Catalyzing Inclusive Agricultural Transformation in Africa

A Machine Learning Approach

Sam Fraiberger (WB)
Marelize Görgens (WB)
Clara Ivanescu (WB)
Andrew Longosz (WB)
Shaffiq Somani (WB)
Tushar Malik (WB)
Theo Hawkins (WB)
Lakshmi Subramanian (WB/NYU)
Ananth Balashankar (NYU)
Eric Deregt (NYU)
David Wilson (BMGF)
# Key Questions concerning agricultural transformation

**AGRICULTURAL TRANSFORMATION**

Use agricultural transformation inputs to define clusters of households of farmers that are associated with differences in productivity and income growth

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are clusters consistent over time?</td>
<td>YES</td>
</tr>
<tr>
<td>How can agricultural transformation within a cluster be optimized?</td>
<td>DETERMINED</td>
</tr>
<tr>
<td>Are there pathways for progress between clusters?</td>
<td>YES</td>
</tr>
<tr>
<td>Do these differ within and between countries (Ethiopia and Tanzania)?</td>
<td>YES</td>
</tr>
</tbody>
</table>

*Additional, not requested by Ag PST.*
What have we learned?

► Distinct set of actionable and non-actionable variables exist that are most correlated with agricultural transformation and social service outcomes.

► Input variables are correlated with one another, forming distinct groups (clusters) of farmers with similar profiles.

► Can use machine learning to identify variables that are most predictive of increase in income, and predict increases in income within and across clusters.

► Can identify what is most important to move between clusters and how changes to specific input variables changes the probability of moving between clusters.
What have we learned?

- Can identify how to maximize the income of a farmer within a cluster

- Results between countries show stark differences:
  - Different types of clusters
  - Different types of input variables
  - Different levels of correlation
  - Different predictions of the level of transformational outcomes that changes in specific input variables will bring

- Potential for results to be used in programmes to:
  - Develop a country-relevant taxonomy of farmers
  - Prioritize different types of support to different groups of farmers
  - Focus innovations on most important input variables, and
  - Change how results are measured and performance incentivised
Development context in Africa is rapidly changing...

...increases in overall and rural populations unlike in other parts of the world

SSA POPULATION GROWTH

RURAL POPULATION GROWTH

Source: Jayne et al. (2016)
Because of population growth, increased need for food in Africa


Source: FAO: acknowledgements to Holger Matthey/FAO, August 2014
Despite some transformation, agriculture in Africa struggles

Agriculture remains the predominant sector of the economy 25% of GDP in SSA

► Yet, Africa has poorest, most unequal societies
► Most food insecure continent with high malnutrition
► Low levels of agricultural productivity and a worsening food trade balance
► Still high levels of subsistence agriculture with small landholdings
► Urgent need for 21st century jobs, yet a shrinking agriculture labor force:
  ▪ Farming is primary employment for approx. 60% of Africa’s working-age population, down from 80% a decade ago

Source: Jayne et al. (2016)
“In Ethiopia, Malawi, Mozambique, Tanzania, Uganda, and Zambia, the food system is projected to add more jobs than the rest of the economy between 2010 and 2025.”
To achieve agriculture’s potential, **transformation** is essential

Measured through:
- Increases in farmers’ income, competitiveness and productivity
- Better food security
- Better access to social services (education and health)

Stronger agricultural growth facilitates human capital growth and economic growth

*How to achieve such agricultural transformation?*
Multiple theories about agricultural transformation pathways

2011 book by Tsakok (WB) suggests 5 conditions at the system level as necessary for successful agricultural transformation

- Stable framework of macroeconomic and political stability
- Effective technology transfer system
- Access to lucrative markets
- Ownership system
- Employment-creating non-agricultural sectors

IS THIS TRUE?
Multiple theories about agricultural transformation pathways

2016 IDDRI report suggests that pathways are unique and country-specific

Is this true?
Multiple theories about agricultural transformation pathways

McKinsey (2017) suggests that there are six steps for every country to follow in the recipe towards agricultural transformation:

1. Successful farmer changes
2. Productivity extension
3. Seed and input technologies
4. Linking farmers to global markets
5. Multi-country partnerships
6. National policy systems

IS THIS TRUE?
Machine Learning to answer these questions
Longitudinal survey of farmers; links farm and non-farm activities
BMGF funding for its implementation
8 Countries:
- Burkina Faso (1 wave)
- Malawi (2 waves)
- Niger (2 waves)
- Tanzania (4 waves)
- Ethiopia (3 waves)
- Mali (1 wave)
- Nigeria (3 waves)
- Uganda (4 waves)

Initial focus on Ethiopia:
- ~1,500 features per households

WHY ETHIOPIA FIRST?
- Priority for Ag PST
- 3 waves of data
- Real commitment to agricultural transformation (Ethiopian ATA established)

Same approach expanded to Uganda and Tanzania to assess differences between countries
Limitations of LSMS-ISA dataset relevant to this dataset

1. **Zeros and missing values** make it difficult to draw inferences at the subpopulation level:
   - Remove features with less than 5% of non-zero observations
   - Remove features with more than 30% missing values

2. **Inconsistencies** in survey questions across years/countries: Remove features that are not observed across years

3. **“Small Data”**: Constraints the number of feasible subpopulations

4. **GPS coordinates** not available (Uganda)
   - Where GPS coordinates of households in panel data set are available, households are clustered together with one scrambled set of GPS coordinates per cluster (Tanzania)
What can we measure from these data?

OUTCOMES:
- Evidence of agricultural transformation and how they change over time
  - Crop sales, crop sales growth, productivity, household expenditure, food expenditure diversification, and food security
  - Education and health service access

INPUTS...
- through which to achieve agricultural transformation and how they change over time: Household, farmer and farming practices characteristics
  - Some inputs can be modified through short term policy actions (actionable) and others not (non-actionable):

  ACTIONABLE
  - Accessibility (distance to road/market/population center)
  - Agronomic practices (crop diversification, fertilizer, seeds type, irrigation, damage prevention, land certificate, extension program)
  - Equipment (axe, oxen, plough, sickle)
  - Rented factors (credit, hired labor)
  - Shocks (health issues, unexpected price changes)
  - Financial inclusion (access to credit, bank accounts and savings)

  NON-ACTIONABLE
  - Demographics
  - (age, marital status, region of origin)
  - Physical conditions (elevation, temperature, precipitation, rooting conditions, variations in greenness)
Machine Learning Results: Ethiopia
First, unsupervised learning results and change in approach

► **INITIAL APPROACH:** Optimal segmentation of farmers into clusters was done by segmenting the population into sub-groups. Sub-groups were chosen in such a way that it yielded the highest gain with respect to population baseline in accuracy of predicting farmers’ crop sales.
  
  ► Unsupervised learning does increase the model’s predictive power by 10%–15%, BUT variables most predictive of the cluster in which a farmer are non-actionable: *Precipitation, level of greenness and lives In Amhara*
  
  ► As we wanted to develop policy recommendations, chose to focus on actionable variables for which policy interventions can be designed

► **DECIDED:** Need a **semi-supervised ML approach**
  
  a. Look at correlation between outcomes: are they cross-correlated to determine if outcomes should be measured separately or together
  
  b. Look at correlation between outcomes and input variables
  
  c. Identify highly-correlated input variables
  
  d. Cluster farmers using k-mean clustering
     
     **In k-means clustering:** Finds groups of farmers such that the values of the farmers across the 7 selected input variables are similar to others in the group and different to farmers in other clusters, i.e., minimize Euclidian distance to the centre and maximize distance between groups.
     
     **Additional step:** Weight each input by its average correlation across outcomes variables
  
  e. Look at most important variable/s within each cluster
  
  f. Look at pathways and thresholds to move between clusters
Are agric. transformation outcomes in Ethiopia correlated with each other?

<table>
<thead>
<tr>
<th></th>
<th>Children Education</th>
<th>Crop Sales</th>
<th>Crop Sales Growth</th>
<th>Expenditure</th>
<th>Food Expenditure Diversification</th>
<th>Has Medical Assistance</th>
<th>No Food Deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children Education</td>
<td>0.011</td>
<td>-0.044</td>
<td>0.141</td>
<td>0.115</td>
<td>0.054</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td>Crop Sales</td>
<td>0.45</td>
<td>0.273</td>
<td>0.047</td>
<td>0.062</td>
<td>0.174</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crop Sales Growth</td>
<td>0.008</td>
<td>-0.032</td>
<td>-0.023</td>
<td>0.043</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure</td>
<td></td>
<td></td>
<td>0.074</td>
<td>0.068</td>
<td>0.228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food Expenditure Diversification</td>
<td></td>
<td></td>
<td></td>
<td>0.086</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has Medical Assistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>No Food Deficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Varying levels** of correlation between outcomes: mostly low
- **So, need to evaluate each outcome separately** in terms of its correlation with inputs
First, determine cross-correlation between inputs and selected outcomes

- Many inputs are cross-correlated with each other – can choose one input to represent a cluster of closely-correlated inputs

- Cross-correlations between inputs and outputs are low

- Most predictive inputs have a similar directional effect across outcome variables, yet their impact varies

- Similar results hold across years (3 waves of analysis)
Then, cluster farmers by considering inputs that are highly cross-correlated with outcomes and not with other inputs.
K-means clustering results

- K-means clustering achieves desired outcome: clusters farmers based on their own unique set of actionable variables most correlated with outcomes and not with other input variables.
- Clustering consistent over time.
- We pick: number of clusters = 4.
Where are the clusters?
Cluster 2

Legend
- Cluster 2

Average crop sales - 2010 Birr
- <1150
- 1150–1545
- 1546–1898
- 1899–2110
- 2111–8034

River network
(% farmers from survey in that region)
Cluster 3

Legend
- Cluster 3

Average crop sales - 2010 Birr
- <1150
- 1150–1545
- 1546–1898
- 1899–2110
- 2111–8034

River network

(% farmers from survey in that region)
How are the clusters different? *(actionable variables)*

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hired workers</td>
<td>0.32</td>
<td>0.17</td>
<td>0.18</td>
<td>17.69</td>
</tr>
<tr>
<td>Number of oxen owned</td>
<td>0.16</td>
<td>0.96</td>
<td>1.76</td>
<td>2.06</td>
</tr>
<tr>
<td>Number of plough owned</td>
<td>0.32</td>
<td>0.78</td>
<td>1.22</td>
<td>1.24</td>
</tr>
<tr>
<td>Number of crops planted</td>
<td>3.64</td>
<td>2.95</td>
<td>2.84</td>
<td>2.78</td>
</tr>
<tr>
<td>Quantity of chemical fertilizers used (in kg)</td>
<td>22.99</td>
<td>7.45</td>
<td>378.15</td>
<td>343.63</td>
</tr>
<tr>
<td>Quantity of improved seeds used (in kg)</td>
<td>2.12</td>
<td>0.93</td>
<td>11.83</td>
<td>12.77</td>
</tr>
<tr>
<td>% of households who own land certificate</td>
<td>0.43</td>
<td>0.54</td>
<td>0.67</td>
<td>0.62</td>
</tr>
<tr>
<td>% of households using extension programs</td>
<td>0.25</td>
<td>0.06</td>
<td>0.80</td>
<td>0.41</td>
</tr>
<tr>
<td>% of households who saved</td>
<td>0.12</td>
<td>0.15</td>
<td>0.15</td>
<td>0.21</td>
</tr>
<tr>
<td>% of households using credit services</td>
<td>0.12</td>
<td>0.13</td>
<td>0.28</td>
<td>0.26</td>
</tr>
</tbody>
</table>
How are the clusters different? (non-actionable variables and outcomes)

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land surface (in Ha)</td>
<td>0.24</td>
<td>1.65</td>
<td>2.07</td>
<td>2.96</td>
</tr>
<tr>
<td>Crop sales (in Birr 2010)</td>
<td>710.99</td>
<td>1279.19</td>
<td>1524.49</td>
<td>2427.23</td>
</tr>
<tr>
<td>Non-food expenditure (in Birr 2010)</td>
<td>1064.44</td>
<td>1236.68</td>
<td>1773.64</td>
<td>2397.29</td>
</tr>
<tr>
<td>% households using medical assistance</td>
<td>0.2</td>
<td>0.22</td>
<td>0.23</td>
<td>0.28</td>
</tr>
<tr>
<td>% households without food deficiencies</td>
<td>0.47</td>
<td>0.7</td>
<td>0.78</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

First, variation in **non-actionable variables**
Variation in land surface across clusters

- Cluster 1: 0.24 Ha
- Cluster 2: 1.65 Ha
- Cluster 3: 2.07 Ha
- Cluster 4: 2.96 Ha

Clusters of increasing crop sales
Variation in land surface across clusters

- Clusters of increasing crop sales:
  - Cluster 1: 0.24 Ha
  - Cluster 2: 1.65 Ha
  - Cluster 3: 2.07 Ha
  - Cluster 4: 2.96 Ha

- Average crop yield per Ha (Birr, 2010):
  - Cluster 1: 511.577 Birr
  - Cluster 2: 347.658 Birr
  - Cluster 3: 517.361 Birr
  - Cluster 4: 2,500 Birr
Variation in crop sales (in Birr 2010) across clusters

Crop sales and productivity are 72% correlated with each other so can analyze them together.
Variation in % households heads who are female across clusters

Household heads who are female

<table>
<thead>
<tr>
<th>Clusters of increasing crop sales</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28%</td>
<td>10%</td>
<td>15%</td>
<td>11%</td>
</tr>
</tbody>
</table>

World Bank Group
Variation in % widowed households across clusters

- Cluster 1: 0.17
- Cluster 2: 0.08
- Cluster 3: 0.10
- Cluster 4: 0.09

Clusters of increasing crop sales
Variation in **household literacy rate** across clusters

Clusters of increasing crop sales

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Literacy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>0.32</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Proportion of households across clusters who lives in Oromiya
Proportion of households across clusters who lives in SNNPR

Clusters of increasing crop sales:

- Cluster 1: 0.65
- Cluster 2: 0.27
- Cluster 3: 0.12
- Cluster 4: 0.16
Now that we have clustered farmers and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

Second, variation in actionable variables
Variation in hiring workers across clusters

- Clusters of increasing crop sales:
  - No hired workers: 1, 2, 3
  - Average: 17.7

Number of hired workers:
- 0
- 1
- 2
- 3
- 4
Variation in agricultural equipment (oxen) across clusters

Clusters of increasing crop sales

- **0 oxen**, Cluster 1
- **1 oxen**, Cluster 2
- **1.8 oxen**, Cluster 3
- **2.1 oxen**, Cluster 4

Number of oxen owned
The variation in agricultural equipment (ploughs) across clusters

Clusters of increasing crop sales

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of Ploughs Owned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32 ploughs</td>
</tr>
<tr>
<td>2</td>
<td>0.78 ploughs</td>
</tr>
<tr>
<td>3</td>
<td>1.22 ploughs</td>
</tr>
<tr>
<td>4</td>
<td>1.24 ploughs</td>
</tr>
</tbody>
</table>
The variation in number of crops planted (male-headed household) across clusters.
The variation in number of crops planted (female-headed household) across clusters

Female-headed households show a different pattern that warrants further investigation.
The variation in the use of chemical fertilizers across clusters

Clusters of increasing crop sales:

- Cluster 1: 22.99 Kg
- Cluster 2: 7.45 Kg
- Cluster 3: 378 Kg
- Cluster 4: 343 Kg

WORLD BANK GROUP
The variation in **quantity of improved seeds used** across clusters

![Bar chart showing the quantity of improved seeds used in different clusters. Cluster 1: 2 Kg, Cluster 2: 1 Kg, Cluster 3: 12 Kg, Cluster 4: 13 Kg.](chart.png)
The variation in % of households who saved across clusters

Clusters of increasing crop sales

Fraction households who saved

- Cluster 1: 0.12
- Cluster 2: 0.15
- Cluster 3: 0.15
- Cluster 4: 0.21
The variation in **credit used** across clusters
The variation **number of water storage pits owned** across clusters.
The variation % of households who own a land certificate across clusters
Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at **variation of identified variables across clusters** in order to make policy recommendations.

Third, variation in **outcomes**
Variation in households without food deficiencies across clusters

Clusters of increasing crop sales

- Cluster 1: 0.47
- Cluster 2: 0.70
- Cluster 3: 0.78
- Cluster 4: 0.84

World Bank Group
The variation in non-food expenditure across clusters

Non-food expenditure (in Birr, 2010)

Clusters of increasing crop sales

1: 1,064
2: 1,237
3: 1,774
4: 2,397
Variation in % of households using medical assistance across clusters

<table>
<thead>
<tr>
<th>Cluster of increasing crop sales</th>
<th>Fraction Households using medical assistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>0.24</td>
</tr>
<tr>
<td>4</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Initial policy observations

FOR LOW INCOME CLUSTER
Expand equipment (oxen and ploughs) and crop diversification

FOR MIDDLE INCOME CLUSTERS
Improve all the other features

FOR HIGH INCOME CLUSTER
Increase hired workers and increasing savings
Optimizing income in a cluster
## How to maximize income within a cluster

<table>
<thead>
<tr>
<th>Most Impactful Input...</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase farmers’ savings</td>
<td>Increase # of hired workers</td>
<td>Increase # of hired workers</td>
<td>Increase # of hired workers</td>
<td></td>
</tr>
<tr>
<td>No other variables</td>
<td>No other variables</td>
<td>No other variables</td>
<td>No other variables</td>
<td></td>
</tr>
<tr>
<td>Input Coefficient</td>
<td>0.134</td>
<td>0.222</td>
<td>0.6</td>
<td>0.292</td>
</tr>
<tr>
<td>Input Coefficient Interpretation</td>
<td>Every 32% increase in % of farmers in cluster who saves, associated with 157 Birr increase in average income in cluster</td>
<td>Every 1 additional worker hired is associated with 393 Birr increase in average income in cluster</td>
<td>Every 1 additional worker hired is associated with 1424 Birr increase in average income in cluster</td>
<td>Every 19 additional workers hired is associated with 933 Birr increase in average income in cluster</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Impactful Input...</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase # of oxen owned</td>
<td>Obtain Water Storage Pit</td>
<td>Increase quantity of chemical fertilizers Used</td>
<td>Use extension program</td>
<td></td>
</tr>
<tr>
<td>Ownership of plough Use of chemical fertiliser</td>
<td>No other variables Use of extension program</td>
<td>Use of chemical fertilizer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input Coefficient</td>
<td>0.106</td>
<td>0.11</td>
<td>0.263</td>
<td>0.054</td>
</tr>
<tr>
<td>Input Coefficient Interpretation</td>
<td>Every 0.65 increase in # of oxen owned, associated with 124 Birr increase in average income in cluster</td>
<td>Every 1.1 increase in water storage pit ownership, associated with 195 Birr increase in average income in cluster</td>
<td>Every 1093kg average increase in amount of fertilizer used, associated with 624 Birr increase in average income in cluster</td>
<td>Every 45% increase in usage of extension program within cluster, associated with 172 Birr increase in average income in cluster</td>
</tr>
</tbody>
</table>
Pathway analysis
Which *pathways* do we actually observe?

<table>
<thead>
<tr>
<th>Rate of moving over time: % Households that moved to a higher cluster (from 2011 to 2013 or 2013 to 2015)</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.6%</td>
<td>32.9%</td>
<td>17.6%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st most impactful input (from optimisation analysis)</th>
<th>Has saved</th>
<th>Number of hired workers</th>
<th>Number of hired workers</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>LIFT FACTOR 1: By how much an increase in input will be associated with an increase in the probability of moving to a higher cluster</th>
<th>No temporal data available (only collected for 2015 wave)</th>
<th>Farmers in this cluster who increase the hired number of workers have a 34% higher probability of moving to a higher cluster</th>
<th>Farmers in this cluster who increase the hired number of workers have a 32% higher probability of moving to a higher cluster</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Other impactful input (also from optimisation analysis)</th>
<th>Number of oxen owned</th>
<th>Number of water storage pit owned</th>
<th>Quantity of chemical fertilizers used</th>
</tr>
</thead>
</table>

| LIFT FACTOR 2: By how much an increase in input will be associated with an increase in the probability of moving to a higher cluster | Farmers in this cluster who increase the hired number of workers have a 7% higher probability of moving to a higher cluster | Farmers in this cluster who acquire more water storage pits have a 18% higher probability of moving to a higher cluster | Farmers in this cluster who increase the chemical fertilizers that they use have a 12% higher probability of moving to a higher cluster |
K-means clustering results

- K-means clustering achieves desired outcome: clusters farmers based on their own unique set of actionable variables most correlated with outcomes and not with other input variable
- Clustering consistent over time
- We pick: number of clusters = 4
Correlation between Tanzania inputs and outcomes

- Correlations between inputs and outcomes are of lower order of magnitude as Ethiopia’s correlations.

- Strength of correlations for individual inputs are not the same for the 2 countries: point to possible different pathways and optimized input variables.
List of most correlated inputs for *k*-means clustering of households in Tanzania’s LSMS-ISA data

- Whether the farmer has hired workers (but not the number of workers hired)
- The number of animals owned
- Quantity of pesticides used
- Crop diversification
- Household size
- Whether the household head is separated
Where are households in the clusters?

3 to 8 households in close geographic proximity have similar GPS coordinates to protect confidentiality of panel survey participants, hence the heat map approach.
Household Distribution by Cluster and Administrative Zone
Household Distribution by Cluster and Administrative Zone

Legend
- Cluster Level
  - 1
Average crop sales
- no data
- < 335478
- 335478 - 450067
- 450067 - 521581
- 521581 - 919562
- 919562 - 1784082
- River Network

(% farmers from survey in that region)
Household Distribution by Cluster and Administrative Zone

Legend

Cluster Level

- 2

Average crop sales

- no data
- < 335478
- 335478 - 450067
- 450067 - 521581
- 521581 - 919562
- 919562 - 1784082

River Network
Household Distribution by Cluster and Administrative Zone

Legend
Cluster Level
4
Average crop sales
- no data
- < 335478
- 335478 - 450067
- 450067 - 521581
- 521581 - 919562
- 919562 - 1784082
River Network
(% farmers from survey in that region)
How are the clusters different? *(actionable variables)*

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Crops Planted</td>
<td>6.792</td>
<td>4.929</td>
<td>5.377</td>
<td>4.282</td>
</tr>
<tr>
<td>% of Households with Bank Account</td>
<td>0.087</td>
<td>0.116</td>
<td>0.083</td>
<td>0.127</td>
</tr>
<tr>
<td>% of Households Who Have Borrowed</td>
<td>0.357</td>
<td>0.34</td>
<td>0.502</td>
<td>0.467</td>
</tr>
<tr>
<td>Has Hired Workers</td>
<td>0.563</td>
<td>0.794</td>
<td>0.839</td>
<td>0.883</td>
</tr>
<tr>
<td>Number of Months Hiring Workers</td>
<td>0.213</td>
<td>0.543</td>
<td>0.562</td>
<td>0.837</td>
</tr>
<tr>
<td>Number Of Animals Owned</td>
<td>5.688</td>
<td>45.117</td>
<td>45.755</td>
<td>75.885</td>
</tr>
<tr>
<td>Number Of Hired Workers</td>
<td>10.674</td>
<td>11.378</td>
<td>10.628</td>
<td>15.272</td>
</tr>
<tr>
<td>Number Of Ploughs Owned</td>
<td>0.064</td>
<td>0.094</td>
<td>0.212</td>
<td>0.713</td>
</tr>
<tr>
<td>% of Households with Land Certificate</td>
<td>0.034</td>
<td>0.076</td>
<td>0.07</td>
<td>0.117</td>
</tr>
<tr>
<td>Quantity Of Fertilizers Used (in Kg)</td>
<td>179</td>
<td>281</td>
<td>519</td>
<td>983</td>
</tr>
<tr>
<td>Quantity Of Pesticides Used (in Liters)</td>
<td>1820</td>
<td>2520</td>
<td>2532</td>
<td>3698</td>
</tr>
<tr>
<td>% of Households Using Credit Services</td>
<td>0.138</td>
<td>0.185</td>
<td>0.195</td>
<td>0.212</td>
</tr>
<tr>
<td>% of Households Using Irrigation</td>
<td>0.033</td>
<td>0.052</td>
<td>0.021</td>
<td>0.023</td>
</tr>
</tbody>
</table>
## How are the clusters different? (outcomes)

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children Education</td>
<td>0.406</td>
<td>0.407</td>
<td>0.411</td>
<td>0.398</td>
</tr>
<tr>
<td>Crop Sales (in Shilling)</td>
<td>268,759.4</td>
<td>520,175.9</td>
<td>683,827.8</td>
<td>151,234.9</td>
</tr>
<tr>
<td>Expenditure (in Shilling)</td>
<td>295,253.3</td>
<td>319,095.6</td>
<td>347,855.5</td>
<td>536,351.6</td>
</tr>
<tr>
<td>% of Households Using Medical Assistance</td>
<td>0.294</td>
<td>0.274</td>
<td>0.253</td>
<td>0.206</td>
</tr>
<tr>
<td>% of Households Without Food Deficiencies</td>
<td>0.51</td>
<td>0.53</td>
<td>0.566</td>
<td>0.669</td>
</tr>
</tbody>
</table>
Now that we have clustered farmers and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

First, variation in non-actionable variables
Variation in distance to market across clusters.
The variation in households living in Kagera across clusters

![Bar chart showing the fraction of households living in Kagera across different clusters of increasing crop sales.]

2. Medium fraction in cluster 2.
3. Lower fraction in cluster 3.
The variation in households living in Simiyu across clusters
The variation in households living in Tabora across clusters
The variation in % of male-headed households across clusters.
Now that we have clustered farmers and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

Second, variation in actionable variables
The variation in land surface across clusters
The variation in land surface across clusters

- Clusters of increasing crop sales:
  - 1: 1,305.89 Ha
  - 2: 547.47 Ha
  - 3: 681.48 Ha
  - 4: 3,875.06 Ha

Average crop yield per Ha (in Schilling):

- Clusters:
  - 1: 0 Schilling
  - 2: 0 Schilling
  - 3: 0 Schilling
  - 4: 4,500 Schilling
The variation in number of crops planted across clusters
The variation in % of households having a bank account across clusters.

[Bar chart showing the variation in the percentage of households having a bank account across clusters, with clusters of increasing crop sales.]
The variation in % of households who have borrowed across clusters.
The variation in % of households who own a land certificate across clusters.
The variation in **quantity of fertilizers used** across clusters
The variation in number of ploughs owned across clusters.
The variation in **number of animals owned** across clusters.
The variation in number of hired workers across clusters

<table>
<thead>
<tr>
<th>Clusters of increasing crop sales</th>
<th>Number of hired workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10±5</td>
</tr>
<tr>
<td>2</td>
<td>10±5</td>
</tr>
<tr>
<td>3</td>
<td>10±5</td>
</tr>
<tr>
<td>4</td>
<td>15±5</td>
</tr>
</tbody>
</table>
Now that we have clustered farmers and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

Third, variation in outcomes
The variation in crop sales across clusters
The variation in *households without food deficiencies* across clusters
The variation in expenditure across clusters
Optimizing income in a cluster.
## How to maximize income within a cluster

<table>
<thead>
<tr>
<th>Most Impactful Input....</th>
<th>Increase # of ploughs owned</th>
<th>Increase % who borrow money</th>
<th>Increase # of animals owned</th>
<th>Increase hiring of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>...which is highly correlated with these inputs</td>
<td>No other variables</td>
<td>No other variables</td>
<td>No other variables</td>
<td>No other variables</td>
</tr>
<tr>
<td>Input Coefficient</td>
<td>0.228</td>
<td>0.136</td>
<td>0.161</td>
<td>0.209</td>
</tr>
<tr>
<td>Input Coefficient Interpretation</td>
<td>Every 26% increase in % of farmers in cluster owns ploughs, is predicted to have 94 485 shillings increase in average income in cluster</td>
<td>Every 47% increase in % who borrow money is predicted to have 203 928 shillings increase in average income in cluster</td>
<td>Every 46 additional animals owned, on average, is predicted to have 263298 shillings increase in average income in cluster</td>
<td>Every 25 additional workers hired is predicted to result in 474568 shillings increase in average income in cluster</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Impactful Input....</th>
<th>Increase hiring of workers</th>
<th>Increase # of animals owned</th>
<th>Increase # of ploughs owned</th>
<th>Increase irrigation use</th>
</tr>
</thead>
<tbody>
<tr>
<td>.... which is highly correlated with these inputs</td>
<td>No other variables</td>
<td>Increase # of ploughs owned</td>
<td>Increase # of animals owned</td>
<td>No other variables</td>
</tr>
<tr>
<td>Input Coefficient</td>
<td>0.21</td>
<td>0.131</td>
<td>0.118</td>
<td>0.156</td>
</tr>
<tr>
<td>Input Coefficient Interpretation</td>
<td>Every 11 additional workers hired is predicted to result in 87 025 shillings increase in average income in cluster</td>
<td>Every 36 additional animals owned, on average, is predicted to result in 196 423 shillings increase in average income in cluster</td>
<td>Every 62% average increase in proportion of farmers who own ploughs, is predicted to result in 192 967 shillings increase in average income in cluster</td>
<td>Every 15% increase in irrigation usage within cluster, is predicted to result in 354 223 shillings increase in average income in cluster</td>
</tr>
</tbody>
</table>
Correlation between Uganda inputs and outcomes

Correlations between inputs and outcomes are of lower order of magnitude as Ethiopia’s correlations, and similar to Tanzania’s.

Strength of correlations for individual inputs are not the same for the 3 countries: points to possible different pathways and optimized input variables.
List of most correlated inputs for k-means clustering of households in Uganda’s LSMS-ISA data

- Whether the farmer has hired workers (but not the number of workers hired)
- The number of animals owned
- Quantity of pesticides used
- Crop diversification
- Household size
- Whether the household head is separated
## How are the clusters different? *(actionable variables)*

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount Borrowed</td>
<td>1388.501</td>
<td>706.973</td>
<td>1074.525</td>
<td>2566.904</td>
</tr>
<tr>
<td>Number of Crops Planted</td>
<td>4.214</td>
<td>10.716</td>
<td>12.359</td>
<td>15.595</td>
</tr>
<tr>
<td>% of Households Who Borrowed Money</td>
<td>0.271</td>
<td>0.337</td>
<td>0.25</td>
<td>0.215</td>
</tr>
<tr>
<td>Has Hired Workers</td>
<td>0.483</td>
<td>0.464</td>
<td>0.564</td>
<td>0.694</td>
</tr>
<tr>
<td>Number Of Days Hired Workers</td>
<td>4.29</td>
<td>4.184</td>
<td>7.38</td>
<td>13.839</td>
</tr>
<tr>
<td>Number Of Hired Workers</td>
<td>12.285</td>
<td>10.378</td>
<td>18.646</td>
<td>42.467</td>
</tr>
<tr>
<td>Number Of Animals Owned</td>
<td>1.137</td>
<td>0.96</td>
<td>1.321</td>
<td>3.842</td>
</tr>
<tr>
<td>Number Of Bulls Owned</td>
<td>0.158</td>
<td>0.097</td>
<td>0.211</td>
<td>0.492</td>
</tr>
<tr>
<td>Number Of Cows Owned</td>
<td>0.566</td>
<td>0.517</td>
<td>0.66</td>
<td>2.271</td>
</tr>
<tr>
<td>Number Of Ploughs Owned</td>
<td>0.063</td>
<td>0.015</td>
<td>0.09</td>
<td>0.141</td>
</tr>
<tr>
<td>Number Of Tools Owned</td>
<td>3.968</td>
<td>4.566</td>
<td>6.239</td>
<td>13.39</td>
</tr>
<tr>
<td>% of Households Who Own Land Certificate</td>
<td>0.439</td>
<td>0.589</td>
<td>0.694</td>
<td>1.171</td>
</tr>
<tr>
<td>Quantity Of Fertilizers Used (in Kg)</td>
<td>36.891</td>
<td>29.53</td>
<td>161.588</td>
<td>515.747</td>
</tr>
<tr>
<td>Quantity Of Improved Seeds Used (in Kg)</td>
<td>159.752</td>
<td>196.421</td>
<td>270.545</td>
<td>479.099</td>
</tr>
<tr>
<td>Quantity Of Pesticides Used (in Liters)</td>
<td>61.613</td>
<td>80.385</td>
<td>84.693</td>
<td>511.852</td>
</tr>
<tr>
<td>% of Households Using Irrigation</td>
<td>0.039</td>
<td>0.096</td>
<td>0.075</td>
<td>0.076</td>
</tr>
</tbody>
</table>
How are the clusters different? (non-actionable variables and outcomes)

<table>
<thead>
<tr>
<th>NON ACTIONABLE</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of Households Attending School</td>
<td>0.691</td>
<td>0.747</td>
<td>0.786</td>
<td>0.864</td>
</tr>
<tr>
<td>Household Head Age</td>
<td>44.132</td>
<td>46.27</td>
<td>47.399</td>
<td>50.68</td>
</tr>
<tr>
<td>Fraction of Divorced</td>
<td>0.086</td>
<td>0.085</td>
<td>0.056</td>
<td>0.025</td>
</tr>
<tr>
<td>Fraction of Male-headed Households</td>
<td>0.667</td>
<td>0.649</td>
<td>0.732</td>
<td>0.845</td>
</tr>
<tr>
<td>Household Head Is Monogamous</td>
<td>0.838</td>
<td>0.815</td>
<td>0.788</td>
<td>0.776</td>
</tr>
<tr>
<td>Household Head Is Polygamous</td>
<td>0.162</td>
<td>0.185</td>
<td>0.212</td>
<td>0.224</td>
</tr>
<tr>
<td>Fraction of Widow</td>
<td>0.152</td>
<td>0.18</td>
<td>0.127</td>
<td>0.07</td>
</tr>
<tr>
<td>Household Head Never Married</td>
<td>0.026</td>
<td>0.033</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>Number of Household Members</td>
<td>5.174</td>
<td>5.145</td>
<td>6.513</td>
<td>8.543</td>
</tr>
<tr>
<td>Land Surface (in Ha)</td>
<td>1.684</td>
<td>1.723</td>
<td>5.06</td>
<td>9.808</td>
</tr>
<tr>
<td>Literacy Rate</td>
<td>0.519</td>
<td>0.56</td>
<td>0.618</td>
<td>0.727</td>
</tr>
<tr>
<td>Rural Household</td>
<td>0.785</td>
<td>0.888</td>
<td>0.915</td>
<td>0.871</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OUTCOMES</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children Education</td>
<td>0.727</td>
<td>0.779</td>
<td>0.806</td>
<td>0.843</td>
</tr>
<tr>
<td>Crop Sales (in Shilling)</td>
<td>260131.5</td>
<td>371814.5</td>
<td>1075892</td>
<td>2516626</td>
</tr>
<tr>
<td>Expenditure (in Shilling)</td>
<td>363122.9</td>
<td>357314.5</td>
<td>380455.4</td>
<td>537182.8</td>
</tr>
<tr>
<td>Medical Assistance</td>
<td>0.302</td>
<td>0.336</td>
<td>0.28</td>
<td>0.214</td>
</tr>
<tr>
<td>% of Households Without Food Deficiencies</td>
<td>0.691</td>
<td>0.759</td>
<td>0.791</td>
<td>0.892</td>
</tr>
</tbody>
</table>
Now that we have clustered farmers and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

First, variation in non-actionable variables
The variation in male-headed household across clusters
The variation in number of household members across clusters
The variation in land surface across clusters
The variation in attended school across clusters
The variation in literacy rate across clusters

Clusters of increasing crop sales
Now that we have clustered farmers and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

Second, variation in actionable variables
The variation in number of bulls owned across clusters
The variation in **number of cows owned** across clusters
The variation in number of animals owned across clusters

![Bar chart showing the variation in number of animals owned across clusters.](chart.png)
The variation in number of days hired workers across clusters
The variation in crop diversification across clusters
The variation in **households using irrigation** across clusters
The variation in quantity of improved seeds used across clusters
The variation in **number of tools owned** across clusters

![Bar chart showing the number of tools owned across clusters of increasing crop sales.](image)

**Clusters of increasing crop sales**

- **Cluster 1**
- **Cluster 2**
- **Cluster 3**
- **Cluster 4**

**Number of tools owned**

- Cluster 1: 4
- Cluster 2: 5
- Cluster 3: 6
- Cluster 4: 15

**World Bank Group**
The variation in % of households who own a land certificate across clusters

[Bar chart showing the variation in % of households who own a land certificate across clusters. Clusters 1 to 4 show an increasing trend in the fraction of households owning a land certificate.]
The variation in quantity of fertilizers used across clusters
The variation in quantity of pesticides used across clusters

Cluster of increasing crop sales

- Cluster 1: Quantity of pesticides used is low.
- Cluster 2: Quantity of pesticides used is moderate.
- Cluster 3: Quantity of pesticides used is lower.
- Cluster 4: Quantity of pesticides used is significantly higher.
Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at **variation of identified variables across clusters** in order to make policy recommendations.

Third, variation in **outcomes**
The variation in crop sales across clusters
The variation in expenditure across clusters

Expenditure (in Shillings)

Clusters of increasing crop sales

1
2
3
4

800,000
600,000
400,000
200,000
0
The variation in children education across clusters.
The variation in households without food deficiency across clusters

![Bar chart showing the fraction of households without food deficiencies across clusters of increasing crop sales. Clusters 1, 2, 3, and 4 are represented with red, purple, blue, and green bars, respectively. The chart indicates an increase in the fraction of households without food deficiencies as the clusters move from 1 to 4.]

Sources: World Bank Group
Optimizing income in a cluster
# How to maximize income within a cluster

<table>
<thead>
<tr>
<th>Most Impactful Input....</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase number of days for which workers are hired</td>
<td>Decrease crop diversification</td>
<td>Increase number of days for which workers are hired</td>
<td>Decrease crop diversification</td>
<td></td>
</tr>
<tr>
<td><strong>...which is highly correlated with these inputs</strong></td>
<td><strong>Increase number of workers</strong></td>
<td><strong>No other variables</strong></td>
<td><strong>Increase number of workers</strong></td>
<td><strong>No other variables</strong></td>
</tr>
<tr>
<td>Input Coefficient</td>
<td>0.29</td>
<td>0.17</td>
<td>0.136</td>
<td>0.399</td>
</tr>
<tr>
<td><strong>Input Coefficient Interpretation</strong></td>
<td>Every average increase of 9 days for which workers are hired, is predicted to have 135 451 shillings increase in average income in cluster</td>
<td>Every 3.5 average decrease in types of crops planted, is predicted to have 91 925 shillings increase in average income in cluster</td>
<td>Every 17 additional workers hired, on average, is predicted to have 270 470 shillings increase in average income in cluster</td>
<td>Every average decrease of 17 in types of crops planted, is predicted to have 1 369 790 shillings increase in average income in cluster</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Impactful Input....</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase amount of pesticide used</td>
<td>Increase amount of pesticide used</td>
<td>Decrease crop diversification</td>
<td>Increase number of tools used</td>
<td></td>
</tr>
<tr>
<td><strong>.... which is highly correlated with these inputs</strong></td>
<td><strong>Use of chemical fertilizer</strong></td>
<td><strong>Use of chemical fertilizer</strong></td>
<td><strong>Use of chemical fertilizer</strong></td>
<td></td>
</tr>
<tr>
<td>Input Coefficient</td>
<td>0.262</td>
<td>0.153</td>
<td>0.125</td>
<td>0.206</td>
</tr>
<tr>
<td><strong>Input Coefficient Interpretation</strong></td>
<td>Every 1577 kg increase in pesticide used on average, is predicted to have 122 373 shillings increase in average income in cluster</td>
<td>Every 894 kg increase in pesticide used on average, is predicted to have 82 732 shillings increase in average income in cluster</td>
<td>Every 5.5 average decrease in types of crops planted, is predicted to have 248 594 shillings increase in average income in cluster</td>
<td>Every average increase of 4.5 in tools owned, is predicted to result in 707 210 shillings increase in average income in cluster</td>
</tr>
</tbody>
</table>
The variation in households using credit services across clusters.

Cross-country comparisons.
Cross-country comparison is limited by:

- Comparable data availability with low level of missingness across countries
- Difficulty to extract the same indicators across countries

**EXAMPLE:** Number of hired workers strong characteristic of highest cluster for Ethiopia

- Yet, in Tanzania’s LSMS-ISA data set it is only measured as a binary (have you hired workers? yes/no, as opposed to number of workers hired)
- This does not allow us to measure the intensity margins observed with Ethiopia data
Country cluster comparison:
# households per cluster

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>Uganda</td>
<td>Ethiopia</td>
<td>Tanzania</td>
</tr>
<tr>
<td>176</td>
<td>453</td>
<td>467</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Uganda</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>193</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ethiopia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>565</td>
<td>288</td>
</tr>
</tbody>
</table>
Country cluster comparison: distance to road (km)

Ethiopian farmers harder to reach across the board; data for Uganda not available.

- **Cluster 1**: Tanzania - 3, Ethiopia - 12
- **Cluster 2**: Tanzania - 2, Ethiopia - 18
- **Cluster 3**: Tanzania - 3, Ethiopia - 12
- **Cluster 4**: Tanzania - 2, Ethiopia - 12
Country cluster comparison: distance to market (km)

...and also harder for Ethiopian farmers to reach markets; data for Uganda not available
Country cluster comparison: Age of household heads

Cluster 1: Tanzania 47, Ethiopia 47
Cluster 2: Tanzania 46, Ethiopia 49
Cluster 3: Tanzania 46, Uganda 47, Ethiopia 47
Cluster 4: Tanzania 50, Uganda 51, Ethiopia 47
Country cluster comparison:
% households that are male
Country cluster comparison:
% households polygamous

Some differences reflect the broader cultural differences between the 3 countries, such as levels of polygamy.
Country cluster comparison: # household members

Other cluster characteristics are similar across countries with increases at higher clusters.
Country cluster comparison:
% household members who have attended school

- Tanzania: 65%, 62%, 65%, 60%
- Uganda: 69%, 75%, 79%, 87%
- Ethiopia: 32%, 30%, 29%, 36%

Clusters:
1. Cluster 1: Tanzania 65%, Uganda 69%
2. Cluster 2: Tanzania 62%, Uganda 75%
3. Cluster 3: Tanzania 65%, Uganda 79%
4. Cluster 4: Tanzania 60%, Uganda 87%
Country cluster comparison:
Number of crops planted

<table>
<thead>
<tr>
<th>Country</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>6.8</td>
<td>5.0</td>
<td>5.4</td>
<td>2.8</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.3</td>
<td>3.0</td>
<td>2.9</td>
<td>4.3</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>3.7</td>
<td>10.8</td>
<td>12.4</td>
<td>15.6</td>
</tr>
</tbody>
</table>
Country cluster comparison: % who have land certificates

Some inputs point to underlying differences in agricultural practices between countries.

- **Tanzania**: 3% (Cluster 1)
- **Uganda**: 44% (Cluster 1)
- **Ethiopia**: 8% (Cluster 2)
- **Tanzania**: 7% (Cluster 3)
- **Uganda**: 69% (Cluster 3)
- **Ethiopia**: 67% (Cluster 3)
- **Tanzania**: 12% (Cluster 4)
- **Uganda**: 71% (Cluster 4)
- **Ethiopia**: 62% (Cluster 4)
Country cluster comparison: pesticide and fertilizer use

- **Cluster 1**: Tanzania (179), Uganda (37), Ethiopia (23)
- **Cluster 2**: Tanzania (281), Uganda (30), Ethiopia (7)
- **Cluster 3**: Tanzania (519), Uganda (162), Ethiopia (378)
- **Cluster 4**: Tanzania (982), Uganda (516), Ethiopia (344)
Country cluster comparison:
% of households who use irrigation

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>3.3%</td>
<td>5.2%</td>
<td>2.1%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Uganda</td>
<td>3.9%</td>
<td>2.9%</td>
<td>7.5%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2.5%</td>
<td>2.1%</td>
<td>7.6%</td>
<td>2.6%</td>
</tr>
</tbody>
</table>
Country cluster comparison:
% with bank account
Country cluster comparison:
% without food deficiencies

Tanzania: CLUSTER 1 (51%), CLUSTER 2 (53%), CLUSTER 3 (57%), CLUSTER 4 (67%
Uganda: CLUSTER 1 (69%), CLUSTER 2 (69%), CLUSTER 3 (79%), CLUSTER 4 (89%
Ethiopia: CLUSTER 1 (47%), CLUSTER 2 (76%), CLUSTER 3 (77%), CLUSTER 4 (84%)

WORLD BANK GROUP
<table>
<thead>
<tr>
<th>Most impactful input: Comparison across countries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLUSTER 1</strong></td>
</tr>
<tr>
<td>Most Impactful Input in Ethiopia</td>
</tr>
<tr>
<td>Most impactful input in Tanzania</td>
</tr>
<tr>
<td>Most impactful input in Uganda</td>
</tr>
<tr>
<td>Other Impactful Input in Ethiopia</td>
</tr>
<tr>
<td>Other impactful input in Tanzania</td>
</tr>
<tr>
<td>Most impactful input in Uganda</td>
</tr>
</tbody>
</table>

**World Bank Group**
Summary
Summary

We found a robust clustering of farmers in all 3 countries

- Characteristics associated with clustering in each country differ dramatically
- Clusters can be described as different phases of the agricultural transformation process
- Describes a pathway towards agricultural transformation
- Each inputs suggest a prioritized policy action at different phase of the transformation process
Most impactful input differs significantly between countries.

Reasons include:

- Differences in correlations between inputs and outcomes
- Differences in farmer characteristics
- Differences in data
- Differences in underlying characteristics of population
Summary

Cross-country comparisons limited by lack of common measurement of some key inputs.

Yet, some patterns emerge:

► clustering analysis clearly shows that different farmers profiles exist across countries, suggesting to design cluster level policies

► inputs which are the most impactful of an increase in crop sales vary across clusters, supporting the implementation of cluster-level policies, rather than population level policies

► across countries, most predictive variables are hiring workers, usage of fertilizers or pesticides, animals, tools, irrigation, or animals; yet their relative importance across clusters (i.e., along income distribution) vary across countries

► interestingly the impact of crop diversification differs across country. Further analysis is required to show which specific crop leads to an increase in farmers competitiveness across countries