DISCUSSION PAPER SERIES

DP14269
(v. 4)

SELECTION, ABSOLUTE ADVANTAGE, AND THE AGRICULTURAL PRODUCTIVITY GAP

Francisco Alvarez-Cuadrado, Francesco Amodio and Markus Poschke

DEVELOPMENT ECONOMICS AND MACROECONOMICS AND GROWTH

ISSN 0265-8003

SELECTION, ABSOLUTE ADVANTAGE, AND THE AGRICULTURAL PRODUCTIVITY GAP

Francisco Alvarez-Cuadrado, Francesco Amodio and Markus Poschke

Discussion Paper DP14269
First Published 03 January 2020
This Revision 16 February 2025

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
187 boulevard Saint-Germain, 75007 Paris, France
Tel: +44 (0)20 7183 8801
www.cepr.org

SELECTION, ABSOLUTE ADVANTAGE, AND THE AGRICULTURAL PRODUCTIVITY GAP

Abstract

Output per worker is lower in agriculture than in other sectors, especially in poor countries. Worker sorting can explain this fact if comparative and absolute advantage in agriculture are positively correlated. We examine this correlation using representative household-level panel data from four African countries. Approximately one-third of households engage in both agriculture and non-farm entrepreneurship, providing a unique lens to study sorting patterns. We find that more productive farming households are more likely to pursue entrepreneurship, allocate more time to it, and increasingly transition into it over time. These findings suggest that agricultural comparative and absolute advantage are negatively correlated.

JEL Classification: J24, J31, J43, L26, O11, O13, O40

Keywords: Agricultural productivity gap, Selection, entrepreneurship, Africa

Francisco Alvarez-Cuadrado - francisco.alvarez-cuadrado@mcgill.ca
Mcgill University

Francesco Amodio - fsco.amodio@gmail.com
Mcgill University and CEPR

Markus Poschke - markus.poschke@mcgill.ca
Mcgill University

Selection, Absolute Advantage, and the Agricultural Productivity Gap*

Francisco Alvarez-Cuadrado† Francesco Amodio‡ Markus Poschke§

February 16, 2025

Abstract

Output per worker is lower in agriculture than in other sectors, especially in poor countries. Worker sorting can explain this fact if comparative and absolute advantage in agriculture are positively correlated. We examine this correlation using representative household-level panel data from four African countries. Approximately one-third of households engage in both agriculture and non-farm entrepreneurship, providing a unique lens to study sorting patterns. We find that more productive farming households are more likely to pursue entrepreneurship, allocate more time to it, and increasingly transition into it over time. These findings suggest that agricultural comparative and absolute advantage are negatively correlated.

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

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

*This paper previously circulated under the title "Selection and Absolute Advantage in Farming and Entrepreneurship." Yifan Li, Lesley Johnson and Masaya Takano provided excellent research assistance. We are thankful to Editor Chris Tonetti and two anonymous referees for their comments. We also thank Mark Bils, Paco Buera, Lorenzo Casaburi, Rui Castro, Julieta Caunedo, Stefan Dercon, Kevin Donovan, Jean-Marie Dufour, Raphael Godefroy, Doug Gollin, Clément Imbert, Fabian Lange, Sonia Laszlo, Martí Mestieri, Todd Schoellman, Mike Waugh, and all participants in many seminars, conferences and workshops including Carleton, CEMFI, Laval, McGill, National University of Singapore, Oxford, Universidad Autónoma de Madrid, Warwick, York, the 2019 Canadian Macroeconomics Study Group Meeting, the 2019 Firms, Markets and Development Workshop, the 2019 SAEe, the 2020 STEG Workshop, the 2020 WUSTL Macro Development Mini-Conference and the 2021 Meeting of the Society of Economic Dynamics. Errors remain our own. We gratefully acknowledge the financial support provided by the International Growth Centre.

†francisco.alvarez-cuadrado@mcgill.ca, McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada. Declarations of interest: none.

‡francesco.amodio@mcgill.ca, McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada. Declarations of interest: none.

§markus.poschke@mcgill.ca , McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada. Declarations of interest: none.

1 Introduction

It is well known that productivity is lower in agriculture than in other sectors in almost all countries in the world. This *agricultural productivity gap* is larger in poor countries and is only partially explained by differences in observable factors.[1] Since agriculture accounts for the majority of employment in poor countries, this gap has important implications for aggregate differences in output per capita across nations.

A recent influential literature argues that an important source of the agricultural productivity gap is worker *self-selection* or *sorting* according to comparative advantage (Lagakos and Waugh 2013). The intuition is simple: if the distributions of abilities in the population are similar across countries and the best potential farmers choose to farm, then only the very best farmers are active in countries with few farmers. In countries with more farmers, those same top farmers remain active, but they are joined by a larger group of less productive farmers. As a result, the average ability of active farmers is lower in countries with a larger farming population, leading to lower productivity in the sector. In the language of the literature, selection widens the agricultural productivity gap if comparative advantage—which determines individuals' sectoral choice—and absolute advantage—their ability or productivity in a sector—are positively correlated or *aligned*: those who choose to farm are also the best farmers overall.

Providing evidence on this hypothesis is challenging because selection itself shapes what is observable in the data. In a typical cross-section, a farmer's non-agricultural productivity is unknown. The same is true for the farming ability of non-agricultural workers. The literature therefore imposes strict distributional assumptions or relies on the information revealed by those switching sector, since they can be observed in both activities. A shortcoming of this approach is that it only focuses on those at the margin between activities and is therefore not informative of the alignment of comparative and absolute advantage in the population.

We take a new, more direct approach to investigating the alignment of comparative and absolute advantage using household-level panel 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-farm entrepreneurship (de Magalhaes and Santaeulalia-Llopis 2018). 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-farm entrepreneurship at high rates—between 27% in Malawi and 51% in Nigeria. Importantly, around a third of households is active in *both* sectors. We can thus make several useful comparisons across households between and within groups, as well as over time.

Our empirical approach is grounded in a straightforward extension of the Roy (1951) model, which allows households to either specialize in a single sector or divide their time between two sectors. The model predicts that households with a strong comparative advantage in one sector will choose to specialize exclusively in that activity, while those with a weaker comparative advantage will engage in both sectors. In the data, we compare the agricultural productivity of

[1] On the relationship between the agricultural productivity gap and development see e.g. Gollin et al. (2002), Caselli (2005), and Restuccia et al. (2008). On the role of observables in explaining the gap see Young (2013), Gollin et al. (2014), Herrendorf and Schoellman (2015; 2018) and Caunedo and Keller (2020).

households exclusively engaged in farming to those also engaged in non-farm entrepreneurship. This comparison reveals the correlation between comparative advantage in agriculture—which is weaker for those engaged in both activities—and absolute advantage in agriculture, as reflected in agricultural productivity. Crucially, we can directly measure agricultural productivity for both groups.

We find that, among those households in a village who produce some agricultural output, it is the more productive ones—those with high absolute advantage—who also engage in non-farm entrepreneurship, revealing that their comparative advantage in agriculture is weak. This suggests that comparative and absolute advantages are negatively correlated, or *misaligned*, in agriculture. This pattern holds in three of the four African countries analyzed, suggesting that the standard self-selection narrative cannot account for agricultural productivity differences across countries. If advantages are aligned in non-agriculture but misaligned in agriculture, average labor productivity decreases in both sectors as labor shifts from agriculture to non-agriculture. The productivity gap between sectors may either narrow or widen, depending on which sector experiences the greater decline in average productivity. The gap decreases if the drop in average productivity is larger in non-agriculture than in agriculture.

Where does the misalignment between comparative and absolute advantage in agriculture come from? 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 implies that if household productive abilities in agriculture and non-agriculture are strongly positively correlated, and returns from non-agriculture are more dispersed, then the best farming households can on average reap higher returns outside agriculture, and therefore tend to specialize there. Households of intermediate farming ability can still reap relatively high returns outside agriculture, and therefore pursue both activities. Those with the lowest farming ability, in contrast, tend to face very low returns outside agriculture, and therefore only pursue agriculture. This is consistent with the activity choice patterns we observe.

Our interpretation of households' activity choices along the extensive margin as reflecting selection based on comparative advantage could be confounded by several factors. First, factors such as soil quality or climate can influence agricultural productivity, thereby affecting the selection into farming and potentially confounding the identification of the correlation between advantages. Indeed, the finding that households more productive in agriculture are more likely to pursue entrepreneurship holds only when controlling for village fixed effects. We interpret this as evidence that household attributes drive selection *within* locations.

Second, selection is likely shaped by the presence of frictions like entry barriers or fixed operating costs. The fact that wealthier farming households are more likely to also engage in entrepreneurship provides suggestive evidence that such barriers are present. To address this issue, we analyze the activity choices of households along the intensive margin, which are unaffected by entry barriers. We find that among households pursuing both activities, those with higher productivity in agriculture work fewer hours in agriculture relative to non-agriculture. This implies that they have weak comparative advantage in agriculture, and thus provides further evidence that comparative and absolute advantage are misaligned in agriculture. Households with higher productivity in non-agriculture instead work more hours in this sector relative to agriculture, suggesting that comparative and absolute advantage are aligned in non-agriculture. These findings are consistent with 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. We thus conclude that the observed patterns of sectoral choices, both along the extensive and intensive margins, can be explained by a strong positive correlation of abilities across the two sectors, coupled with barriers to entry into non-farm entrepreneurship.

Next, we exploit the panel dimension of the data and look at patterns of sectoral choice over time. We find that among households that initially are only active in agriculture, it is the more productive ones who are most likely to start a non-agricultural enterprise in subsequent years. This is consistent with our interpretation of the cross-sectional evidence. It also suggests that individuals and households sort into sectors in similar ways.

Finally, we investigate the role of non-agricultural wage employment—as opposed to self-employment or entrepreneurship—and examine how it affects our analysis. Guided by an extension of our model to three activities, we find that advantages in agriculture are misaligned not only with respect to non-agricultural self-employment, but also with respect to wage employment. Advantages in non-agricultural self-employment, in contrast, are aligned with respect to wage employment. This is consistent with a setting where all abilities are strongly positively correlated, but dispersion is smallest in agriculture, intermediate in non-agricultural wage employment, and largest in entrepreneurship. The patterns of alignment we find—and their implications for selection—are thus robust to the inclusion of wage employment opportunities.

We conclude our analysis by considering a plethora of alternative mechanisms other than selection on ability, including the presence of frictions that distort the activity of households along the intensive margin. We discuss the extent to which these are consistent with the empirical findings. In general, we find little support for these alternative explanations.

To summarize, the fact that a large fraction of households in rural Africa engages in both agricultural and non-agricultural work allows us to sign the correlation of comparative and absolute advantage in agriculture and non-agriculture, as well as the correlation of absolute advantages across sectors. The fact that the best farmers are more likely to also engage in non-agriculture suggests a negative correlation, or misalignment, between 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; and that, over time, the most productive farming households are systematically more likely to start a non-agricultural enterprise. Taken together, results from both cross-sectional and panel data analysis indicate that a strong positive correlation of productive abilities in the two sectors is responsible for the misalignment of advantages in agriculture.

Our paper is not the first one to analyze the correlation between comparative and absolute advantage or the correlation of productive abilities across sectors. To identify these correlations, earlier work exploits sector-level evidence combined with strict distributional assumptions or information from sector switchers. Lagakos and Waugh (2013) impose dependent Frechet distributions of abilities in the two sectors and calibrate them using average wages across sectors in the United States. Their findings imply a positive correlation of advantages in both sectors. Adamopoulos, Brandt, Leight, and Restuccia (2022) calibrate the same joint distribution of abilities 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 positive correlation of advantages. Using data from Brazil, Alvarez (2020) shows that formal workers who transition out of agriculture experience limited compensation gains when compared to the large overall gap in mean wages between agriculture and other sectors. Hamory, Kleemans, Li, and Miguel (2020) 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. Using different information from the same data set, Pulido and Święcki (2019) find income gains of over 20% for workers who move out of agriculture. They also conduct a structural estimation exercise that suggests that, while self-selection is important, there are also barriers that significantly misallocate workers across sectors. These results are obtained under the identifying assumption of uncorrelated shocks to households' sectoral productivity. What these findings highlight is that unless combined with information on infra-marginal individuals, selection itself makes observational returns to switching sectors or rural-to-urban migration uninformative about the correlation of advantages, the role of sorting, and the scope for worker reallocation (see also Herrendorf and Schoellman 2018; Lagakos, Mobarak, and Waugh 2021; Lagakos, Marshall, Mobarak, Vernot, and Waugh 2020).

Our approach differs from the previous literature in that it requires weaker distributional assumptions and exploits the presence of a large group of households that are simultaneously active in both sectors to sign the correlation of advantages. In several specifications, we consistently find misaligned advantages in agriculture. This casts doubts on the role of self-selection on unobserved ability as a major determinant of larger agricultural productivity gaps in poor countries. The theoretical restrictions that our estimates place on the joint distribution of abilities also provide valuable information about the sign and size of the correlation coefficient of sectoral abilities and their relative dispersion.[2]

The remainder of the paper is organized as follows. Section 2 illustrates the theory that motivates our analysis and the core identification arguments. Section 3 presents the data sources and their summary statistics. Section 4 contains the main results on patterns of selection along the extensive margin. Section 5 shows the empirical results on selection along the intensive margin. Section 6 focuses on selection over time. Section 7 discusses the relationship between household-level and individual-level results. Section 8 presents the extended model that incorporates non-agricultural wage work. Section 9 discusses a final set of possible alternative explanations. Section 10 concludes.

2 A Simple Model of Selection

This section presents a simple, general model of the economy with selection into agricultural and non-agricultural activities. We derive the model's empirical implications and describe how

[2]In this respect, our focus on the distribution of advantages rather than abilities is similar to Adão (2016), who assumes constant-elasticity schedules for comparative and absolute advantage to fully characterize the distribution of advantages.

to use them to identify the correlation between absolute and comparative advantage in each sector using data, particularly on households engaged in both activities simultaneously.

2.1 Environment

Consider an economy with two sectors, agriculture and non-farming entrepreneurship, denoted by a and n respectively. There is a mass 1 of households indexed by i. Each household is endowed with a vector of sector-specific abilities $\{z_i^a, z_i^n\}$ drawn from a joint distribution $G(z^a, z^n)$ with support on the positive reals and finite means μ_j and variances σ_j^2, where $j = \{a, n\}$. We define a household's comparative advantage in agriculture as the ratio of agricultural to non-agricultural abilities, z_i^a/z_i^n, while absolute advantage in agriculture is given by agricultural ability z_i^a. Similarly, entrepreneurial comparative and absolute advantage are given by z_i^n/z_i^a and z_i^n, respectively.[3]

The only restriction we impose on $G(\cdot)$ is the assumption that both $\mathbb{E}(z^a | \, z^a/z^n > x)$ and $\mathbb{E}(z^n | \, z^n/z^a > x)$ are monotone in x. This ensures that the correlation between comparative and absolute advantage in each sector maintains the same sign across the entire support of the ability distribution.[4] Unlike Lagakos and Waugh (2013), we do not impose restrictions on whether these two objects increase or decrease with x, since this is determined by the correlation of advantages in each sector, which is what we are ultimately interested in.[5]

2.2 Technology and Selection

Each household i is endowed with one unit of time that it allocates between agriculture l_i^a and non-agricultural entrepreneurship $l_i^n = 1 - l_i^a$. 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

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

where $f(\cdot)$ and $g(\cdot)$ are increasing and strictly concave functions with bounded derivatives at the origin, and κ captures 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

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

[3]We conduct our analysis at the household level due to data availability. In Section 7, we present results indicating that selection patterns at the individual level mirror those that we find at the household level.

[4]For instance, if $\mathbb{E}(z^a | \, z^a/z^n > x)$ is always increasing in x, stronger agricultural comparative advantage is associated, on average, with higher agricultural absolute advantage. As a result, the correlation of advantages in agriculture is positive. In contrast, if the conditional expectation monotonically decreases in x, the correlation of advantages is negative in agriculture.

[5]Young (2014) and Adão (2016) place similar, although more restrictive, conditions on the distribution of abilities. The former requires sectoral abilities to be independent and the elasticity of the cumulative distribution function for each of the abilities to be decreasing in the level of the ability draw. The latter imposes constant elasticity schedules for both comparative and absolute advantage.

2.2.1 Binary Activity Choice

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 will be active 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 comparative advantage in agriculture, i.e. $z_i^a/z_i^n \geqslant g(1)/(\kappa f(1))$, will engage in farming, while those with a strong entrepreneurial comparative advantage, i.e. $z_i^n/z_i^a > \kappa f(1)/g(1)$, will operate in the non-farm entrepreneurship sector. Combining equation (3) with (1) and the joint density function $g(z^a, z^n)$, we derive mean sectoral labor productivity in both sectors

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

Although comparative advantage determines sectoral allocations, absolute advantage determines sectoral productivity. It follows that the relation between sectoral employment shares and labor 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, so that they both feature positive selection. In this case, an increase in the threshold of comparative advantage required to operate in a sector leads to an increase in the absolute advantage of those who remain active in the sector. It follows that average productivity increases as a sector shrinks, since the least productive leave the sector. The converse is true in expanding sectors: incoming workers have not only a lower comparative advantage but also, on average, a 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.

Figure 1 illustrates this reasoning. In each panel, the first figure on the left shows a scatter plot of the abilities z_i^a and z_i^n in a simulated population of households. In panel (a), these are generated such that comparative and absolute advantage are aligned in both sectors. This is the case just discussed. In panel (b), they are generated such that advantages are aligned in entrepreneurship, but misaligned in agriculture. In each of the figures, the lines emanating from the origin are lines of constant comparative advantage, indicating the threshold that determines selection across sectors. To show the impact of changes in such a threshold, we draw two lines: $z_i^a/z_i^n = g(1)/(\kappa_t f(1))$, $t = 0, 1$, with $\kappa_1 < \kappa_0$. In both panels, fewer households find farming optimal when the comparative advantage threshold to engage in farming is higher (with κ_1).

The central figure in each panel shows a scatter plot of households' comparative advantage in agriculture, z_i^a/z_i^n, against their absolute advantage in agriculture, z_i^a. Naturally, the lines of constant comparative advantage determining selection are now horizontal. Finally, the right figure in each panel shows a scatter plot of households' comparative advantage in non-farming entrepreneurship, z_i^n/z_i^a, against their absolute advantage in the same sector, z_i^n. Again, the lines of constant comparative advantage are horizontal.

Panel (a) illustrates the first case discussed above featuring positive selection in both sectors. As the agricultural sector shrinks and some households switch to non-agriculture, the average absolute advantage of those who remain in agriculture ($\bar{z}_1^a$ in the central figure) exceeds that of those who switch sector ($\bar{z}_S^a$). That is, average agricultural productivity increases as the sector shrinks. Panel (b) shows that the opposite is true if advantages are misaligned in agriculture. In this case, the average agricultural absolute advantage of those leaving agriculture exceeds that of those staying. That is, average agricultural productivity decreases as the sector shrinks. Hence, average productivity in agriculture increases as the sector shrinks—or sectoral size and average productivity in agriculture are negatively correlated—only if advantages are aligned in agriculture.

Notice that, in both cases, average productivity in non-agriculture decreases because the average absolute advantage in entrepreneurship of sector switchers ($\bar{z}_S^n$) is lower than that of those already active in that sector ($\bar{z}_0^n$). This occurs because comparative and absolute advantages in non-agriculture are always aligned. It follows that if advantages are also aligned in agriculture, the agricultural productivity gap decreases as the agricultural sector shrinks. If instead advantages are misaligned in agriculture, average labor productivity declines in both sectors as labor reallocates away from agriculture. In this case, the productivity gap between sectors can either decrease or increase depending on which sector sees the larger decline in average productivity, and the productivity gap decreases only if the decline in average productivity is larger in non-agriculture than in agriculture.

2.2.2 Combined Activity Choice

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. First, consider the households who actually do so. These households split their time to equate the marginal value products of labor across the two activities. As a result, their optimal labor allocation $(\tilde{l}_i^a, \tilde{l}_i^n)$ is implicitly defined by

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

Because of diminishing marginal products, relative hours worked in a sector increase with comparative advantage in that sector. More formally,

$$\frac{\partial\left(\tilde{l}_i^a/\tilde{l}_i^n\right)}{\partial\left(z_i^a/z_i^n\right)} = -\frac{\kappa f'\left(l_i^a\right)}{\left(l_i^n\right)^2\left(\frac{z_i^a}{z_i^n}\kappa f''\left(l_i^a\right) + g''\left(1-l_i^a\right)\right)} > 0. \tag{6}$$

We can also use condition (5) to evaluate sectoral choices. Households for whom

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

have such a strong comparative advantage in agriculture that they will engage in farming only. They are at a corner solution of their hours allocation and, accordingly, specialize. At the other end of the spectrum, households for whom

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

have a strong comparative advantage in non-farm entrepreneurship, and thus fully specialize in that sector. Finally, households with intermediate levels of comparative advantage will operate in both sectors. These households have

$$\frac{z_i^a}{z_i^n} \in \left(\frac{g'(1)}{\kappa f'(0)}, \frac{g'(0)}{\kappa f'(1)}\right) \quad \text{or} \quad \frac{z_i^a}{z_i^n} \in (\zeta_n, \zeta_a). \tag{9}$$

The equations above show that when a household is endowed with a pair of relatively similar abilities and thus intermediate levels of comparative advantage, diminishing returns to labor at the sectoral level make it optimal to split the time endowment between the two activities. Still, this intermediate comparative advantage is not informative of absolute advantages, z_i^a and z_i^n. Households operating in both sectors could be high in the marginal distributions of each ability, or could, equally well, be 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.

Figure 2 makes this point graphically. Its structure is identical to that of Figure 1. Again, panel (a) shows the case where advantages are aligned in both sectors, and panel (b) the case where advantages are aligned in entrepreneurship but misaligned in agriculture. In each panel, the left figure shows a scatter plot of abilities, the central figure comparative advantage in agriculture against absolute advantage in agriculture, and the right figure comparative advantage in non-farming entrepreneurship against absolute advantage in that sector. The lines of constant comparative advantage now split the population into three groups: those with strong comparative advantage in agriculture ($z_i^a/z_i^n \geqslant \zeta_a$), those with strong comparative advantage in non-agricultural entrepreneurship ($z_i^a/z_i^n \leqslant \zeta_n$), and those with intermediate comparative advantage ($\zeta_n < z_i^a/z_i^n < \zeta_a$).

2.3 Role of Abilities

What determines the correlation between absolute and comparative advantage? Denoting the correlation between two variables x and y by $\rho(x, y)$, the following proposition holds.

Proposition 1. *The signs of the (approximated) correlations between comparative and ab-*

solute advantage are 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 B.2 for a proof.

Abstracting from trivial cases where the distributions of sectoral abilities coincide or are degenerate in at least one sector, several insights arise from Proposition 1.[6] 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.[7] For the sake of exposition, let us assume $CV\left(z_i^n\right) > CV\left(z_i^a\right)$, so that advantages are always aligned in entrepreneurship. Second, when abilities are not positively correlated, i.e. $\rho\left(z_i^a, z_i^n\right) \leq 0$, advantages are aligned in both sectors.

Third, under positive correlation of abilities, advantages in agriculture will be aligned as long as $\rho\left(z_i^a, z_i^n\right) < CV\left(z_i^a\right)/CV\left(z_i^n\right)$, uncorrelated when $\rho\left(z_i^a, z_i^n\right) = \bar{\rho} \equiv CV\left(z_i^a\right)/CV\left(z_i^n\right)$, and misaligned otherwise. The first equation in (10) thus determines a threshold $\bar{\rho}$ 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 the words of Young, "at worst weakly correlated" (Young 2014). The more different the sectors are 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, $\bar{\rho}$, that ensures that advantages remain aligned in agriculture.

A special case. Proposition 1 clarifies how our findings on the correlation of advantages from this approach are linked to the underlying distribution of abilities. In an important special case, we obtain additional results.

Suppose that $(\ln z^a, \ln z^n)$ are jointly normal with means $\tilde{\mu}_a, \tilde{\mu}_n$ and standard deviations $\tilde{\sigma}_a, \tilde{\sigma}_n$, and correlation $\tilde{\rho}$. Then we can write

$$\ln z^n = n_0 + \theta \ln z^a + u, \tag{11}$$

with $\theta = \tilde{\rho}\tilde{\sigma}_n/\tilde{\sigma}_a$ and $u \sim N(0, \sqrt{\tilde{\sigma}_n^2 - \theta^2\tilde{\sigma}_a^2})$. In this case, advantages are aligned in agri-

[6] 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 $\rho\left(z_i^a, z_i^n\right) = 1$, advantages are uncorrelated in both sectors, $\rho\left(z_i^a/z_i^n, z_i^a\right) = \rho\left(z_i^n/z_i^a, z_i^n\right) = 0$. 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 likely that these cases are not empirically relevant.

[7] Advantages can never be misaligned in both sectors: assume that advantages in agriculture are misaligned, i.e. those with low agricultural comparative advantage z^a/z^n have high agricultural absolute advantage z^a. If entrepreneurial advantages were also misaligned, those same households have not only high entrepreneurial comparative advantage z^n/z^a, but also low entrepreneurial absolute advantage z^n. But then these households have high z^a and low z^n, which contradicts the assumption on their comparative advantage.

culture – and households with a greater absolute advantage in agriculture are more likely to specialize in that sector – only if $\theta < 1$. If $\theta > 1$, advantages are misaligned in agriculture, so that households with greater farming ability are less likely to specialize, and more likely to pursue non-farming entrepreneurship.[8] The threshold of 1 for θ coincides with the one for $\rho(z^a, z^n)$ given in Proposition 1.[9] It is also consistent with Heckman and Sedlacek (1985) showing that positive selection in agriculture arises if and only if $\tilde{\sigma}_a^2 - \tilde{\rho}\tilde{\sigma}_a\tilde{\sigma}_n > 0$ or $\theta < 1$. The result also applies when the difference of log abilities is log-concave (Heckman and Honoré 1990). Although widely used in the literature on selection of migrants (see for instance Borjas 1987), this result seems to have been overlooked in the recent work on selection and sectoral productivity differences.

Summarizing, the model predicts that households with a greater absolute advantage in agriculture also have a greater comparative advantage, indicating alignment of advantages in agriculture only if $\theta < 1$ for jointly log-normal abilities, or more generally under the conditions in Proposition 1.

2.4 Identification

We now demonstrate how the proposed theoretical framework can be leveraged to develop identification strategies that allow us to empirically estimate the correlation between absolute and comparative advantage in each sector.

2.4.1 Extensive Margin

Figure 2 illustrates how to achieve identification by leveraging information about activity choice along the extensive margin. Consider the central panels: they illustrate that observing absolute advantage in agriculture not only among specialized farmers but also among households active in both sectors enables us to determine the sign of the correlation between advantages in agriculture. Panel (a) shows that a positive correlation of advantages in agriculture implies that specialized farmers have on average *higher* absolute advantage in agriculture ($\bar{z}_F^a$) than households engaged in both activities ($\bar{z}_B^a$), whereas panel (b) shows that a negative correlation of advantages in agriculture implies that those who specialize in farming have on average *lower* absolute advantage in agriculture than those who do both. A similar reasoning applies to non-farm entrepreneurship (right panels).

In other words, the model predicts that the probability that a farming household also pursues non-farm entrepreneurship increases with absolute advantage in agriculture only if advantages in agriculture are misaligned. This arises if $\theta > 1$ for jointly log-normal abilities, or more generally under the conditions in Proposition 1. This is illustrated in Figure 3, which plots the

[8] Misalignment in agriculture arises when greater z^a implies a greater probability that comparative advantage in farming, z^a/z^n, lies between ζ_a and ζ_n, rather than above ζ_a. In the joint log-normal case, this probability is $Prob(z^a/z^n \in (\zeta_n, \zeta_a)|z^a)/Prob(z^a/z^n > \zeta_n|z^a) = [\Phi((1-\theta)\ln z^a - n_0 - \ln\zeta_n) - \Phi((1-\theta)\ln z^a - n_0 - \ln\zeta_a)]/\Phi((1-\theta)\ln z^a - n_0 - \ln\zeta_n)$, where Φ is the c.d.f. of u. This increases in z^a only if $\theta > 1$. See Appendix B.4 for details.

[9] For bivariate normal $(\ln z^a, \ln z^n)$, the coefficient of variation of z^k is $(e^{\tilde{\sigma}_k^2} - 1)^{1/2}$, and $\rho(z^a, z^n)$ is $\frac{\exp(\tilde{\rho}\tilde{\sigma}_a\tilde{\sigma}_n)-1}{[\exp(\tilde{\sigma}_a^2)-1]^{1/2}[\exp(\tilde{\sigma}_n^2)-1]^{1/2}}$. Using these expressions reveals that the correlation exceeds the ratio of coefficients of variation, so that agriculture is misaligned following Proposition 1, if $\tilde{\rho}\tilde{\sigma}_n/\tilde{\sigma}_a = \theta > 1$.

probability that a farming household also pursues non-farm entrepreneurship for varying levels of the correlation of abilities.

As a result, a straightforward analysis of how comparative advantage—reflected in activity choices—varies with absolute advantage can reveal the sign of the correlation between advantages. If households with greater absolute advantage in a sector also exhibit higher comparative advantage in that sector, as evidenced by a lower likelihood of participating in the other sector, the advantages are positively correlated. Conversely, if these households are more likely to engage in the other sector, the advantages are negatively correlated.[10] We pursue this identification strategy in Section 4.

2.4.2 Intensive Margin

Our second identification strategy exploits the model's prediction for hours worked in each sector by households that simultaneously engage in both activities. Equations (5) and (6) demonstrate that the relative supply of hours to a sector increases with comparative advantage in that sector.

Therefore, we can use relative hours as a proxy for comparative advantage and infer the correlation between absolute and comparative advantage by examining how this proxy varies with absolute advantage in each sector. We implement this strategy in Section 5.[11]

2.4.3 Sector Switchers

An identification strategy widely used in the literature focuses on sectoral switchers, and it is also applicable in our context. Consider how changes in selection cutoffs influence the activity choices shown in Figure 2. For example, shifts that reduce the fraction of households in agriculture—raising ζ_a and thereby shifting the top horizontal line upward in the middle panel—will primarily cause high-absolute-advantage farmers to transition to non-farm entrepreneurship if advantages in agriculture are misaligned. Conversely, low-productivity farmers are more likely to switch to non-farm entrepreneurship when agricultural advantages are aligned. Therefore, examining whether high- or low-productivity farmers are more likely to transition into non-farm entrepreneurship over time provides insight into the correlation between absolute and comparative advantage. Moreover, this approach allows us to account for time-invariant unobserved factors, such as differences in wealth or access to technology, that are not captured by observable variables. We apply this strategy in Section 6.

2.5 Other Drivers of Selection

Thus far, our theoretical investigation has focused exclusively on comparative advantage, neglecting other factors that could influence selection. We now turn to these additional factors,

[10] A simple comparison of means of absolute advantage across the two groups—specialized farmers and those doing both activities—can also reveal the sign of the correlation of advantages in agriculture. However, this approach is less amenable to empirical analysis in the presence of other, observable dimensions of household heterogeneity.

[11] In the special case of bivariate log-normal abilities discussed in Section 2.3, the log ratio of optimal hours, $\ln(\tilde{l}_i^a/\tilde{l}_i^n)$, equals $\frac{1}{1-\alpha}[(1-\theta)\ln z_i^a + \ln\kappa - n_0 - u_i]$ if the sectoral production functions f and g have a common elasticity α with respect to labor. That is, among households active in both sectors, relative hours in agriculture increase in absolute advantage z^a only if $\theta < 1$.

discussing their potential impact and outlining how our analysis accounts for them.

2.5.1 Complementary Inputs

Suppose that production also uses land and capital, as captured by the production functions

$$y_i^a = \kappa z_i^a F(l_i^a, k_i^a, T_i) \tag{12}$$

$$y_i^n = z_i^n G(l_i^n, k_i^n), \tag{13}$$

where k_i^j denotes capital used by household i in activity j, and T_i denotes land. In this case, the thresholds for comparative advantage in agriculture become

$$\chi_a = \frac{G'(0, k_i^n)}{\kappa F'(1, k_i^a, T_i)} \tag{14}$$

$$\chi_n = \frac{G'(1, k_i^n)}{\kappa F'(0, k_i^a, T_i)}. \tag{15}$$

Like above, a household specializes in agriculture if $z_i^a/z_i^n \geq \chi_a$, in non-farm entrepreneurship if $z_i^a/z_i^n \leq \chi_n$, and pursues both activities if z_i^a/z_i^n lies between the two thresholds.

In this more general setting, the two thresholds depend on the quantities of land and capital that each household can use in production. Use of more or better land—an increase in T—reduces both thresholds. Capital affects the threshold if the two sectors differ in capital intensity. In the empirical analysis, we take these effects of complementary inputs on sectoral choice into account by including controls for land size and proxies for wealth.

2.5.2 Location-Specific Productivity

Consider now a scenario where

$$y_i^a = \kappa A_g z_i^a f(l_i^a), \tag{16}$$

where A_g is location-specific agricultural productivity, reflecting factors like climate and soil characteristics in a location, like a village, that includes several households. In this case, choice thresholds vary systematically across locations. In areas with high agricultural productivity, the agricultural comparative advantage threshold for choosing non-farm entrepreneurship is lower for all households. As a consequence, more of them would choose agriculture in such locations, even if the ability distribution was common across locations.

This discussion makes it clear that location-specific productivity influences both productivity in agriculture and selection, potentially confounding the identification of the correlation between advantages. To address this, we incorporate location fixed effects throughout our analysis, ensuring comparisons are made among households within the same location.

2.5.3 Sector-Specific Fixed Costs

Finally, sectoral choices can also be influenced by the presence of fixed operating costs or entry costs. We can extend the model accordingly and let τ^j capture fixed costs in sector j.[12] These costs, if negative, should be interpreted as amenities and, in principle, may be correlated with abilities.

As before, households take prices as given and allocate labor to maximize income net of operating costs

$$y_i = \kappa z_i^a f\left(l_i^a\right) - \tau^a \mathbf{1}(y_i^a > 0) + z_i^n g\left(1 - l_i^a\right) - \tau^n \mathbf{1}(y_i^n > 0), \tag{17}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Household i will operate in both sectors as long as

$$\kappa z_i^a f\left(\tilde{l}_i^a\right) - \tau^a + z_i^n g\left(1 - \tilde{l}_i^a\right) - \tau^n \geqslant \max\left[\kappa z_i^a f\left(1\right) - \tau^a, z_i^n g\left(1\right) - \tau^n\right], \tag{18}$$

where $\tilde{l}_i^a$ is the optimal labor allocation. In terms of comparative and absolute advantage, this becomes

$$\kappa \frac{z_i^a}{z_i^n} f\left(\tilde{l}_i^a\right) - \frac{\tau^a}{z_i^n} + g\left(1 - \tilde{l}_i^a\right) - \frac{\tau^n}{z_i^n} \geqslant \max\left[\kappa \frac{z_i^a}{z_i^n} f\left(1\right) - \frac{\tau^a}{z_i^n}, g\left(1\right) - \frac{\tau^n}{z_i^n}\right]. \tag{19}$$

or equivalently

$$\kappa f\left(\tilde{l}_i^a\right) - \frac{\tau^a}{z_i^a} + \frac{z_i^n}{z_i^a} g\left(1 - \tilde{l}_i^a\right) - \frac{\tau^n}{z_i^a} \geqslant \max\left[\kappa f\left(1\right) - \frac{\tau^a}{z_i^a}, \frac{z_i^n}{z_i^a} g\left(1\right) - \frac{\tau^n}{z_i^a}\right]. \tag{20}$$

The first expression provides information about the sign of the correlation of advantages in non-farm entrepreneurship $\rho\left(z_i^n/z_i^a, z_i^n\right)$, while the second is informative about the same correlation in agriculture, $\rho\left(z_i^a/z_i^n, z_i^a\right)$.[13]

These final expressions illustrate that, with fixed costs, sectoral choices are influenced not only by comparative advantage but also by the levels of τ^a and τ^n. This threatens identification of the correlation of advantages based on activity choices along the extensive margin or sector switchers. However, it does not affect identification when using the second strategy, which links relative labor supply to productivity in each sector.

We illustrate their effect through an example. Assume the correlation of abilities is positive but weak, in particular, $\rho\left(z_i^a, z_i^n\right) = \bar{\rho} = CV\left(z_i^a\right)/CV\left(z_i^n\right) \in (0, 1)$. According to Proposition 1, in such a scenario advantages will be aligned in entrepreneurship, $\rho\left(z_i^n/z_i^a, z_i^n\right) > 0$, and uncorrelated in farming, $\rho\left(z_i^a/z_i^n, z_i^a\right) = 0$. Panel (a) of Figure 4 reproduces this scenario in the absence of barriers. On the one hand, since abilities are positively correlated and they are more dispersed in entrepreneurship, the most able entrepreneurial households fully specialize in this activity. It follows that comparative and absolute advantage are aligned in entrepreneurship. On the other hand, those with a high comparative advantage in agriculture specialize in farming. For some of them agricultural comparative advantage is high because they are good farmers, but for some others it is high because they are very poor entrepreneurs. Those engaged in both activities have weak comparative advantage in agriculture. Some of them are relatively good at both activities, while others are relatively bad at both. As a result, the average agricultural

[12] While we model the costs as fixed operating costs, fixed costs of entry would have a similar effect in our setting. For recent evidence on dispersion in financing costs, see Cavalcanti et al. (2021).

[13] In Appendix B.2 we show that $sign\left[\rho\left(z_i^a/z_i^n, z_i^n\right)\right] = -sign\left[\rho\left(z_i^n/z_i^a, z_i^n\right)\right]$.

ability of those specialized in farming turns out to coincide with that of those engaged in both activities: $\bar{z}^a_A = \bar{z}^a_B$, and advantages are correctly measured as uncorrelated in agriculture.

Consider now the introduction of a fixed cost to enter non-farm entrepreneurship ($\tau^n > 0$). This situation is illustrated in panel (b) of Figure 4, with the broken line indicating indifference between farming only and both activities—the choice between non-farm entrepreneurship and both activities is unaffected, as the fixed cost is due in both cases. The fixed cost pushes some of the households that in the absence of this cost would choose to engage in both activities to do only farming. These households have relatively low entrepreneurial ability. Given the positive correlation of abilities, these households also have relatively low agricultural ability. It follows that the group of households that remains active in both sectors has, on average, higher ability in both activities. This reduces the inferred correlation of advantages in both sectors. In agriculture it turns from zero to negative (as $\bar{z}^a_F < \bar{z}^a_B$). In entrepreneurship, it is reduced, and for large enough τ^n can change from positive to zero (if $\bar{z}^n_B = \bar{z}^n_E$). Large enough fixed costs could thus confound identification the identification of advantages based on activity choices along the extensive margin or sector switchers. Selection along the intensive margin, in contrast, is not affected by the presence of fixed costs (or amenities), since they do not affect the optimal allocation of labor to activities for those households who engage in both activities—equation (5).

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. Our final dataset combines the information on four countries—Ethiopia, Malawi, Nigeria, and Uganda. The number of survey rounds or waves per country varies from 2 (Malawi) to 4 (Uganda), covering the years from 2009 to 2016. In this section, we describe the main variables we use while referring to Appendix C for detailed information on sampling frame, survey design, and definitions.[14]

Value Added and Hours Worked First, we compute for each household in each wave a measure of value added in agriculture.[15] 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.[16] A common issue in using this data is that of assigning a monetary value to unsold agricultural production, which represents the majority of total household production. To determine its market value, we either use the price at which the household sold that same crop or, if not available, the one reported by households in the same location that sold that crop, or the price recorded in the community-level survey.[17] We then calculate value added as the sum

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

[15] Both agricultural and non-agricultural activities are typically household enterprises, which implies that value added can only be measured at the household level.

[16] As we show later, the empirical results are robust to excluding livestock and livestock products in the definition of farming activity and value added in agriculture.

[17] As reported later in Section 4.1, according to our calculations the fraction of market revenues from agriculture over the total value of agricultural output ranges from 20% in Ethiopia to 37% in Uganda.

across seasons of each household's revenues from selling each product plus the market value of products that were not sold minus the associated production costs.[18]

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

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. Each household member is asked about the hours worked in the last 7 days on the household farm, in any of the household non-farming enterprises, and outside the household, in the form of paid or unpaid work, temporary or not, apprenticeship, etc. We calculate the total number of hours worked in each activity at the household level by aggregating the hours worked across all household members.

We derive measures of value added in agriculture and profits from non-farm entrepreneurship using information on production, sales, and costs over the entire year. This contrasts with the information on hours worked, which pertains to the last 7 days before the interview is conducted. This information belongs to the time use section of the household questionnaire, which is typically administered together with the post-harvest one.[19] The seasonality of farming and non-farming activities may induce measurement error in these records of time use and their relationship with value added and profits across sectors. Notice however that this would be problematic for our empirical analysis only insofar as such measurement error correlates systematically with the variables of interest, a possibility that we discuss in detail in Section 9. Moreover, 77% (86%) of households for which we can derive value added in agriculture (profits from non-farm entrepreneurship) report a positive number of hours worked in that sector in the last 7 days. Finally, as we show later, 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.

Measuring Absolute Advantage We measure absolute advantage in agriculture using value added and value added per hour. Similarly, we measure absolute advantage in non-farm entrepreneurship using profits and profits per hour. We use two measures per sector to correct for the fact that hours choices naturally differ systematically between households active in both sectors and specialized households. Given a production function that is increasing and strictly concave in hours, households that specialize in a sector invest more hours, making a comparison based on total value added overstate their absolute advantage relative to households engaged in both sectors. Conversely, using value added per hour tends to understate their absolute advantage. Taken together, these measures provide bounds on the absolute advantage of specialized households relative to those active in both sectors.

[18] de Magalhaes and Santaeulalia-Llopis (2018) also discuss other measurement issues such as income underreporting and seasonality of reported consumption. Upon investigating the former, they conclude that this is not a major issue in the LSMS-ISA data as for instance the reported agricultural production and the reported annualized self-farmed consumption yield very similar quantities. They further conclude that seasonality affects measures of consumption—for which data are collected with recalls of the past 3 months, past month, or even past week—but not income—for which the recall period is 1 year. Seasonality may still matter for information on hours worked, an issue we discuss below and in Section 9.

[19] See Appendix C for detailed information on the timing of such questionnaire in each country and wave.

To make these measures comparable across countries and waves, we compute for each measure the percentile the household belongs to in the corresponding country-wave distribution. These are our preferred measures of absolute advantage. They are comparable across countries and waves, even in the presence of differences in currency used and inflation rates over time.[20] For robustness, we also conduct our analysis directly using value added and value added per hour as measures of absolute advantage, coupled with country-wave fixed effects. Finally, we also estimate the production function in both sectors and use the estimated productivity term as a measure of absolute advantage.

Measuring Comparative Advantage Section 2 shows that the activity of each household along the extensive margin can be 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 active in farming if we can derive information on value added in agriculture. Similarly, we say that a household is active in non-farm entrepreneurship if we can derive information on profits from that sector. Through the lens of the model, households that only do farming have a high comparative advantage in agriculture; households that only do non-farm entrepreneurship have a high comparative advantage in this sector; households that are active in both sectors have a weak comparative advantage in both sectors.

For those households that are active in more than one sector, 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. We can thus use the ratio between 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, from which we derive the total area of cultivated land. The survey also asks whether each plot of land is owned vs. assigned by the 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 combine the available information in an asset index that counts the number of assets the household reports to have, which is specific to each country.[21] 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.

[20] Another advantage of using percentiles is that they are robust to the lack of information on hired labor: if the amount of hired labor on a household's farm increases with farming productivity, using value added per se would overestimate the level of absolute advantage in hiring households. But, this does not affect the corresponding ranking of households, leaving the percentile measure unaltered.

[21] For this reason, in our empirical analysis, we allow the correlations of this asset index with the variables of interest to vary flexibly across countries.

3.1 Summary Statistics

Table 1 shows the summary statistics for the variables we employ in the empirical analysis.[22] For each variable, the table reports the sample average, its estimated standard error, and the number of observations. It does so separately across three groups of households: those active in farming only, those active in non-farm entrepreneurship only, and those active in both sectors. The final dataset counts around 35,000 household observations across all countries and waves. Overall, 59% of households are active in farming only and 12% do only non-farm entrepreneurship. The remaining 30% of households in the sample are active in both sectors. This number is large in all countries, ranging from 24% in Ethiopia to 38% in Nigeria.

For Ethiopia and Malawi, household-run enterprises are further classified into industries. We can use this information to get a better sense of the kind of non-farming enterprises run by households in these countries. Among the most represented, 28% of household enterprises in Ethiopia provide a non-agricultural service from home or a household-owned shop (such as carwash, metal processing, mechanic, carpenter, tailor, barber, etc.); 25% process or sell agricultural by-products (flour, local beer, seed, etc., but excluding livestock by-products and fish); 15% of enterprises belong to the category of trading business on a street or market, while 12% offer services or sell anything on a street or market (including firewood, home-made charcoal, construction timber, woodpoles, traditional medicine, mats, bricks, cane furniture, weave baskets, thatch grass, etc.). These numbers are quite similar in Malawi, where 25% of household enterprises provide a non-agricultural service from home or a household-owned shop, 15% process or sell agricultural by-products, 29% are trading businesses, and 16% offer services or sell anything on a street or in a market.

Households that are active in both farming and non-farm entrepreneurship differ from the others along a number of characteristics. First, these households are significantly larger, with an average of 0.6 more members than households engaged solely in farming and 1.2 more members than those focused exclusively on entrepreneurship. Second, the total number of hours worked by all members combined is higher in households active in both sectors, totalling 90 hours per week, compared to 75 hours for entrepreneurship-only households and 66 hours for farming-only households. However, households participating in both activities allocate fewer hours to each individual activity than those exclusively dedicated to either farming or entrepreneurship.[23] As we show later, 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 households active in both sectors, 50% have at least one member reporting a positive number of hours worked in both sectors, while 23% have more than one member doing so. On average, one household member reports positive hours in both sectors. This indicates that, in general, there is limited specialization across household members. We discuss in Section 7 the extent to which our empirical results at the household level are informative of the correlation of advantages at the individual level.

While agricultural work and non-agricultural self-employment are the dominant forms of

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

[23] Table 1 also shows that households that report no profits from non-farm entrepreneurship—which we classify as active in farming only—still report an average of 4 hours a week in total of work in that sector. Similarly, households reporting no output in agriculture report positive hours worked in that sector on average.

work in the settings we analyze, some households do have members who contribute some income from non-agricultural wage work. 18% of households in the data have members who held a non-agricultural wage job in the year preceding the survey. This is especially prevalent among households without any farming activities, where the proportion rises to 41%. We will discuss what this implies for our analysis from the theoretical and empirical standpoint in Section 8.

Table 1 also shows that the size of cultivated land is significantly higher for households active in both sectors than for households active in farming only, and that only 10% of households active in non-farm entrepreneurship have land. The asset index value suggests that households in this last group have on average more assets than others.

Evidence so far shows that around one third of households in our sample are active in both farming and non-farm entrepreneurship. It also shows that significant differences exist between these households and those active only in one sector. In the analysis that follows, we probe the robustness of the results by including household characteristics as controls and changing the definitions of sectoral activity whenever appropriate.

4 Selection Along the Extensive Margin

Our first identification strategy investigates the sign of the correlation between absolute and comparative advantage in each sector using information on activity choices along the extensive margin.

Agriculture We begin by restricting the sample to farming households and assign each a productivity percentile from the national distribution. The top left graph in Figure 5 plots the fraction of farming households involved in non-farm entrepreneurship per bin of 5 percentiles of the national distribution of value added in agriculture, together with a linear fit. The bottom left graph uses percentiles of value added per hour instead. As discussed in Section 3, the two lines in the top and bottom left graphs in Figure 5 bound the true relationship between comparative and absolute advantage in agriculture. They indicate that those with stronger absolute advantage are more likely to be specialized, indicating stronger comparative advantage, and thus a positive correlation of advantages in agriculture.

The two right panels of Figure 5 mirror the left ones, after netting out location fixed effects.[24] They both show a clear positive relationship between absolute advantage in agriculture and the probability of non-agricultural entrepreneurship, implying a negative correlation of advantages. Contrasting this to the left panels implies that average differences across locations confound the relationship between agricultural value added and entrepreneurship at the household level. Entrepreneurship rates are systematically higher in locations with lower agricultural value added, indicating the importance of location-specific factors as discussed in Section 2.5.2 above. Comparing households across the entire distribution of agricultural value added is thus misleading unless average differences across locations are netted out. Doing so reveals that among households within locations, households with higher agricultural productivity are the ones who are

[24] We regress a dummy equal to one if the household is active in non-farm entrepreneurship over the full set of location (enumeration area) fixed effects, and plot the corresponding estimated residuals, averaged by bin.

more likely to also pursue entrepreneurship, suggesting that advantages are misaligned in agriculture.

To be able to account for further confounding factors, like wealth and land holdings, we implement the regression specification

$$Y_{igct} = \beta\, P_{ict} + \mathbf{X}'_{icgt}\gamma + \lambda_g + \delta_{ct} + \varepsilon_{igct}, \tag{21}$$

where Y_{igct} is the outcome of interest for household i surveyed in location g, country c, and wave t. P_{ict} is the percentile (divided by 10) the household belongs to in the distribution of absolute advantage in country c and wave t. In our first set of results discussed next, Y_{igct} is a dummy variable that equals 1 if a household is active in non-farm entrepreneurship, and P_{ict} is the percentile (divided by 10) the household belongs to in the distribution of absolute advantage in agriculture. Our coefficient of interest is β, which captures any systematic relationship between absolute advantage in agriculture and likelihood to engage in non-farm entrepreneurship. $\mathbf{X}_{icgt}$ is a vector of household-level characteristics. λ_g and δ_{ct} indicate location and country-wave fixed effects respectively. These capture and net out time-invariant location characteristics and differential trends in the likelihood of engaging in non-farm entrepreneurship between different countries and survey rounds. We allow the residual unobserved determinants of entrepreneurship ε_{igct} to be correlated among household-level observations that belong to the same location by clustering standard errors at the same level.

Panel A of Table 2 shows the corresponding coefficient estimates. In column 1, we implement a regression specification that includes the household's percentile in the distribution of value added in agriculture (divided by 10), $P(VA_a)$, as a regressor, conditioning on the full set of location fixed effects. In column 2, we instead use the percentile (divided by 10) in the distribution of value added per hour, $P(VA_a/h_a)$. Point estimates are consistent with the top and bottom right graphs in Figure 5. The point estimate of β is zero when considering agricultural value added, and positive and significant when considering value added per hour. Households that are higher in the distribution of agricultural value added per hour are more likely to engage in non-farm entrepreneurship.[25] In columns 3 and 4, we include as regressors the full set of country-wave fixed effects as well as a number of household-level characteristics.[26] According to the results in column 4, households in the top percentile of the distribution of agricultural value added per hour are 7 percentage points more likely to engage in non-farm entrepreneurship than households in the bottom percentile. The corresponding coefficient estimate is significant at the 1% level.

Non-farm entrepreneurship We next analyze households that are active in non-farm en-

[25]Table A.2 of Appendix A shows the coefficient estimates obtained without conditioning on location fixed effects, which are consistent with the top and bottom left graphs in Figure 5.

[26]These include: total number of hours worked by all household members, total number of household members, total number of female household members, country-specific asset index. The latter is obtained by allowing the coefficient of the asset index to vary flexibly across countries by including its interaction with the four country dummies and is meant to address the differences across countries in the way assets are recorded. The estimated coefficients of these control variables show that entrepreneurship rates are systematically higher among households that work more overall and have more female members. Entrepreneurship is also more likely among households that have more land and more assets, which is suggestive of the presence of fixed costs to start a non-farming enterprise combined with credit constraints. We address this possibility in the next Section.

trepreneurship. We derive the percentile they belong to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. We identify with a dummy equal to 1 those households that are also active in farming and implement the regression specification from equation (21) using the farming dummy as the dependent variable and the household's percentile in the distribution of profits from non-farm entrepreneurship as the main independent variable.[27] Columns 5 to 8 in Panel A of Table 2 show the corresponding coefficient estimates. Households that are higher in the distribution of profits from non-farm entrepreneurship or profit per hour are no more likely to engage in farming than households that are lower.

Summary The results so far suggest that absolute and comparative advantage are negatively correlated in agriculture, and uncorrelated in non-farm entrepreneurship. Importantly, differences across locations confound these correlations when estimated by comparing households across locations. Figures A.2 to A.5 in Appendix A 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 the possible mechanisms underlying this pattern. Yet, no matter what causes it, the correlation shown here is the one that determines the relationship between sectoral size and productivity. These findings cast doubt on the hypothesis that self-selection based on household unobservables plays a significant role in the low average agricultural productivity observed in poor countries.

4.1 Robustness

Alternative Definitions This section investigates the robustness of this first set of results. We start by employing alternative definitions of a household's activity along the extensive margin. First, instead of classifying households' activity using information on value added in agriculture and profits from entrepreneurship, we use information on hours worked. In Table A.4 in Appendix A, we report the coefficients we obtain when regressing a dummy equal to 1 if any household member reports any hours worked in any of the household non-farming enterprises on the household's percentile (divided by 10) in the distribution of value added in agriculture. Similarly, in Table A.5, we report the coefficient estimates from a regression of a dummy equal to 1 if any household member reports any hours worked in the household farm on the household's percentile in the distribution of profits from entrepreneurship. The results are very similar to those we report in Panel A of Table 2.

Second, we adopt a stricter definition and label a household as active in non-farm entrepreneurship (farming) only if we can retrieve information on business profits (value added in agriculture) and if it devotes at least 15% of the total hours worked by household members to that activity. Tables A.6 and A.7 in Appendix A report the coefficient estimates when using as dependent variable the corresponding dummy. Results are once again similar to the ones obtained at baseline.

[27] Figure A.1 of Appendix A mirrors Figure 5 and illustrates the relationship between these variables. The top and bottom left graphs 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. But, the top and bottom right graphs show that this relationship disappears when comparing individuals within locations. This is consistent with the estimates reported in columns 5 to 8 in Panel A of Table 2.

Third, we adopt a definition of farming activities that excludes livestock and related activities. We do so because live animals can also be considered assets and activities related to them are very different from the ones related to crops. We redefine value added in agriculture by excluding revenues and costs from growing and selling livestock and livestock products. This means that we also relabel households as active in farming if we can derive value added in agriculture excluding these activities. Tables A.8 and A.9 in Appendix A show the results when doing so, once again very similar to the ones in Panel A of Table 2.

Specialization Even if the household as a whole is active in both farming and non-farm entrepreneurship, it could still be the case that each household member is fully specialized in only one of these activities. We defer a fuller discussion of this issue to Section 7, but already investigate the extent to which our results could be driven by full specialization within households. In Table A.10 (A.11) in Appendix A, we show the results obtained when adopting an alternative definition for extensive margin activity. We implement the same regression as in equation (21) but use as dependent variable a dummy that equals 1 if any household member reports any hours worked in a household non-farming enterprise (farm) *and* if at least one household member reports hours worked in both activities. The results are quantitatively and qualitatively similar to the ones obtained before.

Hours Worked for Others Next, we take into account the possibility that households supply hours of work outside the household in the form of paid or unpaid work, temporary or not, apprenticeship, etc. The average number of total hours worked by household members outside their household is equal to 15 in our data, which is not negligible. Later in the paper, we will extend our theoretical framework to allow for the possibility of non-agricultural wage work and derive new empirical implications. For now, we assess the robustness of the extensive margin results by repeating the analysis separately for the subsample of households in which any member reports positive hours worked outside the household and the subsample in which this is not the case. Tables A.12 to A.15 in Appendix A show the results. No meaningful differences emerge with the results reported in Panel A of Table 2.

Alternative Measures of Absolute Advantages We have thus far measured absolute advantage using percentiles of value added and value added per hour in agriculture, and percentiles of profits and profits per hour in non-farm entrepreneurship. We can also use these variables directly as measures of absolute advantage. In order to reduce the sensitivity of results to extreme values, we apply a cube root transformation. This reduces the skewness of the distribution but, unlike the logarithm transformation, can also be applied to zero and negative values.[28] Table A.16 and A.17 in Appendix A show the corresponding coefficient estimates. These are consistent with those reported in Panel A of Table 2.

In addition, we can use information on revenues and hours worked to directly estimate z^a and z^n. Specifically, we regress the log value of agricultural production on the log of hours worked in that sector, together with the full set of location and wave fixed effects. We take the

[28] Specifically, we apply the cube root transformation sign $(x) \times |x|^{1/3}$. In order to ease the interpretation of coefficient estimates, we also normalize the transformed variable by its standard deviation.

residuals of this regression and derive the percentile the household belongs to in the corresponding country-wave distribution. We do the same in the other sector by using the log of annual sales of household-run non-farm enterprises and the log of hours worked in that sector.[29] We then use percentiles of these newly obtained measures of absolute advantage as the main regressor in equation (21). Table A.18 and A.19 in Appendix A show the corresponding coefficient estimates. These are consistent with those reported in Panel A of Table 2. The only exception is the coefficient reported in the last column of Table A.19, which suggests that comparative and absolute advantages are positively correlated in the entrepreneurship sector, a result we will return to later on.

Subsistence vs. Market Production To conclude, we check whether systematic differences exist between households doing only farming and those that also engage in entrepreneurship in the split of agricultural output between internal consumption and market production. The fraction of market revenues from agriculture over the total value of agricultural output ranges from 20% in Ethiopia to 37% in Uganda. This shows that the majority of agricultural output is consumed within the household. Importantly for our analysis, these numbers are not meaningfully different between the group of households that only do farming and the one of households that also engage in non-farm entrepreneurship. The fraction of market revenues from agriculture over the total value of agricultural output ranges from 22 to 35% for the first group, and from 13 to 38% for the second group.

To summarize, our results are robust to several alternative definitions of a household's activity and an alternative, estimated measure of absolute advantage. They hold not only in the entire sample but also separately for the subsamples of households without internal specialization, those who supply hours worked outside the household, and those who do not. Finally, they are unlikely to be driven by differences in market production across groups. All of these findings point to a misalignment of advantages in agriculture.

5 Selection Along the Intensive Margin

When interpreted through the lens of the theory in Section 2, the evidence indicates that advantages are misaligned in agriculture. Proposition 1 suggests that this could be generated by a joint distribution of abilities that is more dispersed in non-farm entrepreneurship than in agriculture, combined with a *strong* positive correlation of abilities across sectors. Yet, the same proposition implies aligned advantages in non-farm entrepreneurship, while the empirical evidence so far suggests none. This reflects the influence of additional factors in driving selection, particularly

[29]The following caveats apply to these newly obtained measures of absolute advantage. First, the previous measures were based on value added and profits, thus taking into account the costs associated with each activity. We are here using the value of agricultural production and sales respectively, since this allows us to retain all observations, including those households that have negative values for profits. Second, hours worked are endogenous to absolute advantage, biasing the estimated coefficient and thus the residual we derive from these regressions. In the case of agriculture, we address this bias by adopting a control function approach and including a third-degree polynomial of lagged production expenditures in the production function regression. We do not have comparable information for non-farming activities. Notice however that the bias induced by the endogeneity of hours worked does not affect the derived percentile measures insofar as it does not change the ranking of estimated absolute advantage across households.

sector-specific fixed costs. Indeed, we showed in Section 2.5 how the presence of a fixed cost to enter non-farm entrepreneurship can push the (observed) correlation of advantages in that sector from positive towards zero.

To gain insights into this issue, we implement our second identification strategy, which exploits the theoretical result in equation (6) and identifies the correlation of advantages using the relative hours supplied to each sector as a measure of comparative advantage. For households who engage in both farming and non-farm entrepreneurship, this measure is unaffected by fixed costs of entry. We thus restrict the sample to these households and test whether a systematic relationship between value added and relative labor supply in a sector exists.

We start again with agriculture. We implement the regression specification given in equation (21), with relative labor supply in agriculture—the ratio of total hours worked in agriculture over those in non-farm entrepreneurship—as the dependent variable. We start by including as the only regressor the household's percentile (divided by 10) in the distribution of agricultural value. Column 1 of Panel B of Table 2 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 and significant 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, highly significant, and much larger in absolute value than the coefficient in column 1. 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 a benchmark, the estimate in column 2 of Panel B of Table 2 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, or 4.5 hours.

In columns 3 and 4, we include the full set of household-level controls together with country-wave fixed effects. The relative supply of hours worked in agriculture is higher for households with more land, and for those that work more hours in total. We also find some evidence that the amount of hours worked in agriculture relative to non-farm entrepreneurship is lower for households with more assets. Perhaps more importantly, the estimates of our main coefficient of interest in columns 3 and 4 support the same conclusion as those in columns 1 and 2: Households that are more productive in agriculture supply relatively *fewer* hours in that sector, a sign of weak comparative advantage. We conclude that absolute and comparative advantages are negatively correlated in agriculture. This result suggests that the patterns we found in Section 4 for the agricultural sector cannot be entirely due to the presence of fixed costs, and must at least in part be due to selection on ability.

Proposition 1 implies that, in the absence of fixed costs, misalignment of advantages in agriculture implies alignment of advantages in entrepreneurship. Columns 5 to 8 of Panel B of Table 2 provide evidence of the latter.[30] We test whether there is a systematic relationship between relative labor supply to non-farm entrepreneurship and profits in that sector. In column 5, we

[30]Note that this is not directly implied by the findings in columns 1 to 4 Panel B of Table 2.

show coefficients from a regression of the relative labor supply in non-farm entrepreneurship—the ratio of total hours worked in non-farm entrepreneurship over those in agriculture—on the household's percentile (divided by 10) in the distribution of profits from entrepreneurship. The regression models and estimates in columns 6 to 8 are ordered as in columns 2 to 4. Once again, the ranking of the estimated coefficients is consistent with the bounding argument outlined above, with the estimate in columns 5 (and 8) being a lower bound for the true correlation between absolute and comparative advantage in entrepreneurship, and the estimate in column 5 (and 7) being an upper bound. Households with higher profits from non-farm 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. Taking again the time allocation of the average household active in both sectors as benchmark, the estimate in column 5 of Panel B of Table 2 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 points, or about 2 hours. We conclude that absolute and comparative advantage are positively correlated in non-farm entrepreneurship.

Abilities, Frictions, and Selection Summarizing, our analysis finds a negative correlation of advantages in agriculture at both the extensive and intensive margin, a positive correlation of advantages in non-farm entrepreneurship at the intensive margin, and no significant correlation of advantages in non-farm entrepreneurship at the extensive margin. The theory in Section 2 indicates a unique setting that is consistent with these observations. First, the patterns of alignment at the intensive margin indicate that the coefficient of variation is higher for the distribution of ability in non-farm entrepreneurship relative to agriculture, $CV\left(z_i^n\right) > CV\left(z_i^a\right)$, and that the correlation of abilities is high, i.e. $\rho\left(z_i^a, z_i^n\right) > \bar{\rho} = CV\left(z_i^a\right)/CV\left(z_i^n\right)$. Second, the absence of correlation of advantages at the extensive margin in non-farm entrepreneurship indicates the presence of fixed costs to enter that sector.

Robustness As in Section 4.1, we verify whether the results we obtain in this section are robust to alternative definitions of activity along the extensive margin. In Tables A.20 and A.21 in Appendix A, we restrict the sample to those households that report positive hours worked in both sectors, with no meaningful changes to the results. In Tables A.22 and A.23 we restrict the sample to households that devote at least 15% of the total hours worked by household members to each activity. In this case, the results are less conclusive because of the reduced variation in the dependent variable. This is not the case when we adopt a stricter definition of farming activities that excludes livestock and related activities. Results in Tables A.24 and A.25 are again very similar to those presented in Panel B of Table 2.

In Tables A.26 and A.27 in Appendix A, we restrict the sample to households that are not fully specialized, i.e. where at least one household member reports hours worked in both the household non-farming enterprise and the household farm. Coefficient estimates are similar to the ones reported in Panel B of Table 2. Finally, results in Tables A.28 to A.31 show that, with the exception of Table A.31, estimates are consistent with the baseline results discussed above both in the subsample of households with positive hours worked outside the household and in the subsample without.

6 Selection Over Time

We begin by reporting in Table 3 the fraction of households in each wave that are active exclusively in agriculture, exclusively in non-farm entrepreneurship, or in both sectors. The data reveal a notable increase in the proportion of households engaged in both sectors, rising from 26% in 2009 to 37% in 2016, a trend observed across all countries in our sample except Uganda.[31] The share of households involved solely in farming has declined in Malawi and Nigeria while remaining stable in Ethiopia and Uganda. Transitions between being active exclusively in one sector to the other are minimal, whereas shifts from farming (or entrepreneurship) to participation in both sectors—and vice versa—are more frequent, involving approximately 10% (2%) of households in the sample.[32]

In light of these non-trivial transition probabilities, we complement the cross-sectional analysis above with a systematic analysis of sectoral transitions. The key advantage of this approach lies in its ability to identify the position of switchers—households with the weakest comparative advantage—within the distribution of absolute advantage and estimate the correlation of advantages conditional on household fixed effects. These capture time-invariant unobserved differences in e.g. wealth and access to technologies beyond those captured by observables.

We implement a panel data regression analysis. We restrict the sample to households that in wave 1 are only active in farming and investigate their probability of being active in non-farm entrepreneurship through wave 3. We implement the following regression specification

$$\mathit{Entrep}_{igct} = \sum_{t=2}^{3} \beta_t \, \mathit{Wave}_t \times \mathit{Rank}_{ig} + \mathbf{X}'_{icgt}\gamma + \lambda_i + \delta_{ct} + \varepsilon_{igct}, \tag{22}$$

where Entrep_{igct} is a dummy variable that equals 1 if household i surveyed in location g, country c, and wave t is active in non-farm entrepreneurship. Wave_t is a wave dummy identifier. Rank_{ig} is defined according to where the household stands in the location-specific ranking of agricultural value added and agricultural value added per hour in the first wave of data. That is, Rank_{ig} is time-invariant and takes a value of 1 if household i is the most productive farming household in its location g in the first wave of the data, 2 if it is the second most productive, etc. $\mathbf{X}_{icgt}$ is a vector of household-level characteristics. λ_i and δ_{ct} capture household and country-wave fixed effects respectively, which allow to control for and net out both time-invariant household-level characteristics and country-specific time trends. As before, we allow the residual unobserved determinants of entrepreneurship ε_{igct} to be correlated among household-level observations that belong to the same location by clustering standard errors at the same level. The coefficient β_t captures whether the likelihood of taking up non-farm entrepreneurship in wave 2 or 3 is correlated with the household's absolute advantage in agriculture.

Table 4 reports the estimated coefficients across different regression specifications, from one

[31] Table A.32 provides the corresponding figures for each country and wave.

[32] Transition matrices detailing these changes between waves 1 and 2, as well as 2 and 3, can be found in Table A.33 in Appendix A.

that includes only household and wave fixed-effects to the fully saturated one. As in the previous analysis, 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 significantly more likely to take up non-farm 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. This pattern is remarkably consistent across countries, as indicated by the coefficient estimates reported in Table A.34 in Appendix A, and despite the fact that the time interval between waves is different across countries.

We can exploit the panel dimension of the data to also investigate the role played by changes in household composition. In survey waves other than the first, we can identify household members that were previously listed but moved out in the time between the previous and the current interview. We define for each household a dummy equal to 1 if any household member moved out since the last interview and, similarly to the analysis in Section 4, we regress it on the household's percentile in the distribution of value added (or value added per hour) in agriculture, and again on the household's percentile in the distribution of profits from entrepreneurship (or profits per hour). Tables A.35 and A.36 in Appendix A report the corresponding coefficient estimates. We find some evidence that the exit of members is systematically more likely to occur among more productive households. If these members were to migrate for work outside of agriculture, this piece of evidence would be once again consistent with misalignment of advantages in that sector. However, coefficient estimates are no longer statistically significant when we control for household characteristics and location fixed effects.

The evidence in this section is consistent with the one presented in Section 4 and 5. 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 (Hamory, Kleemans, Li, and Miguel 2020). 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.

7 Households vs. Individuals

Our findings indicate that comparative and absolute advantage are negatively correlated in the agricultural sector, and positively correlated in entrepreneurship. Given our unit of observation in the data, these findings apply to households, not individuals. This motivated our assumption that production and economic choices occur at the household level. In our theoretical framework, ability or productivity are household-level attributes or, alternatively, the attributes of a single household member who acts as manager and makes production decisions on behalf of all members. A natural question is whether our findings at the household level can also be informative of the correlation of advantages and abilities at the individual level. Specifically, we aim to rule out the possibility that comparative and absolute advantage in agriculture are negatively correlated at the household level but positively correlated at the individual level. The results on sector switchers from the previous section help us to address this concern.

To see this, consider an alternative model that endows each individual in a household with a vector of sector-specific ability or productivity. Households then choose an allocation of individuals and their working hours to activities. Abstracting from productivity interactions across individuals in a household, individuals will sort into activities based on their individual comparative advantage. In this model, a household will specialize in an activity if all its members have strong comparative advantage in that activity. It will engage in both activities either if household members have strong comparative advantage in different activities, or if one or more members have weak comparative advantage and therefore do not specialize.

This is illustrated in Figure 6, where now every dot represents an individual. In the figure, individuals with strong comparative advantage specialize in either agriculture or entrepreneurship, while those with intermediate comparative advantage pursue both. The lines of indifference are drawn to yield proportions of individuals engaged in each activity similar to the data.

A negative correlation of advantages in agriculture at the household level can come about in only two ways. The first one is shown in panel (a): if the correlation of advantages at the individual level is negative and entrepreneurial abilities are more dispersed, then the correlation at the household level is also negative if individuals within a household are relatively "similar." (Some illustrative households are labelled A, B, etc.) Only then is it the case that the households with the lowest agricultural productivity are more likely to specialize in farming. This situation is akin to multi-dimensional positive assortative matching as defined in Lindenlaub (2017).[33] In this case, individual-level productivity ranks are similar to household ones.

The second possibility is shown in panel (b): if the correlation of advantages at the individual level is positive, the correlation at the household level is negative if individuals within a household are relatively "different" (multi-dimensional negative assortative matching). Only then is it the case that the households with the highest agricultural productivity are active in both sectors (e.g. household A), while those with low agricultural productivity are active only in agriculture (e.g. household C).[34]

Importantly, the two settings yield different predictions regarding which households will first enter entrepreneurship as the threshold determining selection across sector changes over time. Consider a decline in κ, driven by any factor that makes agricultural work relatively less attractive. This change makes the indifference lines pivot counterclockwise. In panel (a), it is clear that this prompts the most productive farmers to take up entrepreneurship. When individuals match with similar individuals, these farmers will come from households with high agricultural productivity. In panel (b), in contrast, it is clear that the least productive specialized farmer will switch.

In the previous section, we demonstrated that households having higher agricultural productivity at baseline are significantly more likely to take up non-farm entrepreneurship later on. Based on the reasoning outlined here, this suggests that the likely scenario corresponds to panel

[33] Multi-dimensional sorting problems are very challenging and the literature studying them is in its infancy. Therefore we do not study a full model in this section, but resort to a graphical representation.

[34] This is similar if the correlation of advantages at the individual level is negative and the dispersion of agricultural abilities is larger.

(a) of Figure 6. This suggests that the correlation of advantages observed at the household level is driven by a similar correlation at the individual level.

8 Selection Into Wage Work

As mentioned before, about one fifth of all households in our data, and about 40% of those not doing any farming, have members who held a non-agricultural wage job in the year preceding the survey. This raises the question of whether the presence of wage work affects our finding—that advantages are misaligned in agriculture. To investigate this, we extend our theoretical framework accordingly, derive its empirical implications, and investigate the correlation of advantages empirically in this more general setting.

Consider the model introduced in Section 2, but with a third activity option: wage work in the non-agricultural sector. Income from this activity is given by

$$y_i^w = z_i^w l_i^w = \omega \cdot (z_i^n)^\gamma \; l_i^w, \quad \gamma > 0, \tag{23}$$

where the household's return to an hour of wage work, z_i^w, consists of a common component ω and a household-specific component $(z_i^n)^\gamma$. For tractability, we assume that the latter is a function of the household's non-agricultural entrepreneurial ability. This assumption reflects the similarity in occupations between these two choices in the data.

This formalization nests three different cases. First, when $\gamma = 0$, all households face the same wage rate, so the distribution of abilities for wage employment is effectively degenerate. It follows from Proposition 1 that comparative (relative to wage work) and absolute advantage are positively correlated in both agriculture and non-farm entrepreneurship, i.e. $\rho\left(z_i^n/z_i^w, z_i^n\right) > 0$ and $\rho\left(z_i^a/z_i^w, z_i^a\right) > 0$. Second, when $\gamma < 1$, abilities for wage work are strongly positively correlated with those for non-agricultural entrepreneurship, but less dispersed. As a result, advantages are misaligned in wage work relative to non-farm entrepreneurship, i.e. $\rho\left(z_i^w/z_i^n, z_i^w\right) < 0$, and the most able entrepreneurs choose entrepreneurship. Finally, when $\gamma > 1$, abilities for wage work are strongly positively correlated with those in non-agricultural entrepreneurship, but more dispersed. In this case, advantages are misaligned in non-farm entrepreneurship relative to wage work, i.e. $\rho\left(z_i^n/z_i^w, z_i^n\right) < 0$, and the most able entrepreneurs choose wage work.

To distinguish between these three scenarios, we again employ our second identification strategy, which exploits choices at the intensive margin, and use the relative hours supplied to each sector as an empirical measure of pairwise comparative advantages. Consider households engaged in all three activities. These households split their time to equate marginal returns across these activities. Their optimal time allocations thus satisfy

$$z_i^a \kappa f'(\tilde{l}_i^a) = z_i^n g'(1 - \tilde{l}_i^a - \tilde{l}_i^w) = \omega \left(z_i^n\right)^\gamma. \tag{24}$$

For each pair of activities, the optimal ratio of hours is an increasing function of the relevant comparative advantage:

$$\begin{aligned}
\frac{\partial\left(\tilde{l}_i^a/\tilde{l}_i^n\right)}{\partial\left(z_i^a/z_i^n\right)} &= -\frac{\left(l_i^a+l_i^n\right)}{\left(l_i^n\right)^2}\frac{\kappa f'\left(l_i^a\right)}{\left(\frac{z_i^a}{z_i^n}\kappa f''\left(l_i^a\right)+g''\left(l_i^n\right)\right)} > 0,\\
\frac{\partial\left(\tilde{l}_i^a/\tilde{l}_i^w\right)}{\partial\left(z_i^a/z_i^w\right)} &= -\frac{\left(l_i^a+l_i^w\right)}{\left(l_i^w\right)^2}\frac{f'\left(l_i^a\right)}{\frac{z_i^a}{z_i^w}f''\left(l_i^a\right)} > 0,\\
\frac{\partial\left(\tilde{l}_i^n/\tilde{l}_i^w\right)}{\partial\left(z_i^n/z_i^w\right)} &= -\frac{\left(l_i^n+l_i^w\right)}{\left(l_i^w\right)^2}\frac{g'\left(l_i^n\right)}{\frac{z_i^n}{z_i^w}g''\left(l_i^n\right)} > 0,
\end{aligned} \tag{25}$$

where the first expression in (25) reduces to that in the baseline model, equation (6), when $l_i^w = 0$. See Appendix B.5 for a complete analysis of the model with wage work that parallels that of the baseline model.

Similarly to the analysis in Section 5, we then investigate, for each sector, the relationship between comparative advantage in that sector relative to wage work, proxied by relative hours, and absolute advantage in that sector. We first restrict the sample to households who are active in both farming and non-agricultural wage work, excluding those doing any non-farm entrepreneurship. We implement the regression specification given in equation (21), with the ratio of total hours worked in agriculture over those in non-agricultural wage work as the dependent variable. Columns 1 to 4 of Table 5 report the corresponding coefficient estimates. Among households engaged in both activities, households with higher agricultural value added per hour work significantly fewer hours in this sector relative to non-agricultural wage work, implying that advantages are misaligned in agriculture relative to wage employment. In columns 5 to 8, we restrict the sample to households who are active in both non-farm entrepreneurship and non-agricultural wage work, excluding those doing any farming. We find that households with higher profits from non-farm entrepreneurship work significantly more hours in this sector relative to non-agricultural wage work, implying that advantages are aligned in non-agricultural entrepreneurship relative to wage work.

Taken together, these results are consistent with the specification with $\gamma < 1$. Wage work appears to occupy an intermediate position, with the best farmers more likely to engage in wage work as well as non-agricultural entrepreneurship, but the best entrepreneurs specializing. This implies that including wage work in the analysis not only does not change our conclusions regarding the relationship between advantages in agriculture and non-agricultural entrepreneurship, but broadens our finding: advantages in agricultural work are misaligned with respect to both entrepreneurship and wage work outside agriculture.[35] Overall, our findings are consistent with a setting where all abilities are strongly positively correlated, but dispersion is smallest in agriculture and largest in entrepreneurship.

[35] An alternative modelling approach would have been to assume that the return to wage work is a function of agricultural ability, $z_i^w = (z_i^a)^\psi$. The finding of misaligned advantages in agriculture with respect to wage work then implies $\psi > 1$. The finding of misaligned advantages in both agriculture and wage work with respect to non-agricultural entrepreneurship again is consistent with strong positive correlation of z_i^a and z_i^n and greater dispersion of z_i^n than z_i^w and thus z_i^a. Hence, this alternative approach does not imply different conclusions.

9 Alternative Explanations

In this section, we explore several mechanisms other than selection on ability, grouped into a few distinct categories, and discuss to what extent they are consistent with the data.

Distortions Along the Intensive Margin The presence of constraints along the intensive margin may affect the allocation of hours worked across sectors within the household. For example, it could be the case that the effective marginal cost of agricultural inputs or capital is higher for some farming households and that this induces them to allocate more of their time to non-farm entrepreneurship. However, the results in Section 4 show that it is the most productive farming households who are systematically more likely to engage in non-farm entrepreneurship, both at the extensive and intensive margin. Constraints at the intensive margin would reduce input use and weaken the absolute advantage of these households—but not overturn it. Perhaps more importantly, all results are not sensitive to controlling for various household characteristics such as size of land, fraction of land rented, asset index, etc., which we would expect to correlate with constraints—or size-dependent distortions—to both agricultural and non-agricultural activities along both the extensive and intensive margin.

Diversification as Insurance The choice of the household may be driven by considerations other than joint profit maximization across activities. In particular, farming households may turn to non-farm entrepreneurship in response to negative shocks to agricultural output. This is consistent with the notion of *necessity entrepreneurs*, see De Giorgi and Di Falco (2018) among others. Yet, this appears once again inconsistent with our findings, as households affected by a negative shock and therefore turning to entrepreneurship should have lower agricultural value added. We find instead that entrepreneurship rates are higher among the most productive farming households.

Alternatively, households may choose *ex ante* to diversify, to reduce risk. For this to drive our finding that more productive farming households are more likely to engage in entrepreneurship, it would be necessary for these to be more risk averse, which is possible but seems implausible. Moreover, it is not clear how this would explain the intensive margin results showing that more productive farmers allocate more of their time to non-farm entrepreneurship.

Seasonality Returns to agricultural work have a strong seasonal component. During low agricultural season, households may allocate less of their time to agriculture and more of it to non-farm entrepreneurship. Our finding that more productive farmers allocate more of their time to non-farm entrepreneurship is based on information from the post-harvest questionnaire. Seasonality in agriculture can explain this result only insofar as, at the time of the interview, returns from agricultural work are differentially lower for more productive farmers, prompting them to increase the time they allocate to entrepreneurship. This could be the case if, for instance, more productive farmers grow a set of crops that require less post-harvest work. Yet, this does not explain why these farmers are also highly productive entrepreneurs unless abilities are strongly positively correlated across sectors.

Heterogeneous Fixed Costs One possible reason why entrepreneurship rates are higher among the most productive farming households is that they face lower costs to enter entrepreneurship. For this factor to drive our findings, these costs would need to be orthogonal to all household characteristics we control for. But then again, our analysis in Section 5 is robust to the presence of fixed entry costs.

Missing Land Market If land endowments were fixed and there was no way to sell or rent out land, households with a high comparative advantage in non-farm entrepreneurship would still use this land and thus remain active in farming. This could explain why these households are not systematically different in terms of profits from entrepreneurship. It would also be consistent with the evidence that 90% of the households in our sample who only pursue non-farm entrepreneurship report to have no land. Yet it cannot explain the observed negative correlation of advantages in agriculture, nor the allocation of hours among those pursuing both activities.

Suppose that, in addition, an exogenous production capacity constraint puts a strict upper bound on agricultural output. The most productive farming households hit such a constraint earlier and are pushed into non-farm entrepreneurship. This would be consistent with the results in columns 1 to 4 of Table 2 showing that entrepreneurship rates are not systematically different across households at different percentiles of the distribution of value added in agriculture while they increase systematically with the percentile the household belongs to in the distribution of agricultural value added per hour. Among households active in both activities, the same limit on output implies that the most productive farming households are left with more hours to allocate to non-farm entrepreneurship, thus have higher profits and —if in addition the production function in non-farm entrepreneurship is close to linear—no different profits per hour. This would be consistent with the results presented in Panel B of Table 2.

However, this scenario not only features very strong assumptions, but is also incompatible with some of the other empirical results. First, there is substantial variation in agricultural output, also among those active in both activities. This would require heterogeneity in the bound on agricultural output. Second, in Panel B of Table 2, the main coefficient in column 5 is very similar to the one in column 7. This indicates that, in a scenario where the relationship between value added in entrepreneurship and supply of labor to that sector is uniquely determined by the upper bound to agricultural production, the latter needs to be orthogonal to all other household characteristics that we include as controls, in particular assets. This is unlikely.

Finally, the mechanisms we are considering here still do not explain why, among households active in both sectors, absolute advantages across sectors are significantly positively correlated. This is empirically true using the various measures considered thus far, i.e. agricultural value added and value added per hour, profits from non-farm entrepreneurship and profits per hour, and estimated absolute advantages. In other words, it does not explain why those same highly productive farming households that hit the production capacity constraint earlier and do more non-farm entrepreneurship are also highly productive entrepreneurs. This finding is instead fully consistent with our main interpretation of results and the selection mechanism we propose.

10 Conclusions

Agricultural productivity is lower than that of other sectors. The agricultural productivity gap is particularly large in poor countries. A recent influential literature argues that an important source of this difference is worker self-selection. This mechanism relies on a positive correlation of comparative advantage and absolute advantage in the agricultural sector. We test this hypothesis using household-level data from Ethiopia, Malawi, Nigeria, and Uganda. Our empirical analysis delivers four sets of results. First, around one third of households engage in both agriculture and non-farm entrepreneurship. Second, those households active in both sectors have systematically higher agricultural productivity than those doing only farming. Third, 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. Fourth, over time, households starting a non-farming enterprise have higher baseline agricultural productivity than those who remain only farmers. These findings remain robust across a wide range of checks and, most importantly, hold true when examined within a model framework that incorporates non-agricultural wage work as a third activity option.

These results all imply that comparative and absolute advantages are misaligned in agriculture, casting doubt on the importance of worker self-selection as a root cause of the agricultural productivity gap. The literature suggests other possible explanations such as distortions to the land market (Adamopoulos and Restuccia 2014), or to the use of intermediate inputs (Donovan 2020). Yet, some of our results suggest that selection may still play a role, but along a different margin: land quality (Adamopoulos and Restuccia 2021). When comparing households across locations, the evidence in Section 4 shows that non-farm entrepreneurship rates are higher in places where agricultural productivity is lower. The reason for this could be differences in land quality. It might be the case that, as the agricultural sector shrinks, average agricultural productivity could increase not because the worst farmers switch to non-agriculture—as the worker self-selection story would argue—but because the worst agricultural land is converted to other uses or abandoned. Across countries, only the most productive land would be devoted to agriculture in rich countries, while in poor countries, less suitable land would also be used for farming. We are exploring this hypothesis in separate work.

References

Adão, R. (2016). Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil. Mimeo.

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia (2022). Misallocation, selection, and productivity: A quantitative analysis with panel data from china. *Econometrica 90*(3), 1261–1282.

Adamopoulos, T. and D. Restuccia (2014). The size distribution of farms and international productivity differences. *American Economic Review 104*(6), 1667–97.

Adamopoulos, T. and D. Restuccia (2021). Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data. *The Review of Economic Studies 89*(4), 1629–1653.

Alvarez, J. A. (2020). The Agricultural Wage Gap: Evidence from Brazilian Micro-data. *American Economic Journal: Macroeconomics 12*(1), 153–173.

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.

Caunedo, J. and E. Keller (2020, 12). Capital Obsolescence and Agricultural Productivity*. *The Quarterly Journal of Economics 136*(1), 505–561.

Cavalcanti, T. V., J. P. Kaboski, B. S. Martins, and C. Santos (2021). Dispersion in financing costs and development. *NBER Working Paper 28635*.

Chen, C., D. Restuccia, and R. Santaeulalia-Llopis (2022). Land Misallocation and Productivity. *American Economic Journal: Macroeconomics*. forthcoming.

De Giorgi, G. and S. Di Falco (2018). Farmers to Entrepreneurs. Mimeo.

de Magalhaes, L. and R. Santaeulalia-Llopis (2018). The consumption, income, and wealth of the poorest: An empirical analysis of economic inequality in rural and urban Sub-Saharan Africa for macroeconomists. *Journal of Development Economics 134*(C), 350–371.

Donovan, K. (2020). The Equilibrium Impact of Agricultural Risk on Intermediate Inputs and Aggregate Productivity. *The Review of Economic Studies 88*(5), 2275–2307.

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.

Hamory, J., M. Kleemans, N. Y. Li, and E. Miguel (2020, 11). Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata. *Journal of the European Economic Association 19*(3), 1522–1555.

Heckman, J. J. and B. E. Honoré (1990). The empirical content of the roy model. *Econometrica 58*(5), 1121–1149.

Heckman, J. J. and G. Sedlacek (1985). Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy 93*(6), 1077–1125.

Herrendorf, B. and T. Schoellman (2015). Why is Measured Productivity so Low in Agriculture? *Review of Economic Dynamics 18*(4), 1003–1022.

Herrendorf, B. and T. Schoellman (2018). Wages, Human Capital, and Barriers to Structural Transformation. *American Economic Journal: Macroeconomics 10*(2), 1–23.

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.

Lagakos, D., S. Marshall, A. M. Mobarak, C. Vernot, and M. E. Waugh (2020). Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*.

Lagakos, D., A. M. Mobarak, and M. E. Waugh (2021). The welfare effects of encouraging rural-urban migration. *Econometrica*. forthcoming.

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

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *The Review of Economic Studies 84*(2), 718–789.

Pulido, J. and T. Święcki (2019). Barriers to Mobility or Sorting? Sources and Aggregate Implications of Income Gaps Across Sectors and Locations in Indonesia. Mimeo.

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.

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 Entrepreneurship	Both	Full Sample
Observations	20622	4101	10375	35098
	59%	12%	30%	100%
Household Size	5.274	4.689	5.866	5.381
	(0.019)	(0.042)	(0.027)	(0.015)
	20537	4093	10360	34990
Female HH Members	2.105	1.942	2.108	2.087
	(0.012)	(0.027)	(0.020)	(0.010)
	20537	4093	10360	34990
Hours in Agriculture	47.282	4.141	36.566	39.068
h_a	(0.385)	(0.269)	(0.460)	(0.276)
	19851	3940	10175	33966
Hours in Entrepreneurship	18.540	70.744	53.083	34.944
h_n	(0.270)	(0.856)	(0.510)	(0.264)
	19851	3940	10175	33966
Total Hours	65.662	75.004	90.121	73.984
$h_a + h_n$	(0.501)	(0.904)	(0.730)	(0.384)
	20622	4101	10375	35098
Hours in Agriculture	59.195		52.263	48.406
$h_a > 0$	(0.434)		(0.563)	(0.320)
	15856		7119	27076
Hours in Entrepreneurship		76.407	63.943	25.028
$h_n > 0$		(0.858)	(0.543)	(0.249)
		3648	8447	32717
HH Members with			0.938	0.277
$h_a, h_n > 0$			(0.014)	(0.005)
			10360	35083
Female HH Members with			0.211	0.048
$h_a, h_n > 0$			(0.006)	(0.001)
			7284	32007
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	9.433	13.538	12.041	10.682
	(0.073)	(0.167)	(0.112)	(0.058)
	20530	4053	10354	34937

Notes. Unit of observation is the household 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. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farm entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farm 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-farm entrepreneurial profits.

Table 2: Selection Along the Extensive and Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A. Extensive Margin							
	Any Entrepreneurship				Any Farming			
$P(VA_a)$	0.001		-0.000					
	(0.001)		(0.001)					
$P(VA_a/h_a)$		0.003**		0.007***				
		(0.001)		(0.001)				
$P(VA_n)$					-0.001		-0.001	
					(0.002)		(0.001)	
$P(VA_n/h_n)$						0.001		0.002
						(0.001)		(0.001)
Hours in Agriculture			-0.004***					
			(0.000)					
Hours in Entrepreneurship							-0.002***	
							(0.000)	
Land Size (ha)			-0.000	-0.000				
			(0.002)	(0.002)				
Fraction Rented			0.010	0.031*				
			(0.013)	(0.016)				
Observations	30931	22890	27419	21486	14376	11963	13957	11909
R^2	0.247	0.247	0.337	0.292	0.515	0.539	0.572	0.570
	Panel B. Intensive Margin							
	h_a/h_n				h_n/h_a			
$P(VA_a)$	0.027*		-0.002					
	(0.015)		(0.018)					
$P(VA_a/h_a)$		-0.123***		-0.116***				
		(0.021)		(0.023)				
$P(VA_n)$					0.132***		0.120***	
					(0.035)		(0.034)	
$P(VA_n/h_n)$						-0.037		-0.046
						(0.029)		(0.030)
Land Size (ha)			0.011***	0.018***				
			(0.003)	(0.004)				
Fraction Rented			-0.211	-0.225				
			(0.240)	(0.347)				
Observations	8268	5702	7118	5236	6911	5702	6899	5691
R^2	0.336	0.355	0.348	0.363	0.274	0.265	0.286	0.280
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country $\times$ Wave FE	No	No	Yes	Yes	No	No	Yes	Yes
Controls	No	No	Yes	Yes	No	No	Yes	Yes

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. In Panel A, the dependent variable in columns 1 to 4 is a dummy equal to 1 if we can derive information on profits from non-farm entrepreneurship. The dependent variable in columns 5 to 8 is a dummy equal to 1 if we can derive information on value added in agriculture. In Panel B, the sample is restricted to those households for which we can derive information on both value added in agriculture and profits from non-farm entrepreneurship. The dependent variable in columns 1 to 4 is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. The dependent variable in columns 5 to 8 is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. Control variables include: total number of hours worked by all household members, total number of household members, total number of female household members, country-specific asset index.

Table 3: Activity Choice Over Time

	Only Agriculture	Only Entrepreneurship	Both	Full Sample
Wave 1	63.44%	10.88%	25.68%	100%
	7606	1304	3079	11989
Wave 2	61.36%	9.56%	29.08%	100%
	7228	1126	3425	11779
Wave 3	50.98%	15.35%	33.67%	100%
	4922	1482	3251	9655
Wave 4	51.64%	11.28%	37.07%	100%
	865	189	621	1675

Notes. Unit of observation is the household 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-farm entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farm 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-farm entrepreneurial profits.

Table 4: Transitions Into 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)
Hours in Agriculture			-0.002***	
			(0.000)	
Land Size (ha)			-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			0.006	0.011**
			(0.005)	(0.005)
Females			0.012*	0.002
			(0.006)	(0.007)
Household FE	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	n.a.	n.a.
Asset Index	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Observations	18723	14748	16511	13680
R^2	0.547	0.544	0.589	0.574

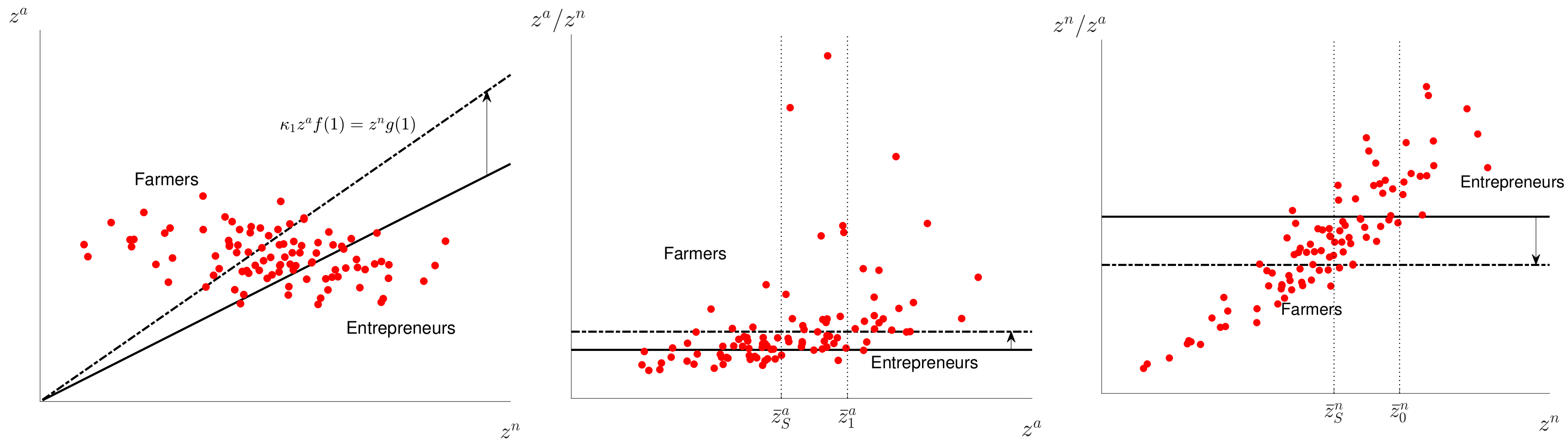
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted 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.

Table 5: Hours Relative to Non-Agricultural Wage Work

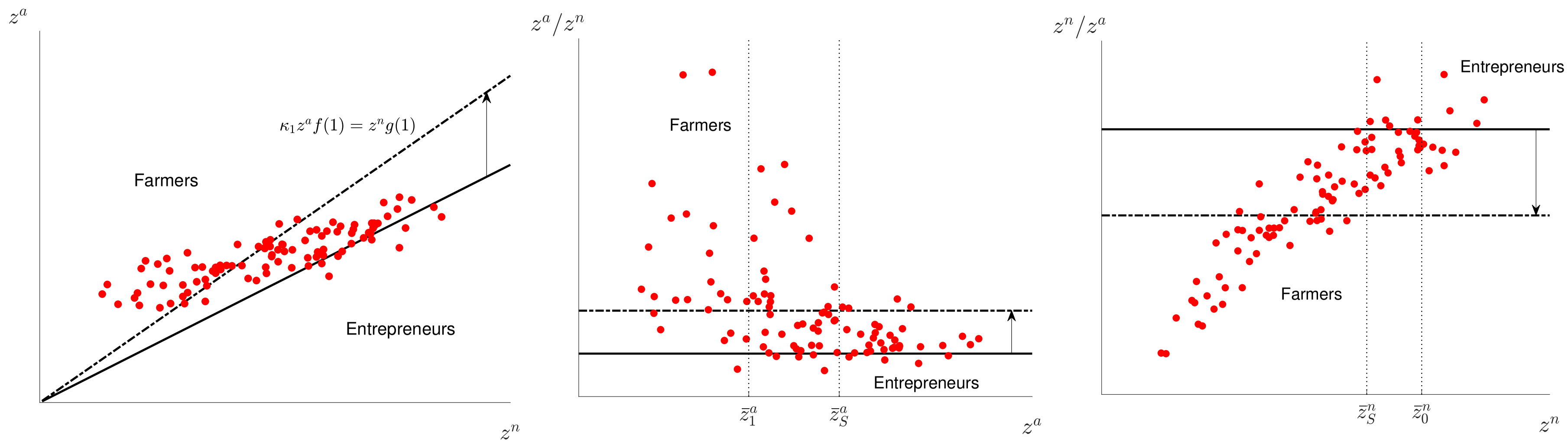
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		h_a/h_w				h_n/h_w		
$P(VA_a)$	0.032		0.018					
	(0.032)		(0.040)					
$P(VA_a/h_a)$		-0.110**		-0.133**				
		(0.050)		(0.063)				
$P(VA_n)$					0.044**		0.040**	
					(0.019)		(0.018)	
$P(VA_n/h_n)$						-0.012		-0.018
						(0.019)		(0.018)
Land Size (ha)			2.286	1.098				
			(1.984)	(1.886)				
Fraction Rented			-0.033	-0.416				
			(0.253)	(0.439)				
Observations	1809	987	1527	842	890	834	884	828
R^2	0.362	0.395	0.373	0.411	0.380	0.382	0.400	0.417
Village FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country $\times$ Wave FE	No	No	Yes	Yes	No	No	Yes	Yes
Controls	No	No	Yes	Yes	No	No	Yes	Yes

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The sample in columns 1 to 4 is restricted to those households for which we can derive information on value added in agriculture. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-agricultural wage work. The sample in columns 5 to 8 is restricted to those households for which we can derive information on profits from non-farm entrepreneurship. The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. non-agricultural wage work. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. Control variables include: total number of household members, total number of female household members, country-specific asset index.

Figure 1: Standard Roy model: Choice of a Single Sector



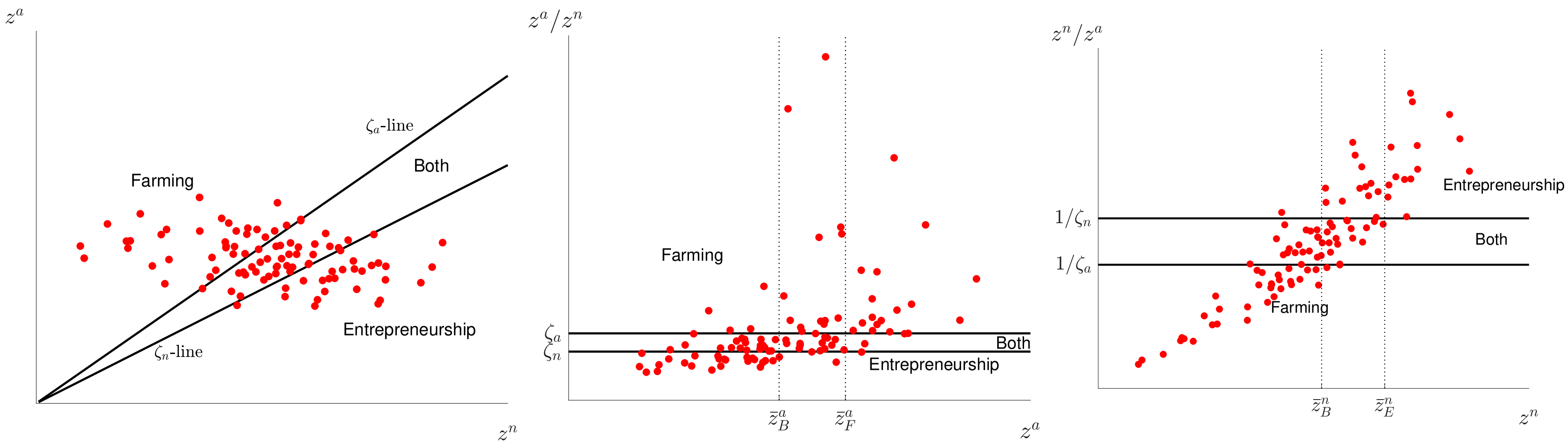
(a) Comparative and absolute advantages aligned in both sectors



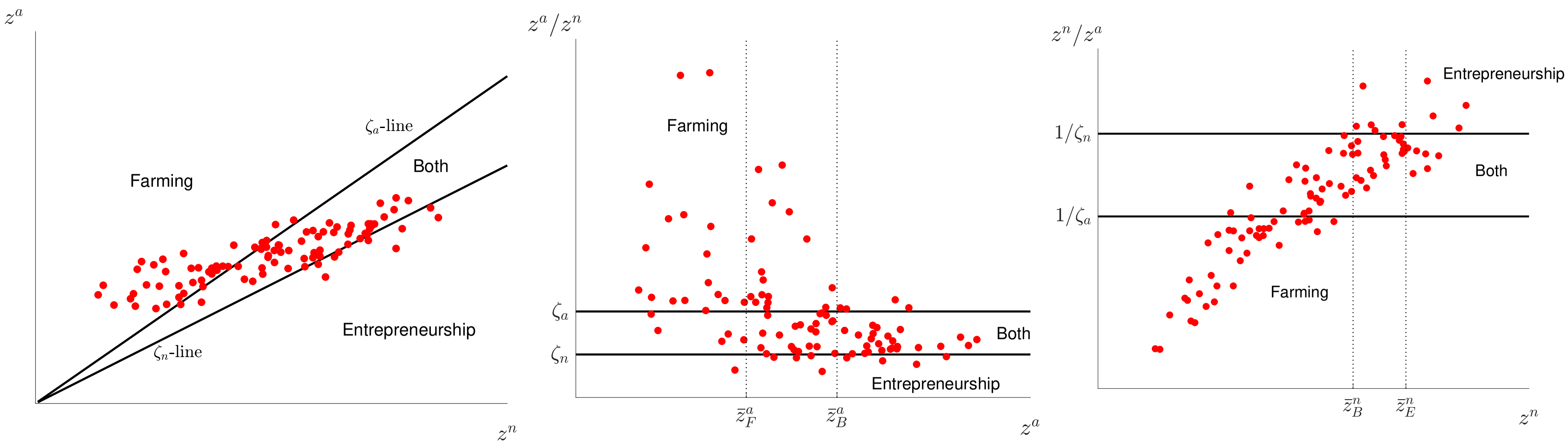
(b) Comparative and absolute advantages misaligned in agriculture

Notes. Simulated data. In both panels, the sample size is 100 and outliers are omitted. In both panels, $\mu_a = \mu_n = 1$, $\sigma_n = 1/3$, $\sigma_a = 1/6$. The correlation between z^a and z^n is -0.2 in panel (a) and 0.85 in panel (b).

Figure 2: Extended Roy Model: Continuous Choice of Hours



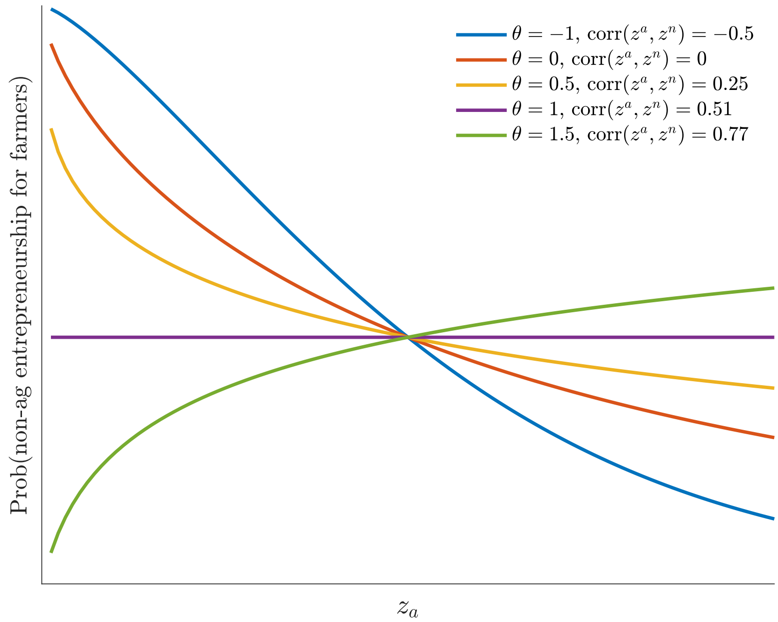
(a) Comparative and absolute advantages aligned in both sectors



(b) Comparative and absolute advantages misaligned in agriculture

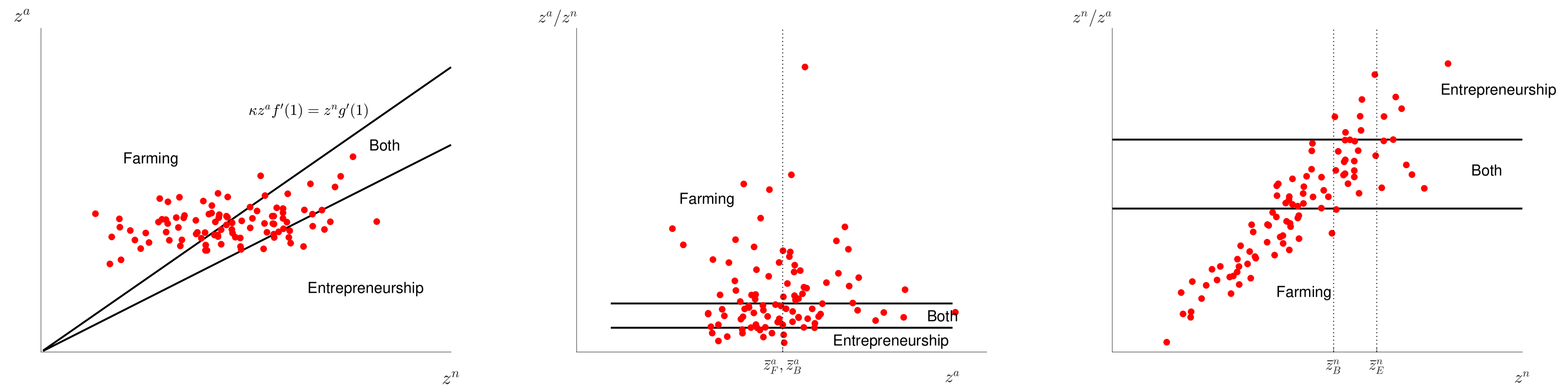
Notes. Same samples as in Figure 1.

Figure 3: Probability of Non-Farm Entrepreneurship for Farmers

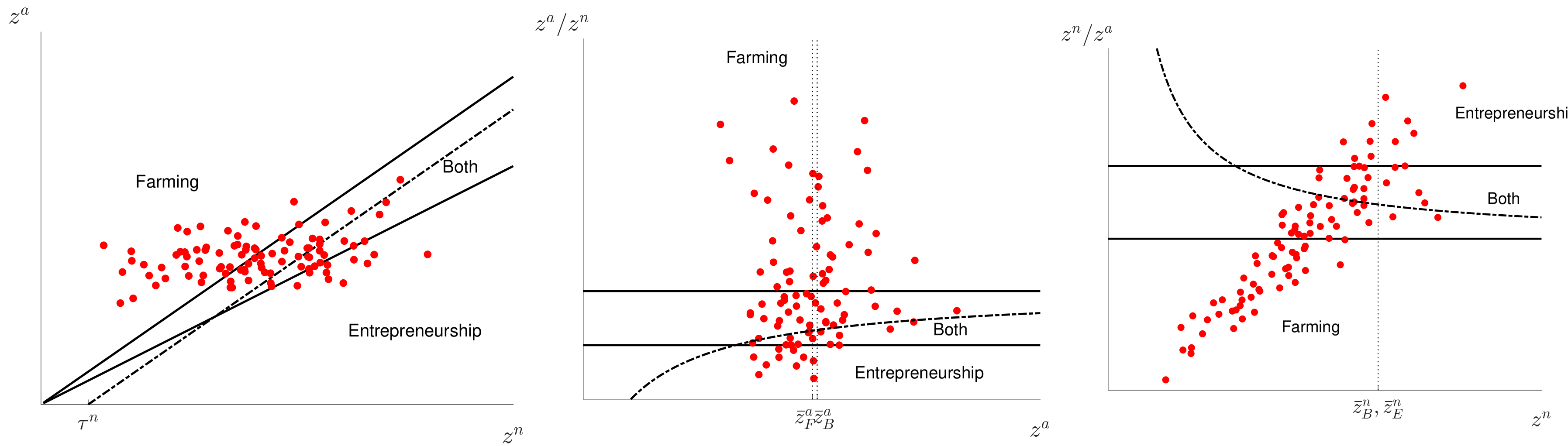


Note: Figure 3 shows the probability of non-agricultural entrepreneurship for farmers for jointly log-normal z_a and z_n, with $\tilde{\mu}_n = 1, \tilde{\mu}_a = 1.23, \tilde{\sigma}_n = 0.29$ and $\tilde{\sigma}_a = 0.15$, for varying values of θ and thus the correlation $\tilde{\rho}$. These figures imply coefficients of variation of .296 for z_n and .15 for z_a. Their ratio is 0.51.

Figure 4: Extended Roy Model: Effect of Sector-Specific Fixed Costs



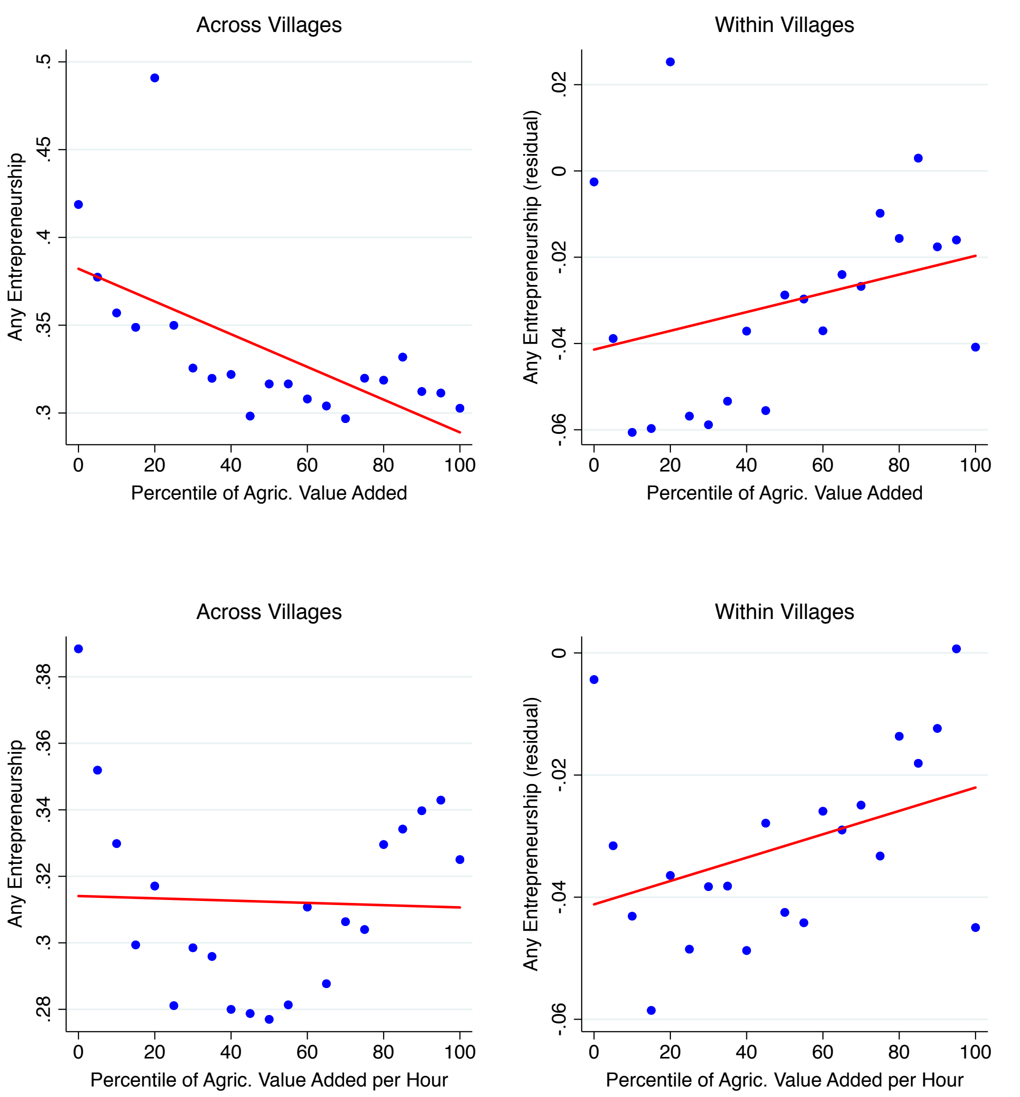
(a) Correlation of advantages with no fixed cost: unaligned in agriculture, aligned in non-agriculture



(b) Correlation of advantages with fixed cost in non-agriculture: misaligned in agriculture, unaligned in non-agriculture

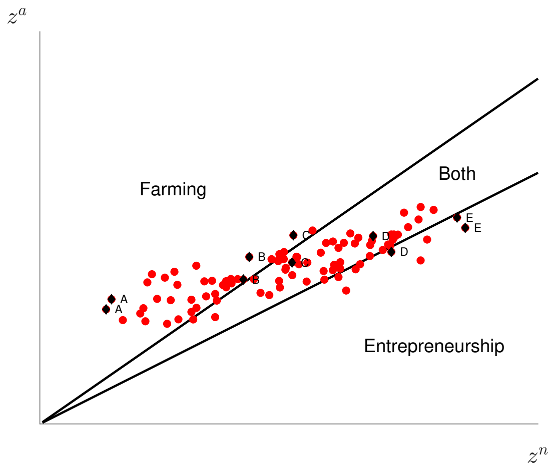
Notes. Simulated data. Same sample in both panels. The sample size is 100 and outliers are omitted. In both panels, $\mu_a = \mu_n = 1$, $\sigma_n = 1/3$, $\sigma_a = 1/6$. The correlation between z^a and z^n is $(\sigma_a/\mu_a)/(\sigma_n/\mu_n) = 1/2$.

Figure 5: Agricultural Value Added and Entrepreneurship

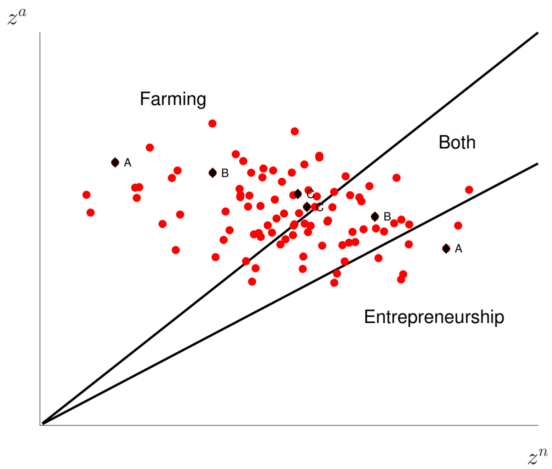


Notes. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The top figures show the fraction of households involved in non-farm entrepreneurship per bin of 5 percentiles of the distribution of value added in agriculture as derived in each country and wave. The bottom figures show the same number per bin of 5 percentiles of the distribution of value added in agriculture per hour as derived in each country and wave. The right figures plot the averaged residuals of the probability of doing entrepreneurship after netting out location (enumeration area) fixed effects.

Figure 6: Choice of sector by households vs individuals



(a) Individual comparative and absolute advantages misaligned in agriculture



(b) Individual comparative and absolute advantages aligned in both sectors

Notes. Same samples as in Figure 1(a) (panel b) and Figure 1(b) (panel a).

Online Appendix

Selection, Absolute Advantage, and the Agricultural Productivity Gap

Francisco Alvarez-Cuadrado, Francesco Amodio and Markus Poschke*

This Appendix is organized as follows. Section A contains the additional tables and figures discussed in detail in the text. Section B provides the proofs and derivations of the the theoretical results. Section C contains further details on the data sources and variable derivations.

*Alvarez-Cuadrado: francisco.alvarez-cuadrado@mcgill.ca, McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada. Amodio: francesco.amodio@mcgill.ca, McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada. Poschke: markus.poschke@mcgill.ca , McGill University and CIREQ, Leacock Building, 855 Sherbrooke St. West, Montreal QC H3A 2T7, Canada.

A Additional Tables and Figures

Table A.1: Summary Statistics by Country

	Only Agriculture	Only Entrepreneurship	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	1421	5384
	73%	1%	26%	100%
Household Size	4.279	4.111	4.479	4.331
	(0.034)	(0.561)	(0.055)	(0.029)
	3934	27	1421	5382
Hours in Agriculture	21.522	2.333	12.336	19.000
	(0.517)	(1.539)	(0.637)	(0.417)
	3934	27	1421	5382
Hours in Entrepreneurship	1.725	49.741	30.763	9.633
	(0.151)	(7.436)	(0.916)	(0.322)
	3934	27	1421	5382
		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. Unit of observation is the household 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. Households doing only agriculture are those for which we can derive information on value added in agriculture, but not on profits from non-farm entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farm 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-farm entrepreneurial profits.

Table A.2: Agricultural Value Added and Entrepreneurship
Estimates Without and With Village FE

	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	30997	22975	27486	21573	30931	22890	27419	21486
R^2	0.003	0.000	0.179	0.080	0.247	0.247	0.337	0.293

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.3: Entrepreneurial Profits and Farming
Estimates Without and With Village FE

	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	12095	14057	12041	14376	11963	13957	11909
R^2	0.012	0.005	0.270	0.154	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, clustered at the location (enumeration area) level. Unit of observation is the household 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 (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.4: Agricultural Value Added and Entrepreneurship
Alternative Definition Based on Hours Worked

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	-0.006***		-0.003***		-0.000		-0.001	
	(0.002)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		0.001		0.005***		0.001		0.007***
		(0.002)		(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	30026	22975	27486	21573	29960	22890	27419	21486
R^2	0.001	0.000	0.559	0.452	0.469	0.533	0.622	0.575

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 any member of the household reports any hour worked in household business. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.5: Entrepreneurial Profits and Farming
Alternative Definition Based on Hours Worked

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.022***		-0.010***		-0.005***		-0.004***	
	(0.002)		(0.001)		(0.002)		(0.001)	
$P(VA_n/h_n)$		-0.017***		-0.008***		-0.006***		-0.000
		(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	14115	12095	14057	12041	14015	11963	13957	11909
R^2	0.017	0.009	0.410	0.152	0.447	0.466	0.597	0.518

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 any member of the household reports any hour worked in household farm. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.6: Agricultural Value Added and Entrepreneurship
Stricter Definition Based on Hours Worked

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	-0.011***		-0.006***		-0.001		-0.000	
	(0.001)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		0.002		0.005***		0.004***		0.008***
		(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	30997	22975	27486	21573	30931	22890	27419	21486
R^2	0.005	0.000	0.260	0.131	0.263	0.262	0.380	0.304

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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-farm entrepreneurship and the household as a whole reports that at least 15% of total hours worked are dedicated to the household business. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.7: Entrepreneurial Profits and Farming
Stricter Definition Based on Hours Worked

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.026***		-0.009***		-0.009***		-0.004***	
	(0.002)		(0.001)		(0.002)		(0.001)	
$P(VA_n/h_n)$		-0.014***		-0.005***		-0.003*		0.002
		(0.002)		(0.002)		(0.002)		(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	12095	14057	12041	14376	11963	13957	11909
R^2	0.022	0.006	0.407	0.123	0.422	0.442	0.575	0.486

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 and the household as a whole reports that at least 15% of total hours worked are dedicated to the household farm. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.8: Agricultural Value Added and Entrepreneurship
Excluding Livestock-related Activities

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	-0.006***		-0.006***		0.002**		0.000	
	(0.001)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		0.000		0.002		0.002**		0.006***
		(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	30615	22965	27486	21573	30549	22878	27419	21486
R^2	0.001	0.000	0.179	0.080	0.242	0.247	0.337	0.292

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) the household belongs to in the distribution of value added in agriculture (excluding livestock-related activities) as derived in each country and wave. $P(VA_a/h_a)$ is the percentile (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.9: Entrepreneurial Profits and Farming
Excluding Livestock-related Activities

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.019***		-0.011***		-0.002		-0.002	
	(0.002)		(0.002)		(0.002)		(0.001)	
$P(VA_n/h_n)$		-0.013***		-0.009***		0.000		0.001
		(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	12095	14057	12041	14376	11963	13957	11909
R^2	0.014	0.006	0.263	0.143	0.540	0.565	0.586	0.586

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 (excluding livestock-related activities). $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.10: Agricultural Value Added and Entrepreneurship
Non-specialized Households

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	0.003***		-0.000		0.006***		0.002*	
	(0.001)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		-0.000		0.003***		0.002		0.006***
		(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	30997	22975	27486	21573	30931	22890	27419	21486
R^2	0.000	0.000	0.164	0.136	0.187	0.242	0.267	0.282

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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-farm entrepreneurship and at least one household member reports hours worked in both the household farm and the household business. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.11: Entrepreneurial Profits and Farming
Non-specialized Households

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.012***		-0.008***		0.001		-0.001	
	(0.002)		(0.001)		(0.001)		(0.001)	
$P(VA_n/h_n)$		-0.013***		-0.005***		-0.002		0.003**
		(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	12095	14057	12041	14376	11963	13957	11909
R^2	0.005	0.006	0.283	0.163	0.396	0.440	0.483	0.483

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 and at least one household member reports hours worked in both the household farm and the household business. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.12: Agricultural Value Added and Entrepreneurship
Restricted Sample of Households with Any Hours Worked for Others

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	-0.010***		-0.006***		0.000		0.002	
	(0.002)		(0.002)		(0.002)		(0.002)	
$P(VA_a/h_a)$		0.001		0.000		0.004		0.005*
		(0.003)		(0.003)		(0.003)		(0.003)
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	6778	4445	5817	4105	6563	4174	5601	3827
R^2	0.004	0.000	0.157	0.085	0.318	0.368	0.417	0.416

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report any hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.13: Entrepreneurial Profits and Farming
Restricted Sample of Households with Any Hours Worked for Others

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.018***		-0.011***		-0.003		-0.003	
	(0.003)		(0.003)		(0.003)		(0.003)	
$P(VA_n/h_n)$		-0.015***		-0.012***		-0.004		-0.002
		(0.004)		(0.003)		(0.003)		(0.003)
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	3410	2892	3344	2883	3086	2571	3022	2562
R^2	0.011	0.008	0.328	0.194	0.616	0.628	0.663	0.652

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report any hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). 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 (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.14: Agricultural Value Added and Entrepreneurship
Restricted Sample of Households with No Hours Worked for Others

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	-0.010***		-0.007***		0.002		-0.001	
	(0.002)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		0.001		0.004**		0.002		0.006***
		(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	24219	18530	21669	17468	24150	18422	21591	17364
R^2	0.003	0.000	0.201	0.086	0.276	0.266	0.369	0.312

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report no hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.15: Entrepreneurial Profits and Farming
Restricted Sample of Households with No Hours Worked for Others

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	-0.017***		-0.010***		-0.002		-0.002	
	(0.002)		(0.002)		(0.002)		(0.001)	
$P(VA_n/h_n)$		-0.011***		-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	11066	9203	10713	9158	10919	9027	10562	8982
R^2	0.012	0.005	0.260	0.161	0.516	0.545	0.576	0.576

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report no hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). 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 (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.16: Agricultural Value Added and Entrepreneurship
No Percentile Transformation

	Any Entrepreneurship							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VA_a	0.012***		-0.014***		0.003		-0.002	
	(0.004)		(0.004)		(0.003)		(0.003)	
VA_a/h_a		0.030***		0.007*		0.011***		0.013***
		(0.004)		(0.004)		(0.004)		(0.004)
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	30997	22975	27486	21573	30931	22890	27419	21486
R^2	0.001	0.004	0.178	0.080	0.247	0.247	0.337	0.292

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report no hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farm entrepreneurship. VA_a is value added in agriculture and VA_a/h_a is value added per hour. A cube root transformation $\mathrm{sign}(x) \times |x|^{1/3}$ and standard deviation normalization is applied to both variables. 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.

Table A.17: Entrepreneurial Profits and Farming
No Percentile Transformation

	Any Farming							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VA_n	-0.018***		-0.025***		0.002		-0.002	
	(0.007)		(0.006)		(0.004)		(0.004)	
VA_n/h_n		0.002		-0.021***		0.005		0.004
		(0.007)		(0.006)		(0.004)		(0.004)
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	12095	14057	12041	14376	11963	13957	11909
R^2	0.002	0.000	0.268	0.154	0.515	0.540	0.572	0.570

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report no hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is a dummy equal to 1 if we can derive information on value added in agriculture. VA_n is profits from non-farm entrepreneurship and VA_n/h_n is profits per hour. A cube root transformation $\mathrm{sign}(x) \times |x|^{1/3}$ and standard deviation normalization is applied to both variables. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Table A.18: Estimated Agricultural Advantage and Entrepreneurship

	Any Entrepreneurship			
	(1)	(2)	(3)	(4)
$P(\hat{z}^a)$	0.006***	0.004***	0.005***	0.003**
	(0.002)	(0.002)	(0.001)	(0.002)
Controls	No	Yes	No	Yes
Country-Wave FE	No	Yes	No	Yes
Village FE	No	No	Yes	Yes
Observations	12080	11587	12080	11563
R^2	0.002	0.066	0.295	0.319

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report no hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is a dummy equal to 1 if we can derive information on profits from non-farm entrepreneurship. $P(\hat{z}^a)$ is the percentile (divided by 10) the household belongs to in the distribution of estimated agricultural productivity as derived in each country and wave. $\hat{z}^a$ is estimated as the residual from a regression of the log of value of agricultural production over the log of hours worked in that sector, together with the full set of location and wave fixed effects. Control variables include: cultivated area, fraction of land that is rented, country-specific asset index.

Table A.19: Estimated Entrepreneurial Advantage and Farming

	Any Farming			
	(1)	(2)	(3)	(4)
$P(\hat{z}^n)$	-0.001	-0.000	-0.002	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Controls	No	Yes	No	Yes
Country-Wave FE	No	Yes	No	Yes
Village FE	No	No	Yes	Yes
Observations	12791	12738	12791	12738
R^2	0.000	0.064	0.518	0.532

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report no hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is a dummy equal to 1 if we can derive information on value added in agriculture. $P(\hat{z}^a)$ is the percentile (divided by 10) the household belongs to in the distribution of estimated productivity in non-farm entrepreneurship as derived in each country and wave. $\hat{z}^n$ is estimated as the residual from a regression of the log of value of sales associated to the household-run enterprise over the log of hours worked in non-farm entrepreneurship, together with the full set of location and wave fixed effects. Control variables include the country-specific asset index.

Table A.20: Agricultural Value Added and Time Allocation
Alternative Definition Based on Hours Worked

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.009		-0.008	
	(0.022)		(0.025)	
$P(VA_a/h_a)$		-0.123***		-0.116***
		(0.021)		(0.023)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	5702	5702	5236	5236
R^2	0.349	0.355	0.359	0.363

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.21: Entrepreneurial Profits and Time Allocation
Alternative Definition Based on Hours Worked

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.149***		0.150***	
	(0.041)		(0.044)	
$P(VA_n/h_n)$		-0.037		-0.047
		(0.029)		(0.032)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	5702	5702	5236	5236
R^2	0.269	0.265	0.261	0.257

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business. The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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.

Table A.22: Agricultural Value Added and Time Allocation
Stricter Definition Based on Hours Worked

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.011**		0.009*	
	(0.005)		(0.005)	
$P(VA_a/h_a)$		-0.041***		-0.040***
		(0.005)		(0.006)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	4996.000	4996.000	4581.000	4581.000
R^2	0.389	0.399	0.400	0.409

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and devote at least 15% of their total hours worked to each activity. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.23: Entrepreneurial Profits and Time Allocation
Stricter Definition Based on Hours Worked

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.030***		0.034***	
	(0.007)		(0.008)	
$P(VA_n/h_n)$		-0.024***		-0.029***
		(0.007)		(0.007)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	4996	4996	4581	4581
R^2	0.310	0.308	0.333	0.332

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and devote at least 15% of their total hours worked to each activity. The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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.

Table A.24: Agricultural Value Added and Time Allocation
Excluding Livestock-related Activities

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.006		-0.013	
	(0.016)		(0.019)	
$P(VA_a/h_a)$		-0.127***		-0.119***
		(0.022)		(0.023)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	7516	5523	6946	5129
R^2	0.334	0.355	0.348	0.364

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 can derive information on both value added in agriculture (excluding livestock-related activities) and profits from non-farm entrepreneurship. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) the household belongs to in the distribution of value added in agriculture (still excluding livestock-related activities) as derived in each country and wave. $P(VA_a/h_a)$ is the percentile (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.25: Entrepreneurial Profits and Time Allocation
Excluding Livestock-related Activities

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.148***		0.142***	
	(0.036)		(0.037)	
$P(VA_n/h_n)$		-0.024		-0.036
		(0.029)		(0.032)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	6692	5530	6267	5129
R^2	0.269	0.260	0.265	0.257

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household 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 can derive information on both value added in agriculture (excluding livestock-related activities) and profits from non-farm entrepreneurship. The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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.

Table A.26: Agricultural Value Added and Time Allocation
Non-specialized Households

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.014		-0.000	
	(0.024)		(0.027)	
$P(VA_a/h_a)$		-0.106***		-0.102***
		(0.021)		(0.025)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	4606	4606	4207	4207
R^2	0.362	0.366	0.370	0.373

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and having at least one household member reporting hours worked in both. The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.27: Entrepreneurial Profits and Time Allocation
Non-specialized Households

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.161***		0.161***	
	(0.051)		(0.056)	
$P(VA_n/h_n)$		-0.031		-0.041
		(0.034)		(0.040)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	4606	4606	4207	4207
R^2	0.255	0.251	0.266	0.263

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and having at least one household member reporting hours worked in both. The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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.

Table A.28: Agricultural Value Added and Time Allocation
Restricted Sample of Households with Any Hours Worked for Others

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.004		-0.013	
	(0.018)		(0.027)	
$P(VA_a/h_a)$		-0.114**		-0.128**
		(0.044)		(0.052)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	1408	721	1084	638
R^2	0.506	0.519	0.521	0.544

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and also report any hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.29: Entrepreneurial Profits and Time Allocation
Restricted Sample of Households with Any Hours Worked for Others

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.098		0.144*	
	(0.089)		(0.080)	
$P(VA_n/h_n)$		-0.050		-0.089
		(0.107)		(0.108)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	943	721	846	638
R^2	0.526	0.505	0.560	0.547

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and also report any hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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.

Table A.30: Agricultural Value Added and Time Allocation
Restricted Sample of Households with No Hours Worked for Others

	h_a/h_n			
	(1)	(2)	(3)	(4)
$P(VA_a)$	0.040**		0.017	
	(0.016)		(0.017)	
$P(VA_a/h_a)$		-0.099***		-0.085***
		(0.019)		(0.020)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	6497	4629	5664	4262
R^2	0.358	0.384	0.377	0.395

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and do not report any hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is the ratio of total hours worked by the household in agriculture vs. non-farm entrepreneurship. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.31: Entrepreneurial Profits and Time Allocation
Restricted Sample of Households with No Hours Worked for Others

	h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.124***		0.131***	
	(0.036)		(0.037)	
$P(VA_n/h_n)$		-0.023		-0.039
		(0.028)		(0.029)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	5589	4629	5201	4262
R^2	0.292	0.299	0.288	0.299

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. Sample is restricted to those households that report positive hours worked in both the household farm and the household business and do not report any hours worked outside the household (paid or unpaid, temporary or not, apprenticeship). The dependent variable is the ratio of total hours worked by the household in non-farm entrepreneurship vs. agriculture. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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.

Table A.32: Activity Choice 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	71.18%	3.74%	25.09%	100%
	2457	129	866	3452
		Panel B. Malawi		
Wave 1	76.86%	.39%	22.75%	100%
	2176	11	644	2831
Wave 2	68.94%	.63%	30.43%	100%
	1760	16	777	2553
		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. Unit of observation is the household 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-farm entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farm 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-farm entrepreneurial profits.

Table A.33: Transition Matrices

Wave 1 to 2	Only Agriculture	Both	Only Entrepreneurship
Only Agriculture	52.64	10.97	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 Entrepreneurship
Only Agriculture	44.62	12.29	1.43
Both	6.67	20.15	2.04
Only Entrepreneurship	0.82	2.47	9.52

Notes. Unit of observation is the household 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-farm entrepreneurship. Households doing only entrepreneurship are those for which we can derive information on profits from non-farm 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-farm entrepreneurial profits.

Table A.34: Transitions Into 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	3694	3694	3512	2364	2364	2298
R^2	0.561	0.561	0.558	0.556	0.556	0.556
			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, clustered at the location (enumeration area) level. 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.

Table A.35: Agricultural Value Added and Migration

	Household Member Moved Out							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_a)$	0.005***		0.002*		0.005***		0.001	
	(0.002)		(0.001)		(0.001)		(0.001)	
$P(VA_a/h_a)$		0.004**		0.002*		0.003**		0.001
		(0.002)		(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	20229	15199	18096	14303	20153	15096	18026	14205
R^2	0.002	0.001	0.333	0.331	0.376	0.388	0.412	0.420

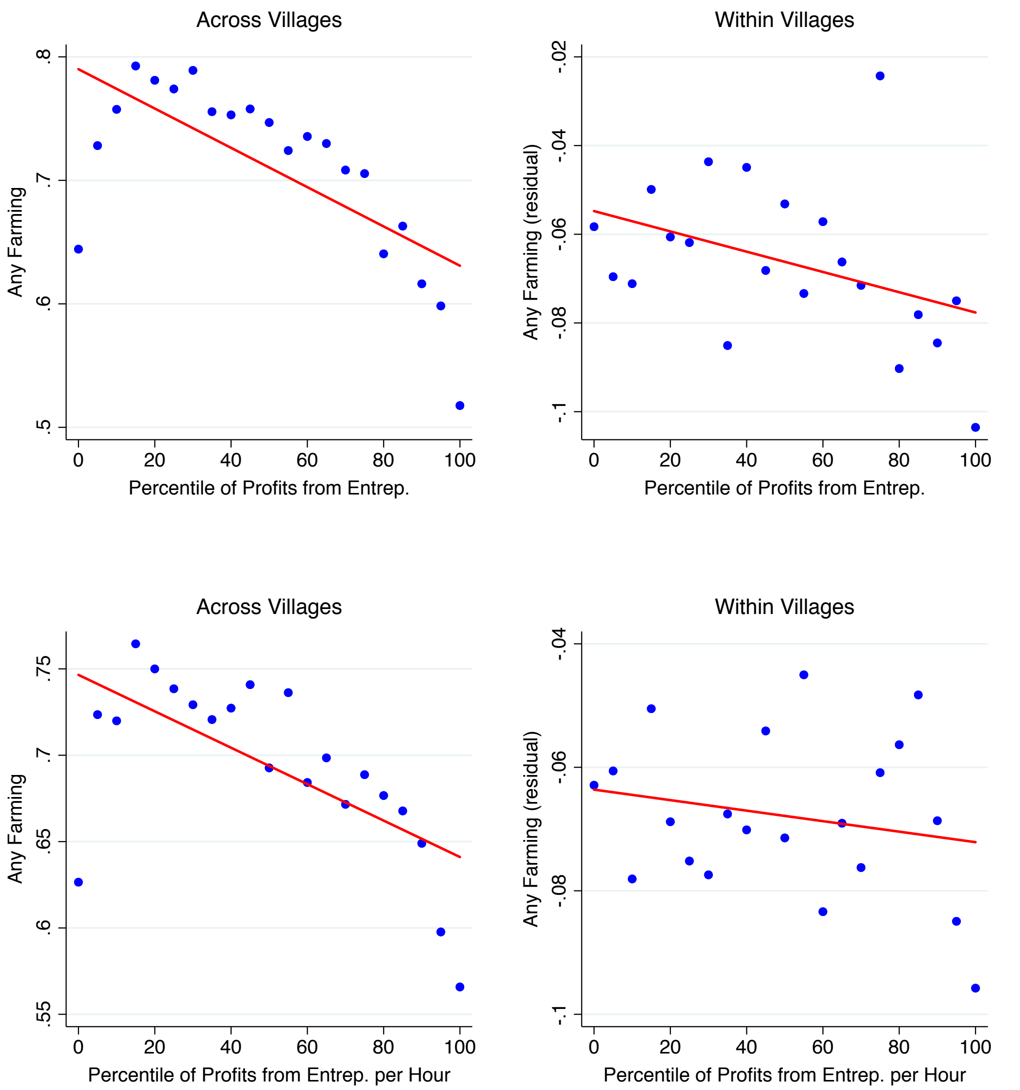
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each but the first wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if any household member moved out of the household since the last interview. $P(VA_a)$ is the percentile (divided by 10) 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 (divided by 10) the household belongs to in the distribution of value added per hour. 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.

Table A.36: Entrepreneurial Profits and Migration

	Household Member Moved Out							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$P(VA_n)$	0.000		-0.000		0.000		-0.001	
	(0.002)		(0.001)		(0.001)		(0.001)	
$P(VA_n/h_n)$		-0.000		0.000		-0.001		-0.001
		(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	10070	8430	9775	8378	9929	8254	9631	8202
R^2	0.000	0.000	0.332	0.303	0.432	0.422	0.459	0.445

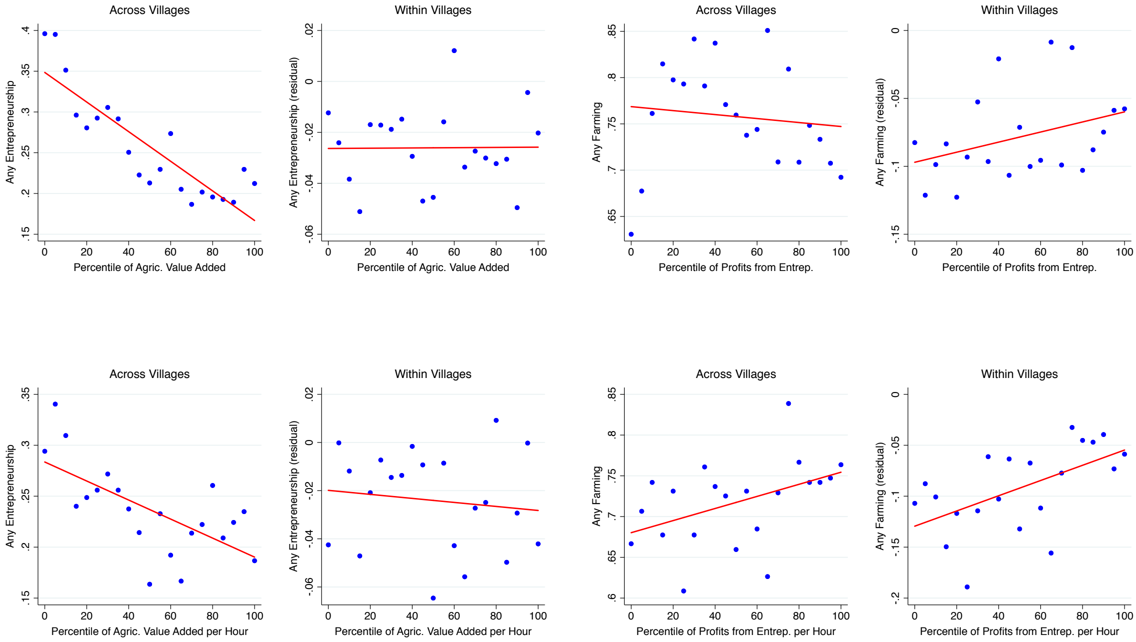
Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. Standard errors in parenthesis, clustered at the location (enumeration area) level. Unit of observation is the household surveyed in each but the first wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The dependent variable is a dummy equal to 1 if any household member moved out of the household since the last interview. $P(VA_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship as derived in each country and wave. $P(VA_n/h_n)$ is the percentile (divided by 10) the household belongs to in the distribution of profits from non-farm entrepreneurship per hour. 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-farm entrepreneurship (columns 3 and 7 only), country-specific asset index.

Figure A.1: Profits from Entrepreneurship and Farming



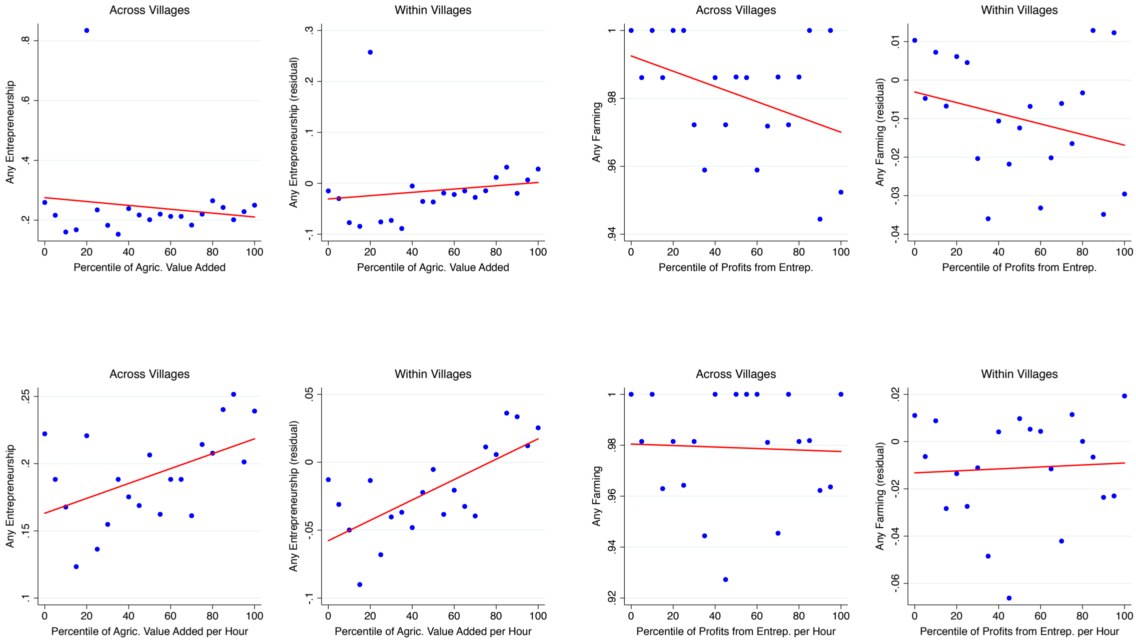
Notes. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia, Malawi, Nigeria, and Uganda. The top figures show the fraction of households involved in farming per bin of 5 percentiles of the distribution of profits from non-farm entrepreneurship as derived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

Figure A.2: Figures by Country - Ethiopia



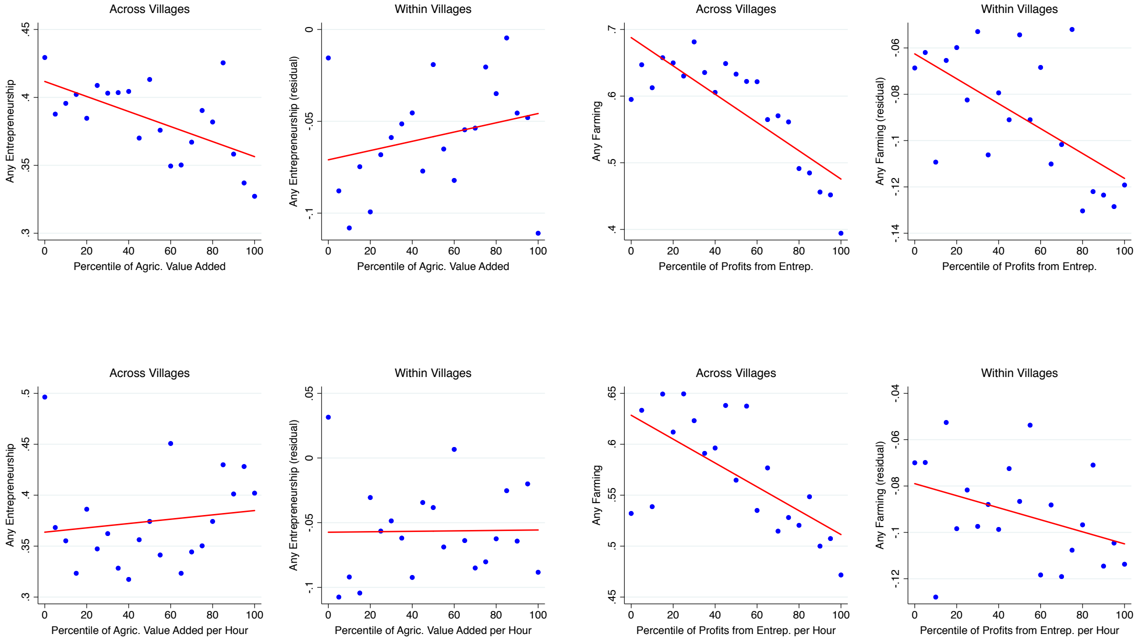
Notes. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Ethiopia. The top figures show the fraction of households involved in non-farm entrepreneurship (left) or farming (right) per bin of 5 percentiles of the distribution of value added in agriculture (left)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

Figure A.3: Figures by Country - Malawi



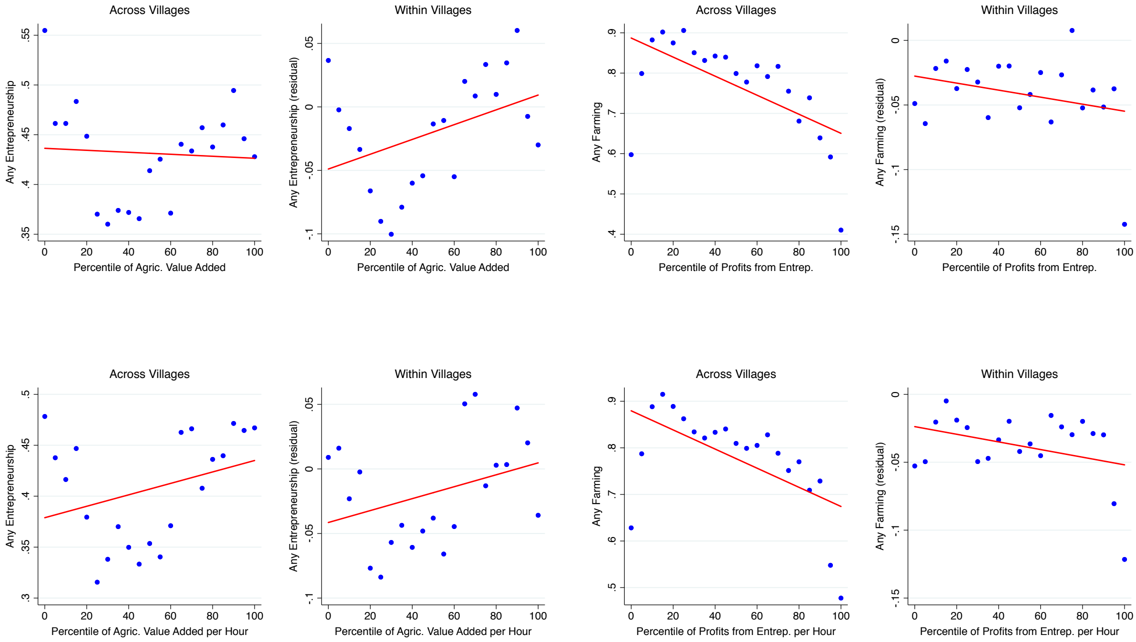
Notes. Unit of observation is the household surveyed in each wave of the LSMS-ISA panel dataset for Malawi. The top figures show the fraction of households involved in non-farm entrepreneurship (left) or farming (right) per bin of 5 percentiles of the distribution of value added in agriculture (left)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

Figure A.4: Figures by Country - Nigeria



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

Figure A.5: Figures by Country - Uganda



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

B Proofs and Derivations

B.1 Derivation of Condition (6)

The optimal allocation of hours under combined choice is implicitly defined by

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

Since $\tilde{l}_i^a/\tilde{l}_i^n = (1-\tilde{l}_i^n)/\tilde{l}_i^n$, the response of optimal relative hours to agricultural comparative advantage is given by

$$\frac{\partial(\tilde{l}_i^a/\tilde{l}_i^n)}{\partial(z_i^a/z_i^n)} = \frac{\partial((1-\tilde{l}_i^n)/\tilde{l}_i^n)}{\partial\tilde{l}_i^n}\frac{\partial\tilde{l}_i^n}{\partial(z_i^a/z_i^n)} = -\frac{1}{(l_i^n)^2}\frac{\kappa f'(l_i^a)}{\frac{z_i^a}{z_i^n}\kappa f''(l_i^a)+g''(1-l_i^a)} > 0. \tag{2}$$

where we obtain the second factor differentiating condition (1).

Similarly, the response of relative hours in entrepreneurship, $\tilde{l}_i^n/\tilde{l}_i^a$, to non-agricultural comparative advantage, z_i^n/z_i^a , is given by

$$\frac{\partial\left(\tilde{l}_i^n/\tilde{l}_i^a\right)}{\partial\left(z_i^n/z_i^a\right)} = -\frac{g'\left(1-l_i^a\right)}{(l_i^a)^2\left(\frac{z_i^n}{z_i^a}g''\left(1-l_i^a\right)+\kappa f''\left(l_i^a\right)\right)} = \frac{g'\left(1-l_i^a\right)}{\kappa f'\left(l_i^a\right)}\left(\frac{1-l_i^a}{l_i^a}\right)^2\frac{z_i^a}{z_i^n}\frac{\partial\left(\tilde{l}_i^a/\tilde{l}_i^n\right)}{\partial\left(z_i^a/z_i^n\right)} > 0.$$

B.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

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

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

$$
\begin{aligned}
Cov\left(\frac{z_i^a}{z_i^n}, z_i^a\right) &= \mathbb{E}\left(\frac{z_i^{a2}}{z_i^n}\right) - \mathbb{E}\left(\frac{z_i^a}{z_i^n}\right)\mu_a \\
&\approx \mathbb{E}\left(\begin{array}{c} \frac{\mu_a^2}{\mu_n} + \frac{2\mu_a}{\mu_n}\left(z_i^a - \mu_a\right) - \frac{\mu_a^2}{\mu_n^2}\left(z_i^n - \mu_n\right) + \\ \frac{1}{2}\left[\frac{2}{\mu_n}\left(z_i^a - \mu_a\right)^2 + 2\frac{\mu_a^2}{\mu_n^3}\left(z_i^n - \mu_n\right)^2 - 4\frac{\mu_a}{\mu_n^2}\left(z_i^a - \mu_a\right)\left(z_i^n - \mu_n\right)\right] \end{array}\right) \\
&\quad -\mathbb{E}\left(\begin{array}{c} \frac{\mu_a}{\mu_n} + \frac{1}{\mu_n}\left(z_i^a - \mu_a\right) - \frac{\mu_a}{\mu_n^2}\left(z_i^n - \mu_n\right) + \\ \frac{1}{2}\left[\frac{2\mu_a}{\mu_n^3}\left(z_i^n - \mu_n\right)^2 - \frac{2}{\mu_n^2}\left(z_i^a - \mu_a\right)\left(z_i^n - \mu_n\right)\right] \end{array}\right)\mu_a \\
&= \frac{\mu_a^2}{\mu_n} + \frac{\sigma_a^2}{\mu_n} + \frac{\mu_a^2\sigma_n^2}{\mu_n^3} - 2\frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right) - \left(\frac{\mu_a}{\mu_n} + \frac{\mu_a\sigma_n^2}{\mu_n^3} - \frac{1}{\mu_n^2}Cov\left(z_i^a, z_i^n\right)\right)\mu_a \\
&= \frac{\sigma_a^2}{\mu_n} - \frac{\mu_a}{\mu_n^2}Cov\left(z_i^a, z_i^n\right),
\end{aligned}
$$

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.

B.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

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

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.4 Details on Footnote 8

Misalignment in agriculture arises when greater z^a implies a greater probability that comparative advantage in farming, z^a/z^n, lies between ζ_a and ζ_n, rather than above ζ_a. In the joint log-normal case, this probability is

$$
\begin{aligned}
P &\equiv \frac{Prob(z^a/z^n \in (\zeta_n, \zeta_a)|z^a)}{Prob(z^a/z^n > \zeta_n|z^a)} &(3)\\
&= \frac{Prob(\exp(\ln z^a)/\exp(n_0 + \theta \ln z^a + u) \in (\zeta_n, \zeta_a)|z^a)}{Prob(\exp(\ln z^a)/\exp(n_0 + \theta \ln z^a + u) > \zeta_n|z^a)} &(4)\\
&= \frac{Prob(\exp((1-\theta)\ln z^a - n_0 - u) \in (\zeta_n, \zeta_a))}{Prob(\exp((1-\theta)\ln z^a - n_0 - u) > \zeta_n)} &(5)\\
&= \frac{\Phi((1-\theta)\ln z^a - n_0 - \ln \zeta_n) - \Phi((1-\theta)\ln z^a - n_0 - \ln \zeta_a)}{\Phi((1-\theta)\ln z^a - n_0 - \ln \zeta_n)}. &(6)
\end{aligned}
$$

Let

$$
\begin{aligned}
\tilde{\bar{z}} &= (1-\theta)\ln z^a - n_0 - \ln \zeta_n &(7)\\
\underline{\tilde{z}} &= (1-\theta)\ln z^a - n_0 - \ln \zeta_a. &(8)
\end{aligned}
$$

Then

$$
\begin{aligned}
\frac{\partial P}{\partial \ln z_a} &= \frac{\Phi(\tilde{\bar{z}})\left[\phi(\tilde{\bar{z}})(1-\theta) - \phi(\underline{\tilde{z}})(1-\theta)\right] - (\Phi(\tilde{\bar{z}}) - \Phi(\underline{\tilde{z}}))\phi(\tilde{\bar{z}})(1-\theta)}{\Phi(\tilde{\bar{z}})^2}\\
&= (1-\theta)\frac{\phi(\tilde{\bar{z}})}{\Phi(\tilde{\bar{z}})}\left[(1-P) - \frac{\phi(\underline{\tilde{z}})}{\phi(\tilde{\bar{z}})}\right]\\
&= (1-\theta)\frac{\phi(\tilde{\bar{z}})}{\Phi(\tilde{\bar{z}})}\left[\frac{\Phi(\underline{\tilde{z}})}{\Phi(\tilde{\bar{z}})} - \frac{\phi(\underline{\tilde{z}})}{\phi(\tilde{\bar{z}})}\right]\\
&= (\theta-1)\frac{\Phi(\underline{\tilde{z}})}{\Phi(\tilde{\bar{z}})}\left[\frac{\phi(\underline{\tilde{z}})}{\Phi(\underline{\tilde{z}})} - \frac{\phi(\tilde{\bar{z}})}{\Phi(\tilde{\bar{z}})}\right]. &(9)
\end{aligned}
$$

The term in square brackets is the difference of the Mill's ratio of u, evaluated at $\tilde{\bar{z}}$ vs $\underline{\tilde{z}}$, where by definition, $\tilde{\bar{z}} > \underline{\tilde{z}}$ since $\zeta_n < \zeta_a$. Defined this way, the Mill's ratio is monotonically decreasing in its argument. Hence, the term in square brackets is always positive. As a result, $\partial P/\partial \ln z_a$ is positive iff $\theta > 1$.

B.5 Model with Wage Work

The optimal time allocation for households engaged in the three activities is implicitly defined by condition (24), which we reproduce here for convenience:

$$z_i^a \kappa f'(\tilde{l}_i^a) = z_i^n g'(1 - \tilde{l}_i^a - \tilde{l}_i^w) = \omega\,(z_i^n)^\gamma .$$

We can use this condition to characterize employment choices. Households with high agricultural comparative advantage will engage in farming only. The threshold is given by

$$\frac{z_i^a}{z_i^n} \geq \frac{1}{\kappa f'(1)} \max\left[g'(0), \omega\,(z_i^n)^{\gamma-1}\right] \equiv \zeta_a.$$

Similarly, households with low agricultural comparative advantage will not do any farming. This threshold is given by

$$\frac{z_i^a}{z_i^n} \leq \frac{1}{\kappa f'(0)} \max\left[g'(1), \omega\,(z_i^n)^{\gamma-1}\right] \equiv \zeta_n.$$

Additionally, households with intermediate levels of agricultural comparative advantage, $z_i^a/z_i^n \in (\zeta_n, \zeta_a)$, will engage in agriculture and at least one of the non-agricultural activities.

Next, we need to characterize the choice between the two non-agricultural activities. This choice is determined by comparative advantage in entrepreneurship relative to wage work. Given our specification of wage income, this only depends on non-agricultural absolute advantage. The thresholds for this choice are given by

$$z_i^n = \left(\frac{\omega}{g'(0)}\right)^{\frac{1}{1-\gamma}} = \zeta_w$$

and

$$z_i^n = \left(\frac{\omega}{g'(1)}\right)^{\frac{1}{1-\gamma}} = \zeta_s.$$

When abilities in wage employment are non-degenerate but less dispersed than those in self-employment, which occurs when $0 < \gamma < 1$, then $\zeta_w < \zeta_s$, households with low non-agricultural ability (below ζ_w) do not engage in any self-employment, while those with high non-agricultural ability (above ζ_s) do not engage in any wage employment. Households with intermediate levels of non-agricultural ability, $z_i^n \in (\zeta_w, \zeta_s)$, will engage in both non-agricultural activities. This case is depicted in the middle panel of Figure A.6.

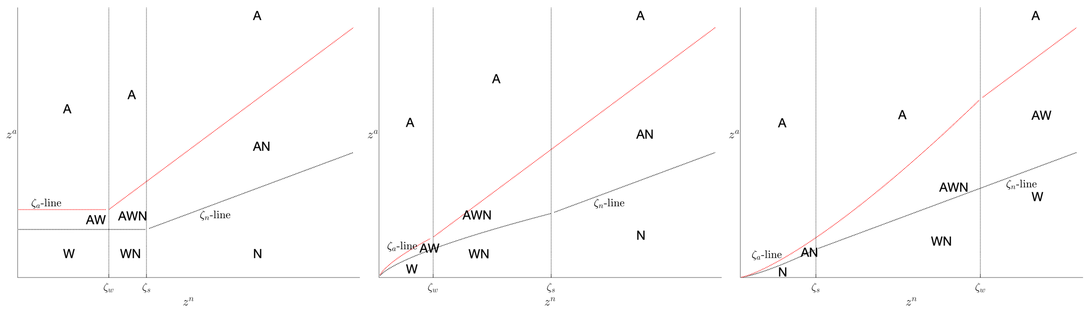


Figure A.6: A graphical representation of choices with three activities

The case where the ability for wage work is common across households, $\gamma = 0$, is depicted in the first panel. The results are very similar to the previous case, but the concave segments of the cone become horizontal, since wage employment does no longer depend on non-agricultural ability. Finally, the third panel of Figure 1 depicts the case where $\gamma > 1$. In this case, returns to wage employment are more dispersed than those in non-agricultural self-employment and therefore the ranking of the thresholds is reversed, i.e. $\zeta_s < \zeta_w$. As a result, households with low non-agricultural ability (below ζ_s) do not engage in any wage employment, while those with high non-agricultural ability (above ζ_w) do not engage in any non-agricultural entrepreneurship, just the opposite as in the case where $\gamma < 1$. As before, those with intermediate levels of non-agricultural ability engage in both non-agricultural activities.

C Data Appendix

Our main source of data is the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA).[1] 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. 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. For our purpose, we use data from the following countries, waves, and number of observations

- Ethiopia - Socioeconomic Survey (ESS)
  - Wave 1—Year 2011/12—$N = 3,969$
  - Wave 2—2013/14—$N = 3,804$
  - Wave 3—2015/16—$N = 3,726$
- Malawi - Third Integrated Household Survey (IHS3), Integrated Household Panel Survey (IHPS)
  - Wave 1—2010/11—$N = 3,247$
  - Wave 2—2013—$N = 2,872$
- Nigeria - General Household Survey (GHS)
  - Wave 1—2010/11—$N = 4,928$
  - Wave 2—2012/13—$N = 4,716$
  - Wave 3—2015/16—$N = 4,575$
- Uganda - National Panel Survey (UNPS)
  - Wave 1—2009/10—$N = 2,975$
  - Wave 2—2010/11—$N = 2,703$
  - Wave 3—2011/12—$N = 2,748$
  - Wave 4—2013/14—$N = 1,832$.

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 information 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. 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.

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

C.1 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 de Magalhaes and Santaeulalia-Llopis (2018) and Chen et al. (2022), we calculate the value added from product c of agricultural sector z as the sum across seasons of each household i's revenues 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 $Q_{c,s,i}^{z}$, we let $P_{c,s,i,r}^{z}$=$Rev_{c,s,i}^{z}/Q_{c,s,i}^{z}$;
- Otherwise, we attribute the average price of the crop sold by other households in the same region if available, meaning $P_{c,s,i}^{z} = \overline{P}_{c,s,j}^{z}$ with j being in the same region as i;
- Otherwise, we attribute the regional community price reported in community section, meaning $P_{c,s,i}^{z} = P_{c,s,com}^{z}$.

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

$$Cost_{c,s,i}^{z} = \sum_{v} Cost_{c,s,i,v}^{z}$$

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 $z = FS$.

C.2 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.

C.3 Labor Hours

In order to study individual labor supply and the intra-household allocation of time, we use information on hours worked by each individual for the household farm or the household non-farming business. This information belongs to the time use module of the household questionnaire which asks for the amount of hours spent in each activity over 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 own consumption. We compute working hours in entrepreneurship as the sum of hours spent in any kind of non-farming household business. We then aggregate this information at the household level within and across the two activities.

The household questionnaire was always administered together with the post-harvest questionnaire in all survey waves in Ethiopia and Nigeria. In Malawi, half of households in the sample received the household questionnaire together with the post-planting one during the first visit, while the remaining half received it together with the post-planting one. In Uganda, the documentation provides no information on when the household questionnaire was administered.

C.4 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.

C.5 Household Characteristics

The data provide individual demographic characteristics of household members including sex and birth year. We derive the total number of household members and the total number of female household members. In all waves following the first, the questionnaire asks if any household member left the household since the previous interview, which we use to capture migration. We also derive for each household an index of asset ownership by counting the number of assets the household reports to have. The list of assets is country-specific, therefore so is the index we derive.