Selection and Absolute Advantage in Farming and Entrepreneurship

Francisco Alvarez-Cuadrado Francesco Amodio
Markus Poschke

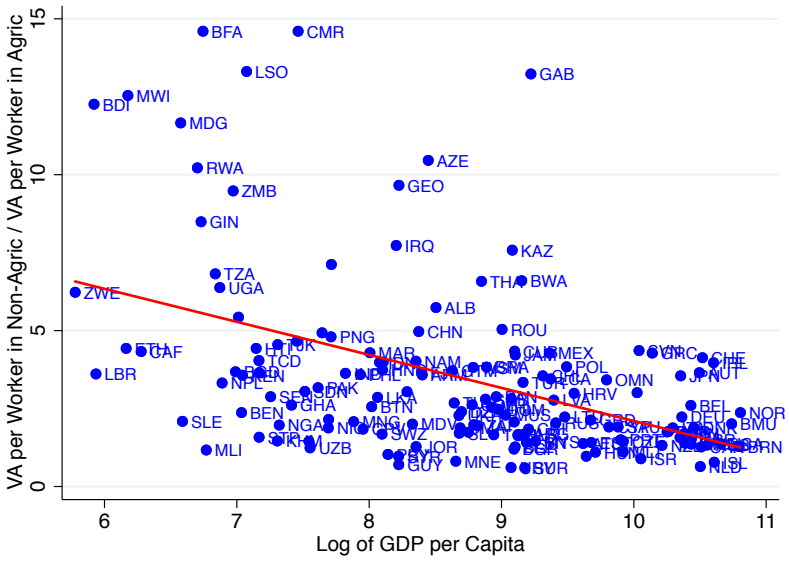
McGill University and CIREQ

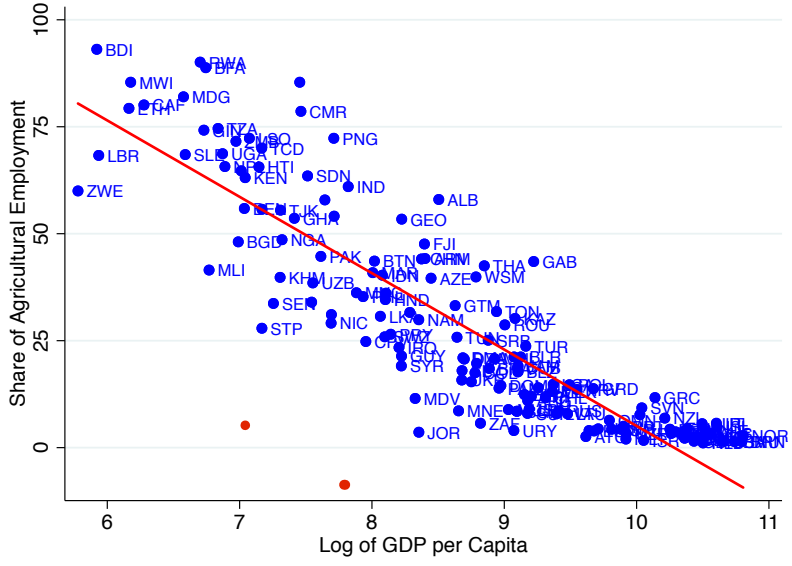
National University of Singapore

May 3/4, 2021

The Agricultural Productivity Gap (APG)

Poor countries: low agricultural productivity, high agr. employment.





(Gollin, Lagakos and Waugh, *QJE* 2014)

VA/worker in agriculture relative to non-agriculture:

- 1/4.4 in Ethiopia versus 1/1.3 in the US/Canada.
- Observables (hours worked, HK, K, land) account for only about 1/3 of the difference.

A Story of Selection

- Heterogeneity in the population
  - Different abilities/skills in agriculture and other activities: heterogeneity in **absolute** advantage
- Sorting according to **comparative** advantage:
  - Relative abilities/payoffs across activities determine choices.

  $\Rightarrow$ Farmers reveal high comparative advantage in agriculture.
- If absolute and comparative advantage are **positively** correlated:
  - The few remaining farmers in US/Canada are the very best.
  - In Ethiopia, less skillful farmers are also active.

$\Rightarrow$ **Average** productivity increases as the share of agricultural employment decreases.

(Lagakos and Waugh 2013)

This paper

- Objectives:
  - Identify the **correlation between absolute and comparative advantage** in agriculture and non-farm entrepreneurship
  - Clarify its relationship with the underlying distributions of absolute advantages

- Challenge: Identification of a Roy model
  - Generally impossible without distributional assumptions
  - Can measure individual productivity in only one activity

(Heckman and Sedlacek 1985, Heckman and Honoré 1990)

- New approach: We consider an extended version of Roy
  - Allow individuals to pursue either one or *both* activities
  - Take the model implications to household-level data
  - Ethiopia, Malawi, Nigeria, Uganda.

Preview of Results

1. Around 1/3 of households engage in both agriculture and non-farming entrepreneurship.
   - ⇒ They have **weak comparative** advantage.
2. *Extensive margin:* These households have systematically higher agricultural productivity than those doing only farming.
   - ⇒ They have **high absolute** advantage in agriculture.
3. *Intensive margin:* Among those doing both, those with higher agricultural productivity supply relatively fewer hours in that sector.
4. *Switchers:* Over time, households starting a non-farming enterprise have higher agricultural productivity than those who remain only farmers.

Implications

- Evidence suggests a **negative correlation of comparative and absolute advantage** in agriculture.
  - $\Rightarrow$ Evidence from within villages shows little support for a selection story driving the APG.
- What could generate the observed patterns?
  - Strong positive correlation between sectoral abilities
  - Higher dispersion of returns to entrepreneurship
  - $\Rightarrow$ Best farmers have comparative advantage in entrepreneurship; choose that.

Outline

1. A Simple Model of Selection
2. Data and Descriptives
3. Selection along the Extensive Margin
4. Interpretation and Discussion
5. Choices on the Intensive Margin
6. Selection Over Time
7. Alternative Explanations
8. Conclusions

An Extended Roy Model

- Two sectors: agriculture and non-agriculture $j = \{a, n\}$
- Continuum of households indexed by i
- Each household is endowed with a vector of abilities $\{z_i^a, z_i^n\}$
- Distributed according to $G(z^a, z^n)$ with means μ_j, variance σ_j^2
- **Absolute** advantage in agriculture: z_i^a
- **Comparative** advantage in agriculture: z_i^a / z_i^n.

Household's Problem

- Value added in the two sectors

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

with $f'(\cdot), g'(\cdot) > 0$ and $f''(\cdot), g''(\cdot) < 0$ and $f'(0), g'(0) < \infty$

- κ captures economy-wide productivity and price differences
- The household allocates one unit of time across activities $\{l_i^a, l_i^n\}$
- Household chooses $\{l_i^a, l_i^n\}$ that maximizes

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

Benchmark Case

- $l_i^j = \{0, 1\}$, household operates in one sector only
- Engages in farming if and only if $\kappa\, z_i^a\, f(1) \geq z_i^n\, g(1)$ or

$$\frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa f(1)} = \text{constant} \tag{3}$$

- Sectoral choice determined by **comparative** advantage. Not informative of absolute advantage!

Benchmark Case

- $l_i^j = \{0, 1\}$, household operates in one sector only
- Engages in farming if and only if $\kappa\, z_i^a\, f(1) \geq z_i^n\, g(1)$ or

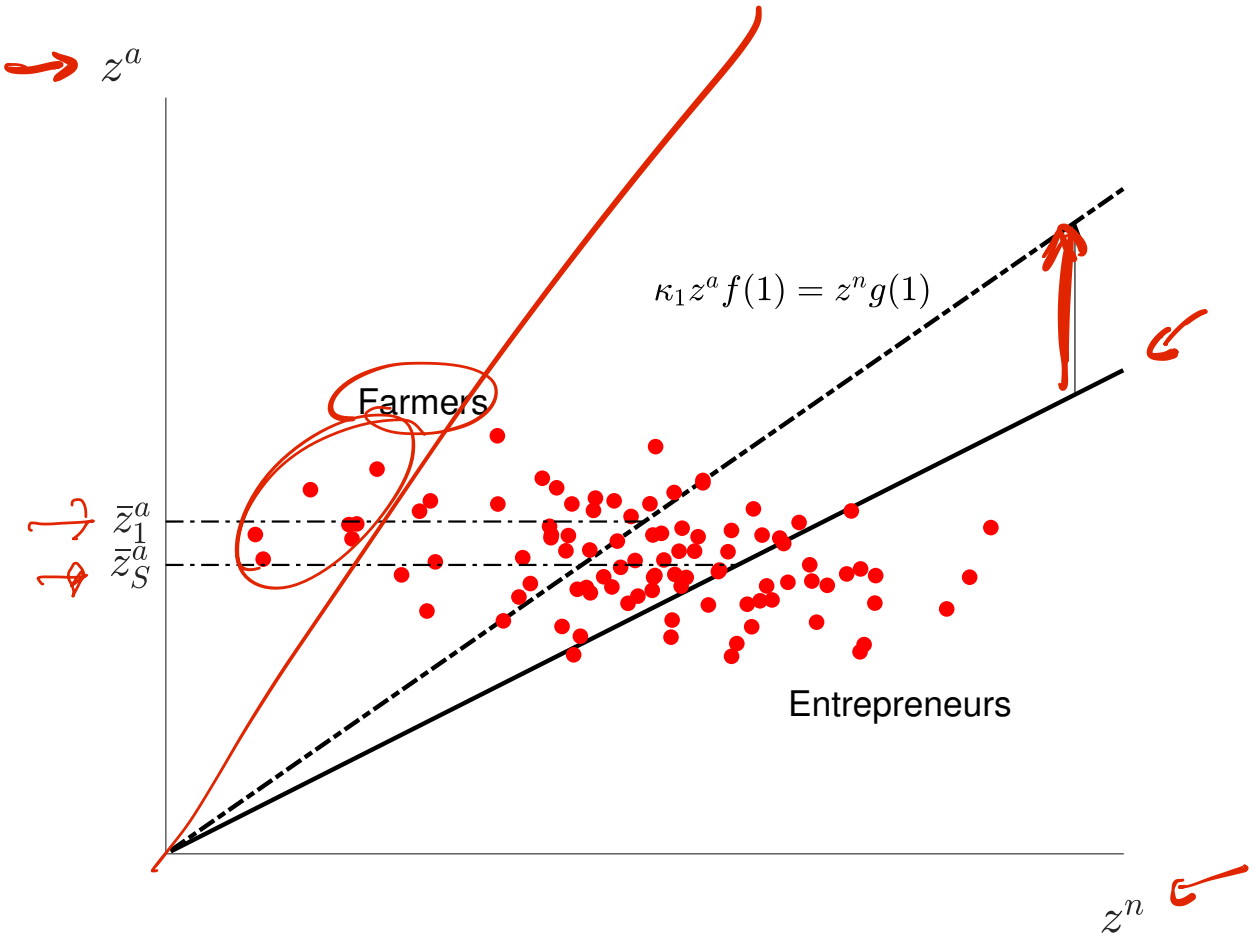
$$\frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa f(1)} = \text{constant} \tag{3}$$

- Sectoral choice determined by **comparative** advantage. Not informative of absolute advantage!
- Mean sectoral productivity in agriculture

$$\bar{y}^a \equiv \mathbb{E}\left(y_i^a \middle| \frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa f(1)}\right) = \frac{\kappa f(1) \int_{\frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa\, f(1)}} z_i^a\, dGi}{\int_{\frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa\, f(1)}} dGi} \tag{4}$$

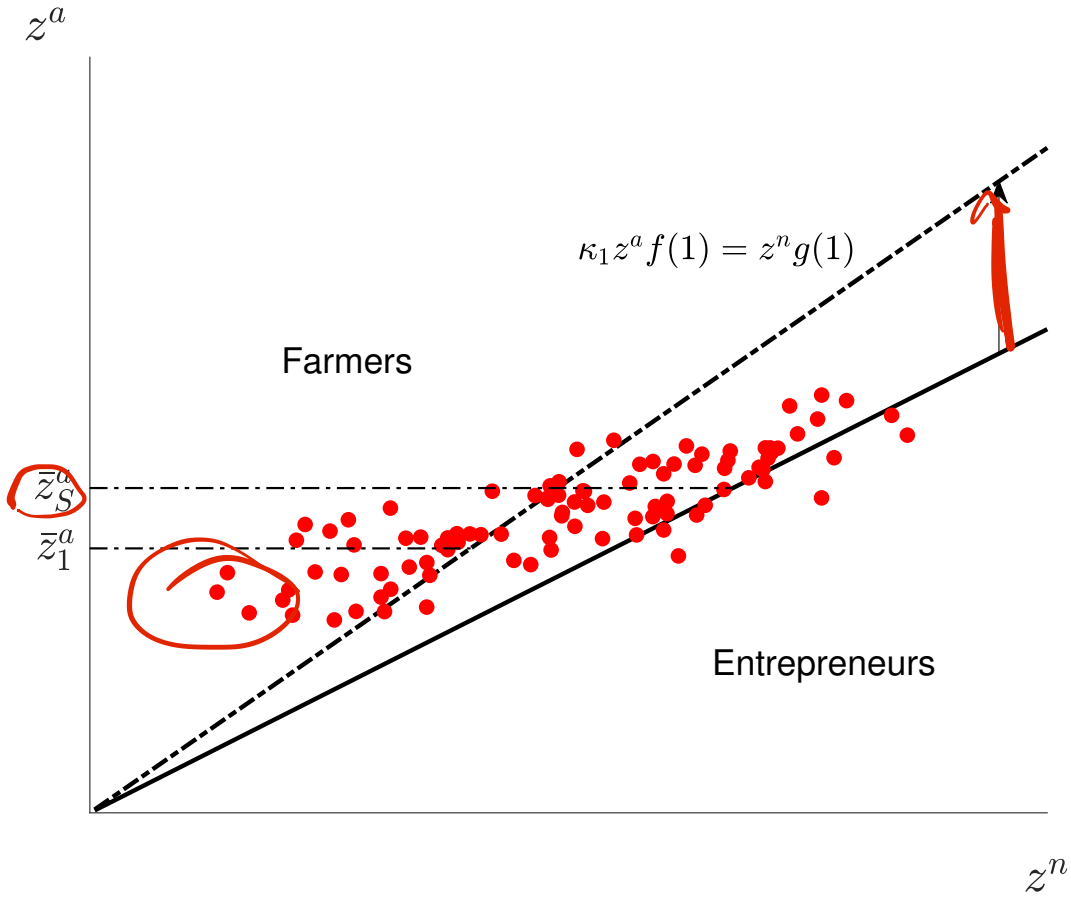
- Sectoral productivities determined by **absolute** advantages.

Positive Correlation of Advantages in Both Sectors



In this case: as agricultural sector shrinks, $\bar{z}^a \uparrow$.

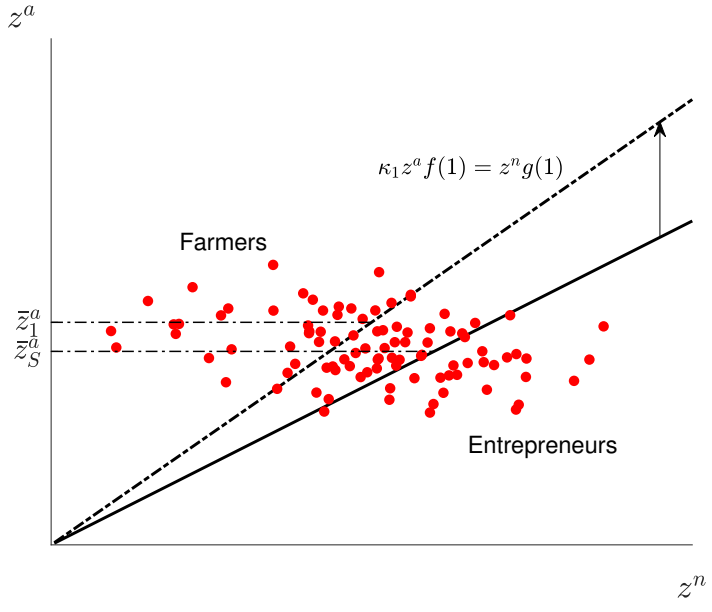
Negative Correlation of Advantages in Agriculture

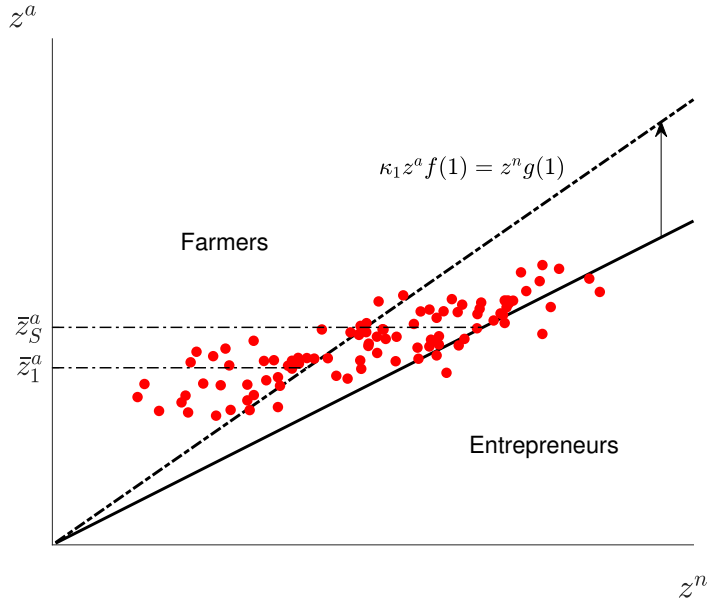


In this case: as agricultural sector shrinks, $\bar{z}^a \downarrow$.

The Correlation of Advantages in Agriculture

The **correlation of advantages** $\rho\left(z_i^a/z_i^n, z_i^a\right)$ determines the relationship between sectoral size and productivity.





Positive correlation: $\bar{z}^a \uparrow$

Negative correlation: $\bar{z}^a \downarrow$

as agricultural sector shrinks.

Its sign is an empirical question.

General Case: $l_i^j \notin [0,1]$, household can operate in both sectors

- Equate MPL across activities:

$$\kappa z_i^a f'(l_i^a) = z_i^n g'(l_i^n)$$

- Attention: Corner solutions $\Rightarrow$ specialization.

$\Rightarrow$ Engage **only** in farming iff strong **comparative advantage**:

$$\frac{z_i^a}{z_i^n} \geq \frac{1}{\kappa} \frac{g'(0)}{f'(1)} \tag{5}$$

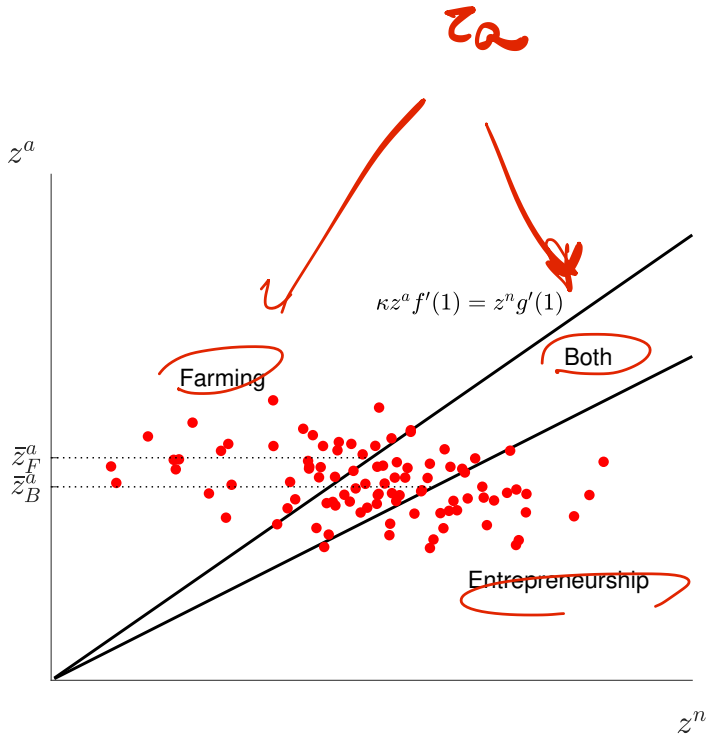
- Operate in **both** sectors iff

$$\frac{z_i^a}{z_i^n} \in \left[\frac{1}{\kappa} \frac{g'(1)}{f'(0)}, \frac{1}{\kappa} \frac{g'(0)}{f'(1)} \right]. \tag{6}$$

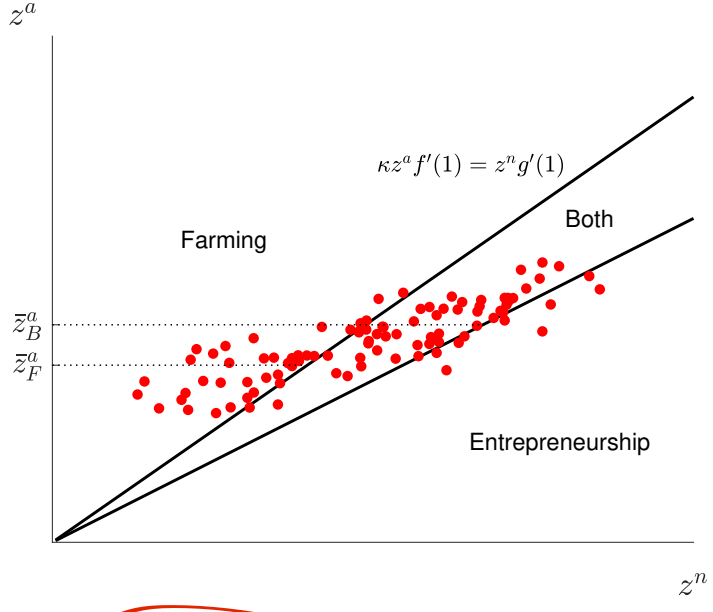
  Weaker **comparative advantage**.

- Sectoral choice is not informative of **absolute advantage**.

What is the correlation of absolute and comparative advantages?

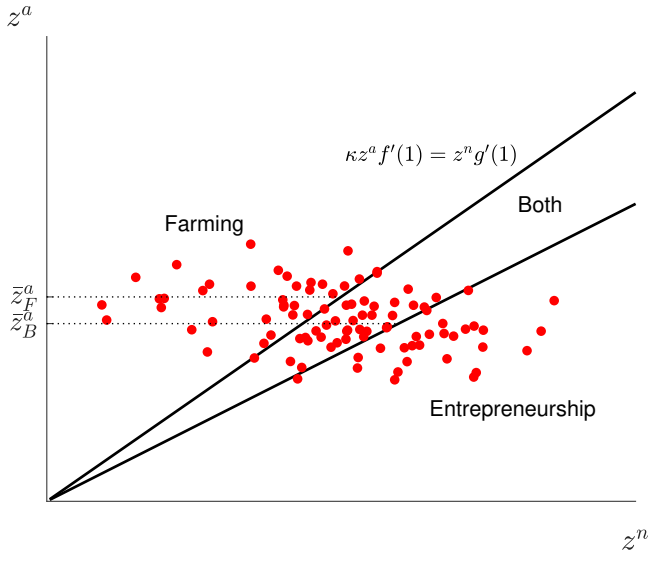


Positive correlation:
only-farmers the best farmers

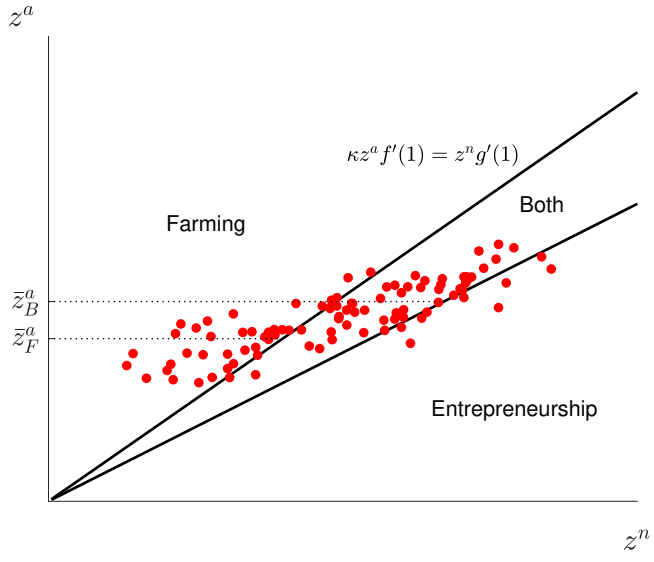


Negative correlation:
only-farmers the worst farmers

What is the correlation of absolute and comparative advantages?



Positive correlation:
only-farmers the best farmers



Negative correlation:
only-farmers the worst farmers

- In each sector, compare households that only work in that sector with households who work in both.
- Those who specialize have higher **comparative advantage**.
- Is their **absolute advantage** higher or lower?

$\Rightarrow$ **Correlation** of advantages in that sector.

Data

- World Bank Living Standards Measurement Study (LSMS-ISA)
  - Nationally representative household panel survey
  - Ethiopia, Malawi, Nigeria, Uganda
  - 2 to 4 waves, 2009 to 2016.
- Value Added in Agriculture
  - Sum of market revenues plus market value of product that was not sold minus production costs

(Santaeulalia-Llopis and Magalhaes 2014)

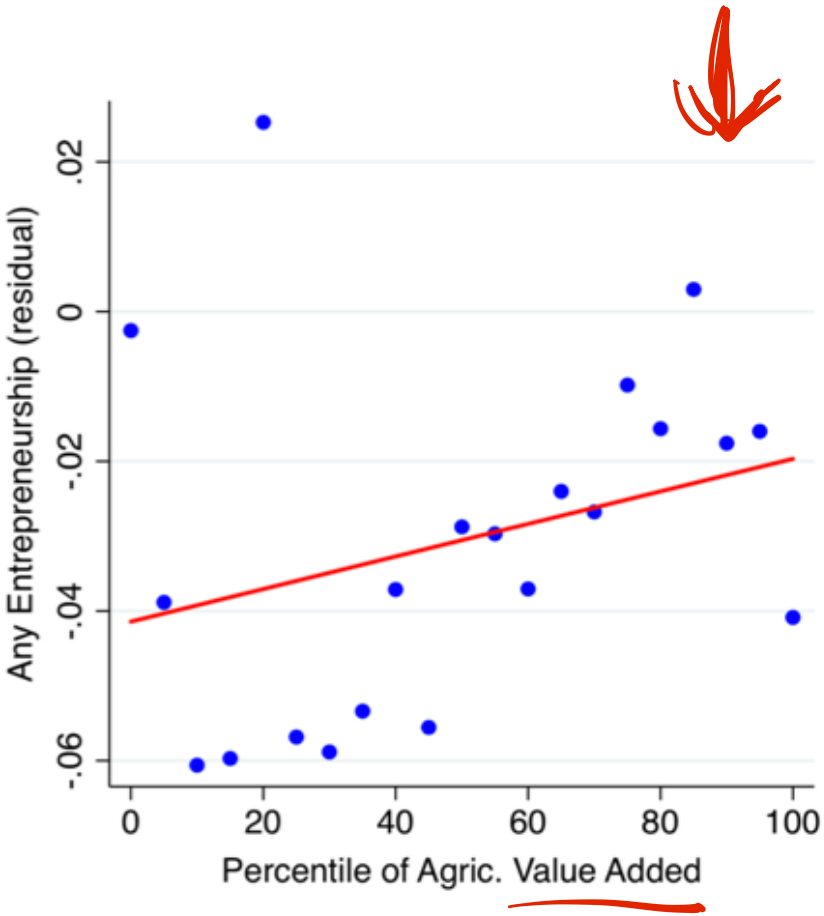
- Value Added in Non-farming Entrepreneurship
  - Enterprises owned by any household member in the 12 months before the interview
  - Difference between total annual sales and associated costs.
- Hours Worked
  - Asked about hours worked per sector in the last 7 days.

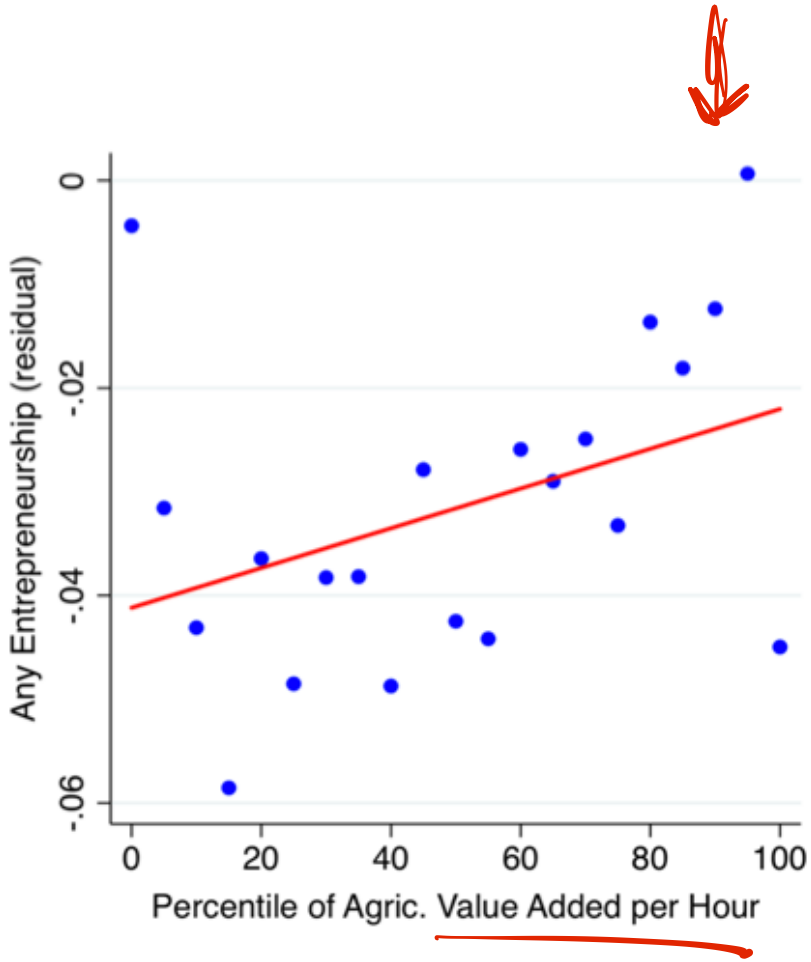
Measuring Advantages

- Mapping from Value Added to z_i^j
  - Production function is increasing and concave
  - Household doing both devote a fraction of time to each
  - For them, **VA** is a downward biased measure of z_i^j
  - For them, **VA per Hour** is an upward biased measure of z_i^j
  - We take percentiles from country-wave distribution
- Mapping from Activities and Hours to z_i^j/z_i^k
  - Households engaging in **one activity** only have higher comparative advantage than those **doing both**
  - Among households doing both, high comparative advantage in a sector maps into relatively more **hours worked** in that sector
- Additional Variables
  - Land, land tenure status, assets, etc.

▸ Summary Statistics

Agriculture: Within Villages





Upper and lower bound for $\rho\left(\frac{z_i^a}{z_i^n}, z_i^a\right)$

▸ across villages

Correlation of Advantages in Agriculture

	Any Entrepreneurship			
	(1)	(2)	(3)	(4)
$Dec(VA_a)$	0.001		0.000	
	(0.001)		(0.001)	
$Dec(VA_a/h_a)$		0.003***		0.007***
		(0.001)		(0.001)
Village FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Observations	30930	22892	27418	21488
R^2	0.247	0.247	0.338	0.293

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis. $Dec(VA_a)$ is the decile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $Dec(VA_a/h_a)$ is the decile 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 (column 3 only), total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.

▸ across villages

Results

- Entrepreneurship rates are higher among the more productive farming households
- Suggest that absolute and comparative advantage are **negatively** correlated in agriculture.
- Households at the margin of leaving agriculture are the most productive, not the least ones
  - Casts doubt on the validity of selection story behind the APG.
- Robustness
  - Alternative definition of activity based on hours worked
  - Consider only households not fully specialized within
  - Hours worked outside the household
  - Subsistence vs. market production

What Drives the Correlation Between Advantages?

- Correlation between abilities and relative dispersions

(Roy 1951, Heckman and Sedlacek 1985, Borjas 1987, Young 2014)

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

$$sign\left[\rho\left(\frac{z_i^a}{z_i^n}, z_i^a\right)\right] = sign\left[\frac{CV\left(z_i^a\right)}{CV\left(z_i^n\right)} - \rho\left(z_i^a, z_i^n\right)\right] \tag{7}$$

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.*

What Drives the Correlation Between Advantages?

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

Implications:

- In the sector with larger dispersion, advantages always positively correlated.
- In the other sector, correlation of advantages depends on correlation of abilities.

Negative correlation of advantages in agriculture consistent with

1. higher dispersion of non-agricultural productivity and
2. strong positive correlation of abilities.

$\Rightarrow$ The best farmers are excellent entrepreneurs; choose that.

The Intensive Margin

- Other factors, like fixed costs or entry costs, could also affect choices.
- **Solution**: look at allocation of hours within group of household doing both – already paid entry cost.
- Optimal allocation of hours:

$$\frac{f'(l_i^a)}{g'(1-l_i^a)} = \frac{1}{\kappa}\frac{z_i^n}{z_i^a}$$

LHS decreases in l_i^a.

$\Rightarrow$ Optimal to spend more time working in sector with **comparative advantage**.

fixed cost graph

Correlation of Advantages in Agriculture

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

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

entrepreneurs

Discussion

- Within households doing both, those with **higher** value added in agriculture put relatively **fewer** hours in agriculture.
- Consistent with a scenario were
  - Abilities are highly positively correlated across sectors.
  - Higher relative dispersion in returns to entrepreneurship.
- Intuition: the good farmers put in fewer farming hours because they are even better entrepreneurs.

Transitions to Entrepreneurship

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

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis. 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*(·) is the within-village ranking of agricultural value added or agricultural value added per hour in Wave 1 among these households. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total number of hours in agriculture (column 3), total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.

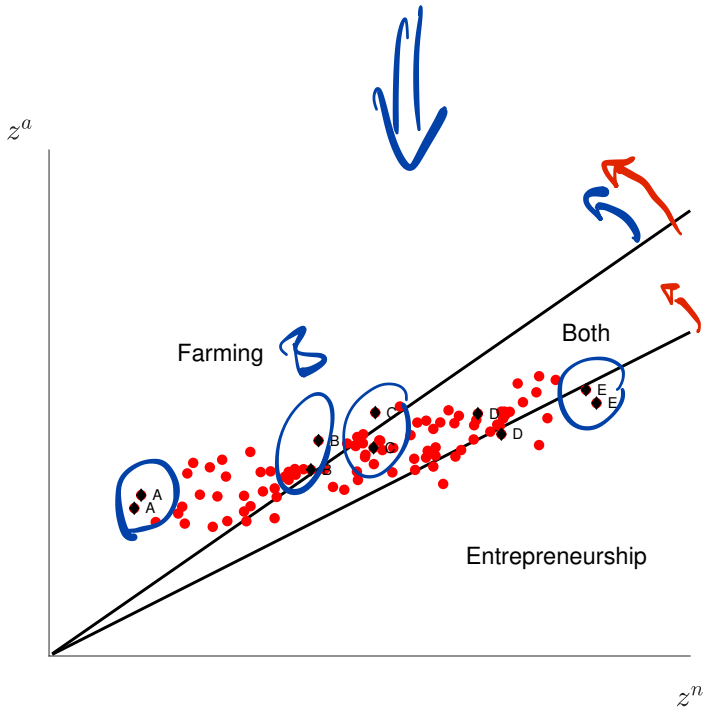
counts

Discussion

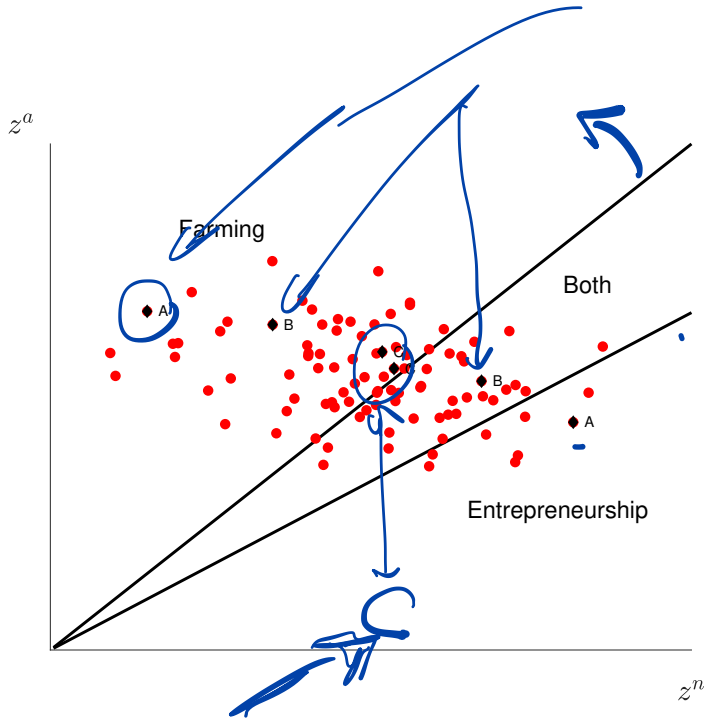
- Over time, it is the households with higher agricultural value added that start a non-farming business.
- Consistent pattern across the four countries.
- Once again suggestive of **negative** correlation between comparative and absolute advantage in agriculture.
- Also suggests that correlation has the same sign for households and for individuals.

Households versus individuals

Two ways of generating **negative** correlation of comparative and absolute advantage among households:



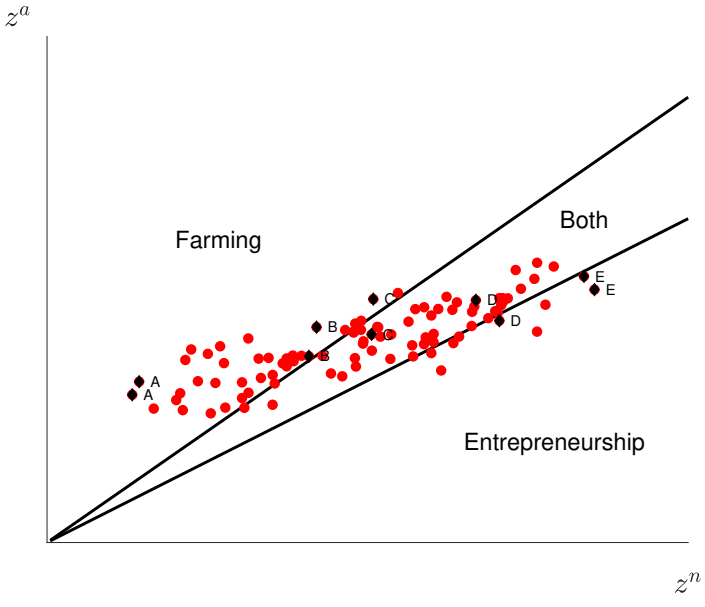
(a) negative corr. among individuals + positive assortative matching

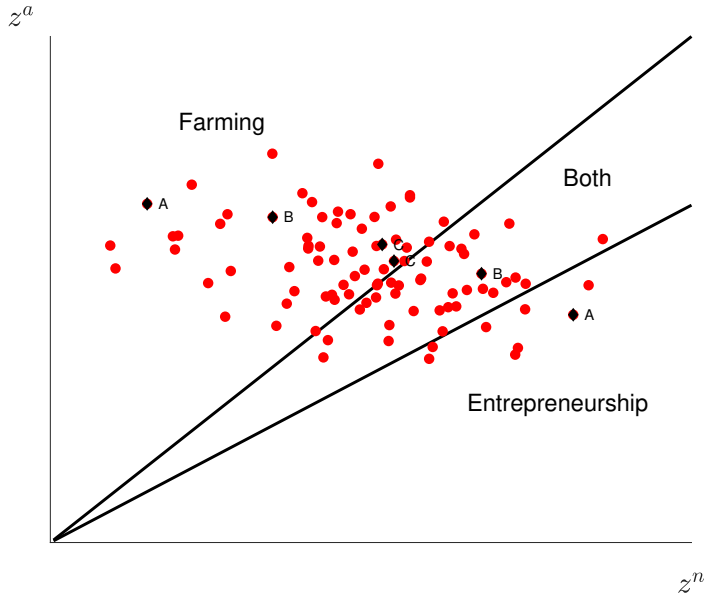


(b) positive corr. among individuals + negative assortative matching

Households versus individuals

Two ways of generating **negative** correlation of comparative and absolute advantage among households:





(a) negative corr. among individuals + positive assortative matching

$\Rightarrow$ best farmers marginal, enter non-ag first

(b) positive corr. among individuals + negative assortative matching

$\Rightarrow$ worst farmers marginal, enter non-ag first

Alternative Explanations

- Distortions along the intensive margin
  - More constrained farmers do more non-farming entrepreneurship.
  - For this alone to drive results, need $\varepsilon((1-\tau), z^a) < -1$: extreme.
- Missing markets
  - Distortions in input use $\Rightarrow$ first point.
  - Distortions along the extensive margin: does not affect analysis restricted to households that do both.
- Heterogeneous fixed costs
  - Productive farmers have lower fixed cost to start business.
  - Reminder: regressions include wealth controls
  - Does not affect analysis restricted to households that do both.
- Diversification as insurance
  - Farmers turn to entrepreneurship when bad shock hits.
  - Would work against finding that these have higher VA in agric.

Conclusion

- Identification of sign of the correlation between absolute and comparative advantage in agriculture and entrepreneurship
  - Exploit presence of households that simultaneously engage in both agriculture and non-agriculture
- Evidence suggests
  - Negative correlation of advantages in agriculture
  - High positive correlation of abilities
  - Higher dispersion in returns from entrepreneurship
  - Little support for a selection story behind the APG.

Thank You!

Summary Statistics

	VA from Only Agriculture	VA from Only Entrep.	VA from Both	Full Sample
Observations	20622 59%	4101 12%	10374 30%	35097 100%
Household Size	5.066 (0.018)	4.625 (0.041)	5.726 (0.027)	5.210 (0.014)
Hours in Agriculture h_a	47.280 (0.385)	4.141 (0.269)	36.569 (0.460)	39.068 (0.276)
Hours in Entrepreneurship h_n	18.540 (0.270)	70.744 (0.856)	53.085 (0.510)	34.944 (0.264)
Total Hours $h_a + h_n$	65.661 (0.501)	75.004 (0.904)	90.126 (0.730)	73.984 (0.385)
HH Members with $h_a, h_n > 0$	n.a. n.a.	n.a. n.a.	0.938 (0.014)	0.277 (0.005)
Land Size (ha)	1.488 (0.087)	0.516 (0.086)	2.464 (0.899)	1.782 (0.289)
Fraction Rented	0.068 (0.002)	0.115 (0.016)	0.070 (0.002)	0.070 (0.001)
Asset Index	9.434 (0.073)	13.538 (0.167)	12.043 (0.112)	10.683 (0.058)

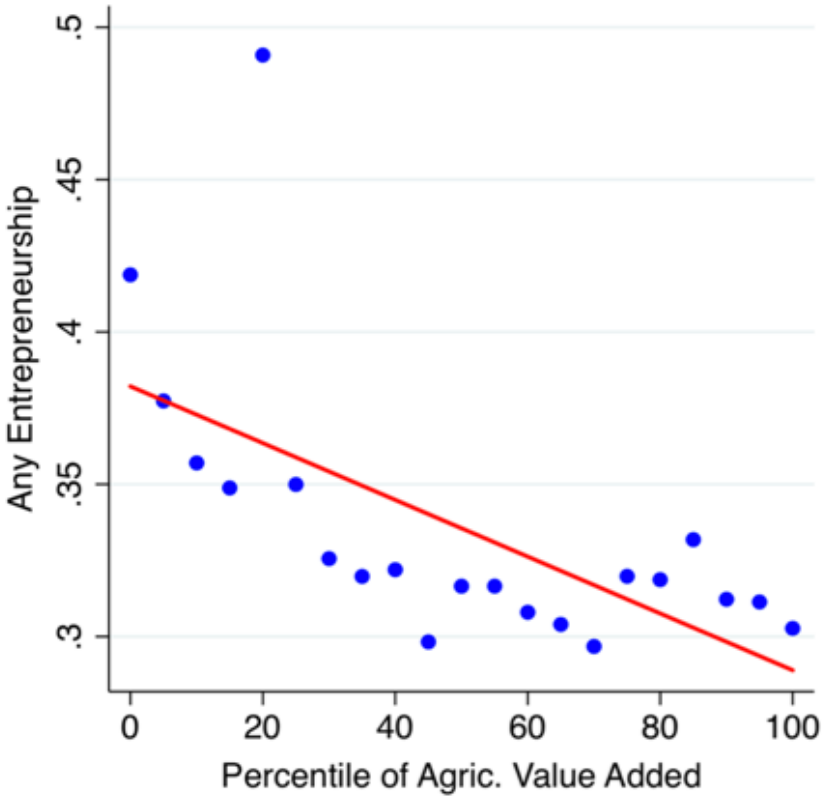
back

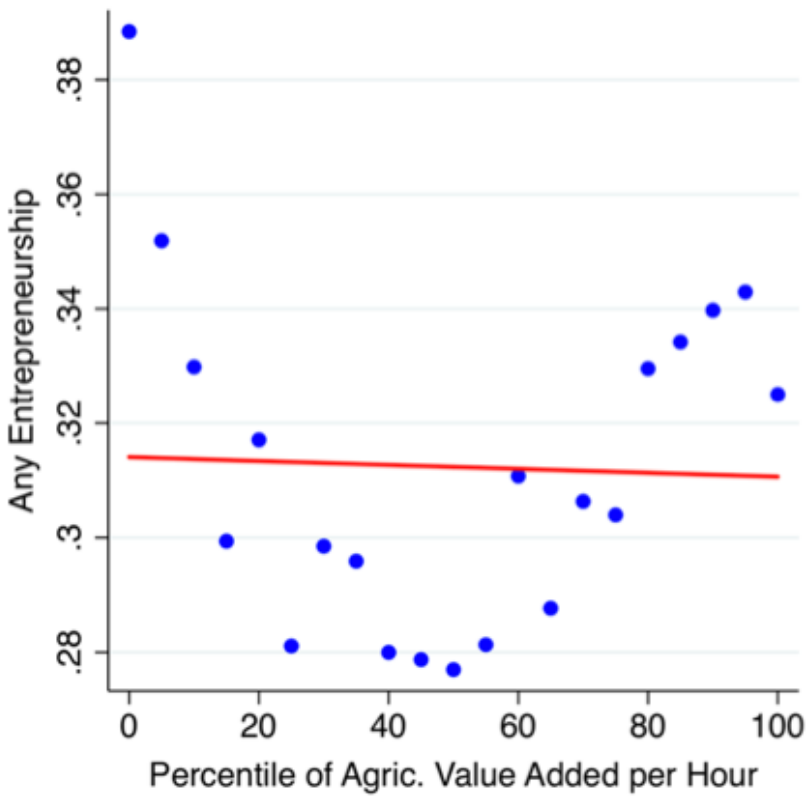
What do non-farm entrepreneurs do?

	Ethiopia	Malawi
non-agricultural service (e.g. mechanic, carpenter, tailor, barber, carwash etc.)	0.28	0.25
process or sell agricultural by-products (flour, local beer, seed, etc., excl. livestock by-products and fish)	0.25	0.15
street or market trading	0.15	0.29
street or market sales (e.g. firewood, home-made charcoal, construction timber, traditional medicine, mats, bricks, baskets, etc.)	0.12	0.16

back

Agriculture: Across Villages





Upper and lower bound for $\rho\left(\frac{z_i^a}{z_i^n}, z_i^a\right)$

back

Correlation of Advantages in Agriculture

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

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis. $Dec(VA_a)$ is the decile the household belongs to in the distribution of value added in agriculture as derived in each country and wave. $Dec(VA_a/h_a)$ is the decile 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. Standard errors are clustered at the level of enumeration area.

back

Correlation of Advantages in Entrepreneurship

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

Notes. * p-value< 0.1; ** p-value<0.05; *** p-value<0.01. Standard errors in parenthesis. $Dec(VA_n)$ is the decile the household belongs to in the distribution of profits from non-farming entrepreneurship as derived in each country and wave. $Dec(VA_n/h_n)$ is the decile the household belongs to in the distribution of profits from non-farming 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-farming entrepreneurship (columns 3 and 7 only), country-specific asset index. Standard errors are clustered at the level of enumeration area.

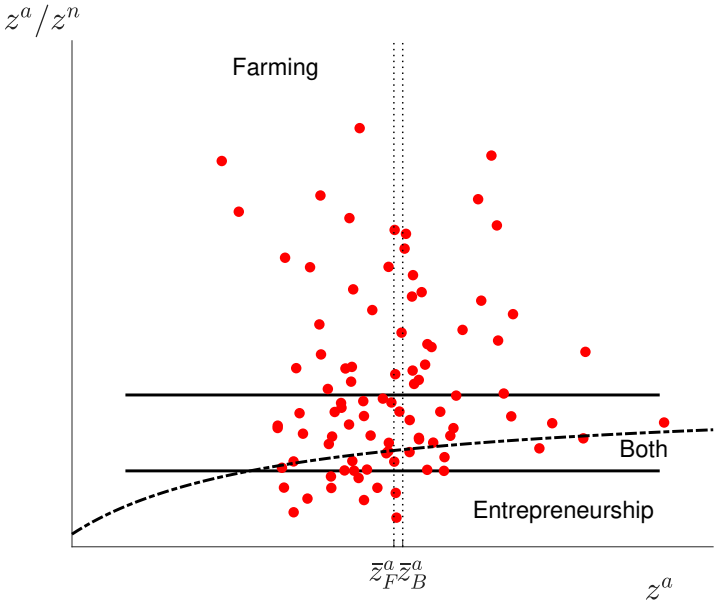
Correlation of Advantages in Entrepreneurship

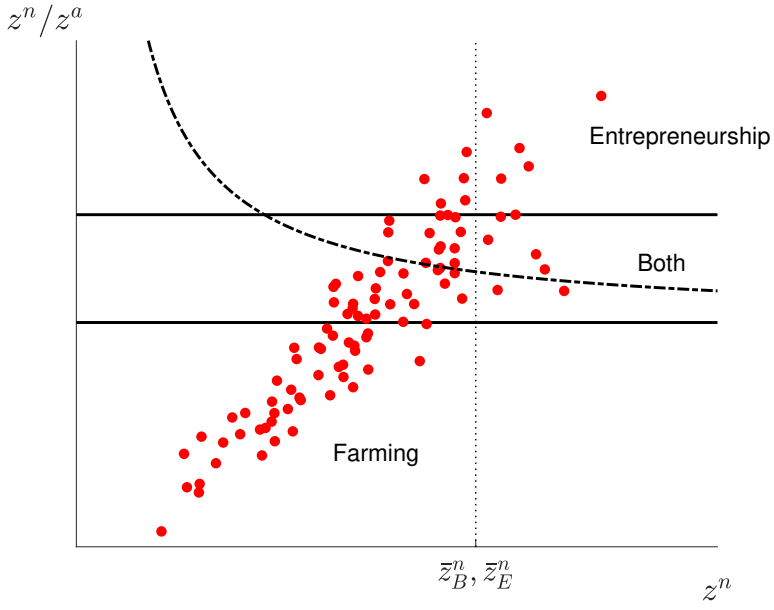
	Time Allocation h_n/h_a			
	(1)	(2)	(3)	(4)
$P(VA_n)$	0.131***		0.128***	
	(0.034)		(0.036)	
$P(VA_n/h_n)$		-0.039		-0.050
		(0.029)		(0.032)
Controls	No	No	Yes	Yes
Country-Wave FE	No	No	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Observations	6913	5702	6416	5236
R^2	0.274	0.265	0.264	0.257

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

back

The effect of fixed costs





Fixed costs in non-agriculture reduce the correlation between absolute and comparative advantage in both sectors. In this example, from 0 to < 0 in agriculture and from > 0 to 0 in non-agriculture.

▸ back

Activities Over Time

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

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

back