
Chenyue Lei and Kexin Zhang

BU

Sept 20, 2018
Motivation

Three stylized facts:

1. Farming in low-income countries - small-scale
   Farming in high-income countries - large-scale

Figure 1
Motivation

Three stylized facts:

1. Farming in low-income countries - small-scale
   Farming in high-income countries - large-scale

2. The productivity of developed-country agriculture is substantially higher than it is in low-income countries
Motivation

Three stylized facts:

1. Farming in low-income countries - small-scale
   Farming in high-income countries - large-scale

2. The productivity of developed-country agriculture is substantially higher than it is in low-income countries

3. A **U-shaped pattern** b/w productivity and farm/plot size
Motivation (Cont’d)

Figure 6. Relationship Between Real Average Profits per Acre and Farm Size (Acres) (ICRISAT VLS 2009-14)

- Decreasing returns to scale
- Increasing returns to scale
Main Question

- Given the global pattern of farm productivity, why is there a **U-shape** relation b/w farm productivity and scale?
  - Why are the smallest farms more productive than less small farms?
  - Why in the developed world, the larger-scale farms are more productive and that productivity increases with the farm scale?
This Paper

- Explains the U-shaped relationship b/w farm productivity and farm scale from two factors:
  1. Transaction costs in the labor market
     - A large % of low-hour workers (≤ 8 hours/day)
     - ↑ hourly wages to lower-hour workers ⇒ fixed transaction costs for hiring workers (transportation costs)
     - Can explain the U-shape, but cannot alone account for the higher productivity of larger farms compared to the smallest farms
  2. Economies of scale in machine capacity
     - The cost per horsepower (-) related to the total horsepower
Overview of the Presentation

1. Introduction
2. Literature
3. Data
4. Fact
5. Model I
6. Empirics I
7. Model II
8. Empirics II
9. Conclusions
An inverse relation b/w farm prod. and size in low-income countries
- Asia & Latin America (Hazell, 2011; Vollrath, 2007; Kagin et al., 2015)
- Africa (Larson et al., 2013; Carletto et al., 2013)

Explanations for the inverse relationship
- Superior incentives, lower supervision costs, and lower unit-labor costs
  - Yotopoulos and Lau, 1973; Carter and Wiebe, 1990; Binswanger-Mkhizee, et al., 2009; Hazel et al., 2010
  - Cannot explain why large-scale farms are more productive

Two prior studies that finds evidence of a U-shape
- Kimhi (2006):
  - Dis-economies of scale in small maize farms in Zambia, but economies of scale above a threshold
- Muyanga and Jayne (2016)
  - medium-sized and small farmers in Kenya in the same villages
  - Neither provides evidence on the mechanisms behind the U-shape
Data

Six latest rounds of the India ICRISAT VLS panel survey

- Covers the agricultural years 2009-2014
- Contains
  - a census of all households in 18 villages in five states
  - a panel survey of the households in those villages (819 farmers)
- Contains in equal numbers landless households, small-farm households, medium-farm households, and large-farm households
  - Could examine both small and larger farms in a common environment
- Also provides information on input quantities and prices; market input prices for workers, machinery, and animal traction; measurement of the power and capacities of machines
Establishing the Fact: the U-shape

Profits are from the main growing season and are measured in 1999 rupees.
Establishing the Fact: the U-shape (Cont’d)

- Ruling out the possibility of a spurious correlation
  - Measurement error?
    - Use the total farm size from the Census elicitation to IV for the total farm size from the survey ⇒ not the main cause
  - Land quality, credit constraints & farmer ability
Establishing the Fact: the U-shape (Cont’d)

All graphs (with soil characteristics, farmer FE) exhibit a U-shape ⇒ neither farmer wealth/ability nor plot/soil quality could explain the U-shape
Labor-only Model

- One-stage agricultural production
- The production of output requires land and nutrients; CRS production function $g(a, e)$
- The production process for nutrients requires only labor $e = l_f + l_h$
- Households choose b/w family labor ($l_f$), hiring outside labor ($l_h$) for their own-farm production
  - Hired labor: has a fixed cost; $w(l_h) = 1(l_h > 0)w_0 + w_1 l_h$
  - Family labor: no fixed cost; $w_1 l_f$
- With the time endowment, households could either work on farm or enter the labor market; $l = l_o + l_f$
- There is a fixed cost $f$ if they enter the labor market
- HH income comes from: working on farm & entering the labor market
- Household cost comes from: hired labor cost & family labor cost
Labor-only model

Farmer maximizes:

$$\pi = g(a, l_h, l_f) - \begin{cases} \hat{w}(l_h) & \text{if hire a worker} \\ w_1 l_f & \text{family labor cost} \\ 1(l_o > 0)(w_0 - f) + w_1 l_o & \text{if work off-farm} \end{cases}$$

subject to the constraint: $l_o + l_f = l$

- Three regimes:
Labor-only model

Farmer maximizes:

\[
\pi = g(a, l_h, l_f) - w(l_h) - w_1 l_f + \mathbb{1}(l_o > 0)(w_0 - f) + w_1 l_o
\]

subject to the constraint: \( l_o + l_f = l \)

- Three regimes:
  - I. \( a < a^* \): family members work both on and off farm
Labor-only model

Farmer maximizes:

$$\pi = g(a, l_h, l_f) - w(l_h) - w_1 l_f + \mathbb{I}(l_o > 0)(w_0 - f) + w_1 l_o$$

subject to the constraint: $l_o + l_f = l$

- Three regimes:
  - I. $a < a^*$: family members work both on and off farm
  - II. $a^* < a < a^{**}$: households neither hire nor work off-farm (autarky)
Labor-only model

Farmer maximizes:

\[ \pi = g(a, l_h, l_f) - w(l_h) - w_1 l_f + \mathbb{1}(l_o > 0)(w_0 - f) + w_1 l_o \]

subject to the constraint: \( l_o + l_f = l \)

- Three regimes:
  - I. \( a < a^* \): family members work both on and off farm
  - II. \( a^* < a < a^{**} \): households neither hire nor work off-farm (autarky)
  - III. \( a > a^{**} \): hire workers
Farmer maximizes:

\[ \pi = g(a, l_h, l_f) - \underbrace{w(l_h)}_{\text{if hire a worker}} - \underbrace{w_1 l_f}_{\text{family labor cost}} + \underbrace{1(l_o > 0)(w_0 - f) + w_1 l_o}_{\text{if work off-farm}} \]

subject to the constraint: \( l_o + l_f = l \)

- Three regimes:
  - I. \( a < a^* \): family members work both on and off farm
  - II. \( a^* < a < a^{**} \): households neither hire nor work off-farm (autarky)
  - III. \( a > a^{**} \): hire workers
- Two thresholds: \( a^*, a^{**} \)
Labor-only model

Farmer maximizes:

\[
\pi = g(a, l_h, l_f) - w(l_h) - w_1 l_f + \mathbb{1}(l_o > 0)(w_0 - f) + w_1 l_o
\]

subject to the constraint: \( l_o + l_f = l \)

- Three regimes:
  - I. \( a < a^* \): family members work both on and off farm
  - II. \( a^* < a < a^{**} \): households neither hire nor work off-farm (autarky)
  - III. \( a > a^{**} \): hire workers

- Two thresholds: \( a^*, a^{**} \)
  - \( a^* \): Households are indifferent b/w entering and not entering the labor market
Labor-only model

Farmer maximizes:

\[
\pi = g(a, l_h, l_f) - w(l_h) - w_1 l_f + \mathbb{1}(l_o > 0)(w_0 - f) + w_1 l_o
\]

if hire a worker  
family labor cost  
if work off-farm

subject to the constraint: \( l_o + l_f = l \)

- Three regimes:
  - I. \( a < a^* \): family members work both on and off farm
  - II. \( a^* < a < a^{**} \): households neither hire nor work off-farm (autarky)
  - III. \( a > a^{**} \): hire workers

- Two thresholds: \( a^*, a^{**} \)
  - \( a^* \): Households are indifferent b/w entering and not entering the labor market
  - \( a^{**} \): Households are indifferent b/w hiring and not hiring workers
Simulation: profits per acre by farm-scale

- On the smallest farms, farm size has no effect on farm profits.
- At 2.5 acres, farms become autarchic, and profitability per acre ↓ in land size due to the ↑ marginal cost of $l_f$.
- At 11.8 acres, the per-acre farm profits increase in farm size.
Simulation: input costs per acre by farm-scale

I: On the smallest farms, farm size has no effect on input costs
II: Per-acre input costs fall due to constant family labor and ↑ farm size
III: Input costs rise discontinuously due to the fixed labor costs and decrease as acreage rises
Comments and Critiques 1:

What is the constraint of $I_h$?

- If $0 < I_h < I$
  - land per labor $\uparrow$ with size $\Rightarrow$ profits per acre would eventually decrease in Regime 3 X

- If $0 < I_h < \infty$ (could hire infinite number workers)
  - should see repeated jumps for input costs per acre each time a new worker is hired X
Identifying Scale Dis-Economies due to Labor Market Transaction Costs

- Testing for varying $\beta_1$ coefficients in the profit function by land size using LWFCM (Locally Weighted Functional Coefficient Model)

$$y_{ijt} = \beta_0(a_{ij}) + \beta_1(a_{ij})a_{ij} + \sum \beta_n(a_{ij})X_{ijt} + \delta_{jt}(a_{ij}) + \eta_{ijt}(a_{ij})$$

- $y_{ijt}$ - total profits over the kharif season for a farmer $i$ in village $j$ in year $t$
- $X_{ij}$ - soil characteristics
- $\delta_{jt}$ - village/time fixed effects
- $\eta_{ijt}$ - time-varying land specific iid errors

A Priori:
- $a_{ij}$ very small $\rightarrow \beta_1(a_{ij})$ does not vary w.r.t. $a_{ij}$
- $a_{ij}$ small $\rightarrow \beta_1(a_{ij}) \downarrow$ in $a_{ij}$
- $a_{ij}$ large $\rightarrow \beta_1(a_{ij}) \uparrow$ in $a_{ij}$
Identifying Scale Dis-Economies due to Labor Market Transaction Costs (Cont’d)
Comments and Critiques 2:

- $\beta_1(a_{ij})$ captures the marginal profits to size, which seems to be DRTS at small farms, and IRTS at large farms
  - But interested in the average profits per acre instead of the marginal profits of size $\Rightarrow$ not a perfect proxy
- Use the level of profits instead of the logarithms in the regression
  - Taking logs of profits could more clearly show the economies-of-scale patterns (e.g., IRTS: $\beta_1 > 1$; DRTS: $\beta_1 < 1$; CRTS: $\beta_1 = 1$)
Moving from the smallest farms to the largest, the avg. hourly wage\(^1\) firstly rises and then falls at some threshold.

\[\text{Figure 18. Average Hourly Wage Paid for Male Labor, by Farm Size (ICRISAT VLS 2009-14)}\]

\(^1\)Family labor is priced at the marginal or eight-hour wage, while hired labor is priced at the wage actually paid.
Comments and Critiques 3:

- What the model predicts: labor costs do not vary with size at small land sizes (stage I), decrease with land sizes at small-to-medium farms (stage II), and increase with land sizes at large farms (stage III).
  - The direct test here only shows the pattern in stage III, but do not show the first two stages.

![Graph showing stages of labor costs with respect to acres](image-url)
Rainfalls: The marginal land size effect on unit labor costs

<table>
<thead>
<tr>
<th>Variable</th>
<th>Profits</th>
<th>Hours Employed</th>
<th>Average Hourly Wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input type</td>
<td>-</td>
<td>Hired Male Labor</td>
<td>Hired Tractor</td>
</tr>
<tr>
<td>Rainfall (mm)</td>
<td>38.1 (17.1)</td>
<td>.182 (.0701)</td>
<td>.00362 (.00316)</td>
</tr>
<tr>
<td>Rainfall squared x10^3</td>
<td>-21.2 (8.59)</td>
<td>-.107 (.0377)</td>
<td>-.00214 (.00161)</td>
</tr>
<tr>
<td>Year and plot FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>H0: Rain and rain squared</td>
<td>3.09 [.0504]</td>
<td>4.18 [.0183]</td>
<td>0.99 [.3742]</td>
</tr>
<tr>
<td>Number of observations</td>
<td>5,291</td>
<td>3,987</td>
<td>4,016</td>
</tr>
</tbody>
</table>

- ↑ rainfalls are associated w/ ↑ productivity
- ↑ rainfalls are associated w/ ↑ input hours and lower average input costs
Rainfalls, input usage, and average input costs by plot size

- At small plot sizes
  - $\uparrow$ rainfalls are associated with $\uparrow$ usage of low-hour labor
  - $\uparrow$ average hourly labor costs
- At larger plot sizes
  - $\uparrow$ rainfalls are associated with $\downarrow$ usage of low-hour labor
  - $\downarrow$ average hourly labor costs
Model w/ heterogeneous machine

Limitation of a labor-only model:

- Smallest farms have the highest per-acre profits, contradicting the empirical fact that large farms are most productive
Limitation of a labor-only model:
- Smallest farms have the highest per-acre profits, contradicting the empirical fact that large farms are most productive

Solution:
- Include machine capacity scale economies in farm production
Limitation of a labor-only model:
- Smallest farms have the highest per-acre profits, contradicting the empirical fact that large farms are most productive

Solution:
- Include machine capacity scale economies in farm production

To have scale economies in farm production:
- Larger farms use higher machine capacity
- Smaller farms use lower machine capacity
Model w/ heterogeneous machine

Additional assumptions:

- add another input, machine: $q$ (machine capacity), $m$ (machine time)
- allow machine time and labor time to be substitutes, the nutrient fn. has a CES form:

$$e(l, q, m) = \left[\omega(\xi l)^\delta + (1 - \omega)(1 - \frac{q}{\Phi(a)})qm\right]^\frac{1}{\delta}$$

  effective machine capacity

- $\Phi'(a) > 0$: inefficient to use large capacity on small farms
- total cost of using a machine per unit of time:

$$p_q q^\nu + w^\theta$$

  rental cost labor operating machine

$0 < \nu < 1$, economies of scale in machinery capacity
Model w/ heterogeneous machine

So the farmer now maximizes the following profit function over $m$, $q$, $l_h$, $l_f$ given farm size $a$:

$$
\pi(a, l_h, l_f, q, m) = g(a, e(l_h + l_f, q, m)) - w(l_h) - w_1l_f - (w\theta + p_qq'')m \\
+ \mathbb{1}(l_o > 0)(w_0 - f) + w_1l_o
$$

subject to the constraint: $l_o + l_f = l$
Simulation: profits per acre by farm-scale

- I: no machine ⇒ profits per acre does not vary w/ size
- II: substitute family labor w/ machine ⇒ profits decrease with size but are higher than the labor-only case
- III: use higher-capacity machine ⇒ profits per acre increase with size and are higher than smallest farms
Case of sprayer

- Data provides info on
  - hours of sprayer usage and the flow rate of a spray → know machine time and capacity
  - labor weeding hours → can measure labor savings from spraying
- the most commonly used technology
- substantial differences in sprayer capacities
  - power sprayer: higher-capacity
- capacity ↑, per unit of capacity sprayer price ↓
Empirical evidence

- Weeding labor cost per acre ↓ w/ size: substitute machine for labor
- Total sprayer cost per acre ↑: use more expensive, higher-capacity machine as farm size goes up
Reduced-form evidence: machine use

- Farmers with more land area are more likely to own a machine
- Farmer are more likely to own a power sprayer if they own any sprayer
Reduced-form evidence: time spent

<table>
<thead>
<tr>
<th>Variable</th>
<th>Any sprayer use</th>
<th>Weeding hours per acre</th>
<th>Sprayer hours per acre</th>
<th>Sprayer log price per hour</th>
<th>Sprayer flow rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation procedure</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td>Owned area</td>
<td>0.006197</td>
<td>-0.5631</td>
<td>-0.4063</td>
<td>0.01335</td>
<td>0.01360</td>
</tr>
<tr>
<td></td>
<td>(0.0009879)</td>
<td>(0.1286)</td>
<td>(0.0853)</td>
<td>(0.00669)</td>
<td>(0.00667)</td>
</tr>
<tr>
<td>All land characteristics</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Village/year fixed effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>N</td>
<td>3,374</td>
<td>3,374</td>
<td>1,219</td>
<td>1,219</td>
<td>1,219</td>
</tr>
</tbody>
</table>

Standard errors in parentheses clustered at the village/year level.

- Large farms are more likely to use more machine
- Larger farms are more likely to reduce labor
- Larger farms are more likely to use pricier and higher capacity sprayers

Chenyue Lei and Kexin Zhang (BU)
Foster and Rosenzweig (2017 WP)
Comments and Critiques 4:

- A hump-shaped curve for per-acre machine hours by farm size

![Figure 22. Per-Acre Equipment Hours for Tractors and Sprayers, by Farm Size](image)

- OLS with only one variable cannot capture the curvature
- better to add a quadratic term,
  or alternatively run linear regressions on a subset of observations
Direct Testing: structural estimation

To test directly for scale economies in spraying and the limits:
estimate the machine price parameter $\nu$ (recall rental cost $p_q q^\nu$) and the
effective capacity fn. $\Phi(a)$

- Method: GMM
- Moment conditions: same wage and price across a pair of randomly
  selected households in each village
- Parameterize $\Phi(a) = b_0 + b_1 a + b_2 a^2$
- IV: $a, a^2$
Direct Testing: structural estimation

Model implies: \( \frac{dq}{da} > 0 \) if \( \Phi'(a) > 0 \)
Direct Testing: structural estimation

Model implies: \( \frac{dq}{da} > 0 \) if \( \Phi'(a) > 0 \)

\( \Rightarrow \) if \( b_1 > 0 \) and \( b_2 < 0 \), \( \Phi(a) \) has a maximum at \( a^* \): \( \Phi(a) \uparrow \) first, then \( \downarrow \)
Direct Testing: structural estimation

Model implies: \( \frac{dq}{da} > 0 \) if \( \Phi'(a) > 0 \)

\[ \Rightarrow \text{if } b_1 > 0 \text{ and } b_2 < 0, \Phi(a) \text{ has a maximum at } a^*: \Phi(a) \uparrow \text{ first, then } \downarrow \]

\[ \Rightarrow a \uparrow, q \uparrow \text{ until } \Phi(a) \text{ is maximized } (a = a^*) \]
Direct Testing: structural estimation

Model implies: \( \frac{dq}{da} > 0 \) if \( \Phi'(a) > 0 \)

\( \Rightarrow \) if \( b_1 > 0 \) and \( b_2 < 0 \), \( \Phi(a) \) has a maximum at \( a^* \): \( \Phi(a) \uparrow \) first, then \( \downarrow \)

\( \Rightarrow \) \( a \uparrow, q \uparrow \) until \( \Phi(a) \) is maximized (\( a = a^* \))

\( \Rightarrow \) further \( \uparrow \) in size \( a \) will not lead to a higher machinery capacity \( q \)
Direct Testing: structural estimation

Model implies: \( \frac{dq}{da} > 0 \) if \( \Phi'(a) > 0 \)

\[ \Rightarrow \text{if } b_1 > 0 \text{ and } b_2 < 0, \Phi(a) \text{ has a maximum at } a^*: \Phi(a) \uparrow \text{ first, then } \downarrow \]

\[ \Rightarrow a \uparrow, q \uparrow \text{ until } \Phi(a) \text{ is maximized } (a = a^*) \]

\[ \Rightarrow \text{further } \uparrow \text{ in size } a \text{ will not lead to a higher machinery capacity } q \]

\[ \Rightarrow \text{an equilibrium trap - no single farmer would have an incentive to expand land size beyond } a^* \]
Direct Testing: structural estimation

Model implies: \( \frac{dq}{da} > 0 \) if \( \Phi'(a) > 0 \)

\( \Rightarrow \) if \( b_1 > 0 \) and \( b_2 < 0 \), \( \Phi(a) \) has a maximum at \( a^* \): \( \Phi(a) \uparrow \) first, then \( \downarrow \)

\( \Rightarrow \) \( a \uparrow, q \uparrow \) until \( \Phi(a) \) is maximized \( (a = a^*) \)

\( \Rightarrow \) further \( \uparrow \) in size \( a \) will not lead to a higher machinery capacity \( q \)

\( \Rightarrow \) an equilibrium trap - no single farmer would have an incentive to expand land size beyond \( a^* \)

\( \Rightarrow \) but if land consolidation \( \uparrow \) num. of farms above \( a^* \), high demand from large farms can support a market for higher-capacity machines
Direct Testing: structural estimation

- Most farms are below the max (24.5) → too many small farms
- There are other barriers to land consolidation
- Yes, there is an excess number of farms 🤔
Comments and Critiques 5:

- $\Phi(a)$ quadratic form $\rightarrow$ Too many (small) farms
- No explanation for the functional choice
  - Are the results robust?
  - better to try different specifications of $\Phi(a)$
Direct Testing: comparative statics

Table 12
Elasticities for Changes in Area, \( n \) and Wage Rates on Sprayer Capacity (\( q \)), Sprayer Hours (\( m \)) and Weeding Labor Hours (\( l \)) for a Farm of Median Size (3 acres), from the Calibration and GMM Estimates

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Point Estimate</th>
<th>Robust SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( dq/dn )</td>
<td>-0.0498</td>
<td>0.0728</td>
</tr>
<tr>
<td>( dm/dn )</td>
<td>-0.233</td>
<td>0.113</td>
</tr>
<tr>
<td>( dl/dn )</td>
<td>0.0299</td>
<td>0.130</td>
</tr>
<tr>
<td>( dq/da )</td>
<td>0.297</td>
<td>0.0124</td>
</tr>
<tr>
<td>( dq/dw )</td>
<td>0.0292</td>
<td>0.0399</td>
</tr>
<tr>
<td>( dm/dw )</td>
<td>-0.756</td>
<td>0.977</td>
</tr>
<tr>
<td>( dl/dw )</td>
<td>-1.365</td>
<td>0.112</td>
</tr>
</tbody>
</table>

- Cost advantage \( \uparrow (n \downarrow) \): \( \uparrow \) capacity, \( \uparrow \) machine time, \( \downarrow \) labor time
- Wage \( w \) \( \uparrow \): \( \uparrow \) capacity, \( \downarrow \) machine time (labor operating), \( \downarrow \) labor time
Conclusions

- Revisit the U-shaped pattern b/w operation scale and farm productivity in agriculture
- Labor-market transaction costs can explain slightly larger farms are least efficient
- Economies of scale in machine capacity can explain the rising upper tail of the U of high-income countries
- There are too many (small-scale) farms, insufficient to exploit locally-available equipment capacity scale-economies.
Percentage of Small-sized Landholders by Country

Figure 1. Percent of Households with Operational Landholdings Below 10 Acres, by Country