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

Francisco Alvarez-Cuadrado       Francesco Amodio
Markus Poschke

McGill University and CIREQ

Oxford

November 4, 2019
Question: Are farmers better at farming than others?

Why it matters:

1. Recent work argues that *selection* can explain part of the *Agricultural Productivity Gap*.
   - This channel requires that active farmers are indeed the best at farming.
Question: Are farmers better at farming than others?

Why it matters:

1. Recent work argues that selection can explain part of the Agricultural Productivity Gap.
   - This channel requires that active farmers are indeed the best at farming.

2. It is a hard question, because of selection.
   - We propose a new idea that allows answering it with little structure.
Motivation 1: The Agricultural Productivity Gap (APG)

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

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.

(Gollin, Lagakos and Waugh, QJE 2014)
Motivation 1: A Story of Selection for the APG

- 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.
  - 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.

⇒ **Average** productivity increases as the share of agricultural employment decreases.

(Lagakos and Waugh 2013)
Motivation 2: Identification

- 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

- 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)

- 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, Tanzania, Uganda.
Preview of Results

1. Around 1/3 of households engage in both agriculture and non-farming entrepreneurship.
   ⇒ They have **weak comparative** advantage.

2. These households have systematically higher agricultural productivity than those doing only farming.
   ⇒ They have **high absolute** advantage in agriculture.

3. Among those doing both, those with higher agricultural productivity supply relatively fewer hours in that sector.

4. Over time, households starting a non-farming enterprise have higher agricultural productivity than those who remain only farmers.
Evidence suggests a negative correlation of comparative and absolute advantage in agriculture.

⇒ Evidence from within villages shows little support for a selection story driving the APG.

What could generate the observed patterns?

▷ Strong positive correlation between abilities
▷ Higher dispersion of returns to entrepreneurship
Literature

- Selection and sorting according to comparative advantage
  - Roy (1951), Borjas (1987)
  - Sectoral size and average productivity: Young (2014)

- Selection and the APG

- Calibration of joint distribution of abilities
  - Lagakos and Waugh (2013) - US wage data: moderately positive correlation
  - Adamopoulos et al. (2017) - panel data from China: negative correlation

- Rural to urban migration
  - Hicks et al. (2018) - panel data from Indonesia: positive selection of migrants.
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 \( \{z^a_i, z^n_i\} \)
- Distributed according to \( G(z^a, z^n) \) with means \( \mu_j \) variance \( \sigma^2_j \)
- **Absolute** advantage in agriculture: \( z^a_i \)
- **Comparative** advantage in agriculture: \( z^a_i / z^n_i \).
Household’s Problem

- The household is endowed with one unit of time to allocate across activities \( \{l^a_i, l^n_i\} \)
- Value added in the two sectors
  \[
  y^a_i = \kappa z^a_i f\left(l^a_i\right) \\
  y^n_i = z^n_i g\left(l^n_i\right) = z^n_i g\left(1 - l^a_i\right)
  \]

- with \( f'(\cdot), g'(\cdot) > 0 \) and \( f''(\cdot), g''(\cdot) < 0 \) and \( f'(0), g'(0) < \infty \)
- \( \kappa \) captures economy-wide productivity and price differences
- Household chooses \( \{l^a_i, l^n_i\} \) that maximizes
  \[
  y_i = \kappa z^a_i f\left(l^a_i\right) + z^n_i g\left(1 - l^a_i\right)
  \]
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
  \[
  \frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa f(1)} = \text{constant}
  \] (3)

- Mean sectoral productivity in agriculture
  \[
  \bar{y}^a \equiv \mathbb{E} \left( y_i^a \middle| z_i^a z_i^n \geq g(1) \frac{z_i^a}{z_i^n} \geq \frac{g(1)}{\kappa f(1)} \right) = \frac{\kappa f(1) \int_{z_i^a}^{z_i^n} z_i^a dG_i}{\int_{z_i^a}^{z_i^n} \frac{g(1)}{\kappa f(1)} dG_i}
  \] (4)

- Occupational choice determined by \textbf{comparative} advantage
- Sectoral productivities determined by \textbf{absolute} advantages
- \textbf{Correlation} \( \rho \left( \frac{z_i^a}{z_i^n}, z_i^a \right) \) determines the relationship between sectoral size and productivity.
Positive Correlation of Advantages in Both Sectors

\[ \kappa_0 z^a f(1) = z^g(1) \]

![Diagram showing the positive correlation of advantages in both sectors with equations and labels for farmers and entrepreneurs.](image-url)
Decrease in Size of the Agricultural Sector: $\bar{z}^a \uparrow$
Negative Correlation of Advantages in Agriculture

\[ \kappa_1 z^a f(1) = z^n g(1) \]

Farmers

Entrepreneurs

switchers

\[ z^n, z^a, z_0, z_1 \]
General Case: $l^i_j \in [0, 1]$, household can operate in both sectors

- Equate MPL across activities:
  \[ \kappa z^a_i f'(l^a_i) = z^n_i g'(l^n_i) \]

- Attention: Corner solutions ⇒ specialization.
  ⇒ Engage **only** in farming iff strong **comparative advantage**:
  \[ \frac{z^a_i}{z^n_i} \geq \frac{1}{\kappa f'(1)} g'(0) \]

  (5)

- Operate in **both** sectors iff
  \[ \frac{z^a_i}{z^n_i} \in \left[ \frac{1}{\kappa f'(0)} g'(1), \frac{1}{\kappa f'(1)} g'(0) \right]. \]

  (6)

  Weaker **comparative advantage**.

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

- The **correlation** of advantages is an empirical question.
Positive Correlation of Advantages in Both Sectors

\[ \kappa z^a f'(1) = z^n g'(0) \]

Marginal farmers are the worst farmers.
Negative Correlation of Advantages in Agriculture

\[ \kappa z^a f'(1) = z^n g'(0) \]

Marginal farmers are the best farmers.
From Theory to Empirics

- 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?

⇒ **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.
Agriculture: Across Villages

Upper and lower bound for $\rho\left(\frac{z^a_i}{z_i}, z^a_i\right)$
Agriculture: Within Villages

Upper and lower bound for $\rho\left(\frac{z^a_i}{z_i}, z^a_i\right)$
### Correlation of Advantages in Agriculture

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec($VA_a$)</td>
<td>-0.009***</td>
<td>-0.006***</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec($VA_a$/h_a)</td>
<td>0.001</td>
<td>0.003**</td>
<td>0.003***</td>
<td>0.007***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Village FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Country-Wave FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>30996</td>
<td>22977</td>
<td>27485</td>
<td>21575</td>
<td>30930</td>
<td>22892</td>
<td>27418</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.003</td>
<td>0.000</td>
<td>0.179</td>
<td>0.080</td>
<td>0.247</td>
<td>0.247</td>
<td>0.338</td>
</tr>
</tbody>
</table>

*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.
## Correlation of Advantages in Entrepreneurship

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

*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.
Results

- Entrepreneurship rates are higher among the more productive farming households.
- Suggest that absolute and comparative advantage are negatively correlated in agriculture, not correlated in entrepreneurship.
- 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 correlations between comparative and absolute advantage are approximated by

\[
\text{sign} \left[ \rho \left( \frac{z_a^i}{z_n^i}, z_a^i \right) \right] \approx \text{sign} \left[ \frac{\text{CV} \left( z_a^i \right)}{\text{CV} \left( z_n^i \right)} - \rho \left( z_a^i, z_n^i \right) \right]
\]

\[
\text{sign} \left[ \rho \left( \frac{z_a^i}{z_n^i}, z_n^i \right) \right] \approx \text{sign} \left[ \frac{\text{CV} \left( z_n^i \right)}{\text{CV} \left( z_a^i \right)} - \rho \left( z_a^i, z_n^i \right) \right]
\]

(7)

where \( \text{CV} \left( z_i^j \right) = \sigma_j / \mu_j \) is the coefficient of variation in the population for sector \( j = \{a, n\} \) and \( \rho \left( z_a^i, z_n^i \right) \) is the correlation coefficient of abilities in the population.
What Drives the Correlation Between Advantages?

\[
\text{sign} \left[ \rho \left( \frac{z^a_i}{z^n_i}, z^a_i \right) \right] \approx \text{sign} \left[ \frac{CV \left( z^a_i \right)}{CV \left( z^n_i \right)} - \rho \left( z^a_i, z^n_i \right) \right]
\]

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.
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.

- Optimal allocation of hours:

\[
\frac{f'(l^a_i)}{g'(1 - l^a_i)} = \frac{1}{\kappa} \frac{z^n_i}{z^a_i}
\]

LHS decreases in \( l^a_i \).

\( \Rightarrow \) Optimal to spend more time in sector with **comparative advantage**.


## Correlation of Advantages in Agriculture

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

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. $Dec(VA_a)$ 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 cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.
Correlation of Advantages in Entrepreneurship

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>Time Allocation $h_n/h_d$ (2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Dec(VA_n)$</td>
<td>0.131***</td>
<td></td>
<td>0.128***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td></td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>$Dec(VA_n/h_n)$</td>
<td></td>
<td>-0.039</td>
<td></td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.029)</td>
<td></td>
<td>(0.032)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Country-Wave FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>6913</td>
<td>5702</td>
<td>6416</td>
<td>5236</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.274</td>
<td>0.265</td>
<td>0.264</td>
<td>0.257</td>
</tr>
</tbody>
</table>

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. $Dec(VA_n)$ 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, country-specific asset index. Standard errors are clustered at the level of enumeration area.
Discussion

▶ Within households doing both, those with higher value added in agriculture put relatively less hours in agriculture.

▶ Those with higher value added in entrepreneurship put relatively more hours in that sector.

▶ 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 less hours because they are even better entrepreneurs.
## Activities Over Time

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

**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.
## Transitions to Entrepreneurship

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>Any Entrepreneurship</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Any Entrepreneurship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 2 $\times$ Rank($V A_a$)</td>
<td>-0.007***</td>
<td>-0.007***</td>
<td>-0.007***</td>
<td>-0.007***</td>
<td>-0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Wave 3 $\times$ Rank($V A_a$)</td>
<td>-0.008***</td>
<td>-0.009***</td>
<td>-0.009***</td>
<td>-0.009***</td>
<td>-0.009***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Wave 2 $\times$ Rank($V A_a/h_a$)</td>
<td>-0.009***</td>
<td>-0.007***</td>
<td>-0.007***</td>
<td>-0.007***</td>
<td>-0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Wave 3 $\times$ Rank($V A_a/h_a$)</td>
<td>-0.012***</td>
<td>-0.010***</td>
<td>-0.010***</td>
<td>-0.010***</td>
<td>-0.010***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

- Household FE: Yes
- Wave FE: Yes
- Controls: No
- Country-Wave FE: No
- 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. \( \text{Rank}(\cdot) \) is the within-village ranking of agricultural value added or agricultural value added per hour in Wave 1 among these households. Control variables include: total number of household members, total number of female household members, total number of hours worked by all household members, total number of hours in agriculture (column 3), total cultivated area, fraction of land that is rented, country-specific asset index. Standard errors are clustered at the level of enumeration area.
Discussion

▶ Over time, it is the households with higher agricultural value added that start a non-farming business.

▶ Once again suggestive of **negative correlation** between comparative and absolute advantage in agriculture.

▶ Consistent pattern across the four countries.
Alternative Explanations

- Distortions along the intensive margin
  - More constrained farmers do more non-farming entrepreneurship.
  - Unlikely to explain why these are also the best farmers overall

- Diversification as insurance
  - Farmers turn to entrepreneurship when bad shock hits
  - Would work against finding that these have higher VA in agric

- Heterogeneous fixed costs
  - Productive farmers have lower fixed cost to start business
  - Does not affect analysis restricted to households that do both

- Missing markets
  - Distortions in input use, back to intensive margin
  - Distortions along the extensive margin, does not affect analysis restricted to households that do both.
Conclusion

- Partial identification 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
  - Selection along the extensive and intensive margin in an extended Roy model
  - Focus on Sub-Saharan Africa: implications for the APG

- 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

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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Ethiopia</th>
<th>Malawi</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-agricultural service (e.g. mechanic, carpenter, tailor, barber, carwash etc.)</td>
<td>0.28</td>
<td>0.25</td>
</tr>
<tr>
<td>process or sell agricultural by-products (flour, local beer, seed, etc., excl. livestock by-products and fish)</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>street or market trading</td>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>street or market sales (e.g. firewood, home-made charcoal, construction timber, traditional medicine, mats, bricks, baskets, etc.)</td>
<td>0.12</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Correlation at households versus individual level

Negative correlation at the individual level