area, and l_k is labor, allocated to crop k, and $e_{ik} > 0$ are the environmental regulating services that cannot be controlled by the farmer, whereas x_{ik} can be controlled at marginal costs q. For analytical tractability, we consider a continuum of non-labor inputs into effective nutrient production and assume that their output elasticities (γ) are identical. Non-controlled regulating ecosystem services are indexed in increasing order, i.e. $de_{ik}/di > 0$. The fraction $\mu \in (0, 1)$ of these is considered as fixed for now. We later endogenize the farmer's decision on μ as a trade-off between profitability and fixed costs of access to the extra inputs.

The marginal products for all inputs is positive but decreasing, $\alpha, \beta, \gamma \in (0, 1)$. The assumption of constant returns to scale implies $\alpha + \beta + \gamma = 1$.

The farmer chooses land and labor allocation as well as fertilizing inputs across crops such as to maximize profit of farming

$$\max_{\{a_k, l_k, x_{ik}\}} \int_0^N \left(p_k\, y_k - \int_0^{\mu} q\, x_{ik}\, di\right) dk \quad \text{subject to} \int_0^N a_k dk = A \text{ and } \int_0^N l_k dk = L. \tag{2}$$

For analytical tractability, we consider a whole continuum of crops. Total land endowment (A) and labor endowment (L) are given, we consider the case of a labor market at a later stage.

Using λ to denote the Lagrangian multiplier for the land constraint, and ω for the labor constraint, the first-order conditions read

$$\alpha\, p_k\, \frac{y_k}{a_k} = \lambda \tag{3a}$$

$$\beta\, p_k\, \frac{y_k}{l_k} = \omega \tag{3b}$$

$$\gamma\, p_k\, \frac{y_k}{x_{ik}} = q \qquad \text{for } i \in [0, \mu]. \tag{3c}$$

In the appendix we show that this implies

$$\frac{a_k}{A} = \frac{l_k}{L} = \frac{p_k^{\frac{1}{(1-\mu)\gamma}}\, r_k\, q^{-\frac{\mu}{1-\mu}}}{\int_0^N p_k^{\frac{1}{(1-\mu)\gamma}}\, r_k\, q^{-\frac{\mu}{1-\mu}}\, dk} = \frac{p_k^{\frac{1}{(1-\mu)\gamma}}\, r_k}{\int_0^N p_k^{\frac{1}{(1-\mu)\gamma}}\, r_k\, dk} \tag{4}$$

The share of land (and labor) allocated to crop k depends on the value p_k of that crop relative to some generalized mean value of all N crops the farmer chooses to grow. In that ratio, and in the generalized mean, the value of each crop is taken to the power $1/(1 - \mu\, \gamma) > 1$, which distorts the land allocation in favor of the more valuable crops. The closer μ comes to one, i.e. the larger the fraction of environmental supporting services that the farmer has under control, the larger is the fraction of land and labor allocated to the most valuable crop. As the p_k are in descending order, this is the first crop. This shows that the economic reason for higher diversity is the lack of control over environmental conditions.

In the appendix we further show that aggregate profit on the farm with N crops is given as

$$Y = \int_0^N \pi_k \, dk = (1 - \mu\,\gamma)\, A^{\frac{\alpha}{1-\mu\gamma}}\, L^{\frac{\beta}{1-\mu\gamma}} \left(\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k \, dk \right)^{\frac{(1-\mu)\,\gamma}{1-\mu\,\gamma}} \left(\frac{\gamma}{q} \right)^{\frac{\mu\,\gamma}{1-\mu\,\gamma}}. \tag{5}$$

The number of crops, however, is an endogenous choice of the farmer, that will depend on farm size and labor endowment. We now turn to the question which number of crops the farmer will choose to grow. Growing another type of crop on the farm incurs (quasi-)fixed costs $F > 0$. Thus, the optimal number of crops is determined by

$$\frac{\partial Y}{\partial N} = F. \tag{6}$$

In order to further explore the farmer's choice of crop diversity, we need to specify the distribution of marginal crop value p_k. We think of p_k as a general measure of marginal value, for example in terms of market price of the respective crop, or in terms of caloric value. We assume that the marginal value of crops differ, and that p_k is described by the distribution

$$p_k = p_1\, k^{-\varphi}, \tag{7}$$

i.e. we assume a Pareto distribution of marginal crop value. We assume a rather 'flat' distribution of crop values, i.e.

$$\varphi < (1 - \mu)\,\gamma. \tag{8}$$

Assumption (8) implies that the benefit of specialization – due to the high marginal value of some crops – is limited compared to the benefit of diversification – due to decreasing marginal returns on controlled inputs, captured by the term $(1 - \mu)\,\gamma$. Borrowing from the New Economic Geography literature (Krugman 1991), we refer to Assumption (8) as the 'no black hole condition'. If (8) is violated, the farmer optimally chooses to allocate all resources – land, labor, and all nutrients – to the most crop only, i.e. the problem to choose the optimal crop number collapses into the corner solution of just growing the most profitable crop. Note that the model implies such a corner solution for $\mu = 1$, i.e., for the case that the farmer fully controls environmental conditions of crop production. This shows again that the model explains a large number of crops as the consequence of the farmer's dependency of regulating services provided by the ecosystem.

Furthermore, we assume that the distribution of e_{ik} is independent of k, which implies $r_k = \bar{r}$ for all k. Thus,

$$\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k \, dk = \bar{r} \int_0^N p_1^{\frac{1}{(1-\mu)\,\gamma}}\, k^{\frac{\varphi}{(1-\mu)\,\gamma}} \, dk = \frac{\bar{r}\, p_1^{\frac{1}{(1-\mu)\,\gamma}}}{1 - \frac{\varphi}{(1-\mu)\,\gamma}} N^{1-\frac{\varphi}{(1-\mu)\,\gamma}} \tag{9}$$

Under the given specification, the relative allocation of land to the crops (equation 4) simplifies to (in logarithmic form)

$$\ln\left(\frac{a_k}{A}\right) = \frac{\varphi}{(1-\mu)\,\gamma}\,\ln(k) - \left(1 - \frac{\varphi}{(1-\mu)\,\gamma}\right)\ln(N), \tag{10}$$

and aggregate output is explicitly given as a function of crop number, land and labor endowments as

$$Y = \left(\Omega\, N^{(1-\mu)\,\gamma-\varphi}\, A^{\alpha}\, L^{\beta}\right)^{\frac{1}{1-\mu\,\gamma}} \tag{11}$$

where

$$\Omega := \frac{(1-\mu\,\gamma)^{1-\mu\,\gamma}\,\bar{r}\,p_1}{\left(1-\frac{\varphi}{(1-\mu)\,\gamma}\right)^{(1-\mu)\,\gamma}}\left(\frac{\gamma}{q}\right)^{\mu\,\gamma} \tag{12}$$

depends on the model parameters only. Taking logs of (11), we obtain

$$\ln(Y) = \ln\left(\frac{\Omega}{1-\mu\,\gamma}\right) + \frac{(1-\mu)\,\gamma-\varphi}{1-\mu\,\gamma}\,\ln(N) + \frac{\alpha}{1-\mu\,\gamma}\,\ln(A) + \frac{\beta}{1-\mu\,\gamma}\,\ln(L) \tag{13}$$

We can estimate this model using our data for the farms in Uganda. In particular, the coefficient estimate for $((1-\mu)\,\gamma-\varphi)/(1-\mu\,\gamma)$ gives the percentage change of output with crop number, and $((1-\mu)\,\gamma-\varphi)/(1-\mu\,\gamma)\,Y$ is an estimate of the value of crop diversity for the farms in Uganda.

The crop number N in (13) is endogenous, however, and in particular it depends on the farm's endowments with land A and labor L. From (6) we obtain the following condition that characterizes the optimal number of crops:

$$\ln(N) = \ln(C) + \frac{\alpha}{\alpha+\beta+\varphi}\,\ln(A) + \frac{\beta}{\alpha+\beta+\varphi}\,\ln(L), \tag{14}$$

where $C > 0$ is some positive constant that depends on the model parameters. Equation (14) shows that the optimal crop number positively depends on the farm size both in terms of land and labor endowments. In our empirical study we include farm fixed effects that capture differences in technology and ecological resources across farms. Note that the coefficients for land and labor are independent of these differences, so all effects related to technology choice and resource endowments are adequately captured by the fixed effects.

Our model further allows us to study the effect of farm size on overall output, including both the direct effect and the indirect effect via effect of land on crop number. Using (14) in (13), we obtain the effect of land area on output, combining the direct effect and the indirect effect via crop number:

$$\ln(Y) = C_2 + \frac{\alpha}{\alpha+\beta+\varphi}\,\ln(A) + \frac{\beta}{\alpha+\beta+\varphi}\,\ln(L), \tag{15}$$

where C_2 is a constant that depends on model parameters.

This result further allows us to study how the farm size distribution affects the overall output of farms in Uganda. We assume, that farm sizes A_i and labor endowments L_i are independently distributed across farms i. Assume that both are log-normally distributed, with $\bar{A}$ and σ_A being mean and standard deviation of the farm size distribution, and $\bar{L}$ and σ_L mean and standard deviation of labor endowment distribution. Wit a total mass M of farms, total output is (see, e.g. Baumgärtner et al. 2017)

$$M\,\bar{Y} = M\,const\,\bar{A}^{\frac{\alpha}{\alpha+\beta+\varphi}} \left(1+\frac{\sigma_A^2}{\bar{A}^2}\right)^{-\frac{1}{2}\frac{\alpha}{\alpha+\beta+\varphi}\left(1-\frac{\alpha}{\alpha+\beta+\varphi}\right)} \bar{L}^{\frac{\beta}{\alpha+\beta+\varphi}} \left(1+\frac{\sigma_L^2}{\bar{L}^2}\right)^{-\frac{1}{2}\frac{\beta}{\alpha+\beta+\varphi}\left(1-\frac{\beta}{\alpha+\beta+\varphi}\right)} \tag{16}$$

This allows us to study how aggregate output increases if the distribution of land becomes more equal. The general prediction from the model is that a more equal farm size distribution will increase aggregate output.

5 Estimation

In this section we test the predictions of the model with data from Uganda. We are especially interested in the drivers of crop diversity and its private benefits. Both questions are closely related because high private benefits would incentivize high levels of crop diversity.

In Section 3 we suggested that economies of scope, economies of scale and the returns to specialization du to comparative advantages are the main drivers of crop diversity in Uganda. In section 4 we further proposed that control over the environment and improved market access reduce the economies of scope. The economies scope result from ecological and economic complementarities of inputs while the economies scale depend on the crop specific fix costs. The returns to specialization due to comparative advantages evolve from revenue differences across crops.

In an ideal experiment, we could randomly assign crop specific fix costs, revenue differences between crops as well as complementarities between crops to farmers in Uganda. We would also randomly vary the farmers' control over the environment and then observe the corresponding crop diversity and revenue levels. Such an experiment may not be feasible. Instead, we use our theoretical predictions to guide our empirical estimations using observational data. Our panel of 66,193 plot level observations allows us to implement a differences-in-differences approach, comparing changes in inputs, crop diversity and output across farms or across plots within the same farm. We further employ an instrumental variable approach exploiting variation in household composition and land inheritance to predict labor and land inputs that would otherwise be co-determined with crop diversity.

One key prediction of our model is that crop diversity increases with land and labor inputs; while the gains from economies of scope scale with the inputs, the fix costs that determine the economies of scale remain constant. Equation (14) specifies this relationship. However, farms may differ with respect to their physical and economic environment which could determine the amount of labor and land for crop production in addition to crop diversity. Seasonal and yearly fluctuations could further affect input use, output and crop diversity. We therefore include, household, year and season fixed effects. These fixed effects absorb the composite parameters such that (14) becomes

$$\ln(N_{ist}) = \frac{\alpha}{\alpha+\beta+\varphi}\ln(A_{ist}) + \frac{\beta}{\alpha+\beta+\varphi}\ln(L_{ist}) + \eta_i + \theta_t + \nu_s + \varepsilon_{ist}. \tag{17}$$

Here N_{ist} is the number of crops grown per farm, A_{ist} is the land used for crop production, L_{ist} is the labor used for crop production and Y_{ist} are the revenues of farm i in season s in year t. The last four terms are farmer fixed effects, year fixed effects, seasonal fixed effects and the error term that we cluster at the district level to account for the two stage sampling process of the survey (Abadie et al. 2017).

In addition to the estimation on household level, we also report results for (14) on plot level. All household level variables are replaced in this case by plot level measurements of the same variables. For example, crop diversity in this specification is measured by the number of crops that are grown together on one plot.

Our theory predicts a constant elasticity relationship between crop numbers and inputs. Because in our case N_{ist} is always larger than zero we use log crop numbers to ensure non-negative outcomes (Wooldridge 2010).

Although our measure of crop diversity is directly derived from our theory section, we also present results using the Simpson (Herfindahl–Hirschman) index to measure crop diversity. Crop numbers is one extreme of the common diversity indices placing no weighs on the allocation of inputs among crops. In that sense, the Simpson index is another extreme of the common diversity indices, placing more weight on the allocation of inputs than other common measures of diversity including the Shannon index.

The land variable in our baseline specification is based on self-reported land size. However, in a robustness check we use GPS corrected land measurement based on the description in the previous section. In this case, the land variable is an estimated regressor based on the estimated relationship between GPS measured and farmer-reported land size for a sub-sample of the observations. We therefore bootstrap our standard errors across both stages of the estimation using 500 bootstrap replicates.

Labor and land may be endogenous, and simultaneously determined with crop diversity. We therefore use land inheritance and household composition as instruments for farm size and labor respectively. Land inheritance is by far the most common form of land transfers in Uganda and one third of the households in our sample inherited land during the survey period. We argue here that land inheritance is exogenous to the household. In contrast to land inheritance, family size is not exogenous. However, the timing of the changes in the household composition may be exogenous because the individual contributions to the household labor force are age specific and the age changes exogeneously. We therefore use the OECD measure of adult equivalent units (Haughton and Khandker 2009) as instrument for changes in labor inputs in combination with household fixed effects.

Droughts are a major concern of farmers in Uganda (see Figure 5). Droughts may also affect crop diversity directly (e.g. farmers may not plant drought sensitive crops in dry years) or indirectly (e.g. farmers who experienced droughts may adjust their crop portfolio). Although an impact of our instruments seems unlikely, we present results using different drought measures as additional controls in Appendix D. We regard the estimated impact of lagged droughts on crop diversity as an extension of the discussion on crop diversity and risk from Section 3.

Although the estimates from (17) support our theoretical predictions they do not allow us to draw direct conclusions about the magnitude of the economy of scale and the economies of scope. In contrast to (14) estimating (13) allows us to identify the crop specific fixed costs as well as the input complementarities. After introducing household, year and seasonal fixed effects, (13) becomes

$$\ln(Y_{ist}) = \frac{(1-\mu)\,\gamma - \varphi}{1-\mu\,\gamma}\ \ln(N_{ist}) + \frac{\alpha}{1-\mu\,\gamma}\ \ln(A_{ist}) + \frac{\beta}{1-\mu\,\gamma}\ \ln(L_{ist}) + \eta_i' + \theta_t' + \nu_s' + \varepsilon_{ist}' \tag{18}$$

where Y_{ist} are the revenues of farm i in season s in year t, N_{ist} is the number of crops grown per farm, A_{ist} is the farm size and L_{ist} is the labor used for crop production. The last four terms are farmer fixed effects, year fixed effects, seasonal fixed effects and the error term that we cluster at the district level. In addition to our specifications on farm level we also estimate (18) on crop-plot level adding crop and plot fixed effects. In this specification we can directly measure output in physical quantities without conversion to revenues. We can also directly control for crop area, plot size and farm size to address concerns that crop diversity and measurement error may be correlated with plot or crop area. We also include crop diversity at the plot and the household levels to separate the economies of scope resulting from ecological complementarities from the economies of scope resulting from economic complementarities. Crop diversity at the farm level would capture economies of scope resulting from complementarities of economic inputs such as labor and capital. In contrast, crop diversity at the plot level would capture ecological complementarities. Including additionally plot fixed effects allows us to compare the outcomes of a plot with monocultures to the outcomes of the same plot with intercropping. However, a plot is not a fixed identity but it changes in size as plots are frequently split and merged. We therefore use a plot level regression with household fixed effects as our baseline specification.

The parameter estimate of crop diversity in (18) has several interpretations. First, for the production environment of Uganda with little control over environmental factors (see section

3), it equals the net benefit of crop diversity ($\gamma - \varphi$) i.e. the complementarities in crop production minus the price difference between crops. Second, equation (6) also suggests that the marginal productivity of crop diversity equals the crop specific fix costs. We can therefore use the results of equation (18) to calculate the fix costs parameter F. Estimating both equations therefore allows use to identify all model parameters.

6 Empirical Results

Table 2 presents the results on the impact of farm size and labor on crop diversity. Crop diversity is measured by log number of planted crops in our baseline specifications. We compare the results to a specification using the Simpson index of diversity. The baseline specification (1) with district fixed effects suggests that a 10 percent increase in the farm size increases crop diversity by 1.4 % while a 10 percent increase of labor increases crop diversity by 1.8 %. Including household fixed effects (specification 2) changes these results to 1.6 % and 1.7 % respectively. Using predicted farm size based on GPS measurements to measure farm size (specification 3) has little impact on the results. Also, using household composition and inherited land as instruments for labor and land (specifications 4 to 6) leaves the results largely unchanged. The F statistic for inherited land and household composition is 89 and 78 respectively in specification (6). Lastly, measuring diversity by a standardized Simson index does not affect the qualitative results but increases the contribution of labor relative to the contribution of land to crop diversity. Simpson diversity puts more weight on the allocation of inputs, suggesting that crop diversity may play and important role for labor smoothing.

In addition to these household level results, we present results on plot level in the Appendix C. Although the magnitude of the effect is reduced, the qualitative results are similar: larger plots and higher labor inputs are associated with different crops planted in combination on one plot.

Overall, these results suggest that crop diversity increases with farm size and labor supply. The rural surplus labor in Uganda therefore partly explains the high crop diversity in Uganda. The suggested underlying mechanisms for these results are the complementarities or economies of scope in crop production in combination with crop specific fixed costs. The next results allow us to quantify these two mechanisms.

Table 3 summarizes the results for the impact of crop diversity on revenues. The baseline specification with district, season and year fixed effects suggests output elasticities of land and labor of 0.52 and 0.24 respectively. The results show further that a 10 % increase in crop diversity increases output by 2.8 %. However, these estimates are based on the differences in revenues and crop diversity across farms within districts which may be partly driven by differences in farmer and land characteristics. Including household fixed effects in specification

Table 2: Farm crop diversity

	Log(number)						Simpson
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	ER	IV	IV	IV	OLS
Log(farm size)	0.138***	0.158***	0.166***	0.137***	0.175***	0.170***	0.147***
	(0.007)	(0.011)	(0.007)	(0.033)	(0.044)	(0.050)	(0.032)
Log(labor)	0.184***	0.171***	0.171 ***	0.217***	0.167***	0.185***	0.414***
	(0.009)	(0.010)	(0.006)	(0.069)	(0.019)	(0.061)	(0.026)
Observations	17,837	17,837	17,837	17,837	14,587	14,587	17,837
Instruments				labor	land	labor & land	
Fixed effects	District	Household	Household	Household	Household	Household	Household

Results for regression equation (17). *The main outcome is crop diversity measured in log number of planted crops. The independent variables are log cultivated area and log labor days. The specifications are: (1) district fixed effects, (2) household fixed effects, (3) similar to specification (2) but with predicted land size and bootstrap standard errors, (4) similar to (2) but using household adult equivalent units as instrument for labor, (5) similar to (2) but using inherited land as instrument for farm size, (6) similar to (2) but using both instruments simultaneously, (7) similar to (2) but measuring crop diversity with Simpson diversity. All models include district, year and season fixed effects. Standard errors are clustered at the district level. Significance levels are* $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. *The number of observations for specification (5) and (6) are reduced due to missing data on land acquisition.*

(2) reduces the estimates for the impact of land on production substantially suggesting that farm size is correlated with land quality. In contrast to the large changes in the estimated contribution of land to production, the estimated output elasticities of crop diversity and labor remain relatively stable which implies that the marginal productivity of crop diversity and labor is little affected by the productivity of land. In specification (3) we replace self reported land size with predicted land size, using 500 bootstrap replications across both levels of estimation to correct standard errors but the results remain largely unchanged. In specifications (4) to (6) we introduce household composition and inherited land as instrument for labor and land. Although the estimates are generally less precise, the main result remain qualitatively unchanged. The specification of column (7) is similar to the specification of column (2) but we replace the log crop number with the Simpson index to measure crop diversity. Again, the results are qualitatively similar.

Although, our differences-in-differences specifications from Table 3 suggest a causal relationship between crop diversity and revenues we present further evidence in Table 4 from our

Table 3: Farm revenues

	Log(revenues)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	ER	IV	IV	IV	OLS
	log(N)	log(N)	log(N)	log(N)	log(N)	log(N)	Simpson
Diversity	0.280***	0.335***	0.335***	0.140	0.387***	0.201*	0.099***
	(0.031)	(0.040)	(0.026)	(0.121)	(0.073)	(0.111)	(0.010)
Log(farm size)	0.518***	0.265***	0.279***	0.164***	0.169	0.130	0.303***
	(0.019)	(0.018)	(0.017)	(0.062)	(0.147)	(0.163)	(0.017)
Log(labor)	0.239***	0.208***	0.208***	0.542***	0.221***	0.499***	0.224***
	(0.016)	(0.019)	(0.014)	(0.192)	(0.037)	(0.188)	(0.017)
Observations	17,837	17,837	17,837	17,837	14,587	14,587	17,837
Instruments				Labor	Land	Labor &	
Instruments						Land	
Fixed effects	District	Household	Household	Household	Household	Household	Household

Results for regression equation (18) *at the household level. The main outcome is log revenues. The independent variables are crop diversity measured in log number of planted crops, log cultivated area and log labor days. The specifications are: (1) district fixed effects, (2) household fixed effects, (3) similar to specification (2) but with predicted land size and bootstrap standard errors, (4) similar to (2) but using household adult equivalent units as instrument for labor, (5) similar to (2) but using inherited land as instrument for farm size, (6) similar to (2) but using both instruments simultaneously, (7) similar to (2) but measuring crop diversity with Simpson diversity. All models include district, year and season fixed effects. Standard errors are clustered at the district level. Significance levels are* $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. *The number of observations for specification (5) and (6) are reduced due to missing data on land acquisition.*

crop stand level results. In these specifications we compare the crop specific output in physical quantities per unit of input between monoculture and mixed crop stands of the same farmer. This specification also allow us to include crop diversity at the farm level to measure synergies or economies of scope that operate at the farm level such is labor complementarities.

Table 4 reports the results on crop diversity and harvest quantities on the crop-plot level. The baseline result with district fixed effects suggests that planting a crop in combination with one other crop increases yields by 10 % compared to monocultures everything else constant. Including household fixed effects increases the estimate to 14 percent. This estimate remains largely unaffected by using the GPS corrected crop patch size measurement (3) or by adding plot fixed effects (4). In the latter specification we compare harvests of the same plot from subsequent years. In column (5) we measure crop diversity with the Simpson index while in column (6) we include a dummy for multicropping as a proxy for crop diversity but the results remain qualitatively unchanged. Estimating the impact of crop diversity on the crop stand or crop-plot level has the advantage that we can additionally include crop diversity at the farm level. While crop diversity at the plot level captures the direct interaction among crops (i.e. the niches or ecological complementarities), complementarities between economic farm level inputs are captured by the farm level crop diversity. However, the results show that farm level crop diversity has no significant impact on crop harvest per unit of input suggesting that ecological complementarities drive our results. Lastly, a large literature studying the relationship between farm size, field size and production finds a negative relationship between farm size and plot size and yields. We therefore include plot size and farm size to address the correlation of crop diversity with farm and plot size that we established with the previous results. Our results suggest that the impact of farm size on crop stand level outputs is negative in line with the findings of the previous literature (e.g. Noack and Larsen (2019)).

Overall, these results provide strong evidence for the production supporting role of crop diversity for agriculture in Uganda.

7 The value of crop diversity

What is the value of crop diversity for production? How does this value change with agricultural development and market integration? The theoretical framework in combination with the empirical results allow us to answer these questions.

Similar to the wage share we can estimate the value of crop diversity based on the output elasticities. Our parameter estimates suggest an output elasticity of crop diversity ranging between 0.34 in Table 3 and 0.14 in Table 4. Using the PPP adjusted GDP for 2019 of 96.6 constant 2017 international dollars and the agricultural contribution to GDP of 24 % from the National Population and Housing Census of 2014 yields a value of agricultural production of 23.2 and subsequently a value of crop diversity of 7.9 billion 2017 international dollars.

However, this calculation neglects the cost of crop diversity and it valuates crop diversity at its current marginal productivity. An alternative approach is to compare the current value of agricultural production with a production based on monocultures following Brock and Xepapadeas (2003). To compute the difference in revenues we predict the individual outcomes based on specification (2) of Table 3 for current crop diversity levels and for a counterfactual with only one crop per farm. The results suggest a reduction of total revenues by 39 % or an average reduction of 35 % per farmer.[6] Using the value of agricultural production in Uganda suggests a value of crop diversity of 9 billion 2017 international dollars.

These results focus on the revenues and neglect any cost. Since labor markets are largely

[6]The difference between both values results from the correlation of crop diversity with farm size.

Table 4: Crop harvest quantity

	(1) OLS Log(N)	(2) OLS Log(N)	(3) ER Log(N)	(4) OLS Log(N)	(5) OLS Simpson	(6) OLS 1(Mixed)
Diversity	0.102***	0.138***	0.139***	0.127***	0.058***	0.124***
	(0.035)	(0.039)	(0.015)	(0.043)	(0.018)	(0.026)
Log(crop area)	0.323***	0.346***	0.369***	0.396***	0.315***	0.335***
	(0.030)	(0.029)	(0.015)	(0.038)	(0.023)	(0.025)
Log(labor)	0.196***	0.186***	0.186***	0.147***	0.187***	0.187***
	(0.012)	(0.011)	(0.007)	(0.014)	(0.011)	(0.012)
Farm diversity	−0.049**	0.003	0.003	0.024	0.013	0.013
	(0.023)	(0.025)	(0.016)	(0.030)	(0.025)	(0.025)
Log(plot area)	0.058**	0.029	0.029	−0.101***	0.063***	0.046**
	(0.026)	(0.025)	(0.015)	(0.036)	(0.019)	(0.020)
Log(farm size)	0.104***	−0.125***	−-0.133***	−0.094***	−0.130***	−0.130***
	(0.016)	(0.018)	(0.011)	(0.022)	(0.018)	(0.018)
Observations	66,193	66,193	66,193	66,193	66,031	66,193
Fixed effects	Household	Household	Household	Plot	Household	Household

Results for regression equation (18) *at the crop stand level. The main outcome is log harvest in kgs. The independent variables are crop diversity measured in log number of planted crops, log cultivated area and log labor days. The specifications are: (1) district fixed effects, (2) household fixed effects, (3) similar to specification (2) but with predicted land size and bootstrap standard errors, (4) similar to (2) but using household adult equivalent units as instrument for labor, (5) similar to (2) but using inherited land as instrument for farm size, (6) similar to (2) but using both instruments simultaneously, (7) similar to (2) but measuring crop diversity with Simpson diversity. All models include district, year and season fixed effects. Standard errors are clustered at the district level. Significance levels are* $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. *The number of observations for specification (5) and (6) are reduced due to missing data on land acquisition.*

absent and other variable input use is largely missing we neglect theses costs in this section but focus on labor markets in the following section. However, there are capital costs and other crop specific fix costs that are captured by the fixed cost parameter F in our model. We can calculate these crop specific costs based on the efficiency condition of crop diversity (6) which suggests that the marginal productivity of crop diversity equals the crop specific fix costs. Based on the estimate of (18) we can quantify these crop specific costs as $F = \frac{(1-\mu)\gamma - \varphi}{1-\mu\gamma} \frac{Y}{N}$

where $\frac{(1-\mu)\gamma-\varphi}{1-\mu\gamma}$ is the output elasticity of crop diversity. Using the men values of revenues and crop diversity from table 1 yields $\frac{Y}{N} = \frac{601.1}{4.3} = 139.8$ international dollar. Based on the output elasticities of crop diversity of 0.34 in Table 3, the fix cost per crop are equivalent to 47.5 international dollar. According to the Uganda Annual Agricultural Survey they were 7.4 million agricultural households in 2018 suggesting that the total aggregate cost of crop diversity equals 1.2 billion international dollar.[7] These figures therefore suggest that the net aggregated value of crop diversity in Uganda based on comparison with monocultures is 9 - 1.2 = 7.8 billion international dollars.

The value of crop diversity depends on the control over the environment (μ) and the price differences across crops (φ). Increased control over the environment results for example from irrigation and drainage to regulate water availability, fertilizer to regulate nutrients, pesticide to reduce biological interaction and greenhouses to control the growing climate. Increased control of the environment is closely related to agricultural development. In contrast to the environment, prices cannot be controlled by the farmer but are determined by markets. The price parameter in our theoretical framework is a combination of prices in a strict sense and total factor productivity. They measure the crop specific revenues for a given level of inputs. Trade for example may increase the slope of the this price profile by increasing the prices of those goods for which the economy has a comparative advantage in production and by reducing the price of those goods for which the economy as a comparative disadvantage in production. These changes would translate into a larger φ.

To visualize the relationships between the value of crop diversity, prices and control over the environment we plot the output elasticity of crop diversity against μ and φ. The output elasticity also depends on the parameter γ which we set equal to $\gamma = 0.835$. The rational for this parameter choice is described in Appendix E. Figure 11 illustrates the impact of μ and φ on the output elasticity of crop diversity. The figure suggests that crop diversity is a valuable input in the low input agriculture of Uganda with small price differences across crops. However, the value of crop diversity declines with increasing control over the environment (μ) and increasing price differences across crops (φ). For a production environments with sufficient control over the environment and large price or productivity differences across crops, the elasticity becomes negative and monocultures prevail.

8 Labor markets and crop diversity

What is the impact of rural surplus labor on crop diversity and agricultural production? How do rural labor markets affect crop diversity (not done yet)? A large share of the labor force in developing countries works in agriculture. According to the National Population and Housing

[7] $(4.3-1) \times 47.5 \times 7.4 = 1159.95$ Million

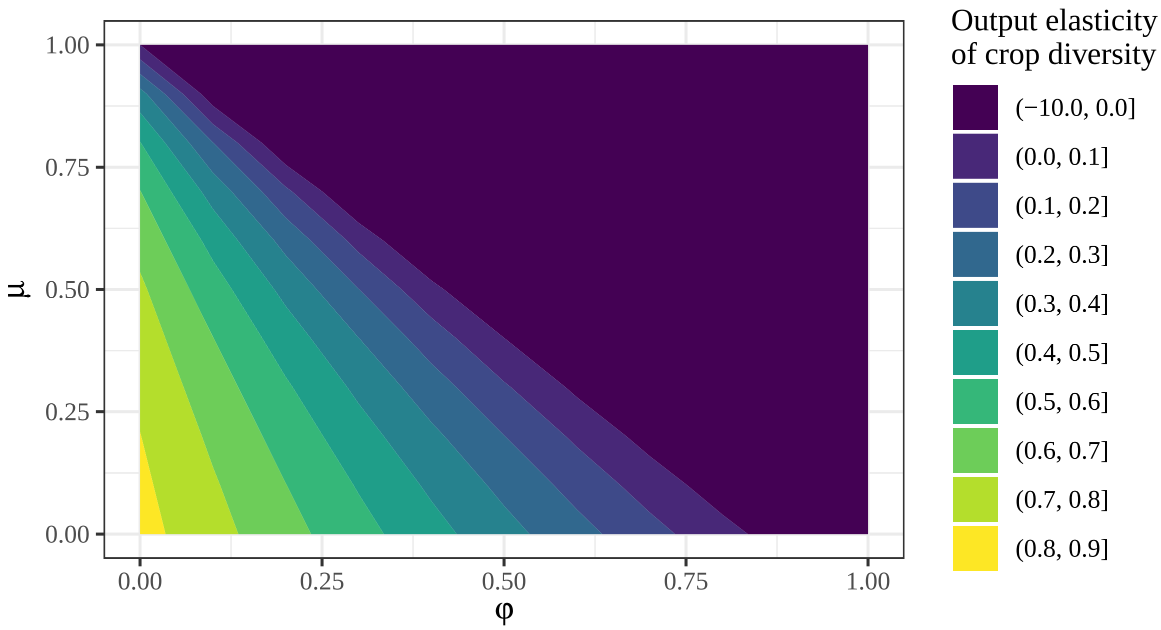

Figure 11: The output elasticity of crop diversity, control over the environment (μ) and price differences across crops (φ). The output elasticity of crop diversity is closely related to the value of crop diversity.

Census of 2014 about 70 % of Uganda's population between 14 and 65 years work in agriculture. The employment in agriculture typically declines during the process of economic development and only a small percentage of the labor force remains in the agricultural sector of developed countries. This decline of the rural labor force affects agricultural production in two ways: It reduces agricultural labor as a direct input into agricultural production and it affects crop diversity. Figure 12 shows the response of crop diversity (blue) and the value of agricultural production (red and orange) to a declining rural labor force. While the orange line highlights the direct effect of labor on production, the red line additionally includes the indirect impact of rural labor decline through the reduction of crop diversity.

9 Discussion and Conclusion

Farms are diverse in Uganda, growing on average more than four crops per hectare. In this article we suggest that this high level of crop diversity is the result of trade-offs between economies of scope and economies of scale. Our findings suggest further that the economies of scope mainly result from ecological complementarities rather than complementarities of other inputs. Based on our estimates we can quantify the value of crop diversity for agricultural production in Uganda. Our estimates suggest that the total annual value of crop diversity equals 7.8 billion dollar. This value captures only the private benefits of crop diversity. The actual economic value may be even higher, as crop diversity additionally generates positive

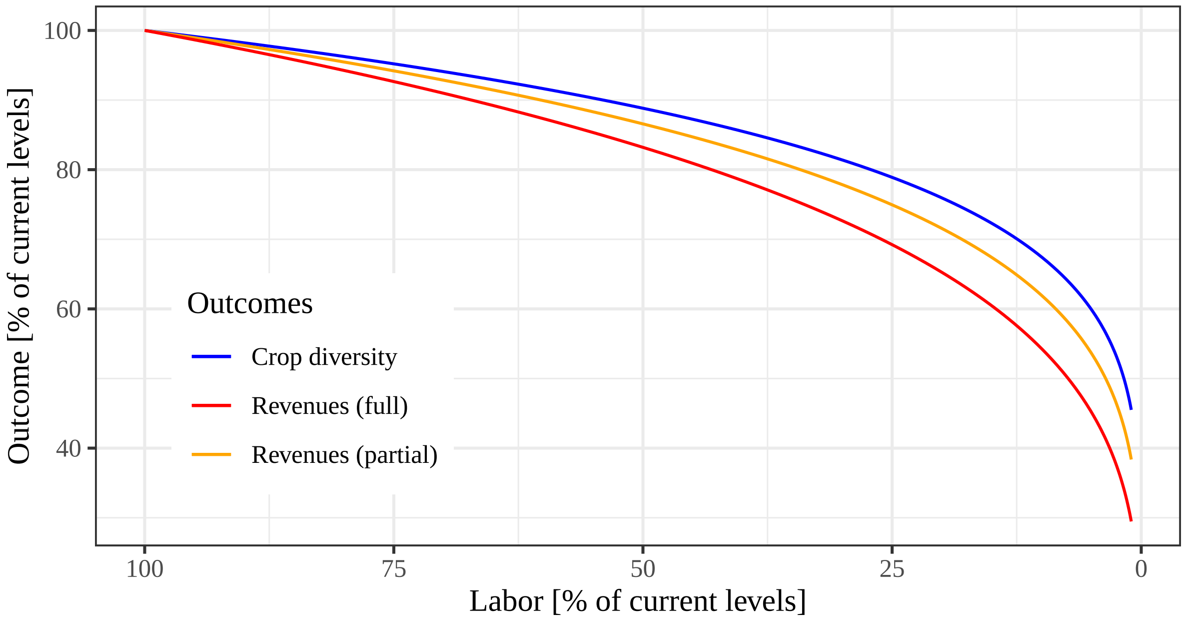

Figure 12: The response of crop diversity and revenues to declining agricultural labor. 'Revenues (partial)' includes only the direct impact of labor on production. 'Revenues (full)' includes the direct effect as well as the indirect effect of labor on production through crop diversity. All variables are expressed in percent of current levels.

externalities (Weitzman 2000, Larsen and Noack 2017, Noack et al. 2019).

Rural to urban migration and structural change is closely related to the process of economic development. During this process, labor is reallocated from agriculture to manufacturing and the service sector. Our results suggest that the high levels of crop diversity in Uganda partly result from the rural surplus labor. In counterfactual simulations we show how a reduction of agricultural labor affect agricultural production directly and indirectly through a reduction of crop diversity.

Similarly to Benin et al. (2004) we find that crop diversity increases with farm size. The impact of farm size on aggregate crop diversity is, however, difficult to predict since it also depend on the similarity of crop portfolios across farms. Similarly, we neglect crop diversity over time, or crop rotations. We leave these extensions for future research.

Lastly, the private benefits of crop diversity largely depend on the ability of the farmer to control the production environment including irrigation and drainage to control water supply, fertilizer to regulate nutrients, glasshouses to stabilize temperatures, pesticides to reduce biological damages etc. Control over the production environment typically increases with economic development. Although this process generally improves the private benefits of agriculture, our finding also show that it reduces the private benefits of crop diversity and therefore incentivizes specialization. The increase in inputs such as fertilizer and pesticides with the simultaneous loss of crop diversity can lead to a loss of ecosystem services and the state of agriculture in developed countries that makes it to one of the largest contributors to global biodiversity loss, water pollution and climate change (Tilman et al. 2017; Poore and Nemecek 2018; Springmann et al. 2018). The loss of crop diversity excercerbates this process further, increasing the need for further environmental control (Larsen and Noack 2017, 2020) to prevent a catastrophic agricultural collapse (Weitzman 2000).

A Precipitation data

Figure 13 compares the seasonal precipitation levels from the TRMM data set and the data set from Willmott and Matsuura of the University of Delaware (UDEL) for each year-season-household level observation from the LSMS household data. The correlation coefficient is 0.73. The precipitation levels in the TRMM data set are generally higher than in the UDEL data set.

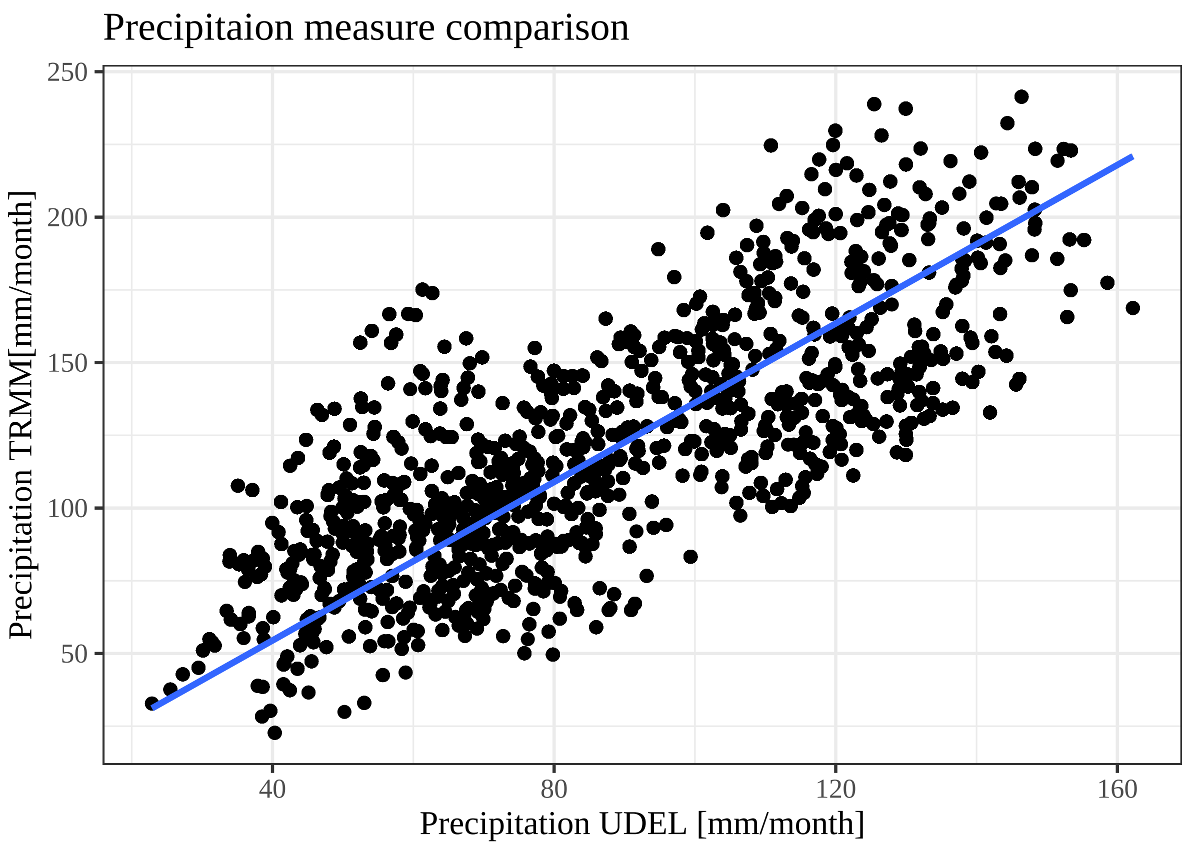

Figure 13: Correlation of TRMM and UDEL data for the LSMS household data. The correlation coefficient is 0.73.

Figure 14 visualizes the relation between mean precipitation levels and precipitation risk. While there is a clear positive relation between mean precipitation levels and the standard deviation of precipitation there is no obvious relationship between mean precipitation levels and the coefficient of variation of precipitation.

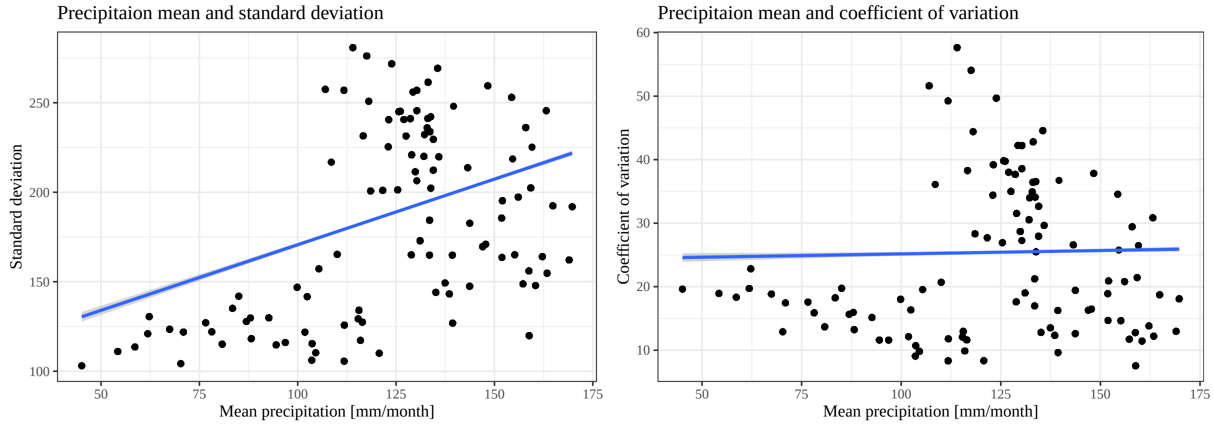

Figure 14: Precipitation levels and measures of precipitation risk.

B Proofs

Dividing (3a) by (3b), and (3a) by (3c),

$$l_k = \frac{\beta}{\alpha}\,\frac{\lambda}{\omega}\,a_k \tag{19}$$

$$x_{ik} = \frac{\gamma}{\alpha}\,\frac{\lambda}{q}\,a_k \tag{20}$$

plugging into (3a)

$$\alpha\,p_k\,a_k^{\alpha-1}\left(\frac{\beta}{\alpha}\,\frac{\lambda}{\omega}\,a_k\right)^{\beta}\left(\frac{\gamma}{\alpha}\,\frac{\lambda}{q}\,a_k\right)^{\mu\,\gamma}\underbrace{\exp\left(\int_{\mu}^{1}\gamma\,\ln(e_{ik})\,di\right)}_{=:\,r_k^{(1-\mu)\,\gamma}} = \lambda. \tag{21}$$

Rearranging,

$$a_k^{1-\alpha-\beta-\mu\,\gamma} = \alpha^{1-\beta-\mu\,\gamma}\,\beta^{\beta}\,\gamma^{\mu\,\gamma}\,p_k\,r_k^{(1-\mu)\,\gamma}\,\lambda^{\beta+\mu\,\gamma-1}\,\omega^{-\beta}\,q^{-\mu\,\gamma} \tag{22}$$

Integrating over crops, and using $\gamma = 1-\alpha-\beta$, we obtain with positive constants Ω_A and Ω_L

$$A = \Omega_A \int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\,r_k\,q^{-\frac{\mu}{1-\mu}}\,dk \tag{23}$$

$$L = \Omega_L \int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\,r_k\,q^{-\frac{\mu}{1-\mu}}\,dk \tag{24}$$

Using this in (22), we obtain (4).

The aggregate output of crop k is obtained by plugging (4) in the production function (1) as

$$y_k = A^\alpha\, L^\beta \left(\frac{p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k}{\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k\, dk} \right)^{\alpha+\beta} \exp\left(\int_0^\mu \gamma \ln\left(\frac{\gamma\, p_k\, y_k}{q} \right) di + \int_\mu^1 \gamma \ln(e_{ik})\, di \right) \tag{25}$$

$$= A^\alpha\, L^\beta \left(\frac{p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k}{\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k\, dk} \right)^{\alpha+\beta} \left(\frac{\gamma\, p_k\, y_k}{q} \right)^{\mu\,\gamma} r_k^{(1-\mu)\,\gamma} \tag{26}$$

Solving yields (27):

$$y_k = A^{\frac{\alpha}{1-\mu\,\gamma}}\, L^{\frac{\beta}{1-\mu\,\gamma}} \left(\frac{p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k}{\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k\, dk} \right)^{\frac{\alpha+\beta}{1-\mu\,\gamma}} \left(\frac{\gamma\, p_k}{q} \right)^{\frac{\mu\,\gamma}{1-\mu\,\gamma}} r_k^{\frac{(1-\mu)\,\gamma}{1-\mu\,\gamma}} \tag{27}$$

Using the previous results, we also find the optimized profit per crop k as:

$$\pi_k = p_k\, y_k - \int_0^\mu q\, x_{ik}\, di = p_k\, y_k - \mu\,\gamma\, p_k\, y_k = (1-\mu\,\gamma)\, p_k\, y_k \tag{28}$$

$$= (1-\mu\,\gamma)\, A^{\frac{\alpha}{1-\mu\,\gamma}}\, L^{\frac{\beta}{1-\mu\,\gamma}} \left(\frac{p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k}{\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k\, dk} \right)^{\frac{\alpha+\beta}{1-\mu\,\gamma}} \left(\frac{\gamma}{q} \right)^{\frac{\mu\,\gamma}{1-\mu\,\gamma}} r_k^{\frac{(1-\mu)\,\gamma}{1-\mu\,\gamma}}\, p_k^{\frac{1}{1-\mu\,\gamma}} \tag{29}$$

$$= (1-\mu\,\gamma)\, A^{\frac{\alpha}{1-\mu\,\gamma}}\, L^{\frac{\beta}{1-\mu\,\gamma}} \frac{p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k}{\left(\int_0^N p_k^{\frac{1}{(1-\mu)\,\gamma}}\, r_k\, dk \right)^{\frac{\alpha+\beta}{1-\mu\,\gamma}}} \left(\frac{\gamma}{q} \right)^{\frac{\mu\,\gamma}{1-\mu\,\gamma}} \tag{30}$$

Aggregate profit thus is (5).

C Crop diversity on plot level

Table 5 reports the results on plot level. It shows that larger plots and higher labor inputs are associated with groing multiple crops per plot. In other words, farmers plant more different crops on larger plots and they also allocate more labor towards those plots with multicropping. However, because we do not have an instrument for labor and land at the plot level, the results lack causal interpretation.

Table 5: Crop diversity on plot level

	log(number)				Simpson
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	ER	OLS	OLS
log(plot_area)	0.023***	0.046***	0.049***	0.017***	0.252***
	(0.007)	(0.009)	(0.002)	(0.003)	(0.024)
log(labor)	0.061***	0.071***	0.071***	0.065***	0.107***
	(0.005)	(0.005)	(0.003)	(0.004)	(0.009)
District FE	✓				
HH FE		✓	✓		✓
Plot FE				✓	
Season FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Observations	46,076	46,076	46,076	46,076	46,076
R^2	0.101	0.246	0.246	0.612	0.280

All models include year and season-region fixed effects. Additional fixed effects are household fixed effects (HH) or district fixed effects (District). Standard errors are clustered at the district level. Significance levels are $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

D Crop diversity and droughts

In this section we test the impact of weather shocks on our results. Weather shocks could affect our outcomes through our main explanatory variables but they could also affect our main outcomes directly. The former case would hint at an omitted variables problem while the latter case can be seen as an extension of our discussion of crop diversity and risk. Overall, we find that including weather shocks in our main regression specification does not affect our estimates from Section 6 which we interpret as the lack of omitted variable bias. However, we do find some evidence that previous weather shocks increase crop diversity. This finding suggests the possibility that risk plays a role in the diversification decision. Although this finding merits further investigation, it does not invalidate any of our results.

Figure 5 shows that households perceive droughts as the most important reason for income shortfalls. We therefore introduce different drought measures to describe weather shocks. Droughts could affect the results instantaneously through e.g. reducing the number of crops that the farmer plants and the resulting revenues but it can also affect future planting decisions by changing risk perceptions or through asset dynamics (e.g. a farmer may not have enough seeds in the next season after a bad harvest). We therefore include drought measures for the current growing season (drought), the previous growing season (drought6), the same growing season but in the previous year (drought12) and the same growing season but two years ago (drought24) using precipitation data based on the TRMM and the SPEI drought index. The advantage of the SPEI index is that it includes temperature data to calculate the severity of droughts though changes in potential evapotranspiration.

Table 6 reproduces the results of Table 2 from the main text but with additional drought controls. The first column is identical to column (1) of Table 2. Column (2) adds the negative of the seasonal precipitation level. We use negative precipitation levels to make the results comparable to the results with drought dummy variables. Column (3) adds a dummy that indicates whether seasonal precipitation was one standard deviation below average levels, column (4) uses the seasonal negative of the SPEI index to measure droughts and column (5) uses a dummy that indicates whether the drought index was one stand deviation below average seasonal drought levels. Overall, adding different measures of droughts does not change the estimated impact of farm size and labor on crop diversity. However, droughts seem to matter for crop diversity but the results are not consistent across the different drought measures. Only droughts two years ago have a consistently positive impact on crop diversity suggesting that farmers respond to experiencing risk by diversifying crop production.

Table 6: Farm size, labor and crop diversity and droughts

	log(number)				
	(1)	(2)	(3)	(4)	(5)
log(farm_size)	0.153***	0.153***	0.153***	0.152***	0.153***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
log(labor)	0.179***	0.177***	0.178***	0.180***	0.180***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
drought		−0.0001***	0.258***	0.005	0.010
		(0.00003)	(0.066)	(0.004)	(0.025)
drought6		0.0001***	0.203*	−0.003	0.019
		(0.0001)	(0.114)	(0.004)	(0.028)
drought12		−0.0001	−0.005	0.006***	0.004
		(0.00004)	(0.060)	(0.002)	(0.021)
drought24		−0.0001*	0.063***	0.013***	0.062***
		(0.00004)	(0.013)	(0.004)	(0.023)
Drought measure	none	TRMM	1(TRMM)	SPEI	1(SPEI)
Observations	17,566	17,566	17,566	17,566	17,566
R^2	0.597	0.600	0.597	0.599	0.598

All specifications are as specification (2) of Table 2 with additional drought variables. The drought variables are for the current growing season (drought), the previous growing season (drought6), the same growing season but in the previous year (drought12) and the same growing season but two years ago (drought24). We measure droughts as the negative of the seasonal precipitation (TRMM), a dummy that indicates whether the seasonal precipitation was one standard deviation below the average seasonal precipitation (1(TRMM)), the negative of the seasonal mean SPEI (SPEI) and a dummy that indicates whether the seasonal SPEI was one standard deviation below the average (1(SPEI)). All models include year, season and household fixed effects. Standard errors are clustered at the district level. Significance levels are $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 7 reproduces the results of Table 3 from the main text but with the same drought measures as described for Table 6. Again, column (1) is identical to column (2) of Table 3 (our main specification). Overall, the impact of crop diversity on revenues is unaffected by adding droughts as controls. Droughts measured based on precipitation from the TRMM data have no consistent impact on revenues while droughts measured by the SPEI index consistently reduce agricultural production. The main difference between both measures is that the SPEI includes temperature to calculate the water demand.

Table 7: Revenues, crop diversity and weather at the household level

	log(revenues)				
	(1)	(2)	(3)	(4)	(5)
log(number)	0.336***	0.343***	0.336***	0.331***	0.339***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.039)
log(farm_size)	0.256***	0.256***	0.257***	0.255***	0.253***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
log(labor)	0.240***	0.238***	0.240***	0.244***	0.245***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
drought		0.0002***	0.549	−0.036***	−0.012
		(0.0001)	(0.492)	(0.008)	(0.062)
drought6		0.0001	−1.013***	−0.062***	−0.269***
		(0.0001)	(0.096)	(0.009)	(0.060)
drought12		0.0003***	0.507***	−0.001	−0.163***
		(0.0001)	(0.062)	(0.007)	(0.040)
Observations	17,566	17,566	17,566	17,566	17,566
R^2	0.606	0.607	0.606	0.609	0.608

All specifications are as specification (2) of Table 2 with additional drought variables. The drought variables are for the current growing season (drought), the previous growing season (drought6), the same growing season but in the previous year (drought12) and the same growing season but two years ago (drought24). We measure droughts as the negative of the seasonal precipitation (TRMM), a dummy that indicates whether the seasonal precipitation was one standard deviation below the average seasonal precipitation (1(TRMM)), the negative of the seasonal mean SPEI (SPEI) and a dummy that indicates whether the seasonal SPEI was one standard deviation below the average (1(SPEI)). All models include year and season-region fixed effects. Additional fixed effects are household fixed effects (HH) or district fixed effects (District). Standard errors are clustered at the district level. Significance levels are $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 8 reproduces the results of specification (2) from Table 8 with additional drought controls. Although the estimated impact of crop diversity on production is robust to including additional drought variables, there is no consistent direct impact of droughts on plot level harvests.

Table 8: Revenues, crop diversity and weather at the crop level

	log(harvest_kg)				
	(1)	(2)	(3)	(4)	(5)
log(crop_area)	0.292^{***}	0.292^{***}	0.292^{***}	0.292^{***}	0.293^{***}
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
log(labor)	0.182^{***}	0.182^{***}	0.182^{***}	0.182^{***}	0.181^{***}
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
log(number)	0.143^{***}	0.143^{***}	0.143^{***}	0.143^{***}	0.143^{***}
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
drought		-0.00003	0.277^{**}	-0.003	0.102
		(0.0001)	(0.130)	(0.009)	(0.064)
drought6		0.0001	-0.510^{***}	0.003	0.189^{***}
		(0.0001)	(0.054)	(0.009)	(0.059)
drought12		0.0001	0.435^{***}	-0.001	0.014
		(0.0001)	(0.057)	(0.003)	(0.028)
Observations	65,851	65,851	65,851	65,851	65,851
R^2	0.523	0.523	0.523	0.523	0.523

All specifications are as specification (2) of Table 4 with additional drought variables. The drought variables are for the current growing season (drought), the previous growing season (drought6), the same growing season but in the previous year (drought12) and the same growing season but two years ago (drought24). We measure droughts as the negative of the seasonal precipitation (TRMM), a dummy that indicates whether the seasonal precipitation was one standard deviation below the average seasonal precipitation (1(TRMM)), the negative of the seasonal mean SPEI (SPEI) and a dummy that indicates whether the seasonal SPEI was one standard deviation below the average (1(SPEI)). All models include year, season and household fixed effects. Standard errors are clustered at the district level. Significance levels are $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

E Parameters for Figure 11

The output elasticity of crop diversity is a function of the control over the environment (μ), the productivity (price) differences between crops (φ) and the output elasticity of ecosystem services or nutrients (γ). To visualize the relation of the output elasticity of diversity, μ and ϕ we need to quantify γ. However, since we have three unknown parameter and one equation we need to solve for the other parameters first.

The results reported in column (2) of Table 4 use crop specific fixed effects in addition to household fixed effects, year fixed effects and season fixed effects. The estimated crop fixed effects equal $\varphi \log(k)$ while $\log(p_1)$ is absorbed by the other fixed effects. Table 9 reports the estimated crop fix effects for crops with more than 100 observations. The table also provides the standard errors, the number of observations (n) and the rank (k). The column 'Price' reports the exponential of the crop fixed effects for a regression through the origin without the other fixed effects (Price $\sim p_k = p_1 k^{-\varphi}$).

Table 9: Crop productivity (crop fixed effects)

Crop name	Estimate	Standard error	Price	n	k
Rice	1.102	0.171	2384.679	290	1
Tomatoes	0.891	0.149	2348.585	311	2
Cotton	0.693	0.146	1472.506	160	3
Irish Potatoes	0.499	0.156	1560.803	1057	4
Banana Food	0.443	0.139	1414.644	12382	5
Cassava	0.437	0.12	1239.502	6287	6
Onions	0.412	0.176	1521.088	139	7
Coffee All	0.395	0.141	1152.09	2046	8
Sweet Potatoes	0.25	0.121	1019.774	5976	9
Finger Millet	0.152	0.136	986.608	1450	10
Simsim	0.118	0.136	921.008	673	11
Sugarcane	0.102	0.196	848.207	169	12
Groundnuts	0.071	0.135	912.676	3338	13
Pumpkins	0.05	0.151	812.393	123	14
Yam	0.016	0.13	782.335	333	15
Pigeon Peas	-0.002	0.164	888.928	220	16
Beans	-0.038	0.122	811.262	13224	17
Sunflower	-0.052	0.206	680.74	207	18
Other	-0.135	0.179	611.025	186	19
Field Peas	-0.177	0.146	620.714	339	20
Sorghum	-0.295	0.143	505.46	1608	21
Maize	-0.414	0.125	535.427	12373	22
Soya Beans	-0.457	0.176	561.578	412	23
Banana Sweet	-0.464	0.138	548.784	643	24
Banana Beer	-0.481	0.164	571.949	1237	25
Cow Peas	-0.581	0.164	437.8	105	26

Figure 15 plots the estimated crop fixed effects against the log rank from Table 9. The slope of the linear regression is $\varphi = 0.5$.

Assuming no control over the environment ($\mu = 0$) simplifies the output elasticity of crop diversity to $\gamma - \varphi$. Using the estimate from the main specification (2) of Table 3 yields $\gamma = 0.835$.

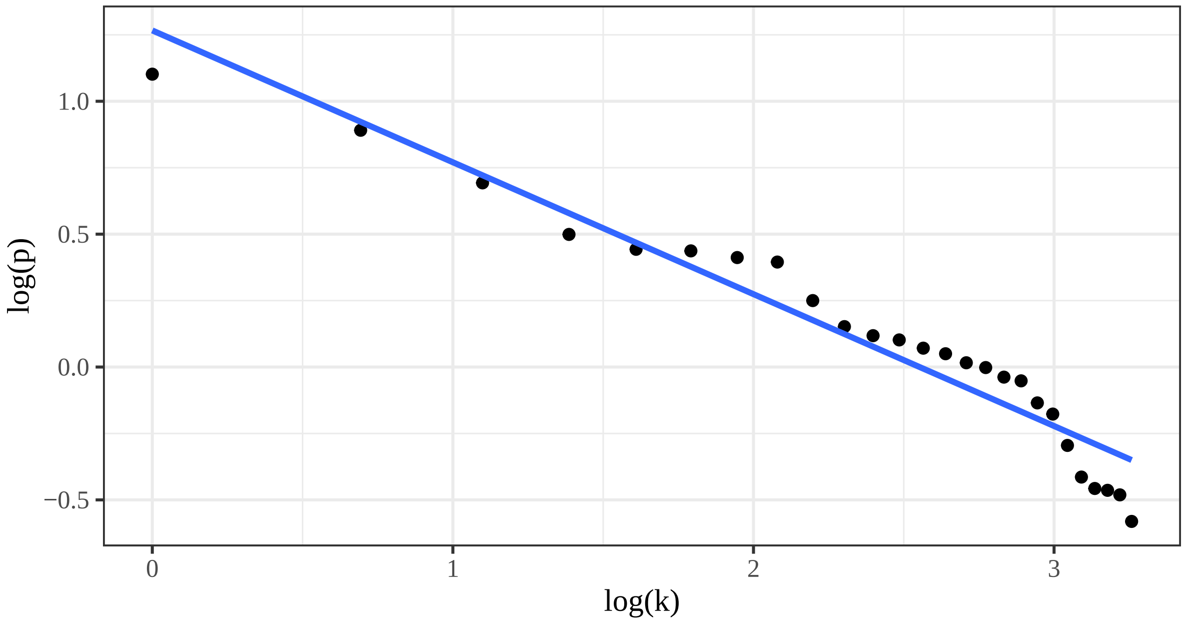

Figure 15: Estimated crop fixed effects and the log rank of the estimated crop fixed effect ($\log(k)$). The blue line is a linear regression with slope $\varphi = -0.5$.

F Labor markets

Efficient labor markets require $\frac{dY}{dL} = w$ where Y are the farm level revenues, L is farm level labor and w is the wage rate. Using equation (11) we can solve for L

$$\frac{dY}{dL} = \frac{\beta}{1-\mu\gamma}\left(\Omega N^{(1-\mu)\gamma-\varphi} A^{\alpha}\right)^{\frac{1}{1-\mu\gamma}} L^{\frac{\beta}{1-\mu\gamma}-1} = w$$

$$\iff$$

$$L = \left[\frac{w(1-\mu\gamma)}{\beta}\left(\Omega N^{(1-\mu)\gamma-\varphi} A^{\alpha}\right)^{\frac{1}{\mu\gamma-1}}\right]^{\frac{1}{1-\frac{\beta}{1-\mu\gamma}}}$$

Next we solve for w using the labor constraint

$$\bar{L} = \sum_i \left[\frac{w(1-\mu\gamma)}{\beta}\left(\Omega N_i^{(1-\mu)\gamma-\varphi} A_i^{\alpha}\right)^{\frac{1}{\mu\gamma-1}}\right]^{\frac{1}{1-\frac{\beta}{1-\mu\gamma}}}$$

$$\iff$$

$$w = \left[\bar{L}^{-1}\frac{(1-\mu\gamma)}{\beta}\sum_i\left(\Omega^{\frac{1}{\mu\gamma-1}} N_i^{\frac{(1-\mu)\gamma-\varphi}{\mu\gamma-1}} A_i^{\frac{\alpha}{\mu\gamma-1}}\right)^{\frac{1}{1-\frac{\beta}{1-\mu\gamma}}}\right]^{\frac{\beta}{1-\mu\gamma}-1}$$

where $\bar{L}$ is the aggregate rural labor supply, i is the individual farm. We use the survey year 2015/16 as baseline with $\bar{L} = 448952$. Using the parameter values of column (2) from Table 2 and 3 yields $\frac{1-\mu\gamma}{\beta} = 4.807692$ or $\frac{\beta}{1-\mu\gamma} = 0.208$ and $\Omega^{\frac{1}{\mu\gamma-1}} N_i^{\frac{(1-\mu)\gamma-\varphi}{\mu\gamma-1}} A_i^{\frac{\alpha}{\mu\gamma-1}} = \frac{Y_i}{L^{\frac{\beta}{1-\mu\gamma}}}$.

References

ABADIE, A., S. ATHEY, G. W. IMBENS, AND J. WOOLDRIDGE (2017): "When should you adjust standard errors for clustering?" Tech. rep., National Bureau of Economic Research.

ABAY, K. A., L. BEVIS, AND C. B. BARRETT (2019): "Measurement Error Mechanisms Matter: Agricultural intensification with farmer misperceptions and misreporting," Tech. rep., National Bureau of Economic Research.

AUFFHAMMER, M. AND T. A. CARLETON (2018): "Regional Crop Diversity and Weather Shocks in India," *Asian Development Review*, 35, 113–130.

BANERJEE, A. V. AND E. DUFLO (2007): "The economic lives of the poor," *Journal of economic perspectives*, 21, 141–168.

BAUMGÄRTNER, S., M. A. DRUPP, J. N. MEYA, J. M. MUNZ, AND M. F. QUAAS (2017): "Income inequality and willingness to pay for public environmental goods," *Journal of Environmental Economics and Management*, 85, 35–61.

BELLORA, C., É. BLANC, J.-M. BOURGEON, AND E. STROBL (2017): "Estimating the Impact of Crop Diversity on Agricultural Productivity in South Africa," *NBER Working Paper*.

BELLORA, C. AND J.-M. BOURGEON (2019): "Food trade and biodiversity effects," *International Economic Review*, 60, 1957–1999.

BENIN, S., M. SMALE, J. PENDER, B. GEBREMEDHIN, AND S. EHUI (2004): "The economic determinants of cereal crop diversity on farms in the Ethiopian highlands," *Agricultural Economics*, 31, 197–208.

BROCK, W. A. AND A. XEPAPADEAS (2003): "Valuing biodiversity from an economic perspective: a unified economic, ecological, and genetic approach," *American Economic Review*, 93, 1597–1614.

BURKE, M. AND D. B. LOBELL (2017): "Satellite-based assessment of yield variation and its determinants in smallholder African systems," *Proceedings of the National Academy of Sciences*, 114, 2189–2194.

CARLETTO, C., S. SAVASTANO, AND A. ZEZZA (2013): "Fact or artifact: The impact of measurement errors on the farm size–productivity relationship," *Journal of Development Economics*, 103, 254–261.

Chuang, Y. (2019): "Climate variability, rainfall shocks, and farmers' income diversification in India," *Economics Letters*, 174, 55–61.

Desiere, S. and D. Jolliffe (2018): "Land productivity and plot size: Is measurement error driving the inverse relationship?" *Journal of Development Economics*, 130, 84–98.

Di Falco, S. and J.-P. Chavas (2009): "On crop biodiversity, risk exposure, and food security in the highlands of Ethiopia," *American Journal of Agricultural Economics*, 91, 599–611.

Díaz, S., U. Pascual, M. Stenseke, B. Martín-López, R. T. Watson, Z. Molnár, R. Hill, K. M. A. Chan, I. A. Baste, K. A. Brauman, S. Polasky, A. Church, M. Lonsdale, A. Larigauderie, P. W. Leadley, A. P. E. van Oudenhoven, F. van der Plaat, M. Schröter, S. Lavorel, Y. Aumeeruddy-Thomas, E. Bukvareva, K. Davies, S. Demissew, G. Erpul, P. Failler, C. A. Guerra, C. L. Hewitt, H. Keune, S. Lindley, and Y. Shirayama (2018): "Assessing nature's contributions to people," *Science*, 359, 270–272.

Dillon, A., S. Gourlay, K. McGee, and G. Oseni (2019): "Land measurement bias and its empirical implications: evidence from a validation exercise," *Economic Development and Cultural Change*, 67, 595–624.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic competition and optimum product diversity," *The American economic review*, 67, 297–308.

Eeckhoudt, L., C. Gollier, and H. Schlesinger (2005): *Economic and financial decisions under risk*, Princeton University Press.

Foster, A. D. and M. R. Rosenzweig (2017): "Are there too many farms in the world? Labor-market transaction costs, machine capacities and optimal farm size," Tech. rep., National Bureau of Economic Research.

Gollier, C. (2001): *The economics of risk and time*, Cambridge, Mass: MIT Press.

Gollin, D., M. Smale, and B. Skovmand (2000): "Searching an ex situ collection of wheat genetic resources," *American Journal of Agricultural Economics*, 82, 812–827.

Gollin, D. and C. Udry (2019): "Heterogeneity, measurement error and misallocation: Evidence from african agriculture," *NBER Working Paper*.

Harpole, W. S. and D. Tilman (2007): "Grassland species loss resulting from reduced niche dimension," *Nature*, 446, 791–793.

Haughton, J. and S. R. Khandker (2009): *Handbook on poverty and inequality*, World Bank Publications.

HILL, M. O. (1973): "Diversity and evenness: a unifying notation and its consequences," *Ecology*, 54, 427–432.

HOOPER, D. U., F. CHAPIN, J. J. EWEL, A. HECTOR, P. INCHAUSTI, S. LAVOREL, J. LAWTON, D. LODGE, M. LOREAU, S. NAEEM, ET AL. (2005): "Effects of biodiversity on ecosystem functioning: a consensus of current knowledge," *Ecological monographs*, 75, 3–35.

KARLAN, D., R. OSEI, I. OSEI-AKOTO, AND C. UDRY (2014): "Agricultural decisions after relaxing credit and risk constraints," *The Quarterly Journal of Economics*, 129, 597–652.

KRUGMAN, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

LARSEN, A. E. AND F. NOACK (2017): "Identifying the landscape drivers of agricultural insecticide use leveraging evidence from 100,000 fields," *Proceedings of the National Academy of Sciences*, 114, 5473–5478.

——— (2020): "Impact of local and landscape complexity on the stability of field-level pest control," *Nature Sustainability*, 1–9.

LOBELL, D. B., G. AZZARI, M. BURKE, S. GOURLAY, Z. JIN, T. KILIC, AND S. MURRAY (2020): "Eyes in the Sky, Boots on the Ground: Assessing Satellite-and Ground-Based Approaches to Crop Yield Measurement and Analysis," *American Journal of Agricultural Economics*, 102, 202–219.

MICHLER, J. D. AND A. L. JOSEPHSON (2017): "To specialize or diversify: agricultural diversity and poverty dynamics in Ethiopia," *World Development*, 89, 214–226.

MILLENNIUM ECOSYSTEM ASSESSMENT (2005): *Ecosystems and Human Well-Being: Synthesis Report*, Washington DC: Island Press.

NOACK, F. AND A. LARSEN (2019): "The contrasting effects of farm size on farm incomes and food production," *Environmental Research Letters*, 14, 084024.

NOACK, F., M.-C. RIEKHOF, AND S. D. FALCO (2019): "Droughts, Biodiversity and Rural Incomes in the Tropics," *Journal of the Association of Environmental and Resource Economists*.

PANZAR, J. C. AND R. D. WILLIG (1981): "Economies of scope," *The American Economic Review*, 71, 268–272.

POORE, J. AND T. NEMECEK (2018): "Reducing food's environmental impacts through producers and consumers," *Science*, 360, 987–992.

QUAAS, M. F. AND T. REQUATE (2013): "Sushi or Fish Fingers? Seafood Diversity, Collapsing Fish Stocks and Multi-species Fishery Management," *Scandinavian Journal of Economics*, 115, 381–422.

RENARD, D. AND D. TILMAN (2019): "National food production stabilized by crop diversity," *Nature*, 571, 257–260.

ROMER, P. M. (1987): "Growth based on increasing returns due to specialization," *The American Economic Review*, 77, 56–62.

SMALE, M., J. HARTELL, P. W. HEISEY, AND B. SENAUER (1998): "The contribution of genetic resources and diversity to wheat production in the Punjab of Pakistan," *American Journal of Agricultural Economics*, 80, 482–493.

SPRINGMANN, M., M. CLARK, D. MASON-D'CROZ, K. WIEBE, B. L. BODIRSKY, L. LASSALETTA, W. DE VRIES, S. J. VERMEULEN, M. HERRERO, K. M. CARLSON, M. JONELL, M. TROELL, F. DECLERCK, L. J. GORDON, R. ZURAYK, P. SCARBOROUGH, M. RAYNER, B. LOKEN, J. FANZO, H. C. J. GODFRAY, D. TILMAN, J. ROCKSTROM, AND W. WILLETT (2018): "Options for keeping the food system within environmental limits," *Nature*, 562, 519–525.

TESFAYE, W. AND N. TIRIVAYI (2020): "Crop diversity, household welfare and consumption smoothing under risk: Evidence from rural Uganda," *World Development*, 125, 104686.

TILMAN, D., M. CLARK, D. R. WILLIAMS, K. KIMMEL, S. POLASKY, AND C. PACKER (2017): "Future threats to biodiversity and pathways to their prevention," *Nature*, 546, 73–81.

VANDERMEER, J. H. (1992): *The ecology of intercropping*, Cambridge University Press.

WEITZMAN, M. L. (2000): "Economic profitability versus ecological entropy," *The Quarterly Journal of Economics*, 115, 237–263.

WOOLDRIDGE, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press.