Selection and Absolute Advantage in Farming and Entrepreneurship: Microeconomic Evidence and Macroeconomic Implications*

Francisco Alvarez-Cuadrado† Francesco Amodio‡ Markus Poschke§

November 1, 2018

Abstract

Labor productivity is lower in poor countries than in rich countries, and relatively more so in the agricultural sector. Self-selection of heterogeneous workers can account for part of these differences if comparative and absolute advantage are aligned in agriculture. In that case, average productivity in agriculture increases when the employment share of the sector decreases. We empirically investigate the correlation between comparative and absolute advantage using representative household-level panel data from four Sub-Saharan African countries. We find that (i) around one third of households engage in both agriculture and non-farming entrepreneurship; (ii) households active in both sectors have systematically higher agricultural productivity than those doing only farming; (iii) among households active in both sectors, those with higher agricultural productivity supply relatively fewer hours in agriculture, while those with higher profits from entrepreneurship supply relatively more hours in this sector; (iv) over time, households starting a non-farming enterprise have higher baseline agricultural productivity than those who remain only farmers. Taken together, these results suggest that comparative and absolute advantage are misaligned in agriculture, casting doubt on the importance of selection as a root cause of the agricultural productivity gap.

Keywords: agricultural productivity gap, selection, entrepreneurship, Africa.

JEL Codes: J24, J31, J43, L26, O11, O13, O40.

*We are thankful to Yifan Li, Lesley Johnson and Masaya Takano for excellent research assistance. Errors remain our own. We gratefully acknowledge the International Growth Centre for financial support.

†francisco.alvarez-cuadrado@mcgill.ca, McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada.

‡francesco.amodio@mcgill.ca, McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada.

§markus.poschke@mcgill.ca , McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada.

1 Introduction

It is well-known that cross-country differences in output per capita are enormous. Output per capita in the country at the 90th percentile of the world income distribution is 22 times that in the country at the 10th percentile (Caselli 2005). These differences are even larger in agriculture, where the factor is 45 (Caselli 2005). A large part of these differences cannot be accounted for by differences in observables (Jones and Romer 2010, Young 2013, Gollin, Lagakos, and Waugh 2014). These differences are particularly important given that in poor countries agriculture accounts for the majority of employment, whereas it accounts for a negligible employment share in the richest countries. What causes these enormous differences?

A recent influential literature (Lagakos and Waugh 2013) argues that an important source is worker *selection*: if farmers in poor countries have particularly low farming ability or productivity, this can explain part of the productivity differences.[1] The basic intuition is simple: if ability distributions are similar across countries and in each country, the best potential farmers actually engage in farming, then in countries with few farmers only the best farmers will be active. In countries with more farmers, farmers from the same top segment of the farming ability distribution are also active, but they are accompanied by a potentially large group of less productive farmers. Therefore, the average ability of *active* farmers is lower in countries with more farmers. In the language of the literature, this occurs if comparative advantage (which determines individuals' sectoral choice) and absolute advantage (their ability or productivity) are positively correlated, or aligned: those who choose farming are also the best farmers. Evidence on this issue in the existing literature generally comes from indirect methods or from sector switchers.

We take a new, more direct approach to investigating whether this pattern holds using household-level data from four African countries: Ethiopia, Malawi, Nigeria, and Uganda. The data we use come from the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) project, which provides uniquely rich data on agricultural production and non-farming entrepreneurship (Restuccia and Santaeulalia-Llopis 2017). The four countries are all poor, have low agricultural productivity, and large shares of employment in agriculture. At the same time, rural households in these countries engage in non-agricultural self-employment at quite high rates – from 27% in Malawi to 51% in Nigeria – and occasionally switch activities.[2] Therefore, we can use data on these households to study patterns of sectoral choices, and in particular, the alignment of comparative and absolute advantage.

In general, this is a difficult undertaking because in a cross-section one only observes data that

[1] More generally, the observed difference across sectors could be due to differences in productivity, or to differences in unobserved factor inputs. Selection on unobserved ability falls into the second category.

[2] 22% of Africa's working-age population are starting new businesses, the highest rate in the world (OECD 2017).

has been shaped by selection. In most settings, it is not known what a farming household's non-agricultural productivity is. The same holds for the farming ability of a household engaged only in non-agricultural work. As a result, the literature has relied on panel data, and in particular on the information revealed by those switching sector, since they can be observed in both activities. A disadvantage of this approach is that it only focuses on households at the margin between activities and is therefore not informative of the alignment of comparative and absolute advantage in the population.

We take a different approach. The key motivating observation is that, in our sample countries, 30% of households engage in *both* agricultural and non-agricultural work. This provides us with a very large group of households for which we observe contemporaneous information from both sectors. As a consequence, we can make several useful comparisons. First, we can compare the agricultural productivity of households only engaged in agriculture to those also engaged in non-agricultural self-employment. This is informative about the correlation of comparative advantage in agriculture (which is weaker for those engaged in both activities) with absolute advantage (their agricultural productivity), which we observe for both groups. We can do the same for those engaged in only non-agricultural self-employment, comparing them to those engaged in both activities. Second, we can also analyze how production decisions of those engaged in both activities align with their productivity in the two sectors.

To do so, we first carefully measure agricultural and non-agricultural outputs and inputs for each household. We then classify households depending on whether they do only agriculture, only non-agricultural self-employment, or both, based on whether they have any returns and costs associated with these activities. With this information in hand, we can check the alignment of advantages. A simple and general theoretical analysis tells us that when households differ in their productivity in two activities, and have the option to pursue either only one of these or both, then those with a strong comparative advantage will choose to pursue a single activity, while those with a weak comparative advantage will pursue both. In the data, we find that among those households in a village who produce some agricultural output, it is the more productive ones – high absolute advantage – who also engage in entrepreneurship – weak comparative advantage. That is, comparative and absolute advantage are not aligned in agriculture. Among those pursuing non-agricultural self-employment, in contrast, there is no clear relationship between productivity and the propensity to also engage in agriculture.

The finding that it is the best active farmers who also engage in non-agriculture is important, and informative about selection patterns and their consequences. The fact that we find misalignment between comparative and absolute advantages in agriculture in three out of the four African countries we analyze suggests that selection may not be central in explaining productivity differences in agriculture.

How can the misalignment between comparative and absolute advantages in agriculture be rationalized? There are several possibilities, most notably the following: household productive abilities in agriculture and non-agriculture are strongly positively correlated, and returns to non-agriculture are more dispersed. Then it is a core prediction of the classical Roy model that higher-ability households will tend to choose the activity with higher return dispersion (Roy 1951; Young 2014). In our setting, this means that the best farmers can reap even higher returns outside agriculture, and therefore specialize there. Intermediate-ability farmers still can reap relatively high returns outside agriculture, and therefore pursue both activities. The lowest-ability farmers, in contrast, face very low returns outside agriculture, and therefore only pursue agriculture. This is consistent with the activity choice patterns we observe for farmers.

We provide further evidence in favor of this scenario by looking at the activity of households along the intensive margin. Among the set of households pursuing both activities, those with higher productivity in agriculture work fewer hours in agriculture relative to non-agriculture. Those households with higher productivity in non-agriculture instead work more hours in this sector relative to agriculture. While puzzling at first, this finding is intuitive in the scenario with strongly correlated abilities: when being a good farmer is associated with even higher returns outside farming, better farmers spend less time farming. Evidence from switchers is also consistent with our main findings. Using the panel dimension of the data, we can follow households over time. Doing so, we find that, over time, among the households that at baseline only engage in farming, it is the most productive ones who are most likely to also start a non-agricultural enterprise.

To summarize, the fact that a large fraction of households in rural Africa engages in both agricultural and non-agricultural work allows us to obtain evidence on the correlation of comparative and absolute advantages in agriculture and non-agriculture as well as on the household-level correlation of productivities. The fact that the best farmers are more likely to also engage in non-agriculture clearly shows a negative correlation, or misalignment, of comparative and absolute advantage in agriculture. The same conclusion can be drawn from the fact that, among those engaged in both activities, the more productive farmers spend fewer hours farming, and more hours in non-agricultural activities. Both results also point towards a situation where it is a strong positive correlation of productive abilities in the two sectors that underlies the misalignment. Finally, the lack of a significant association between non-agricultural productivity and the choice to also engage in farming suggests the presence of barriers to entry or fixed costs in farming.

Our paper is not the first one to analyze the correlation of comparative and absolute advantage or the correlation of productive abilities across sectors. Earlier work has relied on information from sector switchers or on sector-level evidence to back out these correlations. For example, Lagakos and Waugh (2013) calibrate the joint distribution of ability in agricultural and non-agricultural work using micro-level US wage data. They infer a moderately positive correlation of abilities from the ratio of average wages in the two sectors. Adamopoulos, Brandt, Leight, and Restuccia (2017) calibrate the same correlation coefficient using information from sector switchers in Chinese panel data. Through the lens of their model, the observed weak correlation between the agricultural and non-agricultural incomes of switchers implies a negative correlation of abilities. Hicks, Kleemans, Li, and Miguel (2017) use individual-level panel data from Indonesia and Kenya to estimate wage gains from sector switches, conditional on individual fixed effects. They find that wage gains for switchers from agriculture to non-agriculture are much smaller than average earnings differences between the two activities. Their findings are consistent with ours in showing that rural-to-urban migrants are positively selected relative to non-migrants.

Our paper is organized as follows. Section 2 briefly sketches a simple, general theory motivating our analysis. Section 3 describes data sources, and Section 4 shows descriptive statistics. Section 5 contains the main results on patterns of selection. Section 6 extends the simple theory slightly to motivate Section 7, which shows key empirical results on the correlation of abilities. Section 8 provides additional results from switchers. Section 9 concludes.

2 A Simple Model of Selection

This section describes a simple, general model that motivates the following empirical analysis. Consider an economy with two sectors, agriculture and non-agriculture, denoted by a and n respectively. There is a measure 1 of households indexed by i. These households are heterogeneous in terms of their abilities in the two sectors. In particular, each household is endowed with a vector of sector-specific abilities $\{z_i^a, z_i^n\}$. These abilities are drawn from a joint distribution $G(z^a, z^n)$ with support on the positive reals and finite mean μ_j and variance σ_j^2, where $j = \{a, n\}$. We define agricultural comparative advantage as the ratio of agricultural to non-agricultural abilities z_i^a/z_i^n while agricultural absolute advantage is given by agricultural ability z_i^a. Similarly, non-agricultural comparative and absolute advantage are given by z_i^n/z_i^a and z_i^n, respectively.

Household i is endowed with one unit of time that it allocates between agriculture l_i^a and non-agriculture $l_i^n = 1 - l_i^a$. The value added of household i in each sector, y_i^a and y_i^n, is produced combining hours of work with sector-specific abilities as given by

$$
\begin{aligned}
y_i^a &= \kappa z_i^a f\left(l_i^a\right) \\
y_i^n &= z_i^n g\left(l_i^n\right) = z_i^n g\left(1 - l_i^a\right)
\end{aligned}
\tag{1}
$$

where $f(.)$ and $g(.)$ are increasing and strictly concave functions with bounded derivatives at the origin, and κ captures economy-wide sectoral productivity differences and, in particular, the relative price of the agricultural good. It follows that agricultural value added y_i^a is expressed in units of non-agricultural value added y_i^n, which is the numéraire.

Households take the relative price as given and allocate labor to maximize income

$$y_i = \kappa z_i^a f\left(l_i^a\right) + z_i^n g\left(1 - l_i^a\right) \tag{2}$$

In order to fix ideas, we start by focusing on the standard case considered in the selection literature (Roy 1951) in which households operate only in one of the two sectors, i.e. $l_i^j = \{0, 1\}$. The i-th household compares the payoffs of operating in each sector and decides accordingly. This household engages only in farming if and only if

$$\kappa z_i^a f(1) \geqslant z_i^n g(1) \tag{3}$$

As a result, sectoral choices are fully determined by comparative advantage: households with a strong agricultural comparative advantage, $z_i^a/z_i^n \geqslant g(1)/\kappa f(1)$, will engage only in farming, while those with a strong non-agricultural comparative advantage, $z_i^n/z_i^a > \kappa f(1)/g(1)$ will operate in the non-agricultural sector only. Combining equation 3 with 1, and the joint density function $g(z^a, z^n)$ we derive mean sectoral productivities as

$$\begin{aligned}\mathbb{E}\left(y_i^a | z_i^a/z_i^n \geq g(1)/\kappa f(1)\right) &= \frac{\kappa f(1) \int_{z_i^a/z_i^n \geqslant g(1)/\kappa f(1)} z_i^a dGi}{\int_{z_i^a/z_i^n \geqslant g(1)/\kappa f(1)} dGi} \\ \mathbb{E}\left(y_i^n | z_i^a/z_i^n < g(1)/\kappa f(1)\right) &= \frac{g(1) \int_{z_i^a/z_i^n < g(1)/\kappa f(1)} z_i^n dGi}{\int_{z_i^a/z_i^n < g(1)/\kappa f(1)} dGi}.\end{aligned} \tag{4}$$

These expressions increase in absolute advantage.

Although comparative advantage determines sectoral allocations, absolute advantage determines sectoral productivities. It follows that the relation between sectoral employment shares and productivities is determined by the correlation between comparative and absolute advantage in each sector. To understand this, consider first a situation where comparative and absolute advantage are positively correlated – aligned – in both sectors. As the threshold of comparative advantage required to operate in a sector increases, so does the comparative advantage of those that remain in the sector, and given the positive correlation, so does the average absolute advantage in the sector. It follows that average labor productivity increases as a sector shrinks. The converse is true in expanding sectors: incoming workers have not only lower comparative advantage, but also, on average, lower absolute advantage than those already in the sector. As a result, average productivity declines in expanding sectors. This is the intuition developed by Lagakos and Waugh (2013) to rationalize the larger agricultural productivity gap in poor countries and by Young (2014) to understand the lower measured growth in labor productivity in the expanding service sector.

Having illustrated the distinct roles of comparative and absolute advantage in a simple case with full specialization, we return to the general case where households can operate in both sectors simultaneously. Let's begin with the households who operate in both sectors. They need to allocate their labor supply between the two sectors. Optimally, they equate the marginal value products of labor across the two activities. As a result, their optimal labor allocation $\tilde{l}_i^a$ is implicitly defined by

$$\frac{z_i^a}{z_i^n} = \frac{1}{\kappa} \frac{g'\left(1 - \tilde{l}_i^a\right)}{f'\left(\tilde{l}_i^a\right)}. \tag{5}$$

The fraction of time devoted to farming is an increasing function of their agricultural comparative advantage, i.e.

$$\frac{\partial \tilde{l}_i^a}{\partial z_i^a / z_i^n} = \frac{\partial \tilde{l}_i^a}{\partial \kappa} \frac{z_i^a}{z_i^n \kappa} = -\frac{\kappa f'\left(l_i^a\right)}{\frac{z_i^a}{z_i^n} \kappa f''\left(l_i^a\right) + g''\left(1 - l_i^a\right)} > 0. \tag{6}$$

Once we determined the optimal time allocation for those operating both sectors, we can use condition equation 5 to evaluate sectoral choices. Households with a strong agricultural comparative advantage will engage in farming only. These are households for which

$$\frac{z_i^a}{z_i^n} \geqslant \frac{1}{\kappa} \frac{g'(0)}{f'(1)} \tag{7}$$

Households with a high non-agricultural comparative advantage will fully specialize in non-agriculture. For these households we have

$$\frac{z_i^n}{z_i^a} \geqslant \frac{\kappa f'(0)}{g'(1)} \tag{8}$$

Finally, households with intermediate levels of comparative advantage will operate in both sectors. These households have

$$\frac{z_i^a}{z_i^n} \in \left[\frac{1}{\kappa} \frac{g'(1)}{f'(0)}, \frac{1}{\kappa} \frac{g'(0)}{f'(1)}\right] \tag{9}$$

The equations above show that when a household is endowed with a pair of relatively similar abilities and as a result its comparative advantage is neither high nor low in either sector, diminishing returns to labor at the sectoral level make it optimal to split the time endowment between the two activities. Nonetheless, it is worth noticing that these "mediocre" households are only so in terms of their comparative advantage. This mediocrity of comparative advantage is not informative about their absolute advantages, z_i^a and z_i^n. Households operating in both sectors could be fairly high in the marginal distributions of each ability, or could equally well be fairly low. In the same fashion, selection is not informative about the absolute advantage of those who fully specialize in either sector: selection is only informative about the fact that these households have fairly different abilities across sectors, but not about the level of these abilities.

This analysis makes clear that selection together with the correlations between absolute and comparative advantage at the sectoral level are key determinants of the link between sectoral productivity and sector size. However, selection on the basis of comparative advantage places no restrictions on the sectoral correlation of advantages, which are ultimately determined by the underlying distribution of abilities. The empirical analysis that follows aims to identify the sign of the correlation between comparative and absolute advantage.

3 Data

The data we use belong to the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) project. This is led by the World Bank in collaboration with several national statistical offices. In this section, we describe the main features of the dataset and the variables we use. We refer to Appendix B for detailed information on the project, sampling frame, survey design, and variable definitions.[3]

In each partner country, the LSMS-ISA project supports multiple rounds of a nationally representative panel survey with a multi-topic approach designed to gather information on agriculture, socioeconomic status, and non-farm income activities. Our final dataset combines the information on four countries for which we retrieved consistent information: Ethiopia, Malawi, Nigeria, and Uganda. The number of waves of data per country varies from 2 (Malawi) to 4 (Uganda), covering the years from 2009 to 2016. We combine the information available for each country and wave to derive the main variables of interest.

Value Added and Hours Worked First, we compute for each household in each wave the value added produced in agriculture. We follow Gollin, Lagakos, and Waugh (2014), and obtain it by adding value added from non-permanent crops, permanent crops, livestock, livestock products, and fishery. For each product category, we calculate value added as the sum across seasons of each household's revenue from selling each product plus the market value of the product that was not sold minus the associated production costs (Santaeulalia-Llopis and de Magalhaes 2014). To determine the market value of the product that was not sold, we either

[3] See also http://surveys.worldbank.org/lsms [consulted on October 9, 2018].

use the price at which the household sold that same crop or the one reported by households in the same region that sold that crop, or the price recorded in the community-level survey.

Second, we calculate value added from non-farming entrepreneurship. We identify all enterprises owned by any household member in the 12 months before the interview. For each one of them, we calculate value added as the difference between total annual sales and associated costs. We then aggregate these figures to derive total value added from non-farming entrepreneurship for each household in each wave.

To describe the activity of households along the intensive margin, we rely on the information provided on the number of hours allocated to each activity. In the survey, each household member is asked about the hours worked in the last 7 days. We can separately identify hours worked in the household farm and hours worked in any of the household non-farming enterprises. We derive the total number of hours worked in agriculture and entrepreneurship by aggregating the hours worked in each activity across all household members.

Measuring Absolute Advantage We measure absolute advantage using value added and value added per hour in each sector. The theoretical framework presented in Section 2 clarifies the mapping between these measures and the true value of absolute advantage. On the one hand, those households involved in both activities devote less hours to either of the two activities than those households who fully specialize. Since the production function is increasing in hours worked, value added will be a downward biased measure of absolute advantage for those households that are engaged in both activities relative to those that specialize in one of them. On the other hand, the production function is strictly concave in hours worked. It follows that value added per hour will be an upward biased measure of absolute advantage for those households that are engaged in both activities.[4]

Measuring Comparative Advantage Section 2 shows that the activity of each household along the extensive margin is informative of its comparative advantage. We use the information on value added described above to also define the activity of each household along the extensive margin. That is, we say that a household is engaged in any farming if we can derive information on value added in agriculture. Similarly, we say that a household is engaged in any entrepreneurial activity if we can derive information on profits from entrepreneurship. Households that only do farming have high comparative advantage in agriculture. Households that

[4]It is worth further illustrating these biases focusing on one sector, for instance farming. On the one hand, agricultural value added for a household engaged in both activities would always be less than the counterfactual if it was fully devoted to farming, i.e. $z_i^a f(l_i^a) < z_i^a f(1)$. On the other hand, agricultural value added per hour for a household engaged in both activities would always be more than the counterfactual if it was fully devoted to farming, since $z_i^a f'(l_i^a) > z_i^a f'(1)$. As a result, agricultural value added places a lower bound on the agricultural absolute advantage of those doing both activities relative to those that fully specialize in farming, while value added per hour places an upper bound.

only engage in entrepreneurship have high comparative advantage in this sector. Households that are active in both sectors have low comparative advantage in both sectors.

For those households that are active in both farming and non-farming entrepreneurship, we can derive an additional measure of comparative advantage that is informed by their activity along the intensive margin. Given that the production function is strictly concave in both sectors, equation 5 shows that households that have a comparative advantage in one sector also work relatively more hours in that sector. For households engaged in both sectors, we can thus use the ratio between total hours worked in the two sectors as a continuous measure of comparative advantage.

Additional Variables The data provide detailed information on each land plot operated by the household. We can therefore define the total area of land that is cultivated by the household. We complement this with information on ownership. The survey asks whether each plot of land is owned or assigned by decision of the local leader, inherited, or rented. We calculate the fraction of land that is rented, which we also consider as a proxy for local development of land markets. The survey also asks a number of questions about asset ownership. Household members are given a list of durable goods, and asked whether they possess any. This module is not always consistent across countries. We thus combine the available information in an asset index that is specific for each country. Finally, we derive information on the total number of household members and the total number of female household members, which we use as controls to evaluate the robustness of the empirical results.

4 Descriptives

Table 1 shows summary statistics for the variables of interests in our dataset.[5] For each variable, the Table reports the estimated average of each variable, together with the corresponding standard error and the number of observations used. It does so separately across three groups of households: those doing only farming, those engaged in non-farming entrepreneurship only, and those doing both activities.

The final dataset counts around 35,000 household observations across all countries and waves. Overall, 59% of households do only farming and 12% do only non-farming entrepreneurship. 30% of households in our sample are active in both sectors. This number is large in all countries, ranging from 24% in Ethiopia to 38% in Nigeria.

These households differ from the others along a number of characteristics. First, they are significantly larger, counting around half a member more than households doing only farming

[5] Table A.1 in Appendix A.1 shows the summary statistics of main variables by country.

and one more than households doing only entrepreneurship. Second, the total number of hours worked by all members is much larger in these households than those in the other two groups. The number of hours worked in a week in total is 90 for households that are active in both sectors as compared to 75 for those doing only entrepreneurship and 66 for those doing only farming. Yet, the number of hours allocated to each activity by households engaged in both sectors is significantly lower than the one allocated by households doing only one of each activity. Table 1 also shows that households for which we cannot derive value added in entrepreneurship – which we label as doing only farming – still report 4 hours a week in total of work in that sector. The opposite is also true, as households for which we cannot derive value added in agriculture – which we label as doing only entrepreneurship – report positive hours worked in agriculture. Our results are not sensitive to the choice of using value added or hours worked in the definition of households' activity along the extensive margin.

Among those households that are active in both sectors, the fraction of those where at least one member reports to work a positive number of hours in both sectors is 50%. This indicates that there is not full specialization across household members. Indeed, the average number of members reporting positive hours in both sector is one.

Table 1 also shows that the size of land that is cultivated is significantly higher for households active in both sectors than for other households, while the index value suggests that households doing only non-farming entrepreneurship are those with more assets.

Evidence so far shows that around one third of households engages in both farming and non-farming entrepreneurship in the countries we are considering. It also shows that significant differences exist between these households and those doing only farming or entrepreneurship. In the analysis that follows, we report unconditional estimates, but also evaluate the robustness of results by including additional household characteristics as controls whenever applicable.

5 Selection Along the Extensive Margin

Our first objective is to identify the sign of the correlation between absolute and comparative advantage in each sector. As discussed in Section 3, we consider value added and value added per hour in each sector as a measure of absolute advantage, while using household's activity along the extensive margin as a measure of comparative advantage. Households active in both sectors have low comparative advantage. If absolute and comparative advantages are positively correlated, value added or value added per hour in one sector should be negatively correlated with the likelihood of engaging in the other sector. The opposite holds if absolute and comparative advantages are negatively correlated.

5.1 Agricultural Sector

We first analyze households that are active in the agricultural sector. For this, we restrict the sample to households that do any farming, thus excluding those that only do non-farming entrepreneurship. We define a dummy equal to 1 if the household engages in non-farming entrepreneurship. We also derive the percentile $P(VA_a)$ the household belongs to in the distribution of value added in agriculture as derived in each country and wave. The top left graph in Figure 1 illustrates the unconditional relationship between the two variables. It reports the fraction of households involved in non-farming entrepreneurship per bin of 5 percentiles of the distribution of value added in agriculture. The relationship is negative. The likelihood of engaging in non-farming entrepreneurship is around 28% for households in the top percentile of the agricultural value added distribution as compared with the 38% of those in the bottom percentile. The bottom left graph in Figure 1 is drawn in a similar fashion. It plots the relationship between entrepreneurship rates and value added per hour in agriculture. We observe no correlation between the two.

At the country level, households that also engage in entrepreneurship have a lower comparative advantage in agriculture. As discussed in Section 3, value added is a downward biased measure of absolute advantage for those households that are engaged in both activities, while value added per hour is an upward biased measure of absolute advantage for those same households. It follows that the two lines in the top and bottom left graphs in Figure 1 serve as bounds for the true relationship between comparative and absolute advantage in agriculture. The graphs suggest that the correlation between comparative and absolute advantages is positive in the agricultural sector for each country as a whole.

We produced these figures by comparing farmers surveyed in the same country and wave, but across locations (enumeration areas). This may be problematic since average differences exist in the returns from both activities across locations. The top and bottom right graphs in Figure 1 address this issue by considering the *residual* likelihood to engage in entrepreneurship after netting out average differences across locations. In contrast with the right graphs, the relationship between agricultural value added and entrepreneurship becomes positive. These new estimates suggest that the correlation between comparative and absolute advantages is negative in the agricultural sector. These results suggest that average differences across locations are driving the results discussed in the previous paragraph as entrepreneurship rates being higher in those locations where agricultural value added is on average lower.

We investigate these patterns more systematically in a regression framework. In column 1 of Table 2, we regress a dummy equal to 1 if the household engages in non-farming entrepreneurship on the percentile the household belongs to in the distribution of value added in agriculture. In column 2, we instead use the percentile in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the location level. The point estimates are consistent with the top and bottom left graphs in Figure 1. Households in the top percentile of the agricultural value added distribution are 9 percentage points less likely to engage in non-farming entrepreneurship than households in the bottom percentile, while no significant differences emerge across households across the distribution of agricultural value added per hour. In column 3 and 4, we add a number of household level characteristics.[6] Columns 5 to 8 report the results from the same regression specification, but conditioning on the the full set of location fixed effects. Again, the point estimates are consistent with the top and bottom right graphs in Figure 1. Households in the top percentile of the distribution of agricultural value added per hour are 7 percentage points more likely to engage in non-farming entrepreneurship than households in the bottom percentile.

5.2 Entrepreneurship Sector

Figure 2 and Table 3 are informative of the correlation between comparative and absolute advantage in the sector of non-farming entrepreneurship. The top and bottom left graphs of Figure 2 show that the likelihood of doing farming is significantly lower for those households at the top of the distribution of profits from entrepreneurship, suggesting that comparative and absolute advantage are positively correlated in this sector. The top and bottom right graphs show that the relationship disappears when comparing individuals within locations. These results are consistent with the regression coefficient estimates in Table 3.

The results so far suggest that absolute and comparative advantage are negatively correlated in agriculture, and not correlated in non-farming entrepreneurship. In doing so, we show that average differences across locations may confound these correlations when estimated by comparing households across locations in a country. Figures A.1 to A.4 in Appendix A.1 show that this pattern holds consistently in three out of the four countries in our sample, with Ethiopia being the exception. In what follows, we discuss two mechanisms that can generate this main result. Yet, no matter what causes it, the correlation shown here is the one that matters for the relationship between size and productivity, of which the agricultural productivity gap is an example. Our findings cast doubts on selection playing a role on the low average agricultural productivity in developing countries: if anything, the misalignment of advantages in agriculture together with selection should imply that average agricultural productivity in poor countries is relatively *high*, i.e. it is low despite of selection that is biasing it upwards.

[6] Tables A.2 and A.3 in Appendix A.1 report all estimated coefficients for all these variables as added one by one.

6 Mechanisms of Selection Along the Extensive Margin

6.1 Underlying Distribution of Abilities

What determines the correlation between absolute and comparative advantage that we observe in the data? Clearly, the underlying distribution of abilities in the population plays an important role, which we explore first. We turn to the role of entry or operating costs in Section 6.2.

Denoting the correlation of advantages in agriculture and non-agriculture by $\rho\left(z_i^a/z_i^n, z_i^a\right)$ and $\rho\left(z_i^n/z_i^a, z_i^n\right)$ respectively, the following proposition summarizes our results.

Proposition 1. *The signs of the correlations between comparative and absolute advantage are (approximately) given by*

$$\begin{aligned} sign\left[\rho\left(\frac{z_i^a}{z_i^n}, z_i^a\right)\right] &= sign\left[\frac{CV\left(z_i^a\right)}{CV\left(z_i^n\right)} - \rho\left(z_i^a, z_i^n\right)\right] \\ sign\left[\rho\left(\frac{z_i^n}{z_i^a}, z_i^n\right)\right] &= sign\left[\frac{CV\left(z_i^n\right)}{CV\left(z_i^a\right)} - \rho\left(z_i^a, z_i^n\right)\right] \end{aligned} \tag{10}$$

where $CV\left(z_i^j\right) = \sigma_j/\mu_j$ is the coefficient of variation in the population for sector $j = \{a, n\}$ and $\rho\left(z_i^a, z_i^n\right)$ is the correlation coefficient of abilities in the population. See Appendix A.2 for a proof.

Abstracting from trivial cases, where the distributions of sectoral abilities coincide or it is degenerate in at least one sector, several insights arise from Proposition 1.[7] First, given that $\rho\left(z_i^a, z_i^n\right) \leq 1$, the correlation of advantages is always positive in one sector – the sector with higher dispersion of abilities as measured by the coefficient of variation. For the sake of the exposition let's assume $CV\left(z_i^n\right) > CV\left(z_i^a\right)$, so that advantages are always aligned in non-agriculture. Second, when abilities are not positively correlated, $\rho\left(z_i^a, z_i^n\right) \leq 0$, advantages are aligned in both sectors. Third, under perfect positive correlation of abilities, $\rho\left(z_i^a, z_i^n\right) = 1$, advantages are misaligned in the sector with less dispersion in abilities (here, agriculture), i.e. $\rho\left(z_i^a/z_i^n, z_i^a\right) < 0$. Fourth, under imperfect positive correlation of abilities, advantages in agriculture will be aligned as long as $\rho\left(z_i^a, z_i^n\right) < CV\left(z_i^a\right)/CV\left(z_i^n\right)$ and misaligned otherwise. The first equation in 10 determines a threshold for the correlation of abilities below which advantages in agriculture will be aligned, i.e. an upper bound for the correlation of abilities that ensures that they are, in words of Young, "at worst weakly correlated" (Young 2014).

[7] When the coefficients of variation of abilities in both sectors coincide, $CV\left(z_i^a\right) = CV\left(z_i^n\right)$, the correlation of advantages will be positive in both sectors if abilities are not perfectly positively correlated. If $\rho\left(a, n\right) = 1$, advantages are uncorrelated in both sectors, $\rho\left(\frac{n}{a}, n\right) = \rho\left(\frac{a}{n}, a\right) = 0$. When the distribution of abilities in one sector is degenerate, for instance $CV\left(z_i^n\right) = 0$, abilities are uncorrelated in this sector, i.e. $\rho\left(z_i^n/z_i^a, z_i^n\right) = 0$. It is clear that empirically, these cases are not relevant.

This threshold, well known in the literature on selection of migrants (see for instance Borjas 1987), seems to have been overlooked in the recent work on selection and sectoral productivity differences. The more different are the sectors in terms of the dispersion of abilities in the population – as reflected by lower $CV\left(z_i^a\right)/CV\left(z_i^n\right)$ – the lower is the correlation of abilities in the population that ensures that advantages remain aligned in agriculture.

In view of this proposition, our empirical results – advantages misaligned in agriculture and uncorrelated in non-agriculture – suggest the presense of additional factors, besides the underlying distribution of abilities, that shape the correlation of advantages in the data. Next, we explore whether the introduction of fixed costs of entry may interact with selection in generating the observed correlations of advantages. As we will see, the introduction of fixed costs not only allows us to rationalize our empirical results, but also suggests an empirical approach for inferring crucial features of the distribution of abilities from our data.

6.2 Sector-specific Fixed Costs

A prominent explanation for the sectoral differences in labor productivity relies on the presence of some sort of friction or fixed cost that prevents the implementation of the efficient allocation of resources. Along these lines we now extend extend the model discussed in Section 2 to allow for the presence of fixed costs of entering or operating in any of the two sectors, τ^j. These costs, if negative, should be interpreted as amenities. As before, households take prices as given and allocate labor to maximize income net of operating costs[8]

$$y_i = \kappa z_i^a f\left(l_i^a\right) - \tau^a + z_i^n g\left(1 - l_i^a\right) - \tau^n. \tag{11}$$

The *i-th* household compares the payoffs of being only a farmer, only a non-farmer, or operating simultaneously in both sectors and decides accordingly. This household will operate in both sectors as long as

$$\kappa z_i^a f\left(l_i^a\right) - \tau^a + z_i^n g\left(1 - l_i^a\right) - \tau^n \geqslant \max\left[\kappa z_i^a f\left(1\right) - \tau^a, z_i^n g\left(1\right) - \tau^n\right], \tag{12}$$

which in terms of comparative and absolute advantages becomes

$$\kappa \frac{z_i^a}{z_i^n} f\left(l_i^a\right) - \frac{\tau^a}{z_i^n} + g\left(1 - l_i^a\right) - \frac{\tau^n}{z_i^n} \geqslant \max\left[\kappa \frac{z_i^a}{z_i^n} f\left(1\right) - \frac{\tau^a}{z_i^n}, g\left(1\right) - \frac{\tau^n}{z_i^n}\right], \tag{13}$$

[8]While we model the costs as fixed operating costs, fixed costs of entry would have a similar effect.

or equivalently

$$\kappa f\left(l_i^a\right)-\frac{\tau^a}{z_i^a}+\frac{z_i^n}{z_i^a}g\left(1-l_i^a\right)-\frac{\tau^n}{z_i^a}\geqslant\max\left[\kappa f\left(1\right)-\frac{\tau^a}{z_i^a},\frac{z_i^n}{z_i^a}g\left(1\right)-\frac{\tau^n}{z_i^a}\right],\qquad(14)$$

where the first expression provides information about the sign of the correlation of advantages in non-agriculture $\rho\left(z_i^n/z_i^a, z_i^n\right)$, while the second is informative about the same correlation in agriculture, $\rho\left(z_i^a/z_i^n, z_i^a\right)$.[9]

As is clear from the last two expressions, the correlation of advantages is no longer determined only by the the underlying distribution of abilities in the population, since selection combined with fixed costs systematically affects these correlations.

Restricting our analysis to those cases relevant for our empirical exercise, i.e. comparisons between households who are active only in one sector and those who engage in both activities, the impact of these costs on the correlation of advantages depends on their sign – costs or amenities – and on their presence in one or both sectors.

Consider a situation where there is a fixed entry cost in only one sector, agriculture, so that $\tau^a > 0$ and $\tau^n = 0$. Since everyone doing some farming needs to pay the fixed cost, this cost does not affect the choice between fully specializing in farming or doing both activities. It still affects the choice between specializing in non-farming or doing both activities. Those specializing in non-farming will not pay the cost, while those doing both activities will have to pay it. As a result, compared with those that fully specialize in non-agriculture, households engaged in both activities not only have a low comparative advantage in non-agriculture, low z_i^n/z_i^a, but also a high absolute advantage in this sector, high z_i^n, required to pay for the fixed cost.[10] As a result, the presence of a fixed cost to enter agriculture reduces the correlation of advantages in non-agriculture, $\rho\left(z_i^n/z_i^a, z_i^n\right)$ as measured when comparing those that do both activities with those who fully specialize in non-agriculture. The opposite, an increase in the correlation of advantages in non-agriculture when comparing the same two groups, would arise if one considers the presence of amenities only in agriculture, so that $\tau^a < 0$ and $\tau^n = 0$. Overall, when comparing households who fully specialize in a given sector and those engaged in both activities, the introduction of a fixed cost (amenity) in that sector will tend to lower (increase) the observed correlation between comparative and absolute advantage in the other sector.

[9] In Appendix A.3 we show that $sign\left[\rho\left(z_i^a/z_i^n, z_i^n\right)\right]=-sign\left[\rho\left(z_i^n/z_i^a, z_i^n\right)\right]$.

[10] A higher agricultural absolute adantage z_i^a will also allow to pay the fixed cost. This becomes clear from condition (14) that suggests that $\rho\left(z_i^a/z_i^n, z_i^a\right)$ increases when comparing those who do both with those engaged only in non-agriculture. Nonetheless, since z_i^a is never observed for those engaged only in non-agriculture we never estimate $\rho\left(z_i^a/z_i^n, z_i^a\right)$ for this set of households.

6.3 Discussion of the Empirical Results

The results in Section 5 suggest that absolute and comparative advantage are negatively correlated in agriculture, and not correlated in non-farming entrepreneurship. This is consistent with three scenarios:

1. The coefficient of variation is higher in entrepreneurship than in agriculture, $CV\left(z_i^n\right) > CV\left(z_i^a\right)$, there are positive fixed costs in agriculture, $\tau^a > 0$, and the correlation between absolute advantages is (relatively) high, i.e.

$$\rho\left(z_i^a, z_i^n\right) > \frac{CV\left(z_i^a\right)}{CV\left(z_i^n\right)}$$

   In the absence of fixed costs, the observed correlation between absolute and comparative advantage would be negative in agriculture and positive in entrepreneurship. Fixed costs to enter agriculture push the observed correlation between absolute and comparative advantage in entrepreneurship towards zero.

2. The coefficient of variation is higher in agriculture than in entrepreneurship, $CV\left(z_i^n\right) < CV\left(z_i^a\right)$, there are positive fixed costs in entrepreneurship, $\tau^n > 0$, negative fixed costs (amenities) exist to enter agriculture, $\tau^a < 0$, and the correlation between absolute advantages is (relatively) high, i.e.

$$\rho\left(z_i^a, z_i^n\right) > \frac{CV\left(z_i^n\right)}{CV\left(z_i^a\right)}$$

   In the absence of fixed costs, the observed correlation between absolute and comparative advantage would be positive in agriculture and negative in entrepreneurship. Fixed costs to enter entrepreneurship push the observed correlation between absolute and comparative advantage in agriculture towards negative, while the amenities associated with agriculture push the observed correlation between absolute and comparative advantage in entrepreneurship towards zero.

3. There are positive fixed costs in both agriculture and entrepreneurship, $\tau^a, \tau^n > 0$, and the correlation between absolute advantages is (relatively) low, i.e.

$$\rho\left(z_i^a, z_i^n\right) < \min\left\{\frac{CV\left(z_i^n\right)}{CV\left(z_i^a\right)}, \frac{CV\left(z_i^a\right)}{CV\left(z_i^n\right)}\right\}$$

   In the absence of fixed costs, the observed correlation between absolute and comparative advantage would be positive in both sectors. Fixed costs to enter entrepreneurship push the observed correlation between absolute and comparative advantage in agriculture towards negative, while fixed costs to enter agriculture push the observed correlation between absolute and comparative advantage in entrepreneurship towards zero.

Each of these cases is consistent with our empirical findings. Since all of them feature fixed costs, it is clear that fixed costs in one or both sectors are required to rationalize our empirical findings. This analysis also suggests an alternative empirical approach to inferring some of the properties of the underlying distribution of abilities. We proceed with this analysis in what follows.

7 Selection Along the Intensive Margin

In light of the discussion above, we proceed with the empirical analysis by restricting the sample to those households that are engaged in both agriculture and non-agriculture entrepreneurship. Fixed entry costs (or amenities) do not affect the marginal labor supply choice of these households, which is instead determined only by their comparative advantage. As explained in Section 3, equation 5 the ratio between total hours worked in the two sectors is a continuous measure of comparative advantage. For these households, any systematic relationship between value added and relative labor supply in each sector is informative of the correlation between absolute and comparative advantage net of fixed entry costs. From that, we can also infer some features of the underlying distribution of abilities in the population.

As in Section 5, we start by investigating sectoral correlations in agriculture. We begin by regressing the relative labor supply in agriculture – the ratio of total hours worked in agriculture vs. non-farming entrepreneurship – on the percentile the household belongs to in the distribution of agricultural value added in each country and wave. Column 1 of Table 4 reports the corresponding coefficient estimate. We condition on the full set of location fixed effects, and cluster standard errors at the same level. The estimated relationship is positive, but only at the 10% level. Households in the top percentile of the agricultural value added distribution on average work slightly more hours in agriculture relative to entrepreneurship than households in the bottom percentile. In column 2, we instead use the household's percentile in the distribution of agricultural value added per hour as the main regressor. The coefficient of interest is negative and highly significant, indicating that, among households engaged in both activities, households with higher agricultural value added per hour work significantly fewer hours in this sector relative to entrepreneurship. Table 1 shows that the average household active in both sectors allocates 40.7% of total hours worked to agriculture (36.5 hours compared to 53.1 hours in entrepreneurship). Taking this as benchmark, the estimate in column 2 of Table 4 implies that moving up one decile in the distribution of agricultural value added per hour is associated with a reduction in the share of time allocated to agriculture of about 5 percentage points, a reduction of 4.5 hours.

As discussed earlier, decreasing returns to scale imply that value added per hour is an upward biased measure of absolute advantage for households that are engaged in both activities, and more so the lower the absolute amount of hours worked. This implies that the estimate in column 2 is a lower bound for the true correlation between absolute and comparative advantage in agriculture, while the estimate in column 1 is an upper bound. The ordering of the two estimated coefficients is consistent with this reasoning.

In column 3 and 4, we include the full set of household-level controls.[11] Following the same bounding argument, evidence shows that the estimated upper bound for the correlation between absolute and comparative advantage in agriculture is negative and non-significant while the estimated lower bound is negative and highly significant. Households that are more productive in agriculture supply relatively less hours in that sector, a sign of low comparative advantage. We conclude that absolute and comparative advantages are negatively correlated in agriculture. Given our focus on households engaged in both activities, this result rules out the possibility that the patterns we found in Section 5 for the agricultural sector are fully explained by the presence of fixed costs to enter entrepreneurship. That is, we can rule out cases 2 and 3 among those discussed in Section 6.3.

All evidence thus points to case 1 in Section 6.3 as being the relevant one. In the absence of fixed costs, the observed correlation between absolute and comparative advantage should be negative in agriculture, but positive in entrepreneurship. Table 4 provides evidence of the former. Table 5 provides consistent evidence of the latter.[12] We test whether a systematic relationship exists between profits from non-farming entrepreneurship and relative labor supply in that sector. In column 1 of Table 5, we regress the relative labor supply in entrepreneurship – the ratio of total hours worked in non-farming entrepreneurship vs. agriculture – on the percentile the household belongs to in the distribution of profits from entrepreneurship in each country and wave. The regression models and estimates in columns 2 to 4 are ordered as in Table 4, replacing the percentile the household belongs to in the distribution of profits from entrepreneurship per hour as main regressor in column 2, and then adding controls in columns 3 and 4.[13] Once again, the ordering of estimated coefficients is consistent with the bounding argument outlined above, with the estimate in columns 2 (and 4) being a lower bound for the true correlation between absolute and comparative advantage in entrepreneurship, and the estimate

[11] Table A.6 in Appendix A.1 reports all estimated coefficients for all these control variables as added one by one. Table A.8 also reports the same results but restricting the sample to those households that report positive hours worked in both sectors, thus excluding those with zero reported hours worked in agriculture.

[12] Note that this is not directly implied by the findings in Table 4.

[13] As before, Table A.7 in Appendix A.1 reports all estimated coefficients for all these control variables as added one by one. Table A.9 also reports the same results but restricting the sample to those households that report positive hours worked in both sectors, thus excluding those with zero reported hours worked in entrepreneurship.

in column 1 (and 3) being an upper bound. Households with higher profits from non-farming entrepreneurship work significantly more hours in this sector relative to agriculture, while no systematic differences emerge in relative labor supply across percentiles of the distribution of hourly profits from entrepreneurship. Taking again the time allocation of the average household active in both sectors as benchmark, the estimate in column 1 of Table 5 implies that moving by one decile in the distribution of profits from entrepreneurship is associated with an increase in the share of time allocated to entrepreneurship of about 2.3 percentage point, an increase of around 2 hours. We conclude that absolute and comparative advantages are positively correlated in entrepreneurship, consistent with case 1 in Section 6.3.

Following the discussion in Section 6, we can exploit these results to place some restrictions on the main features of the underlying distribution of abilities. We infer that the coefficient of variation is higher for the distribution of ability in entrepreneurship relative to agriculture, $CV\left(z_i^n\right) > CV\left(z_i^a\right)$, and the correlation between absolute advantages is (relatively) high, i.e. $\rho\left(z_i^a, z_i^n\right) > CV\left(z_i^a\right)/CV\left(z_i^n\right)$. This implies a positive correlation of comparative and absolute advantages in entrepreneurship, and a negative correlation in agriculture. The observation that the propensity to take up agricultural work is unrelated to entrepreneurial productivity in the data can then be explained by a cost of entering agriculture.

8 Selection Over Time

Our findings indicate that comparative and absolute advantage are negatively correlated in the agricultural sector, and positively correlated in entrepreneurship. This has implications for the validity of the argument that identifies selection as responsible for the observed agricultural productivity gap between developing and developed countries. As the agricultural sector shrinks, individuals and households with weaker comparative comparative advantage leave the agricultural sector. The negative correlation between comparative and absolute advantage in this sector implies that these households have higher absolute advantage compared to those who stay in agriculture: agricultural productivity should therefore decrease as its employment share decreases.

The evidence presented so far did not consider changes over time. In this section, we exploit the panel dimension of the data and investigate the behavior of households engaged in different activities at different points in time – switchers.

We begin by reporting in Table 6 the fraction of households in each wave that is engaged in agriculture, in non-farming entrepreneurship, or in both. This reveals that the fraction of households engaged in both activities has been growing between 2009 and 2016, from 26% to

37%. This is true in all countries in our sample with the exception of Uganda.[14] The fraction of households doing only farming decreased in Malawi and Nigeria, but remained stable in Ethiopia and Uganda. Table A.11 in Appendix A.1 reports the transition matrices across the different groups between waves 1 and 2, and 2 to 3. This shows that the fraction of households transitioning from doing only one activity to doing only the other is negligible, while transitions from doing only farming (entrepreneurship) to doing both and vice versa are more common, covering around 10% (2%) of households in the sample.

In light of these non-trivial transition probabilities, we can complement the cross-sectional analysis above by a systematic analysis of transitions. Do the farming households that start to engage in non-farming entrepreneurship have high or low absolute advantage in agriculture? We implement a panel data regression analysis. We restrict our sample to all households that in wave 1 are doing only agriculture, and investigate their probability of engaging in non-farming entrepreneurship through wave 3. We implement the following regression specification

$$entrep_{it} = \sum_{t=2}^{3} \beta_t \, wave_t \times rank_i + \theta_i + \delta_t + \varepsilon_{it} \qquad (15)$$

where $entrep_{it}$ is a dummy equal to one if household i engages in non-farming entrepreneurship in wave t. $wave_t$ is a wave dummy identifier. $rank_i$ is defined according to where the household stands in the ranking of agricultural value added and agricultural value added per hour in its location in the first wave of data. That is, $rank_i$ takes a value of 1 if household i is the most productive farming household in its location in the first wave of the data. θ_i and δ_t capture household and wave fixed effects respectively, which allow to control for both time-invariant household-level characteristics and overall wave-specific time trends. The coefficient β_t captures whether the likelihood to take up non-farming entrepreneurship in wave 2 or 3 is systematically correlated with households' absolute advantage in agriculture.

Table 7 reports the coefficient estimates from this specification. As in the previous exercises, we define the ranking position of the household in terms of either agricultural value added or agricultural value added per hour. The estimated β_t is negative and significant for all waves and across all specifications. Columns 1 and 2 are consistent with each other in showing that households having a lower rank, i.e. higher agricultural value added or value added per hour in wave 1 are differentially more likely to take up non-farming entrepreneurship in subsequent waves. The magnitude and significance of coefficient estimates is only marginally affected by the inclusion of time-varying household-level controls in columns 3 and 4.[15] This pattern is remarkably consistent across countries, as indicated by the coefficient estimates reported in Table A.13 in Appendix A.1.

[14] Table A.10 reports the same numbers separately for each country and wave.

[15] Table A.12 in Appendix A.1 reports the estimated coefficients for all included controls.

The evidence in this section is consistent with the one presented in Section 5 and 7. Farming households at the margin of entrepreneurship have a lower comparative advantage in agriculture than inframarginal ones. It is thus natural that their gains from switching sector are limited (Hicks, Kleemans, Li, and Miguel 2017). Yet, evidence shows that they are among the most productive farming households. Results from this panel data analysis provide further indication that absolute and comparative advantage are negatively correlated in agriculture.

9 Conclusions

Labor productivity is lower in poor countries than in rich countries, and relatively more so in the agricultural sector. A recent influential literature argues that an important source is worker selection, suggesting that comparative advantage and absolute advantage are positively correlated in the agricultural sector. We investigate whether this pattern holds using household-level data from Ethiopia, Malawi, Nigeria, and Uganda, and produce four sets of results. First, we find that around one third of households engage in both agriculture and non-farming entrepreneurship. Second, we show that those households active in both sectors have systematically higher agricultural productivity than those doing only farming. Third, we report that among the households active in both sectors, those with higher agricultural productivity supply relatively fewer hours in agriculture, while those with higher profits from entrepreneurship supply relatively more hours in this sector. Finally, we find that, over time, households starting a non-farming enterprise have higher baseline agricultural productivity than those who remain only farmers.

These results altogether suggest that comparative and absolute advantage are misaligned in agriculture, casting doubt on the importance of selection as a root cause of the agricultural productivity gap. These results can also provide new lenses through which to analyze and reconcile existing findings in the literature.

References

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia (2017). Misallocation, Selection and Productivity: A Quantitative Analysis with Panel Data from China. Working Papers tecipa-593, University of Toronto, Department of Economics.

Alvarez, J. A. (2018). The Agricultural Wage Gap: Evidence from Brazilian Micro-Data. Mimeo.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *The American Economic Review 77*(4), 531–553.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1 of *Handbook of Economic Growth*, Chapter 9, pp. 679–741. Elsevier.

Chanda, A. and C.-J. Dalgaard (2008). Dual economies and international total factor productivity differences: Channelling the impact from institutions, trade, and geography. *Economica 75*(300), 629–661.

Gollin, D., D. Lagakos, and M. E. Waugh (2014). The Agricultural Productivity Gap. *The Quarterly Journal of Economics 129*(2), 939–993.

Gollin, D., S. Parente, and R. Rogerson (2002). The role of agriculture in development. *American Economic Review 92*(2), 160–164.

Herrendorf, B. and T. Schoellman (2015). Why is measured productivity so low in agriculture? *Review of Economic Dynamics 18*(4), 1003 – 1022.

Hicks, J. H., M. Kleemans, N. Y. Li, and E. Miguel (2017). Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata. NBER Working Papers 23253, National Bureau of Economic Research, Inc.

Jones, C. I. and P. M. Romer (2010). The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital. *American Economic Journal: Macroeconomics 2*(1), 224–245.

Kuznets, S. S. (1971). *Economic Growth of Nations: Total Output and Production Structure*. Harvard University Press.

Lagakos, D. and M. E. Waugh (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review 103*(2), 948–80.

OECD (2017). *African Economic Outlook 2017*. African Development Bank and OECD and United Nations Development Programme.

Pulido, J. and T. Święcki (2018). Barriers to Mobility or Sorting? Sources and Aggregate Implications of Income Gaps Across Sectors and Locations in Indonesia. Mimeo.

Restuccia, D. and R. Santaeulalia-Llopis (2017). Land misallocation and productivity. Working Paper 23128, National Bureau of Economic Research.

Restuccia, D., D. T. Yang, and X. Zhu (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics 55*(2), 234 – 250.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Santaeulalia-Llopis, R. and L. de Magalhaes (2014). Consumption and Income Inequality in Sub-Saharan Africa: A Lifetime with No Humps and Low Partial Insurance. 2014 Meeting Papers 1287, Society for Economic Dynamics.

Vollrath, D. (2014). The efficiency of human capital allocations in developing countries. *Journal of Development Economics 108*, 106 – 118.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics 128*(4), 1727–1785.

Young, A. (2014). Structural transformation, the mismeasurement of productivity growth, and the cost disease of services. *American Economic Review 104*(11), 3635–67.

Tables and Figures

Table 1: Summary Statistics

	Only Agriculture	Only Entrep.	Both	Full Sample
Observations	20622	4101	10374	35097
	59%	12%	30%	100%
Household Size	5.066	4.625	5.726	5.210
	(0.018)	(0.041)	(0.027)	(0.014)
	20537	4093	10359	34989
Female HH Members	2.032	1.915	2.052	2.024
	(0.012)	(0.027)	(0.019)	(0.010)
	20537	4093	10359	34989
Hours in Agriculture	47.280	4.141	36.569	39.068
h_a	(0.385)	(0.269)	(0.460)	(0.276)
	19851	3940	10174	33965
Hours in Entrepreneurship	18.540	70.744	53.085	34.944
h_n	(0.270)	(0.856)	(0.510)	(0.264)
	19851	3940	10174	33965
Total Hours	65.661	75.004	90.126	73.984
$h_a + h_n$	(0.501)	(0.904)	(0.730)	(0.385)
	20622	4101	10374	35097
Hours in Agriculture	59.189	n.a.	52.261	48.403
$h_a > 0$	(0.434)	n.a.	(0.563)	(0.320)
	15857	n.a.	7119	27077
Hours in Entrepreneurship	n.a.	76.407	63.946	25.028
$h_n > 0$	n.a.	(0.858)	(0.543)	(0.249)
	n.a.	3648	8446	32716
HH Members with	n.a.	n.a.	0.938	0.277
$h_a, h_n > 0$	n.a.	n.a.	(0.014)	(0.005)
	n.a.	n.a.	10359	35082
Female HH Members with	n.a.	n.a.	0.211	0.048
$h_a, h_n > 0$	n.a.	n.a.	(0.006)	(0.001)
	n.a.	n.a.	7283	32006
Land Size (ha)	1.488	0.516	2.464	1.782
	(0.087)	(0.086)	(0.899)	(0.289)
	19298	410	9075	28783
Fraction Rented	0.068	0.115	0.070	0.070
	(0.002)	(0.016)	(0.002)	(0.001)
	19298	410	9075	28783
Asset Index	9.434	13.538	12.043	10.683
	(0.073)	(0.167)	(0.112)	(0.058)
	20529	4053	10354	34936

Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The table reports the estimated average of each variable across the different subsamples, together with the corresponding standard error and the number of observations used. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farming entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farming entrepreneurship, but not on value added in agriculture. Households doing both are those for which we can derive information on both value added in agriculture and non-farming entrepreneurial profits.

Table 2: Agricultural Value Added and Entrepreneurship

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	-0.009***		-0.006***		0.001		0.000	
	(0.001)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		0.001		0.003**		0.003***		0.007***
		(0.001)		(0.001)		(0.001)		(0.001)
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes	No	No	Yes	Yes
Village FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	30996	22977	27485.	21575	30930	22892	27418	21488
R^2	0.003	0.000	0.179	0.080	0.247	0.247	0.338	0.293

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farming entrepreneurship. $P(VA_a)$ is the percentile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $P(VA_a/h_a)$ is the percentile the household belongs to in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total number of hours in agriculture (columns 3 and 7 only), total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.

Table 3: Entrepreneurial Profits and Farming

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.017***		-0.010***		-0.001		-0.001	
	(0.002)		(0.002)		(0.002)		(0.001)	
$P(VA_n)/h_n$		-0.012***		-0.008***		0.001		0.002*
		(0.002)		(0.002)		(0.001)		(0.001)
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes	No	No	Yes	Yes
Village FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	14476	12094	14057	12040	14376	11962	13957	11908
R^2	0.012	0.005	0.270	0.155	0.515	0.539	0.572	0.570

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if we can derive information on value added in agriculture. $P(VA_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship per hour. Both values are rescaled and multiplied by 10. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total number of hours in non-farming entrepreneurship (columns 3 and 7 only), country-specific asset index. Standard errors are clustered at the level of enumeration area.

Table 4: Agricultural Value Added and Time Allocation

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.026*		-0.002	
	(0.015)		(0.018)	
$P(VA_a/h_a)$		-0.123***		-0.116***
		(0.022)		(0.024)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	8267	5701	7117	5236
R^2	0.336	0.354	0.348	0.362

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we derive information on both value added in agriculture and profits from non-farming entrepreneurship. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farming entrepreneurship. $P(VA_a)$ is the percentile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $P(VA_a/h_a)$ is the percentile the household belongs to in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.

Table 5: Entrepreneurial Profits and Time Allocation

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.131***		0.128***	
	(0.034)		(0.036)	
$P(VA_n/h_n)$		-0.039		-0.050
		(0.029)		(0.032)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	6913	5702	6416	5236
R^2	0.274	0.265	0.264	0.257

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we derive information on both value added in agriculture and profits from non-farming entrepreneurship. The dependent variable is the ratio of total hours worked by the household in non-farming entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship per hour. Both values are rescaled and multiplied by 10. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, country-specific asset index. Standard errors are clustered at the level of enumeration area.

Table 6: Activities Over Time

	Only Agriculture	Only Entrep.	Both	Full Sample
Wave 1	63.44%	10.88%	25.68%	100%
	7606	1304	3079	11989
Wave 2	61.37%	9.56%	29.07%	100%
	7228	1126	3424	11778
Wave 3	50.99%	15.35%	33.66%	100%
	4923	1482	3250	9655
Wave 4	51.64%	11.28%	37.07%	100%
	865	189	621	1675

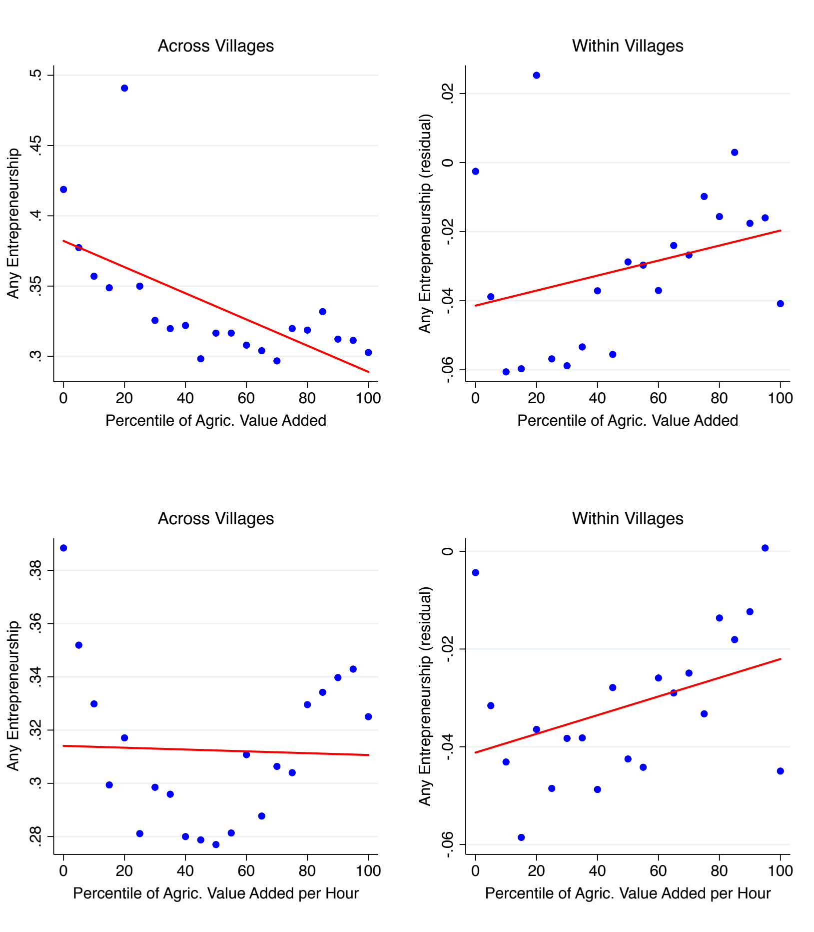
Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The table reports the relative and absolute number of households across the different subsamples over different waves. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farming entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farming entrepreneurship, but not on value added in agriculture. Households doing both are those for which we can derive information on both value added in agriculture and non-farming entrepreneurial profits.

Table 7: Transitions To Entrepreneurship

	Any Entrepreneurship			
	(1)	(2)	(3)	(4)
Wave 2 × $Rank(VA_a)$	-0.007***		-0.007***	
	(0.002)		(0.002)	
Wave 3 × $Rank(VA_a)$	-0.008***		-0.009***	
	(0.003)		(0.003)	
Wave 2 × $Rank(VA_a/h_a)$		-0.009***		-0.007***
		(0.002)		(0.002)
Wave 3 × $Rank(VA_a/h_a)$		-0.012***		-0.010***
		(0.003)		(0.003)
Household FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	n.a.	n.a.
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Observations	18721	14746	16509	13678
R^2	0.547	0.544	0.590	0.574

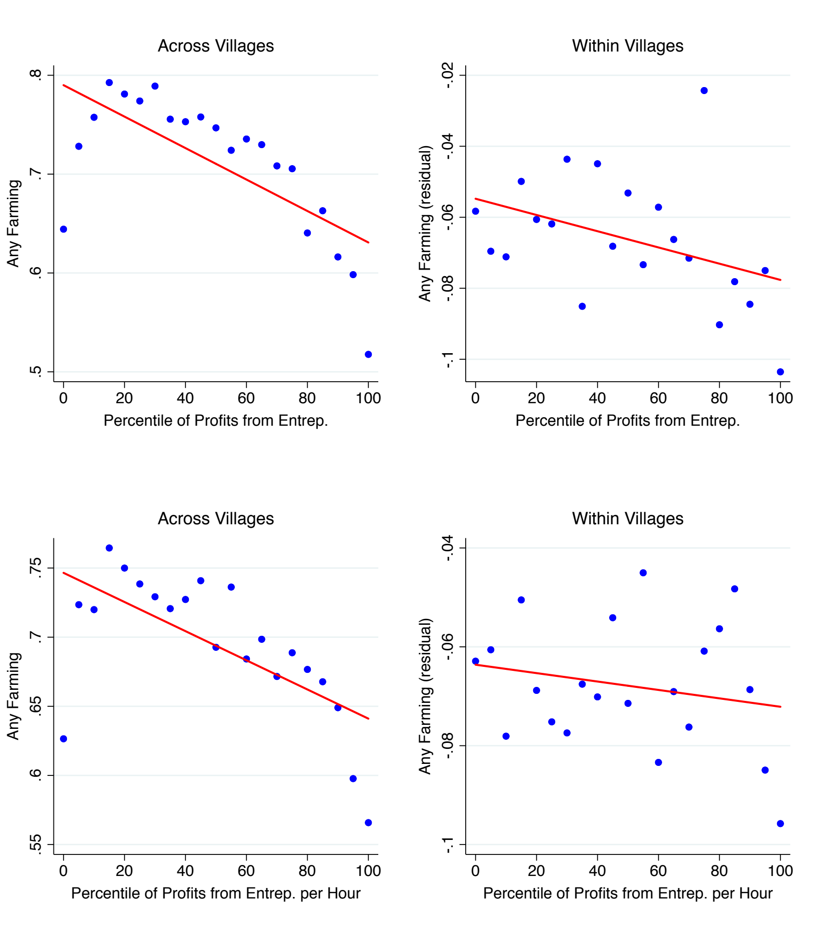
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we cannot derive any information on profits from entrepreneurship in Wave 1, and observed again over time through Wave 3. $Rank(\cdot)$ is the within-village ranking of agricultural value added or agricultural value added per hour in Wave 1 among these households. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total number of hours in agriculture (column 3), total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.

Figure 1: Agricultural Value Added and Entrepreneurship



Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The top figures shows the fraction of households involved in non-farming entrepreneurship per bin of 5 percentiles of the distribution of value added in agriculture. The bottom figures shows the same number per bin of 5 percentiles of the distribution of value added in agriculture per hour. The right figures plot the averaged residuals of the probability of doing entrepreneurship after netting out village fixed effects.

Figure 2: Profits from Entrepreneurship and Farming



Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The top figures shows the fraction of households involved in farming per bin of 5 percentiles of the distribution of profits from non-farming entrepreneurship. The bottom figures shows the same number per bin of 5 percentiles of the distribution of profits from non-farming entrepreneurship per hour. The right figures plot the averaged residuals of the probability of doing farming after netting out village fixed effects.

A Appendix

A.1 Additional Tables and Figures

Table A.1: Summary Statistics by Country

	Only Agriculture	Only Entrep.	Both	Full Sample
		Panel A. Ethiopia		
Observations	6946	738	2371	10055
	69%	7%	24%	100%
Household Size	5.168	3.954	5.450	5.145
	(0.027)	(0.084)	(0.046)	(0.023)
	6924	738	2371	10033
Hours in Agriculture	53.504	10.797	43.316	48.037
	(0.660)	(0.914)	(1.006)	(0.531)
	6697	688	2328	9713
Hours in Entrepreneurship	4.598	43.686	23.986	12.013
	(0.207)	(1.956)	(0.777)	(0.298)
	6697	688	2328	9713
		Panel B. Malawi		
Observations	3936	27	1420	5383
	73%	1%	26%	100%
Household Size	4.280	4.111	4.477	4.331
	(0.034)	(0.561)	(0.055)	(0.029)
	3934	27	1420	5381
Hours in Agriculture	21.515	2.333	12.337	18.997
	(0.516)	(1.539)	(0.637)	(0.417)
	3934	27	1420	5381
Hours in Entrepreneurship	1.725	49.741	30.761	9.628
	(0.151)	(7.436)	(0.916)	(0.322)
	3934	27	1420	5381
		Panel C. Nigeria		
Observations	5605	2474	3492	11571
	48%	21%	30%	100%
Household Size	5.535	4.894	6.536	5.700
	(0.042)	(0.054)	(0.054)	(0.029)
	5603	2474	3492	11569
Hours in Agriculture	63.571	2.446	45.121	44.794
	(0.925)	(0.269)	(0.954)	(0.579)
	5329	2389	3396	11114
Hours in Entrepreneurship	18.936	71.702	57.130	41.949
	(0.547)	(0.937)	(0.766)	(0.459)
	5329	2389	3396	11114
		Panel D. Uganda		
Observations	4135	862	3091	8088
	51%	11%	38%	100%
Household Size	5.007	4.444	5.596	5.173
	(0.040)	(0.090)	(0.047)	(0.029)
	4076	854	3076	8006
Hours in Agriculture	40.307	3.568	33.155	33.554
	(0.632)	(0.633)	(0.652)	(0.431)
	3891	836	3030	7757
Hours in Entrepreneurship	58.997	90.951	81.372	71.181
	(0.775)	(2.286)	(1.044)	(0.631)
	3891	836	3030	7757

Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The table reports the estimated average of each variable across the different subsamples, together with the corresponding standard error and the number of observations used. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farming entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farming entrepreneurship, but not on value added in agriculture. Households doing both are those for which we can derive information on both value added in agriculture and non-farming entrepreneurial profits.

Table A.2: Agricultural Value Added and Entrepreneurship

	Any Entrepreneurship					
	(1)	(2)	(3)	(4)	(5)	(6)
$P(VA_a)$	-0.009***	-0.008***	-0.003**	-0.005***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hours in Agriculture		-0.001***	-0.001***	-0.005***	-0.005***	-0.005***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Land Size			0.009***	0.006***	0.007***	0.006***
			(0.002)	(0.001)	(0.001)	(0.001)
Fraction Rented			-0.002	0.037**	0.030*	0.022
			(0.018)	(0.016)	(0.015)	(0.015)
Total Hours				0.004***	0.004***	0.004***
				(0.000)	(0.000)	(0.000)
Household Size					0.002	-0.002
					(0.002)	(0.002)
Females					0.023***	0.016***
					(0.003)	(0.003)
Asset Index	No	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	No	Yes
Village FE	No	No	No	No	No	No
Observations	30996	30025	27520	27520	27520	27485
R^2	0.003	0.012	0.006	0.140	0.149	0.179

	Any Entrepreneurship				
	(1)	(2)	(3)	(4)	(5)
$P(VA_a/h_a)$	0.001	0.002	0.006***	0.004***	0.003**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Land Size		0.009***	0.007***	0.007***	0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
Fraction Rented		-0.021	-0.008	0.001	0.012
		(0.021)	(0.021)	(0.022)	(0.021)
Total Hours			0.001***	0.001***	0.001***
			(0.000)	(0.000)	(0.000)
Household Size				0.017***	0.002
				(0.002)	(0.002)
Females				-0.013***	0.010***
				(0.003)	(0.004)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	No	No	No	No	No
Observations	22977	21595	21595	21595	21575
R^2	0.000	0.000	0.030	0.035	0.080

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farming entrepreneurship. $P(VA_a)$ is the percentile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $P(VA_a/h_a)$ is the percentile the household belongs to in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.3: Agricultural Value Added and Entrepreneurship

	Any Entrepreneurship					
	(1)	(2)	(3)	(4)	(5)	(6)
$P(VA_a)$	0.001	0.001	0.004***	0.002*	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hours in Agriculture		-0.000***	-0.000***	-0.004***	-0.004***	-0.004***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Land Size			-0.000	0.000	0.000	-0.000
			(0.001)	(0.002)	(0.002)	(0.002)
Fraction Rented			0.010	0.007	0.007	0.010
			(0.014)	(0.013)	(0.013)	(0.013)
Total Hours				0.004***	0.003***	0.004***
				(0.000)	(0.000)	(0.000)
Household Size					0.004*	-0.001
					(0.002)	(0.002)
Females					0.015***	0.014***
					(0.003)	(0.003)
Asset Index	No	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30930	29959	27453	27453	27453	27418
R^2	0.247	0.253	0.251	0.315	0.318	0.338

	Any Entrepreneurship				
	(1)	(2)	(3)	(4)	(5)
$P(VA_a/h_a)$	0.003***	0.003***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Land Size		0.000	-0.000	0.000	-0.000
		(0.001)	(0.002)	(0.002)	(0.002)
Fraction Rented		0.022	0.024	0.024	0.030*
		(0.016)	(0.016)	(0.016)	(0.016)
Total Hours			0.001***	0.001***	0.001***
			(0.000)	(0.000)	(0.000)
Household Size				0.008***	0.002
				(0.002)	(0.002)
Females				0.008**	0.007**
				(0.003)	(0.003)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	22892	21508	21508	21508	21488
R^2	0.247	0.254	0.268	0.270	0.293

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farming entrepreneurship. $P(VA_a)$ is the percentile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $P(VA_a/h_a)$ is the percentile the household belongs to in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.4: Entrepreneurial Profits and Farming

	Any Farming				
	(1)	(2)	(3)	(4)	(5)
$P(VA_n)$	-0.017***	-0.014***	-0.009***	-0.009***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Hours in Entrepreneurship		-0.001***	-0.005***	-0.005***	-0.005***
		(0.000)	(0.000)	(0.000)	(0.000)
Total Hours			0.004***	0.003***	0.003***
			(0.000)	(0.000)	(0.000)
Household Size				0.038***	0.029***
				(0.003)	(0.003)
Females				-0.037***	0.003
				(0.004)	(0.004)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	No	No	No	No	No
Observations	14476	14115	14115	14115	14057
R^2	0.012	0.030	0.144	0.177	0.270

	Any Farming			
	(1)	(2)	(3)	(4)
$P(VA_n/h_n)$	-0.012***	-0.009***	-0.011***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Total Hours		0.001***	0.000***	0.001***
		(0.000)	(0.000)	(0.000)
Household Size			0.037***	0.032***
			(0.003)	(0.003)
Females			-0.027***	0.009**
			(0.005)	(0.005)
Asset Index	No	No	No	Yes
Country-Wave FE	No	No	No	Yes
Village FE	No	No	No	No
Observations	12094	12094	12094	12040
R^2	0.005	0.021	0.051	0.155

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if we can derive information on value added in agriculture. $P(VA_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.5: Entrepreneurial Profits and Farming

	Any Farming				
	(1)	(2)	(3)	(4)	(5)
$P(VA_n)$	-0.001	-0.001	0.000	-0.001	-0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hours in Entrepreneurship		-0.000*	-0.002***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)
Total Hours			0.002***	0.002***	0.002***
			(0.000)	(0.000)	(0.000)
Household Size				0.021***	0.019***
				(0.002)	(0.002)
Females				0.003	0.003
				(0.003)	(0.003)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	14376	14015	14015	14015	13957
R^2	0.515	0.524	0.549	0.561	0.572

	Any Farming			
	(1)	(2)	(3)	(4)
$P(VA_n/h_n)$	0.001	0.003**	0.002	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Total Hours		0.001***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)
Household Size			0.021***	0.019***
			(0.002)	(0.002)
Females			0.006*	0.005
			(0.004)	(0.004)
Asset Index	No	No	No	Yes
Country-Wave FE	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	11962	11962	11962	11908
R^2	0.539	0.547	0.559	0.570

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if we can derive information on value added in agriculture. $P(VA_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.6: Agricultural Value Added and Time Allocation

	h_a/h_n				
	(1)	(2)	(3)	(4)	(5)
$P(VA_a)$	0.026*	0.013	-0.003	-0.005	-0.002
	(0.015)	(0.018)	(0.017)	(0.017)	(0.018)
Land Size		0.010***	0.009***	0.010***	0.011***
		(0.002)	(0.003)	(0.003)	(0.003)
Fraction Rented		-0.220	-0.221	-0.224	-0.190
		(0.240)	(0.238)	(0.238)	(0.239)
Total Hours			0.005***	0.005***	0.006***
			(0.001)	(0.001)	(0.001)
Household Size				0.023	0.023
				(0.018)	(0.019)
Females				-0.001	0.002
				(0.041)	(0.041)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	8267	7123	7123	7123	7117
R^2	0.336	0.334	0.344	0.344	0.348

	h_a/h_n				
	(1)	(2)	(3)	(4)	(5)
$P(VA_a/h_a)$	-0.123***	-0.135***	-0.120***	-0.124***	-0.116***
	(0.022)	(0.024)	(0.024)	(0.024)	(0.024)
Land Size		0.016***	0.017***	0.018***	0.018***
		(0.004)	(0.004)	(0.004)	(0.004)
Fraction Rented		-0.342	-0.357	-0.356	-0.224
		(0.367)	(0.367)	(0.366)	(0.347)
Total Hours			0.004***	0.003***	0.003***
			(0.001)	(0.001)	(0.001)
Household Size				0.046*	0.045*
				(0.026)	(0.025)
Females				-0.006	-0.005
				(0.056)	(0.056)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	5702	5241	5241	5241	5236
R^2	0.354	0.353	0.356	0.356	0.362

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we derive information on both value added in agriculture and profits from non-farming entrepreneurship. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farming entrepreneurship. $P(VA_a)$ is the percentile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $P(VA_a/h_a)$ is the percentile the household belongs to in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.7: Entrepreneurial Profits and Time Allocation

	h_n/h_a				
	(1)	(2)	(3)	(4)	(5)
$P(VA_n)$	0.131***	0.138***	0.136***	0.135***	0.128***
	(0.034)	(0.037)	(0.035)	(0.035)	(0.036)
Land Size		-0.025***	-0.025***	-0.024***	-0.025***
		(0.003)	(0.003)	(0.003)	(0.005)
Fraction Rented		0.129	0.126	0.127	0.093
		(0.227)	(0.228)	(0.227)	(0.232)
Total Hours			0.001	0.000	0.002
			(0.001)	(0.001)	(0.002)
Household Size				0.025	-0.027
				(0.037)	(0.037)
Females				0.009	0.030
				(0.062)	(0.063)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	6913	6422	6422	6422	6416
R^2	0.274	0.253	0.253	0.253	0.264

	h_n/h_a				
	(1)	(2)	(3)	(4)	(5)
$P(VA_n/h_n)$	-0.039	-0.044	-0.045	-0.047	-0.052
	(0.029)	(0.030)	(0.032)	(0.032)	(0.032)
Land Size		-0.018***	-0.018***	-0.016***	-0.018***
		(0.003)	(0.003)	(0.003)	(0.005)
Fraction Rented		0.017	0.018	0.019	-0.054
		(0.290)	(0.291)	(0.291)	(0.297)
Total Hours			-0.000	-0.001	0.000
			(0.002)	(0.002)	(0.002)
Household Size				0.059	-0.006
				(0.042)	(0.042)
Females				-0.021	0.007
				(0.077)	(0.077)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	5702	5241	5241	5241	5236
R^2	0.265	0.244	0.244	0.245	0.257

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we derive information on both value added in agriculture and profits from non-farming entrepreneurship. The dependent variable is the ratio of total hours worked by the household in non-farming entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.8: Agricultural Value Added and Time Allocation

	h_a/h_n				
	(1)	(2)	(3)	(4)	(5)
$P(VA_a)$	0.008	0.004	-0.010	-0.013	-0.010
	(0.021)	(0.023)	(0.024)	(0.024)	(0.024)
Land Size		0.008***	0.010***	0.011***	0.012***
		(0.002)	(0.003)	(0.003)	(0.003)
Fraction Rented		-0.336	-0.356	-0.356	-0.210
		(0.372)	(0.371)	(0.371)	(0.350)
Total Hours			0.004***	0.004***	0.005***
			(0.001)	(0.001)	(0.001)
Household Size				0.030	0.033
				(0.026)	(0.026)
Females				-0.000	-0.002
				(0.056)	(0.057)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	5702	5241	5241	5241	5236
R^2	0.348	0.346	0.351	0.351	0.357

	h_a/h_n				
	(1)	(2)	(3)	(4)	(5)
$P(VA_a/h_a)$	-0.123***	-0.135***	-0.120***	-0.124***	-0.116***
	(0.022)	(0.024)	(0.024)	(0.024)	(0.024)
Land Size		0.016***	0.017***	0.018***	0.018***
		(0.004)	(0.004)	(0.004)	(0.004)
Fraction Rented		-0.342	-0.357	-0.356	-0.224
		(0.367)	(0.367)	(0.366)	(0.347)
Total Hours			0.004***	0.003***	0.003***
			(0.001)	(0.001)	(0.001)
Household Size				0.046*	0.045*
				(0.026)	(0.025)
Females				-0.006	-0.005
				(0.056)	(0.056)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	5702	5241	5241	5241	5236
R^2	0.354	0.353	0.356	0.356	0.362

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we derive information on both value added in agriculture and profits from non-farming entrepreneurship and report positive total working hours in both activities. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farming entrepreneurship. $P(VA_a)$ is the percentile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $P(VA_a/h_a)$ is the percentile the household belongs to in the distribution of value added per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.9: Entrepreneurial Profits and Time Allocation

	h_n/h_a				
	(1)	(2)	(3)	(4)	(5)
$P(VA_n)$	0.148***	0.158***	0.162***	0.160***	0.149***
	(0.041)	(0.045)	(0.042)	(0.042)	(0.043)
Land Size		-0.026***	-0.027***	-0.025***	-0.027***
		(0.003)	(0.003)	(0.003)	(0.006)
Fraction Rented		0.052	0.056	0.056	-0.001
		(0.288)	(0.290)	(0.289)	(0.297)
Total Hours			-0.001	-0.001	-0.000
			(0.001)	(0.002)	(0.002)
Household Size				0.043	-0.015
				(0.042)	(0.042)
Females				0.002	0.027
				(0.076)	(0.077)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	5702	5241	5241	5241	5236
R^2	0.269	0.248	0.248	0.249	0.261

	h_n/h_a				
	(1)	(2)	(3)	(4)	(5)
$P(VA_n/h_n)$	-0.039	-0.044	-0.045	-0.047	-0.052
	(0.029)	(0.030)	(0.032)	(0.032)	(0.032)
Land Size		-0.018***	-0.018***	-0.016***	-0.018***
		(0.003)	(0.003)	(0.003)	(0.005)
Fraction Rented		0.017	0.018	0.019	-0.054
		(0.290)	(0.291)	(0.291)	(0.297)
Total Hours			-0.000	-0.001	0.000
			(0.002)	(0.002)	(0.002)
Household Size				0.059	-0.006
				(0.042)	(0.042)
Females				-0.021	0.007
				(0.077)	(0.077)
Asset Index	No	No	No	No	Yes
Country-Wave FE	No	No	No	No	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Observations	5702	5241	5241	5241	5236
R^2	0.265	0.244	0.244	0.245	0.257

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we derive information on both value added in agriculture and profits from non-farming entrepreneurship and report positive working hours in both activities. The dependent variable is the ratio of total hours worked by the household in non-farming entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile the household belongs to in the distribution of profits from non-farming entrepreneurship per hour. Both values are rescaled and multiplied by 10. Standard errors are clustered at the level of enumeration area.

Table A.10: Trends Over Time by Country

	Only Agriculture	Only Entrep.	Both	Full Sample
		Panel A. Ethiopia		
Wave 1	66.92%	12.12%	20.96%	100%
	2142	388	671	3201
Wave 2	68.96%	6.5%	24.54%	100%
	2346	221	835	3402
Wave 3	71.21%	3.74%	25.06%	100%
	2458	129	865	3452
		Panel B. Malawi		
Wave 1	76.86%	.39%	22.75%	100%
	2176	11	644	2831
Wave 2	68.97%	.63%	30.41%	100%
	1760	16	776	2552
		Panel C. Nigeria		
Wave 1	59.44%	17.92%	22.64%	100%
	2213	667	843	3723
Wave 2	55.03%	18.27%	26.69%	100%
	2066	686	1002	3754
Wave 3	32.39%	27.38%	40.23%	100%
	1326	1121	1647	4094
		Panel D. Uganda		
Wave 1	48.12%	10.65%	41.23%	100%
	1075	238	921	2234
Wave 2	51.01%	9.81%	39.18%	100%
	1056	203	811	2070
Wave 3	54.01%	11%	34.99%	100%
	1139	232	738	2109
Wave 4	51.64%	11.28%	37.07%	100%
	865	189	621	1675

Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The table reports the relative and absolute number of households across the different subsamples over different waves per country. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farming entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farming entrepreneurship, but not on value added in agriculture. Households doing both are those for which we can derive information on both value added in agriculture and non-farming entrepreneurial profits.

Table A.11: Transition Matrices

Wave 1 to 2	Only Agriculture	Both	Only Entrep.
Only Agriculture	52.65%	10.96%	0.85%
Both	9.1%	16.66%	1.04%
Only Entrepreneurship	0.94%	2.5%	5.3%

Wave 2 to 3	Only Agriculture	Both	Only Entrep.
Only Agriculture	44.62%	12.29%	1.43%
Both	6.67%	20.15%	2.04%
Only Entrepreneurship	0.82%	2.47%	9.52%

Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The table reports the relative number of households across the different subsamples and their transitions from Wave 1 (row) to 2 (column) and from Wave 2 to 3. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farming entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farming entrepreneurship, but not on value added in agriculture. Households doing both are those for which we can derive information on both value added in agriculture and non-farming entrepreneurial profits.

Table A.12: Transitions To Entrepreneurship

	Entrepreneurship Dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
Wave 2 × $Rank(VA_a)$	-0.007***	-0.009***	-0.007***			
	(0.002)	(0.002)	(0.002)			
Wave 3 × $Rank(VA_a)$	-0.008***	-0.012***	-0.009***			
	(0.003)	(0.002)	(0.003)			
Wave 2 × $Rank(VA_a/h_a)$				-0.009***	-0.009***	-0.007***
				(0.002)	(0.002)	(0.002)
Wave 3 × $Rank(VA_a/h_a)$				-0.012***	-0.013***	-0.011***
				(0.003)	(0.003)	(0.003)
Hours in Agriculture			-0.002***			
			(0.000)			
Land Size			-0.051***			-0.041**
			(0.015)			(0.020)
Fraction Rented			-0.003			-0.011
			(0.023)			(0.027)
Total Hours			0.002***			0.000**
			(0.000)			(0.000)
Household Size			0.005			0.010**
			(0.005)			(0.005)
Females			0.013*			0.003
			(0.007)			(0.008)
Asset Index	No	No	Yes	No	No	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
Wave FE	Yes	n.a.	n.a.	Yes	n.a.	n.a.
Country-Wave FE	No	Yes	Yes	No	Yes	Yes
Observations	18721	18721	16511	14746	14746	13680
R^2	0.547	0.573	0.590	0.544	0.569	0.574

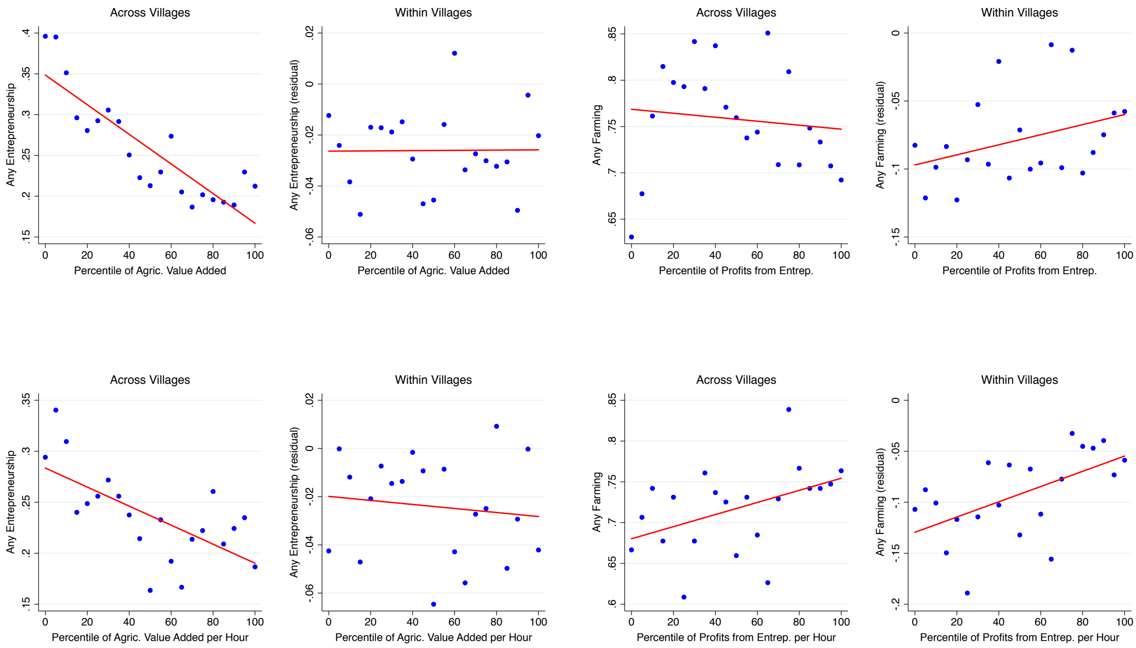
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households for which we cannot derive any information on profits from entrepreneurship in Wave 1, and observed again over time through Wave 3. $Rank(\cdot)$ is the within-village ranking of agricultural value added or agricultural value added per hour in Wave 1 among these households. Standard errors are clustered at the level of enumeration area.

Table A.13: Transitions To Entrepreneurship by Country

	Entrepreneurship Dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A. Ethiopia			
Wave 2 × $Rank(VA_a)$	-0.005**	-0.005**	-0.004*			
	(0.002)	(0.002)	(0.002)			
Wave 3 × $Rank(VA_a)$	-0.006**	-0.006**	-0.005**			
	(0.003)	(0.003)	(0.003)			
Wave 2 × $Rank(VA_a/h_a)$				-0.004*	-0.004*	-0.003
				(0.002)	(0.002)	(0.003)
Wave 3 × $Rank(VA_a/h_a)$				-0.006**	-0.006**	-0.006**
				(0.003)	(0.003)	(0.003)
Observations	6062	6062	5923	5346	5346	5237
R^2	0.524	0.524	0.539	0.517	0.517	0.533
			Panel B. Malawi			
Wave 2 × $Rank(VA_a)$	-0.010***	-0.010***	-0.010***			
	(0.002)	(0.002)	(0.002)			
Wave 2 × $Rank(VA_a/h_a)$				-0.011***	-0.011***	-0.011***
				(0.003)	(0.003)	(0.003)
Observations	3692	3692	3512	2362	2362	2298
R^2	0.561	0.561	0.558	0.556	0.556	0.555
			Panel C. Nigeria			
Wave 2 × $Rank(VA_a)$	-0.008**	-0.008**	-0.004			
	(0.004)	(0.004)	(0.004)			
Wave 3 × $Rank(VA_a)$	-0.014***	-0.014***	-0.008*			
	(0.004)	(0.004)	(0.005)			
Wave 2 × $Rank(VA_a/h_a)$				-0.010**	-0.010**	-0.007
				(0.004)	(0.004)	(0.005)
Wave 3 × $Rank(VA_a/h_a)$				-0.017***	-0.017***	-0.013**
				(0.005)	(0.005)	(0.006)
Observations	6058	6058	5191	4653	4653	4060
R^2	0.582	0.582	0.584	0.582	0.582	0.583
			Panel D. Uganda			
Wave 2 × $Rank(VA_a)$	-0.017**	-0.017**	-0.018*			
	(0.008)	(0.008)	(0.009)			
Wave 3 × $Rank(VA_a)$	-0.017**	-0.017**	-0.018**			
	(0.008)	(0.008)	(0.008)			
Wave 4 × $Rank(VA_a)$	-0.017**	-0.017**	-0.018**			
	(0.009)	(0.009)	(0.009)			
Wave 2 × $Rank(VA_a/h_a)$				-0.010	-0.010	-0.012
				(0.009)	(0.009)	(0.011)
Wave 3 × $Rank(VA_a/h_a)$				-0.019**	-0.019**	-0.021**
				(0.008)	(0.008)	(0.009)
Wave 4 × $Rank(VA_a/h_a)$				-0.013	-0.013	-0.017
				(0.010)	(0.010)	(0.010)
Observations	3547	3547	3036	2907	2907	2571
R^2	0.481	0.481	0.493	0.474	0.474	0.495
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
Wave FE	Yes	n.a.	n.a.	Yes	n.a.	n.a.
Country-Wave FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

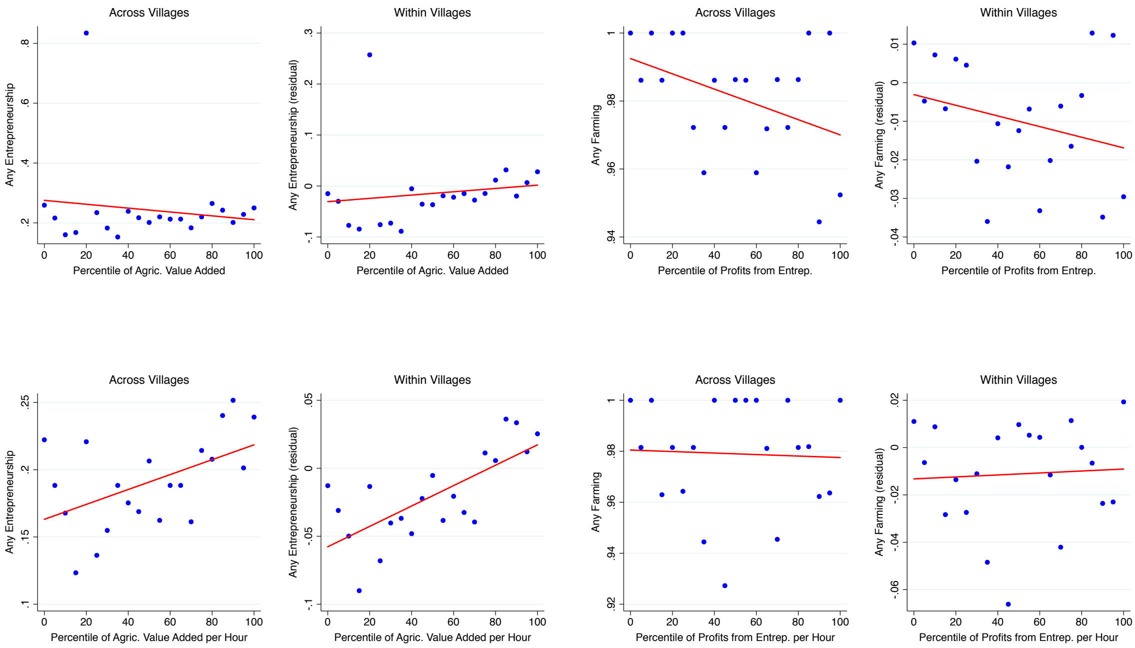
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis. Sample is restricted to those households for which we cannot derive any information on profits from entrepreneurship in Wave 1, and observed again over time through Wave 3. $Rank(\cdot)$ is the within-village ranking of agricultural value added or agricultural value added per hour in Wave 1 among these households. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total number of hours in agriculture (column 3), total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.

Figure A.1: Figures by Country - Ethiopia



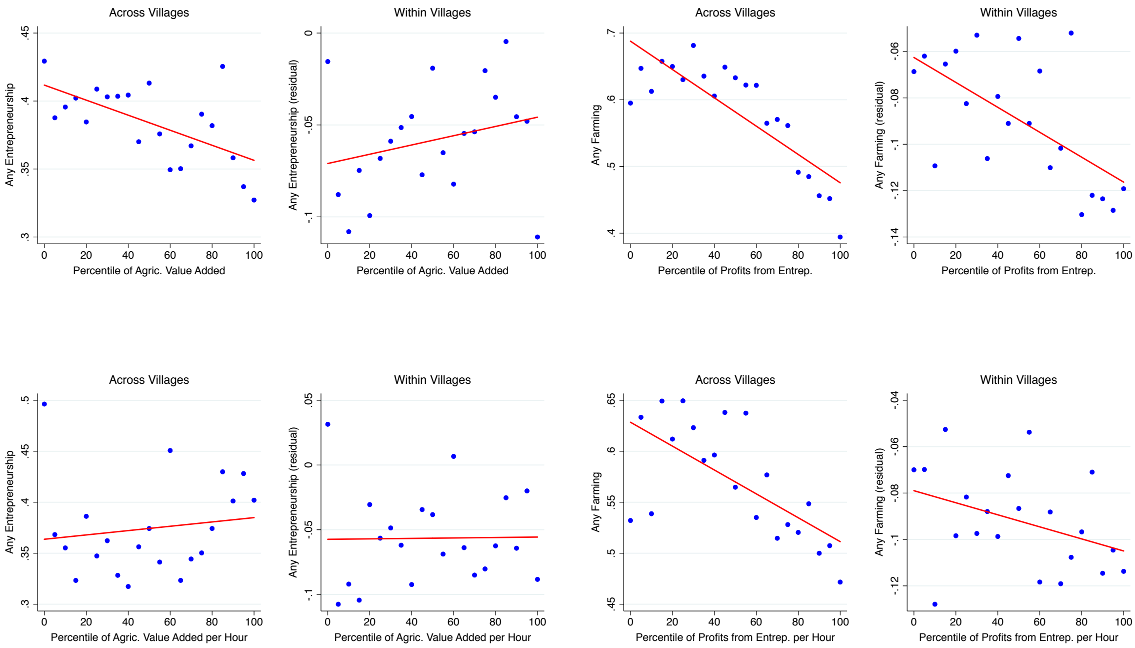
Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia. The top figures shows the fraction of households involved in non-farming entrepreneurship (left) or farming (right) per bin of 5 percentiles of the distribution of value added in agriculture (left) or profits from non-farming entrepreneurship (right). The bottom figures shows the same number per bin of 5 percentiles of the distribution of value added in agriculture per hour (left) or profits from non-farming entrepreneurship per hour (right).

Figure A.2: Figures by Country - Malawi



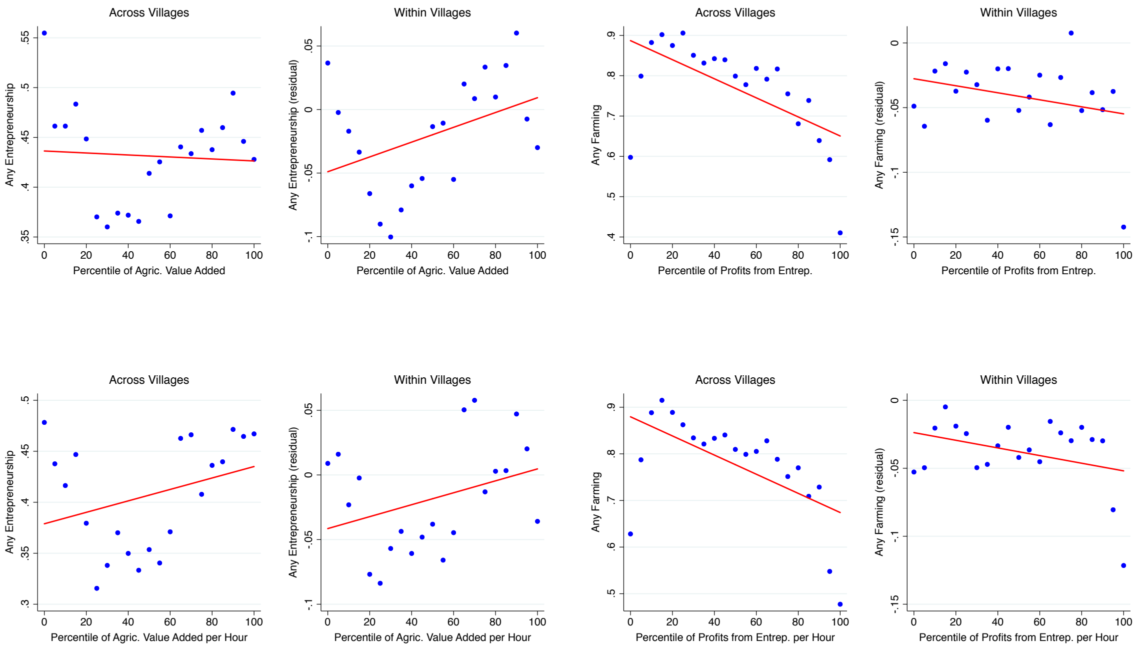
Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Malawi. The top figures shows the fraction of households involved in non-farming entrepreneurship (left) or farming (right) per bin of 5 percentiles of the distribution of value added in agriculture (left) or profits from non-farming entrepreneurship (right). The bottom figures shows the same number per bin of 5 percentiles of the distribution of value added in agriculture per hour (left) or profits from non-farming entrepreneurship per hour (right).

Figure A.3: Figures by Country - Nigeria



Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Nigeria. The top figures shows the fraction of households involved in non-farming entrepreneurship (left) or farming (right) per bin of 5 percentiles of the distribution of value added in agriculture (left) or profits from non-farming entrepreneurship (right). The bottom figures shows the same number per bin of 5 percentiles of the distribution of value added in agriculture per hour (left) or profits from non-farming entrepreneurship per hour (right).

Figure A.4: Figures by Country - Uganda



Notes. The unit of observation is the household as surveyed in each wave of the LSMS-ISA panel dataset for Uganda. The top figures shows the fraction of households involved in non-farming entrepreneurship (left) or farming (right) per bin of 5 percentiles of the distribution of value added in agriculture (left) or profits from non-farming entrepreneurship (right). The bottom figures shows the same number per bin of 5 percentiles of the distribution of value added in agriculture per hour (left) or profits from non-farming entrepreneurship per hour (right).

A.2 Proof of Proposition 1

We start from

$$\rho\left(z_i^a/z_i^n, z_i^a\right) = \frac{Cov\left(z_i^a/z_i^n, z_i^a\right)}{\sigma_{z_i^a/z_i^n}\sigma_{z_i^a}}$$

The sign of this correlation is thus given by the sign of $Cov\left(z_i^a/z_i^n, z_i^a\right)$. This is equal to

$$Cov\left(\frac{z_i^a}{z_i^n}, z_i^a\right) = E\left(\frac{z_i^a}{z_i^n}z_i^a\right) - E\left(\frac{z_i^a}{z_i^n}\right)E\left(z_i^a\right) = E\left(\frac{\left(z_i^a\right)^2}{z_i^n}\right) - E\left(\frac{z_i^a}{z_i^n}\right)\mu_a$$

where, in order to evaluate the two expectations in the last expression, we use the following second order Taylor series approximations around the means

$$\frac{\left(z_i^a\right)^2}{z_i^n} \approx \frac{\mu_a^2}{\mu_n} + \frac{2\mu_a}{\mu_n}\left(z_i^a - \mu_a\right) - \frac{\mu_a^2}{\mu_n^2}\left(z_i^n - \mu_n\right) + \frac{1}{2}\left[\frac{2}{\mu_n}\left(z_i^a - \mu_a\right)^2 + 2\frac{\mu_a^2}{\mu_n^3}\left(z_i^n - \mu_n\right)^2 - 4\frac{\mu_a}{\mu_n^2}\left(z_i^a - \mu_a\right)\left(z_i^n - \mu_n\right)\right]$$

and

$$\frac{z_i^a}{z_i^n} \approx \frac{\mu_a}{\mu_n} + \frac{1}{\mu_n}\left(z_i^a - \mu_a\right) - \frac{\mu_a}{\mu_n^2}\left(z_i^n - \mu_n\right) + \frac{1}{2}\left[\frac{2\mu_a}{\mu_n^3}\left(z_i^n - \mu_n\right)^2 - \frac{2}{\mu_n^2}\left(z_i^a - \mu_a\right)\left(z_i^n - \mu_n\right)\right].$$

Taking expectations we get

$$\begin{aligned}
Cov\left(\frac{z_i^a}{z_i^n}, z_i^a\right) &= E\left(\frac{\left(z_i^a\right)^2}{z_i^n}\right) - E\left(\frac{z_i^a}{z_i^n}\right)\mu_a \\
&\approx E\left(\begin{array}{c} \frac{\mu_a^2}{\mu_n} + \frac{2\mu_a}{\mu_n}\left(z_i^a - \mu_a\right) - \frac{\mu_a^2}{\mu_n^2}\left(z_i^n - \mu_n\right) + \\ \frac{1}{2}\left[\frac{2}{\mu_n}\left(z_i^a - \mu_a\right)^2 + 2\frac{\mu_a^2}{\mu_n^3}\left(z_i^n - \mu_n\right)^2 - 4\frac{\mu_a}{\mu_n^2}\left(z_i^a - \mu_a\right)\left(z_i^n - \mu_n\right)\right] \end{array}\right) \\
&\quad -E\left(\begin{array}{c} \frac{\mu_a}{\mu_n} + \frac{1}{\mu_n}\left(z_i^a - \mu_a\right) - \frac{\mu_a}{\mu_n^2}\left(z_i^n - \mu_n\right) + \\ \frac{1}{2}\left[\frac{2\mu_a}{\mu_n^3}\left(z_i^n - \mu_n\right)^2 - \frac{2}{\mu_n^2}\left(z_i^a - \mu_a\right)\left(z_i^n - \mu_n\right)\right] \end{array}\right)\mu_a \\
&= \frac{\mu_a^2}{\mu_n} + \frac{\sigma_a^2}{\mu_n} + \frac{\mu_a^2\sigma_n^2}{\mu_n^3} - 2\frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right) - \left(\frac{\mu_a}{\mu_n} + \frac{\mu_a\sigma_n^2}{\mu_n^3} - \frac{1}{\mu_n^2}Cov\left(z_i^a, z_i^n\right)\right)\mu_a \\
&= \frac{\sigma_a^2}{\mu_n} - \frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right),
\end{aligned}$$

Since we are only interested in the sign it follows that

$$\begin{aligned} sign\left[Cov\left(\frac{z_i^a}{z_i^n}, z_i^a\right)\right] &= sign\left[\frac{\sigma_a^2}{\mu_n} - \frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right)\right] = \frac{\mu_n^2}{\mu_a}\frac{1}{\sigma_a\sigma_n}sign\left[\left(\frac{\sigma_a^2}{\mu_n} - \frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right)\right)\right] \\ &= sign\left[\frac{\mu_n^2}{\mu_a}\frac{1}{\sigma_a\sigma_n}\left(\frac{\sigma_a^2}{\mu_n} - \frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right)\right)\right] = sign\left[\frac{CV\left(z_i^a\right)}{CV\left(z_i^n\right)} - \rho\left(z_i^n, z_i^a\right)\right], \end{aligned}$$

and therefore

$$sign[\rho\left(\frac{z_i^a}{z_i^n}, z_i^a\right)] = sign\left[Cov\left(\frac{z_i^a}{z_i^n}, z_i^a\right)\right] = sign\left[\frac{CV\left(z_i^a\right)}{CV\left(z_i^n\right)} - \rho\left(z_i^n, z_i^a\right)\right]$$

as stated in Proposition 1.

A.3 On the relation between the signs of $\rho\left(z_i^a/z_i^n, z_i^n\right)$ and $\rho\left(z_i^n/z_i^a, z_i^n\right)$

Notice that

$$sign\left[\rho\left(\frac{z_i^a}{z_i^n}, z_i^n\right)\right] = sign\left[Cov\left(\frac{z_i^a}{z_i^n}, z_i^n\right)\right] = sign\left[E\left(z_i^a\right) - E\left(\frac{z_i^a}{z_i^n}\right)E\left(z_i^n\right)\right],$$

Using a second-order Taylor series approximation around the means this becomes

$$\begin{aligned} sign\left[\rho\left(\frac{z_i^a}{z_i^n}, z_i^n\right)\right] &\approx sign\left[\mu_a - \left(\frac{\mu_a}{\mu_n} - \frac{Cov\left(z_i^a, z_i^n\right)}{\left(\mu_n\right)^2} + \frac{\sigma_n^2\mu_a}{\left(\mu_n\right)^3}\right)\mu_n\right] = sign\left[\frac{Cov\left(z_i^a, z_i^n\right)}{\mu_n} + \frac{\sigma_n^2\mu_a}{\left(\mu_n\right)^2}\right] \\ &= sign\left[\frac{Cov\left(z_i^a, z_i^n\right)}{\sigma_n\sigma_a} + \frac{\sigma_n\mu_a}{\sigma_a\mu_n}\right] = sign\left[\rho\left(z_i^a, z_i^n\right) - \frac{CV\left(z_i^n\right)}{CV\left(z_i^a\right)}\right], \end{aligned}$$

which from Proposition 1 equals $-sign\left[\rho\left(\frac{z_i^n}{z_i^a}, z_i^n\right)\right]$.

B Data Appendix

Our main source of data is the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA).[16] The LSMS-ISA project is a household survey project established with a grant from the Bill and Melinda Gates Foundation. The project is housed within the Survey Unit of the World Bank's Development Data Group. It provides technical assistance to national statistical offices (NSOs) of its eight partner countries in Sub-Saharan Africa in the design and implementation of multi-topic household surveys. Its purpose is to design and implement systems of multi-topic, nationally representative panel household surveys with a strong focus on agriculture. In each partner country, the LSMS-ISA supports multiple rounds of a nationally representative panel survey with a multi-topic approach designed to improve the understanding of the links between agriculture, socioeconomic status, and non-farm income activities. The frequency of data collection is determined on a country-by-country basis, depending on data demand and the availability of complementary funding. For our purpose, we use data from the following countries and waves

- Ethiopia - Socioeconomic Survey (ESS) - Wave 1 (2011/12), 3,969 households; Wave 2 (2013/14), 3,776 households, Wave 3 (2015/16), 3,726 households;
- Malawi - Third Integrated Household Survey (IHS3, 2010/11), 3,247 households; Integrated Household Panel Survey (IHPS, 2013), 3,246 households;
- Nigeria - General Household Survey (GHS) - Wave 1 (2010/11), 4,916 households; Wave 2 (2012/13), 4,716 households; Wave 3 (2015/2016), 4,581 households;
- Uganda - National Panel Survey (UNPS) - Wave 1 (2009/10), Wave 2 (2010/11), Wave 3 (2011/12), Wave 4 (2013/14), 3,123 households.

Each country-year sample follows a stratified two-stage sample design to ensure national representativeness. Enumeration areas (EAs) are selected with probability proportional to size within each district of the country. Random systematic sampling is used to select a certain number of primary households and some replacement households from the household listing for each sample EA. A sub-sample is randomly selected to be visited twice during the first survey to reduce recall associated with different aspects of agricultural data collection. The selected households are then tracked and resurveyed and serve as a baseline for the panel follow-up.

LSMS-ISA surveys typically include three main questionnaires: household (H), agriculture (AG), and community (C). As part of the agriculture questionnaire, fishery questionnaires are sometimes listed independently. In the agriculture questionnaire, households also report infor-

[16] See also http://surveys.worldbank.org/lsms [consulted on October 9, 2018].

mation separately on the last completed rainy and dry seasons, or post-harvest and post-planting season. These surveys collect detailed information at the household (and individual) level on income, health, education, expenditure and consumption, labor allocation, asset ownership, and details on agricultural production, business operation, and other economic activities. The surveys undertaken in different countries do not always follow identical methodologies; nevertheless, efforts have been made to follow the same method as much as possible in generating variables used in the empirical analysis. These micro-data allow us to compute measures of household-level value added by agricultural and non-agricultural activity for the four countries considered.

All of the LSMS surveys are publicly available from the World Bank website. A basic information document is available for each country, as are the survey questionnaires themselves.

Agricultural Value Added The agricultural activities of each household are generally reported separately for non-permanent crop harvested and sold, permanent crop harvested and sold, livestock sales, livestock products sales, and fishery sales. We follow Gollin, Lagakos, and Waugh (2014) and calculate the agricultural value added $VA_{a,i}$ of household i as the sum of value added from non-permanent crops ($VA_{a,i}^{NPC}$), permanent crops ($VA_{a,i}^{PC}$), livestocks ($VA_{a,i}^{LS}$), livestock products ($VA_{a,i}^{LSP}$), and fishery ($VA_{a,i}^{FS}$), i.e.

$$VA_{a,i} = \left(VA_{a,i}^{NPC} + VA_{a,i}^{PC} + VA_{a,i}^{LS} + VA_{a,i}^{LSP} + VA_{a,i}^{FS}\right)$$

Agricultural activities are questioned and reported in the survey in different seasons. Let $z \in \{NPC, PC, LS, LSP, FS\}$ identify the different agricultural activities and let s identify the rainy and dry seasons respectively in the agricultural activities, or representing high or low landing season in the fishery survey. Similarly to Santaeulalia-Llopis and de Magalhaes (2014) and Restuccia and Santaeulalia-Llopis (2017), we calculate the value added from product c of agricultural sector z as the sum across seasons of each household i's revenue from selling product c in season s ($Rev_{c,s,i}^z$), plus the market value of the product c of that was not sold (recorded as stored, lost, saved for seeds, etc) measured as $P_{c,s,i}^z(Output_{c,s,i}^z - Sold_{c,s,i}^z)$, while subtracting the associated costs ($Cost_{c,s,i}^z$), i.e.

$$VA_{a,i}^z = \sum_s Rev_{c,s,i}^z + \sum_s P_{c,s,i,r}^z(Output_{c,s,i}^z - Sold_{c,s,i}^z) - \sum_s Cost_{c,s,i}^z$$

$P_{c,s,i,r}^z$ is the inferred price of the product c in agricultural sector z in season s produced by household i in region r. Prices are imputed as follows:

- If household i sold crop c in season s and reported total sales $Rev_{c,s,i}^z$ and quantity sold

$Q^z_{c,s,i}$, we calculate $P^z_{c,s,i,r}$=$Rev^z_{c,s,i}/Q^z_{c,s,i}$;

- Otherwise, we attribute the average price of the crop sold by other households in the same region if available, meaning $P^z_{c,s,i} = \overline{P}^z_{c,s,j}$ with j being in the same region as i;
- Otherwise, we attribute the regional community price reported in community section, meaning $P^z_{c,s,i} = P^z_{c,s,com}$.

In agricultural production, each household i incurs cost $Cost^z_{c,s,i}$ per season s associated with cost type v. That is

$$Cost^z_{c,s,i} = \sum_v Cost^z_{c,s,i,v}$$

where v ={intermediate goods purchased (fertilizer, seeds, pesticides/herbicides), hired labor, rented capital (and land), transportation} if $z \in \{NPC, PC\}$; v ={intermediate goods purchased (animal feed, vaccinations, other inputs), hired labor, housing equipment, feeding utensils, transportation, veterinary services} if $z \in \{LS, LSP\}$; v ={energy cost (fuel, oil, maintenance), hired labor, rented capital (gears, boats/engines), other cost} if $z = FS$.

Entrepreneurial Profits We define household i's annual non-agricultural value added $VA_{n,i}$ as the sum of profits of all enterprises owned by the household. We identify households engaged in any kind of non-agricultural income-generating activity (owned a non-agricultural business or provided a non-agricultural service, owned a trading business, owned a professional office or offered professional services, etc.) in the last 12 month before the interview. For each household i we compute entrepreneurial profits as the total annual sales minus costs across all enterprises in the household. The value of annual total sales is annualized from the average monthly sales reported by each enterprise. And the value of annual total costs per enterprise is also annualized from the average monthly costs which consist of variable costs including raw materials, inventory, freight/transport, fuel/oil, electricity, water, insurance, etc. and total wages/salaries paid to hired labor. We thus compute

$$VA_{n,i} = \sum_i (Rev_{n,i} - Cost_{n,i})$$

where $Rev_{n,i}$ is imputed annual revenues in non-agricultural businesses n owned by household i, and $Cost_{n,i}$ is the annual aggregation of any intermediate or factor cost incurred in the same non-agricultural business.

Labor Hours In order to study individual labor supply and the intra-household allocation of time, we identify and calculate working hours of each individual by agricultural or non-agricultural activities reported in the last 7 days. We compute working hours in agriculture as the sum of hours spent in agricultural activities (including livestock and fishing-related activities) whether for sale or for household food. We compute working hours in entrepreneurship as the sum of hours spent in any kind of non-agricultural or non-fishing household business, big or small for the individual itself, or providing help in any of the household's non-agricultural or non-fishing household businesses. We then aggregate this information at the household level within and across the two activities.

Land Use The land available to each household is identified as the cumulative area of plots that any member of the household owns or cultivates. The area of the land is measured by farmer estimation and GPS measurement. We identify the ownership status of the plot as acquired by decision of the local leader, inheritance, or rented. We use this information to calculate the total cultivated area, and fraction of land that is rented, which we also consider a proxy for land market development.

Household characteristics The data provide individual demographic characteristics of household members including sex, age, birth year, geographical variables, and migration characteristics. We compute the total number of household members, the total number of female household members, a country-specific index of asset ownership that counts the number of assets listed by the household.