Measurement Considerations for Prioritizing Input Use among Smallholder Farmers: Examples from Ethiopia and Tanzania

TRAVIS W. REYNOLDS¹, C. LEIGH ANDERSON², DIDIER ALIA², PIERRE BISCAYE³, JOSHUA MERFELD⁴

¹ UNIVERSITY OF VERMONT
² UNIVERSITY OF WASHINGTON
³ UNIVERSITY OF CALIFORNIA – BERKELEY
⁴ NEW YORK UNIVERSITY – WAGNER
Many policies and programs seek to promote use of inputs (improved seed, fertilizer, and others) among smallholders in low-income countries.

Prioritizing inputs requires understanding each input’s potential to boost production and overcome yield gaps (Sheahan & Barrett, 2017).

To this end some recent research efforts have used **machine-learning / meta-analyses** to synthesize findings across multiple studies to draw conclusions about relative returns to input use.

• **Silesi et al. (2008):** Estimates from 94 peer-reviewed publications from Sub Saharan Africa suggest positive returns to legume intercropping relative to unfertilized maize.

• **Rusinamhodzi et al. (2011), Corbeels et al. (2014), & Himmelstein et al. (2016):** Estimated returns to conservation agriculture in smallholder farm systems vary depending on climate, soil quality, farm management, and cropping systems that vary across studies.

• **Tonitto et al. (2016):** Published studies on sorghum in Sub Saharan Africa suggests a combination of fertilizer inputs and diversified rotations hold promise for yield gains.


• To what degree do measurement choices – including variable construction and data cleaning choices – impact estimates of smallholder productivity?

• How might measurement decisions affect estimated associations between agricultural interventions of interest, such as seed or fertilizer use, and smallholder productivity?

• Examples use plot- and crop-level microdata from the World Bank LSMS-ISA in Tanzania and Ethiopia (2015/16)
  • Replication codes: https://github.com/EvansSchoolPolicyAnalysisAndResearch/335_Agricultural-Indicator-Curation

Maize Area Cultivated (Hectares)

Select: Season or Season(s)

LRS + SRS (ag2_01 + ag2_12)
12,471 LRS and SRS plots
? How to count multiple harvests on a single plot

LRS only (ag2_01)
9,157 LRS plots

Filter by:
Cultivated plots only (ag3a_03)
Maize plots only (ag3a_07)

Select: Area Planted or Area Harvested

Area Planted → Select Area Measure
4,413 with area planted by farmer measure or GPS

Area Harvested: Farmer Report (ag4a_21)
N=4,138
Min: 0.004  Max: 50.586  Mean: 0.600  Median: 0.404

Construction Decision:

Maize plots only

Area Planted

Area Harvested

Large, uniform maize field

Smallholder maize plot, drought conditions

Area Planted

Area Harvested
Distribution of Maize Yield (kg/ha) for Smallholders by Area Harvested or Area Planted: Tanzania

Outliers differ across the two measures

With implications for mean vs. median estimates

Kg/Ha$_{\text{harv}}$

Kg/Ha$_{\text{plant}}$
Distribution of Maize Yield (kg/ha) for Smallholders by Area Harvested or Area Planted: Tanzania

But how can total production divided by area harvested be less than total production divided by area planted? It cannot. (Biophysical reality).

Suggesting a need for data cleaning.

\[ K_{g/Ha_{harv}} \leq K_{g/Ha_{plant}} \]
Distribution of Maize Yield (kg/ha) for Smallholders by Area Harvested or Area Planted, Capped: Tanzania

Capping area harvested (farmer reported) at total plot size (GPS measured) increases mean and median estimated yield by area harvested

But outliers continue to affect mean estimates
Distribution of Maize Yield (kg/ha) for Smallholders by Area Harvested or Area Planted, Capped, and Winsorized: Tanzania

Winsorizing the top 1% of yield estimates slightly reduces mean yield by area harvested (no impact on the median).

But even with Winsorizing mean yield by area planted remains higher than yield by area harvested...
Maize Area Cultivated (Hectares)

Select: Season or Season(s)

LRS + SRS (ag2_01 + ag2_12)
12,471 LRS and SRS plots

LRS only (ag2_01)
9,157 LRS plots

Filter by:
- Cultivated plots only (ag3a_03)
- Maize plots only (ag3a_07)

Select: Area Planted or Area Harvested

Area Planted → Select Area Measure
4,413 with area planted by farmer measure or GPS

Modify by:
- Entire area (ag4a_01)
- Partial area (ag4a_02)

Area Harvested: Farmer Report (ag4a_21)
N=4,138  Min: 0.004  Max: 50.586
Mean: 0.600  Median: 0.404

Area Harvested: GPS (ag2a_09)
N = 3,736  Min: 0  Max: 27.235
Mean: 0.869  Median: 0.415

Area Planted: Farmer Report (ag2a_04)
N = 4,412  Min: 0.010  Max: 28.328
Mean: 0.791  Median: 0.405

Area Planted: GPS (ag2a_09)
N = 3,736  Min: 0  Max: 27.235
Mean: 0.869  Median: 0.415

Consider: By weather (ag2a_10, ag2b_10)

Further considerations
- Account for intercrops (ag4a_04)
- Account for trees
- Limitations to area planted categories reported in ag4b_02 (1/4, 1/2, 3/4)
- Accounting for multiple harvests on a single plot?
Accounting for Multiple Crops on a Single Plot

Example Area Calculations in the Presence of Multiple Crops (1 ha plot)

<table>
<thead>
<tr>
<th>Method</th>
<th>Maize</th>
<th>Sorghum</th>
<th>Beans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>1 ha</td>
<td>1 ha</td>
<td>1 ha</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.25 ha</td>
<td>0.75 ha</td>
<td>0.75 ha</td>
</tr>
<tr>
<td>Method 3</td>
<td>0.33 ha</td>
<td>0.33 ha</td>
<td>0.33 ha</td>
</tr>
<tr>
<td>Method 4</td>
<td>0.25 ha</td>
<td>0.375 ha</td>
<td>0.375 ha</td>
</tr>
</tbody>
</table>

Distribution of Maize Yield (kg/ha) for Smallholders by Area Harvested or Area Planted, Capped, Winsorized, and Intercropped: Tanzania

Recall that in the raw data median yield by area harvested was less than median yield by area harvested (biophysically impossible, since area_{harv} \leq area_{plant}).

With land area capped at GPS measures, Winsorized for outliers, and adjusted for intercropping, land area and yield estimates by area harvested and planted are consistent with expectations - but embody many researcher decisions.
Defining “Smallholders”


### Smallholder Definitions

#### AGRA (2017) Smallholder Definitions

<table>
<thead>
<tr>
<th>Less than 33% of crop value sold</th>
<th>More than 33% of crop value sold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 2ha farm size</td>
<td>Small Non-Commercial Farm</td>
</tr>
<tr>
<td>More than 2ha farm size</td>
<td>Small Commercial Farm</td>
</tr>
<tr>
<td></td>
<td>Large Commercial Farm</td>
</tr>
</tbody>
</table>

#### Mellor & Malik (2016) Smallholder Definitions

<table>
<thead>
<tr>
<th>Less than 5% of crop value sold</th>
<th>Between 5% and 50% of crop value sold</th>
<th>More than 50% of crop value sold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 33% of income from non-farm sources</td>
<td>Subsistence Farm</td>
<td>Pre-commercial Farm</td>
</tr>
<tr>
<td>More than 33% of income from non-farm sources</td>
<td>Transitioning Farm</td>
<td>Diversified Farm</td>
</tr>
</tbody>
</table>

---

The proportion of smallholders in the Tanzania and Ethiopia LSMS-ISA sample ranges from 7% to 93% of respondents, depending on the definition of “smallholder” used.

Restraining the sample to smallholders as defined by less than 2 ha total landholding increases mean and median yield estimates across different yield measures.
• Many policies and programs seek to promote use of inputs (improved seed, fertilizer, and others) among smallholders in low-income countries.

• Prioritizing inputs requires understanding each input’s potential to increase production and overcome yield gaps.

• However the results of data cleaning and variable choices may have substantial impacts on estimated productivity.
Among smallholders (<2ha) in the Tanzania data, improved seed use is associated with higher yields, with a median yield of 827 kg/ha with improved seed versus 565 kg/ha without improved seed, an additional 262 kg/ha.
## Median Intercropped and Pure Stand Maize Yield (kg/ha) for Smallholders (by Area Planted): Tanzania

### Yield of Smallholders, by Definition

<table>
<thead>
<tr>
<th></th>
<th>Intercropped</th>
<th></th>
<th>Pure Stand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Method 2</td>
<td>Method 3</td>
<td>Method 4</td>
</tr>
<tr>
<td>Farm Size &lt; 2 ha</td>
<td>549</td>
<td>827</td>
<td>896</td>
</tr>
<tr>
<td>Farm Size &lt; 4 ha</td>
<td>509</td>
<td>778</td>
<td>865</td>
</tr>
<tr>
<td>RuLIS Smallholder</td>
<td>494</td>
<td>706</td>
<td>712</td>
</tr>
<tr>
<td>AGRA Subsistence Farm</td>
<td>577</td>
<td>577</td>
<td>751</td>
</tr>
<tr>
<td>Mellor Small Non-Commercial Farm</td>
<td>431</td>
<td>593</td>
<td>624</td>
</tr>
</tbody>
</table>

---

23rd ICABR CONFERENCE: REGULATION AND FINANCE OF INNOVATIONS FOR A SUSTAINABLE ECONOMY. Ravello, Italy. June 4-7, 2019
Among smallholders (<2ha) in the Ethiopia data improved seed is associated with much higher yields, with a median yield of 2,004 kg/ha with improved seed versus 1,044 kg/ha without + 960 kg/ha.
Yield of Smallholders, by Definition

<table>
<thead>
<tr>
<th></th>
<th>Intercropped</th>
<th>Pure Stand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Method 2</td>
<td>Method 3</td>
</tr>
<tr>
<td>Farm Size &lt; 2 ha</td>
<td>1957</td>
<td>2628</td>
</tr>
<tr>
<td>Farm Size &lt; 4 ha</td>
<td>1679</td>
<td>1919</td>
</tr>
<tr>
<td>RuLIS Smallholder</td>
<td>1568</td>
<td>1840</td>
</tr>
<tr>
<td>AGRA Subsistence Farm</td>
<td>1325</td>
<td>1424</td>
</tr>
<tr>
<td>Mellor Small Non-Commercial Farm</td>
<td>1983</td>
<td>2628</td>
</tr>
</tbody>
</table>
Returns to Fertilizer Use (Y/N)

Tanzania:
Median maize yields among smallholder farmers are 1,472 kg/ha with fertilizer use vs. 593 kg/ha without

Ethiopia:
Median maize yields among smallholder farmers are 1,601 kg/ha with fertilizer use vs. 857 kg/ha without
Returns to Smallholder Input Use

- Fertilizer, by unit (kg/ha)

**Returns to Fertilizer Use (By Trimming)**

**Tanzania:**

- Winsorize Only
- 1% Trim
- 5% Trim

**Ethiopia:**

Outliers remain prominent in the Ethiopian sample even after 1% trimming.

At 5% trimming estimates roughly mirror Tanzania.

---

23rd ICABR CONFERENCE: REGULATION AND FINANCE OF INNOVATIONS FOR A SUSTAINABLE ECONOMY. Ravello, Italy. June 4-7, 2019
Returns to Smallholder Input Use

- Estimated returns to seed versus fertilizer further vary depending on trimming decisions

Returns to Fertilizer Use (By Trimming)

Quantile Regression Estimates for Median Maize Yield Gain from Fertilizer Use or Improved Seed Under Different Trimming Rules in Ethiopia

- Return to 100kg Inorganic Fertilizer
- Return to Improved Seed
Discussion: “Measurement with Consequences”

- The results of **data cleaning and variable choices** may have substantial impacts on estimated productivity – in some cases exceeding the estimated effects of productivity-enhancing interventions.

- Meta-analyses and machine learning approaches can provide valuable insights, but may also amplify the effects of researcher decisions.

- More detailed reporting of researcher choices in variable construction and data cleaning – alongside public access to data and analysis code – can help address these challenges.
Public Data and Analysis

- ... alongside public access to data and analysis code...
We thank David Coomes, Terry Fletcher, Isabella Sun and Emma Weaver for their excellent research assistance.

This research was supported by grant OPP1135685 from the Bill & Melinda Gates Foundation. The findings and conclusions presented here are those of the authors and do not necessarily reflect the positions or policies of the foundation.


World Bank Living Standards Measurement Study (LSMS) datasets and documentation are available at http://surveys.worldbank.org/lsms