Assessing misallocation in agriculture: plots versus farms

Fernando Aragon (Simon Fraser University),
Diego Restuccia (U. of Toronto), and
Juan Pablo Rud (Royal Holloway)

EEA-ESEM 2022
U. Bocconi, Milano

Starting point

- Evidence of sizeable factor misallocation, especially in agriculture
  - But, concerns that measures can reflect other factors: **unobserved heterogeneity or measurement error**
- A promising approach to 'purge' these factors require using granular micro-data at **plot level** (Gollin and Udry, 2021)
- Approach contrasts with existing studies that use data aggregated at farm/**household level**

Research question

Is the assessment of factor misallocation in agriculture affected by the choice of micro-data aggregation (plot vs farm)?

Short answer: yes, level of aggregation can lead to quantitatively different conclusions.

Why does this matter?

- Recent work using plot-level data suggests that unobserved heterogeneity and measurement error are quite important
  - Gollin and Udry (2021): these factors can explain as much as 70% of assessed misallocation in Uganda and Tanzania's small-scale agriculture.
- Finding shed doubts on the quantitative importance of factor misallocation as a source of underdevelopment and cross-country differences.
- If this result is driven by the choice of data aggregation, then it can led to misleading assessments of factor misallocation.

What we do and find

1. Use micro-data from Uganda and assess factor misallocation using data at plot- and farm-level
   - Find that estimates are substantially **larger** using plot-level data
2. Examine possible explanations for this discrepancy.
   - Differences in estimates of the production function
   - Greater measurement error in plot-level data
3. Assess measurement error using farm-level panel data
   - Find much smaller role for mismeasurement than using plot-level data

1. Estimates of factor misallocation

Calculate efficiency gains $= \frac{\text{agg. output in efficient allocation}}{\text{agg. output in actual allocation}}$

Efficient allocation $\rightarrow$ max. aggregate output $=$ equalize marginal products of land and labor across production units i

1. Estimates of factor misallocation

Finding the efficient allocation

$$\max \sum_i Y_i = s_i (L_i)^{\alpha_L} (X_i)^{\alpha_X}$$
$$\text{s.t.} \sum T_i = T, \sum L_i = L$$

Solution:

$$T_i^e = \frac{z_i}{\sum_i z_i} T, \qquad L_i^e = \frac{z_i}{\sum_i z_i} L,$$

where $z_i \equiv s_i^{1/(1-\alpha_L-\alpha_X)}$

- Positive relation between input allocation and s_i (returns to scale)
- Need microdata on actual output and input use + production function parameters!

1. Estimates of factor misallocation

Calculating the efficient allocation

Data from 3 waves of Uganda Panel Survey (2009=2014)

Prod. function estimates:

- Plot-level data (Gollin and Udry, 2021): 2SLS
  - Estimates of plot productivity (TFPA) + adjusted by unobserved heterogeneity and measurement error
- Farm-level data (Aragón et al., 2022): panel data with household FE

1. Estimates of factor misallocation

Table 1: Efficiency gain and productivity dispersion in plot- and farm-level analysis

	Plot-level data			Farm-level data
	Plot productivity	Plot productivity (adjusted)	Plot productivity aggregated at farm level	Farm productivity
	(1)	(2)	(3)	(4)
A. Efficiency gains				
Nationwide	23.92	6.66	14.28	2.86
Region	16.38	5.36	8.35	2.48
Parish (Village)	4.05	2.47	2.11	1.57
B. Dispersion				
Variance of log	1.26	0.53	0.78	0.84

Plot-level data: **very large efficiency gains**, even after adjusting by unobs. heterogeneity and measurement error.

2. What explain these differences

1. Different production function estimates
2. Excess measurement error in plot-level data

2. What explain these differences

Different production function estimates

Several identification strategies require panel data (Ackerberg et al., 2015, Shenoy, 2017, 2020)

Panels of farms, but not panels of plots $\rightarrow$ drawback of using plot-level data

2. What explain these differences

Different production function estimates

Table 2: Production function estimates at the plot and farm levels

	IV (2SLS)	Panel data with fixed effects
	(1)	(2)
Land contribution (α_L)	0.69	0.37
Labor contribution (α_X)	0.22	0.34
Returns to scale ($\alpha_L + \alpha_X$)	0.91	0.71
Aggregation level	Plot	Household

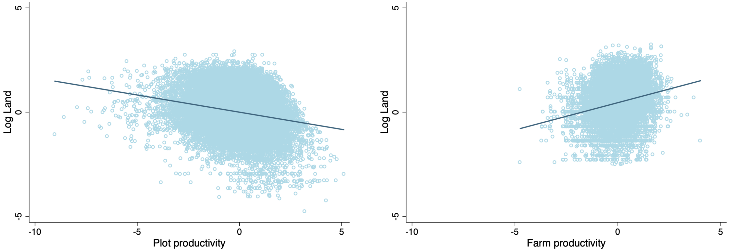
Notes: Column 1 displays 2SLS estimates reported in Table 9 (column 3) in Gollin and Udry (2021). Column 2 display estimates reported in Table A.1 (column 1) in Aragón et al. (2022).

Plot-level data: larger contribution of land and returns to scale (closer to CRS)

2. What explain these differences

Different production function estimates

Figure 1: Land size and productivity across production units



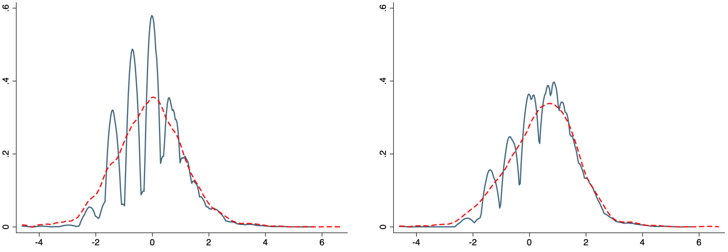
(a) IV estimates (plots) (b) Panel estimates (farms)

Different correlation between productivity and land size

2. What explain these differences

Excess measurement error in plot-level data

Figure 2: Distribution of landholding size, self-reported and GPS measured



(a) Parcel level (b) Farm level

Discrepancies in both measures, specially for smaller units (systematic measurement error) + 'Heaping' (suggestive of rounding-up)

2. What explain these differences

Excess measurement error in plot-level data

Farm > parcel > plot, but GPS data at parcel-level only

Log difference between GPS and self- reported (median):
1.9% (parcel-level) vs 0.45% (farm-level)

Variance of the log of self-reported to GPS land ratio:
0.54 (parcel-level) vs 0.45 (farm-level)

3. Measurement error and misallocation

How to assess the extent of measurement error in measures of misallocation?

Alternative using panel data Bils et al. (2017):

- Purge measures of TFPR from additive measurement error
- λ, fraction of dispersion in TFPR that is due to variation in distortions

We find $\lambda = 0.925 \rightarrow$ measurement error explains 7.5% of assessed misallocation

- Comparable to estimates from China Adamopoulos et al. (2021): 4-10%

Conclusion

- Caution when using plot-level data to study misallocation
- Plot-level analysis exacerbate extent of misallocation and the importance of measurement error
- Main drawbacks: difficult to estimate production function + increase measurement error

References I

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer, "Identification properties of recent production function estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia, "Misallocation, selection and productivity: A quantitative analysis with panel data from China," *Econometrica*, 2021.

Aragón, Fernando M., Diego Restuccia, and Juan Pablo Rud, "Are small farms really more productive than large farms?," *Food Policy*, 2022, *106*, 102168.

Bils, Mark, Peter J Klenow, and Cian Ruane, "Misallocation or mismeasurement?," *Working paper No. 599, Center for International Development, Stanford University*, March 2017.

Gollin, Douglas and Christopher Udry, "Heterogeneity, measurement error, and misallocation: Evidence from african agriculture," *Journal of Political Economy*, 2021, *129* (1), 000–000.

Shenoy, Ajay, "Market failures and misallocation," *Journal of Development Economics*, 2017, *128*, 65–80.

References II

— , "Estimating the Production Function Under Input Market Frictions," *Review of Economics and Statistics*, 2020, pp. 1–45.