Typology Characterization in Ethiopia

April 6th 2016

Sara Signorelli

Introduction

Africa RISING is testing alternative technology options with heterogeneous populations of farmers that will likely respond to the technologies differently. Creating farm typologies is one approach to design targeted interventions that adequately address the needs of different types of farmers. Notably, creating typologies can help:

- **Identify suitable farms to target innovations (ex-ante)**: we assume that not all innovations are appropriate for all farms, and that structuring into groups would support the identification of technology-specific suitable farming systems.
- **Scale out innovations**: on the basis of the heterogeneity in a population we can formulate extension messages, policies and other incentive schemes to further spread the use of designed innovations.
- **Assess agro-economic effects (ex-post)** Explaining trends and farmer 'behavior' (functional characteristics, including sustainable intensification indicators) and verification of the agro-economic effects of the interventions for different farm types.

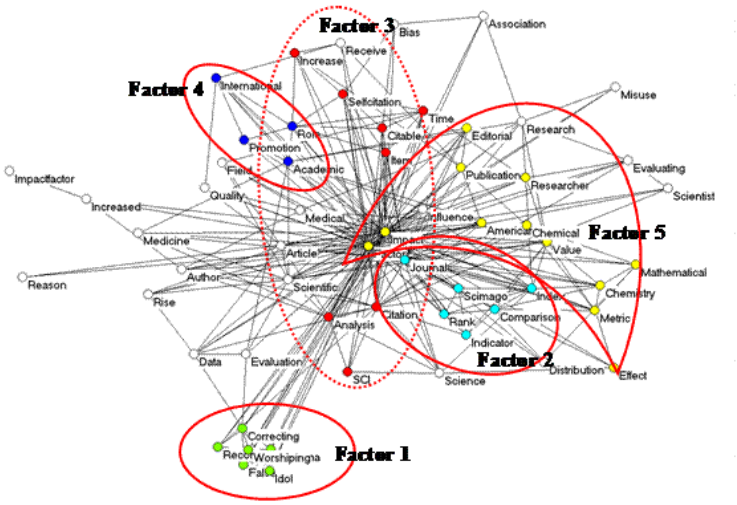
This document presents a summary of a typology study done using quantitative statistical methods (discussed below) applied to micro data from the Ethiopia Africa RISING Baseline Evaluation Survey (EARBES) (conducted in 2014) and secondary data on environmental/biophysical variables from various source. The quantitative approaches have the advantage that they are reproducible and do not impose any ex-ante structure to the clustering process, while more qualitative approaches can potentially incorporate less tangible insights such as cultural patterns. Once the different farm types are identified through systematic quantitative analysis, they need to be validated with input from Africa RISING colleagues (especially working in Ethiopia).

Methodological steps

We apply a combination of factor and cluster analysis to obtain the final groups, or "types" (See Cunningham & Maloney, 1999 for an empirical application). We first use factor analysis to reduce the number of socio-economic variables to characterize the farms by selecting the most relevant ones in differentiating the sample. Factor analysis is often used to discover underlying patterns in data and its aim is to explain the largest portion of the entire dataset variation with the lowest possible number of factors. Factors are unobserved variables that summarize the correlation among several observed variables and factor analysis allows us to divide the dataset into different factors, or dimensions, and categorize each variable into one of the factors. Figure 1 shows an example of how the variables in a dataset are divided into different dimensions to explain the total variation in the data. The analysis

also allows us to rank the factors by their importance in explaining the variation in the data and to further rank each variable by its explanatory power within the factor.

Figure 1: Example of factor analysis[1]



Our factor analysis based on EARBES data involves the following main steps (see for example McDonald; 2014. Basilevsky; 2009. Mulaik; 2009 for a discussion on the methods):

1. We divide the variables in EARBES into the five domains of sustainability that have been identified within Africa RISING to gauge progress: **productivity, economic, environment, social and human**.
2. We perform separate factor analysis on each domain to select the variables that explain the largest portion of the variation in the data.
3. We use scree plots to define the number of factors to look at and, within each of the selected factors, we consider the two variables with the highest absolute values of factor loads, conditional on them being greater than 0.5 (or smaller than -0.5).
4. Finally, we obtain a parsimonious set of socio-economic variables that explain most of the variation in the data and thus are highly relevant in defining the different farm types.

The sub-set of variables obtained using steps (1) to (4) are used to perform a cluster analysis, which divides the total sample into a chosen number of clusters (Kaufman & Rousseeuw; 2009. Romesburg; 2004. Galbraith et Al.; 2002). The numbers of clusters are chosen in order to represent groups that are different enough from each other while ensuring that each group to be included has a sufficient amount of observations. There are several different methods to perform cluster analysis, some hierarchical and some non-hierarchical. We chose the hierarchical method using medians, where the distance between two clusters is calculated as the median distance between all pairs of subjects in the two clusters. The results obtained and the characteristics of each group formed are reported in the next section.

[1] http://www.leydesdorff.net/words/

Results

1. Factor analysis of productivity variables (Sustainability Domain 1)

The scree plot of the factorization of the productivity variables (Figure 2) shows that the first four factors (represented by the first four dots at the top of the line graph) are highly relevant but starting from the 5th factor they start to be less important in explaining the variation (smaller vertical jump).

Figure 2: Scree plot of productivity variables

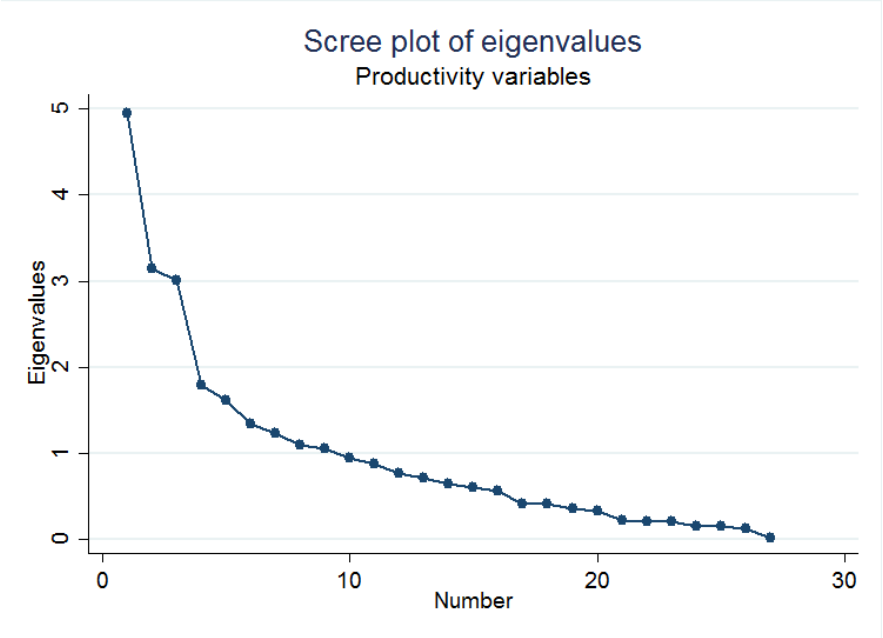


Table 1 shows the rotated matrix of factor loads for the four factors we have chosen, with the relevant variables highlighted (>0.5 or <-0.5). Factor 1 captures elements related to legumes production and number of cultivated plots. Factor 2 captures total land size, production of cereals and breeding of big ruminants. Factor 3 captures intercropping practices and, finally, Factor 4 captures diversity of livestock ownership and poultry TLU. The final selection of variables for the cluster analysis includes the area cultivated with legumes and Kg of legumes' production for factor 1, land size and Kg of cereals' production for factor 2, the share of households practicing intercropping and the average number of intercropped plots for factor 3, and finally the share of households breeding mixed livestock and the number of livestock types possessed for factor 4.

Table 1: Factor loads of productivity variables

Variable	Factor1	Factor2	Factor3	Factor 4
Land size (Ha)	-0.0859	0.8913	-0.0139	0.014
N. parcels	0.4537	0.502	-0.0574	0.0466
Min distance plot	-0.2253	0.2152	-0.0255	-0.2359
Max distance plot	0.021	0.285	-0.0934	-0.0731
N. trees	0.1585	0.0663	0.0065	0.0482
N. crops	0.0921	0.0627	0.2985	0.1432
N. plots	0.6316	0.132	0.1007	0.278
HH does intercropping	0.0248	-0.0006	0.9275	0.0442
HH does intercropping with legumes	0.0436	-0.025	0.2789	0.0187
N. of intercropped plots	0.0198	-0.0069	0.9309	0.0716
Size intercropped land (Ha)	0.0403	0.0619	0.8507	-0.0538
Size legumes-intercropped land (Ha)	0.0337	-0.0224	0.1984	-0.0042
Ownership mixed livestock	0.0638	0.1064	-0.0072	0.7619
N. livestock types owned	0.3239	0.1146	0.0416	0.8259
Cultivation of cereals	0.1037	0.1217	-0.0363	0.141
Cultivation of legumes	0.7193	-0.2707	0.0797	0.1912
Area cultivated with cereals (Ha)	-0.1172	-0.0806	-0.0053	-0.0253
Area cultivated with legumes (Ha)	0.7746	-0.045	0.018	0.0701
Production cereals (Kg)	-0.0663	0.8591	0.0517	-0.0633
Production legumes (Kg)	0.8487	0.0385	0.0033	0.0951
Yield cereals (Kg/Ha)	0.1232	0.2342	0.0593	-0.0285
Yield legumes (Kg/Ha)	0.6531	-0.101	0.0444	0.1555
TLU small ruminants	0.332	0.2878	-0.0526	0.3973
TLU big ruminants	0.0423	0.7605	0.0505	0.4111
TLU poultry	0.1102	-0.1514	0.1753	0.5989
Fertilizer used (Kg)	0.4336	0.0584	0.0033	0.1878
HH does irrigation	-0.0151	-0.2477	0.08	0.0848

Note: "N" stands for number. "HH" stands for household. "TLU" stands for Tropical Livestock Units

2. Factor analysis of economic variables (Sustainability Domain 2)

For the economic variables we considered, the relevant factors seem to be the first three (Figure 3). Table 2 shows that factor 1 captures total harvest and its uses; factor 2 captures agricultural non-labor inputs; and factor 3 captures labor inputs. Dwelling conditions and the wealth indices does not play a significant role in differentiating the sample. The final list of variables considered includes total harvest of grains and Kg of harvest used for other reasons (factor 1), pesticide and fertilizer costs (factor 2), and total and male person days used in agriculture (factor 3).

Figure 3: Scree plot of economic variables

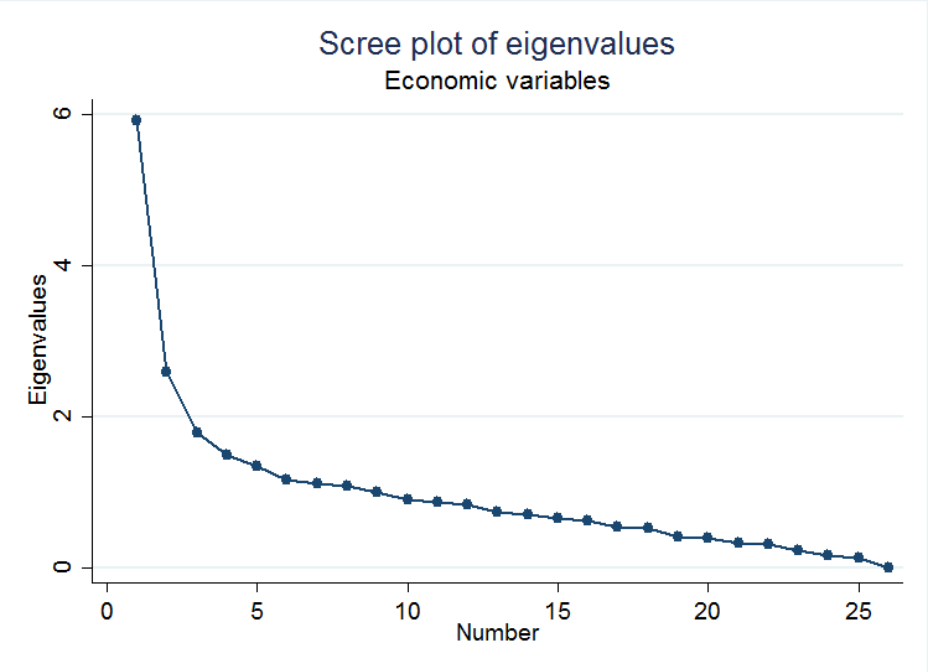


Table 2: Factor loads of economic variables

Variable	Factor1	Factor2	Factor 3
Fertilizer cost	0.1574	0.7759	0.2295
Traditional seeds cost	0.0629	0.0237	-0.0359
Improved seeds cost	0.069	0.6069	0.1066
Pesticide cost	0.2026	0.8525	0.009
Other non-labor cost	0.4473	0.6282	-0.0936
Animal feed cost	0.178	0.5999	0.0002
Agricultural wage	-0.0551	0.1149	-0.0123
HH uses community labor	0.1167	-0.1173	0.3623
HH uses hired labor	0.0601	-0.0245	0.2183
Total PD used for crops	0.1262	0.072	0.9587
Male PD used for crops	0.1812	0.1624	0.8759
Female PD used for crops	-0.0806	-0.1997	0.6944
Total harvest of grains (Kg)	0.8528	0.2324	0.1872
Total harvest of stover (Kg)	0.0113	-0.0512	0.0366
Total harvest used for animal feed (Kg)	0.1334	0.2072	-0.0724
Total harvest used for crop residual (Kg)	0.0501	0.1692	0.0329
Total harvest used for seeds (Kg)	0.7146	0.3495	0.1682
Total harvest used for gifts (Kg)	0.6659	0.0439	-0.0411
Total harvest used for own consumption (Kg)	0.691	0.1046	0.3031
Total harvest sold (Kg)	0.1322	-0.048	0.0135
Total harvest used for other reasons (Kg)	0.7892	0.1161	0.051
Agri wealth index	0.3401	0.4731	0.4789
Non-agri wealth index	0.4143	0.4058	0.061
Good floor material in dwelling	0.026	0.1619	0.0299
Good source of drinking water	-0.0382	0.1804	-0.038
Good lightning source	0.388	0.0519	-0.2437

Note: "HH" stands for household and "PD" refers to person-days.

3. Factor analysis of environment variables (Sustainability Domain 3)

For the environment domain, we identified four relevant factors. The first concerns the characteristics of the soil, the second captures fallowing and the issues related to soil erosion, the third includes crop rotation and the use of manure, and the fourth includes irrigation practices and the use of urea.

Figure 4: Scree plot of environment variables

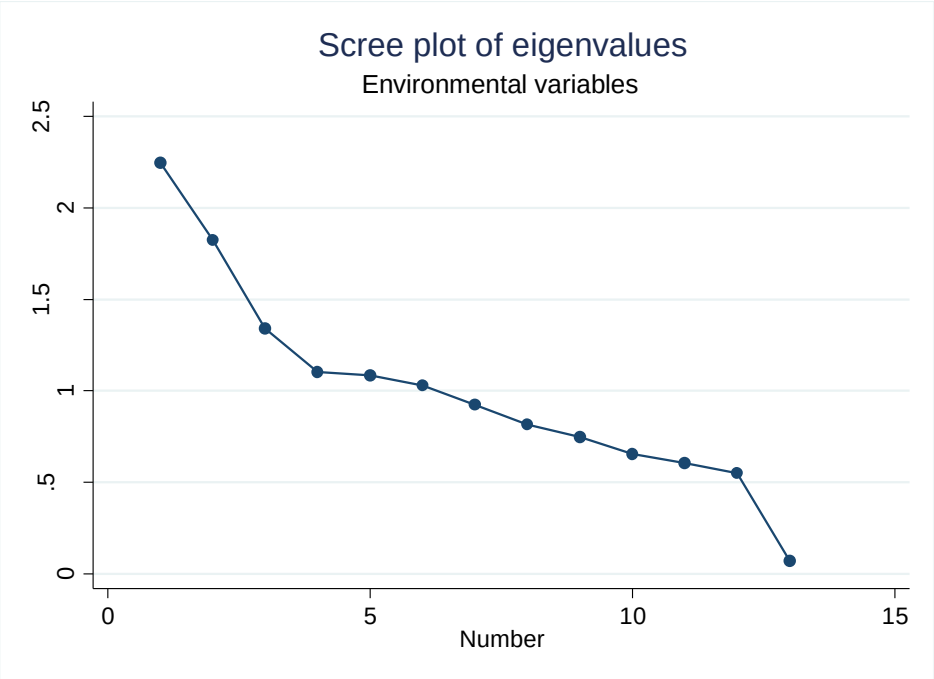


Table 3: Factor loads of environment variables

Variable	Factor1	Factor2	Factor3	Factor4
HH uses irrigation	-0.0296	-0.0224	0.1115	0.8064
HH uses crop rotation	0.1408	0.0309	0.6732	0.0178
HH uses fallowing	0.0917	0.6005	0.228	0.0264
HH uses alternative tillage	0.0092	-0.0075	-0.012	-0.0029
HH uses manure	0.0583	-0.0267	0.7947	0.1248
HH uses urea	-0.0396	0.1487	0.0188	0.6288
HH experiences soil erosion	-0.0716	0.5923	0.2603	0.2772
HH experiences soil erosion and does not takes any preventive measure	-0.0219	0.8249	-0.1619	-0.0455
Share of parcels with clay or loam soil	0.945	-0.0175	0.0629	-0.0506
Share of parcels with black or brown soil	0.5896	-0.0752	-0.2879	0.1862
Share of parcels with incrusted soil	0.9521	0.0222	0.0784	-0.02
Number of leguminous trees owned	-0.1011	0.0298	0.1876	-0.1784
Number of fruit trees owned	0.0178	0.0116	-0.0179	0.2134

4. Factor analysis of social variables (Sustainability Domain 4)

Our dataset has a relatively small set of variables capturing social aspects, focusing on gender disparities. We thus chose only the first factor, which highlights the presence of females-only managed plots and livestock as the main variables of interest.

Figure 5: Scree plot of social variables

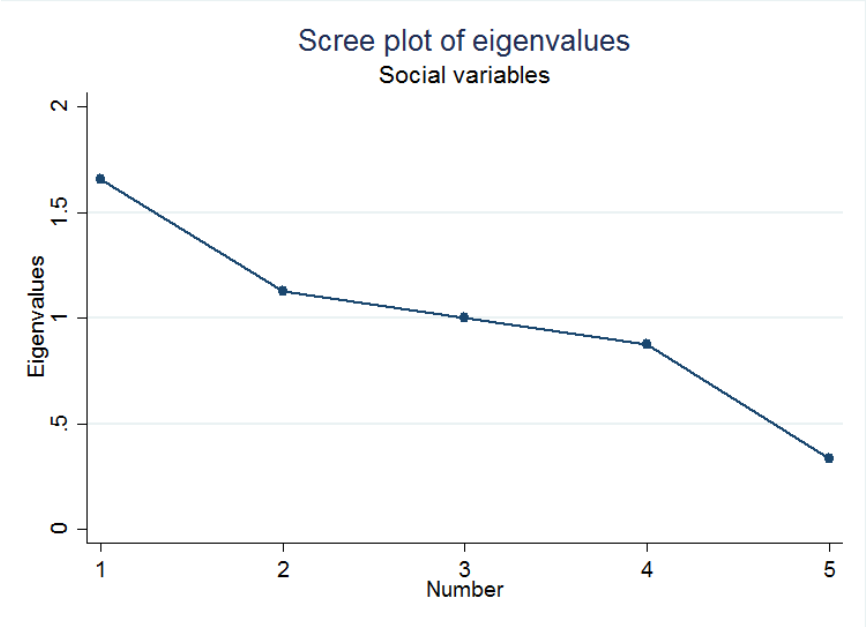


Table 4: Factor loads of social variables

Variable	Factor1
Females also responsible for plots	0.5138
Females only responsible for plots	0.9093
Females also responsible for livestock	-0.0902
Females only responsible for livestock	0.6769

5. Factor analysis of human variables (Sustainability Domain 5)

The final sustainability domain we focus on is human capital. We select the first four factors, which capture the age composition of household members (factor 1) the head's years of education, mean age in the household, total dependency ratio and food insecurity level (factor 2), the main characteristics of the household head (factor 3) and the level of education in the household (factor 4). We finally select old dependency ratio and share of members between 0 and 14 years old (factor 1), number of males adults and mean adult's age in the household (factor 2), whether the household head is married or widow (factor 3), and whether the head is literate and maximum years of education in the household.

Figure 6: Scree plot of human variables

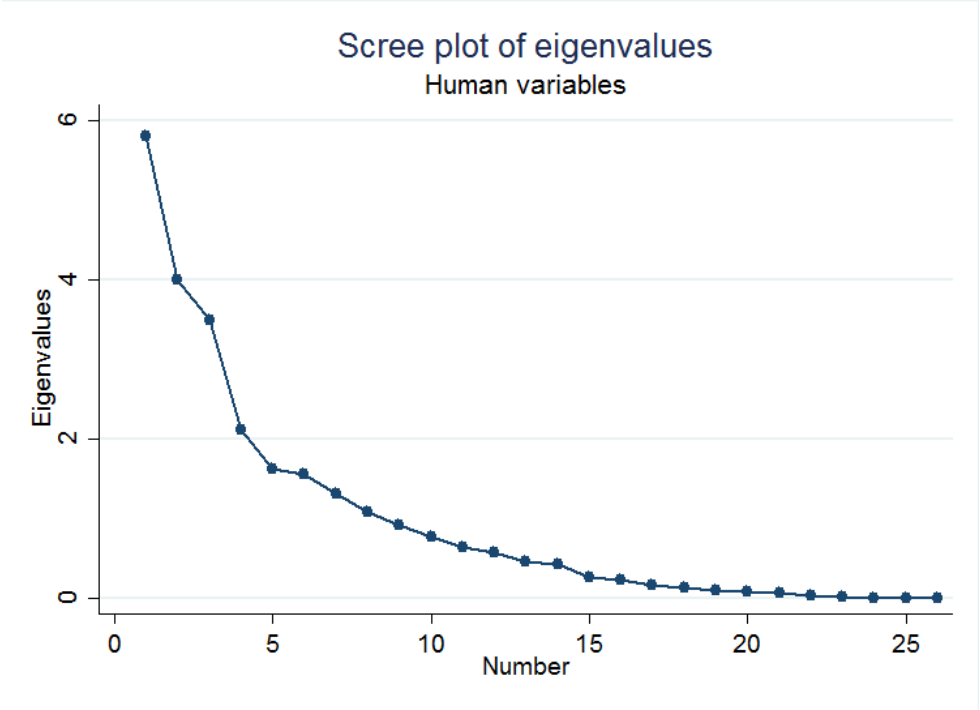


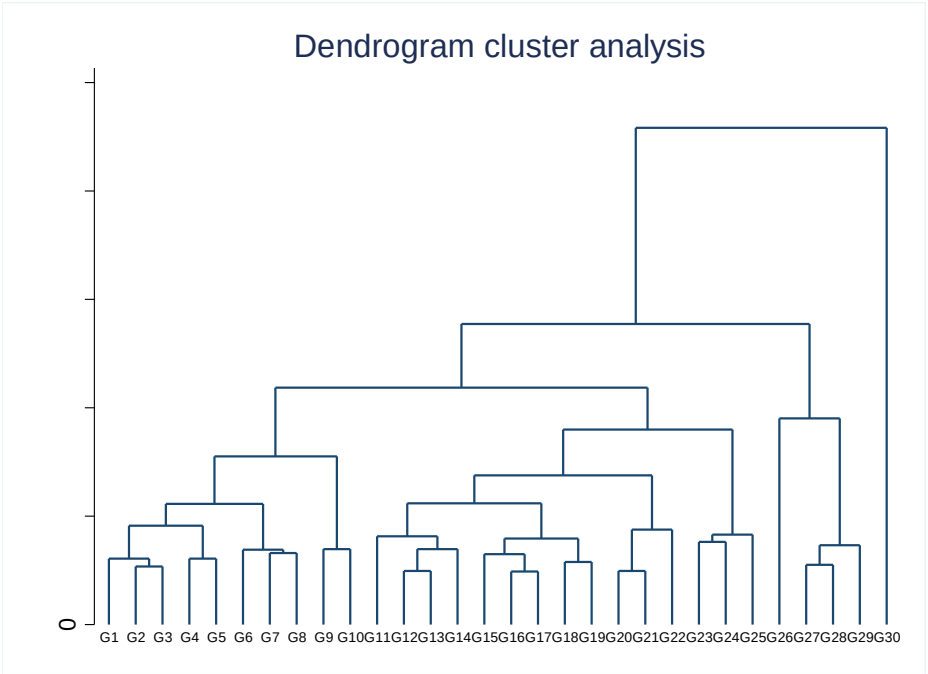
Table 5: Factor loads of human variables

Variable	Factor 1	Factor 2	Factor 3	Factor 4
HH size	0.2877	-0.1748	-0.2961	0.1585
Head is married	0.0571	0.0128	-0.8254	0.0399
Head is widow	-0.0489	-0.027	0.9165	-0.0655
Head is single	-0.0499	-0.0054	0.0962	0.0085
Head is female	-0.0002	-0.0576	0.8021	-0.0752
Head is male and single	-0.0208	-0.0623	0.1022	-0.0558
Head's age	-0.0652	0.1043	0.0108	0.1211
Head's years of educ	-0.3705	0.5736	-0.0547	-0.1477
Head is literate	0.135	-0.1051	-0.1697	0.8313
Mean years of edu.	0.1619	-0.0372	-0.2973	0.6144
Highest years of edu.	-0.2209	-0.2179	0.0521	0.8333
Mean age	-0.3636	-0.1004	-0.0026	0.7073
Mean adult age	-0.5511	0.8057	0.0064	-0.0756
N. of males adults	0.1426	0.9165	-0.097	-0.1535
N. of females adults	-0.3806	-0.0333	-0.3678	0.1112
children	-0.2747	0.1248	0.0863	0.1349
Young dep. Ratio	0.6733	-0.2693	-0.1673	0.0644
Old dep. Ratio	0.9452	-0.1699	0.0087	-0.0671
Total dep. ratio	0.0902	0.7564	0.1156	0.0565
Share of 0-14 y.o.	0.9603	0.0353	0.0397	-0.0513
Share of 15-29 y.o.	0.9023	-0.2925	-0.0605	-0.0427
Share of 30-44 y.o.	-0.75	-0.2984	0.1366	0.0789
Share of > 45 y.o.	0.2163	-0.15	-0.0526	0.193
HH worries for food shortages	-0.3834	0.8	-0.0415	-0.1928
Months experienced food shortages	0.1176	-0.0999	0.0731	0.0111

6. Cluster analysis

The analysis summarized in the preceding section informed the selection of a list of factors that we used in the cluster analysis. These are 8 productivity variables, 6 economic variables, 8 environmental variables, 2 social variables and 8 human variables. Figure 7 shows the dendrogram illustrating how the farm households in our sample can be split into different groups (or types) based on these variables we have identified. The vertical distance between separations illustrates the distance of the different groups to each other.

Figure 7: Dendrogram



Considering the number of observations within each group and differentiation of characteristics between groups, we decided to create three final groups, or "types" of farmers. Tables 6a to 6e illustrates the distribution of characteristics across these types and sustainability domains discussed before. Because the clusters were defined using the variables accounting for most of the data variation, as captured by the factor analysis, most of the characteristics differ significantly across every type. Type 1 includes 188 of the farmers in the sample. Type 2 is the biggest one and defines 229 farmers. Finally, type 3 is the smallest, with 67 farmers.

Table 6a: distribution of characteristics by type in the productivity domain

	Type 1	Type 2	Type 3
Productivity Domain			
Total land size (Ha)	1.23***	2.25*	6.62***
	[0.08]	[0.10]	[0.41]
Share of households doing intercropping	0.03	0.04	0.03
	[0.01]	[0.01]	[0.02]
Area of intercropped plots	0.01	0.02	0.04
	[0.01]	[0.01]	[0.03]
Share of households owning mixed livestock	0.81***	0.95***	0.9
	[0.03]	[0.01]	[0.04]
N. of different livestock types owned	2.66***	3.58***	2.99
	[0.10]	[0.08]	[0.14]
Share of households growing cereals	0.87***	1.00***	1.00**
	[0.02]	[0.00]	[0.00]
Share of households growing legumes	0.53	0.65***	0.22***
	[0.04]	[0.03]	[0.05]
Area of cereals(ha)	0.02	0.01	0.01
	[0.01]	[0.00]	[0.00]
Area of legumes(ha)	0.18	0.24***	0.10***
	[0.02]	[0.02]	[0.03]
Production of cereals(kg)	309.80***	1820.92**	8856.57***
	[26.40]	[68.94]	[581.28]
Production of legumes(kg)	74.04***	257.38***	129.4
	[12.63]	[23.58]	[39.89]
Yield of cereals(kg/ha)	1244.15***	1880.9	2680.18***
	[81.71]	[60.86]	[152.70]
Yield of legumes(kg/ha)	831.94***	1363.34***	1322.15
	[97.48]	[72.93]	[249.26]
TLU small ruminants	0.31***	0.49	0.63***
	[0.03]	[0.04]	[0.13]
TLU big ruminants	1.60***	3.22***	5.61***
	[0.10]	[0.10]	[0.38]
TLU poultry	0.03***	0.04***	0.02***
	[0.00]	[0.00]	[0.00]
Kg fertilizer used	1361.23**	2273.48***	1427.46
	[171.86]	[260.79]	[297.66]
N. of observations	188	229	67

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6b: distribution of characteristics by type in the economic domain

	Type 1	Type 2	Type 3
Economic Domain			
Value of fertilizer used (GHC)	894.66***	2399.77	5887.54***
	[65.21]	[120.15]	[640.97]
Value of traditional seeds purchased (GHC)	201.74	224.72	243.64
	[26.72]	[33.74]	[76.93]
Value of improved seed purchased (GHC)	73.66***	180.73	784.28***
	[13.96]	[28.81]	[204.15]
Value of pesticides used (GHC)	36.78***	161.33***	1855.81***
	[10.18]	[30.89]	[259.02]
Share of households using communal labor	0.54***	0.72***	0.7
	[0.04]	[0.03]	[0.06]
Share of households using hired labor	0.36***	0.50***	0.45
	[0.04]	[0.03]	[0.06]
Total person-days used, male & female	68.89***	123.55***	124.93***
	[3.39]	[4.39]	[8.18]
Total Kg of grains harvested	442.19***	2535.81	7986.89***
	[35.54]	[97.83]	[482.51]
Total Kg harvest used for own consumption	195.37***	847.91***	1577.37***
	[15.75]	[38.81]	[120.31]
Total Kg harvest sold	79.65***	676.24	2812.78***
	[15.80]	[64.32]	[285.32]
Agricultural wealth index	-0.61***	0.20***	1.06***
	[0.05]	[0.05]	[0.16]
Non-agricultural wealth index	-0.42***	0.01	1.11***
	[0.04]	[0.06]	[0.17]
Share of households with good floor in dwelling	0.03	0.04	0.10***
	[0.01]	[0.01]	[0.04]
Share of households with good source of drinking water	0.62	0.59*	0.81***
	[0.04]	[0.03]	[0.05]
Share of households with good source of lighting	0.28**	0.32	0.60***
	[0.03]	[0.03]	[0.06]
N. of observations	188	229	67

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6c: distribution of characteristics by type in the environment domain

	Type 1	Type 2	Type 3
Environment Domain			
Share of households practicing irrigation	0.13	0.14	0.1
	[0.02]	[0.02]	[0.04]
Share of households practicing rotation	0.89	0.93*	0.84*
	[0.02]	[0.02]	[0.05]
Share of households practicing fallowing	0.16	0.18	0.04***
	[0.03]	[0.03]	[0.03]
Share of households practicing alternative tillage	0.01*	0.03	0.06*
	[0.01]	[0.01]	[0.03]
Share of households using manure on (any) plot in either season	0.78	0.78	0.58***
	[0.03]	[0.03]	[0.06]
Share of households using urea on (any) plot in either season	0.06**	0.1	0.15*
	[0.02]	[0.02]	[0.04]
Share of households affected by soil erosion	0.5	0.50*	0.19***
	[0.04]	[0.03]	[0.05]
Share of households with soil erosion but no erosion control measure	0.06	0.07	0.01
	[0.02]	[0.02]	[0.01]
Average share of parcels with clay or loam soil	0.79***	0.86**	0.88
	[0.02]	[0.01]	[0.02]
Average share of parcels with black or brown soil	0.50***	0.59	0.75***
	[0.03]	[0.02]	[0.04]
Average share of parcels with incrusted soil	0.81***	0.88**	0.9
	[0.02]	[0.01]	[0.02]
Number of leguminous trees owned on the land	1.42**	4.19***	0.61**
	[0.35]	[0.74]	[0.30]
Number of fruit trees owned on the land	2	1.67	2
	[0.57]	[0.34]	[1.68]
N. of observations	188	229	67

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6d: distribution of characteristics by type in the social domain

	Type 1	Type 2	Type 3
Social Domain			
Share of HH with female having shared plot responsibility	0.53	0.54	0.71***
	[0.03]	[0.03]	[0.05]
Share of HH with female having exclusive plot responsibility	0.26***	0.10***	0.09*
	[0.03]	[0.02]	[0.03]
Share of HH with female having shared livestock responsibility	0.09***	0.14	0.22***
	[0.01]	[0.01]	[0.02]
Share of HH with female having exclusive livestock responsibility	0.03**	0.03	0.01**
	[0.00]	[0.00]	[0.01]
*Gender wage gap (wage women/wage men*100)*	81.05	126.36	74.55
	[14.65]	[62.62]	[7.49]
N. of observations	188	229	67

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

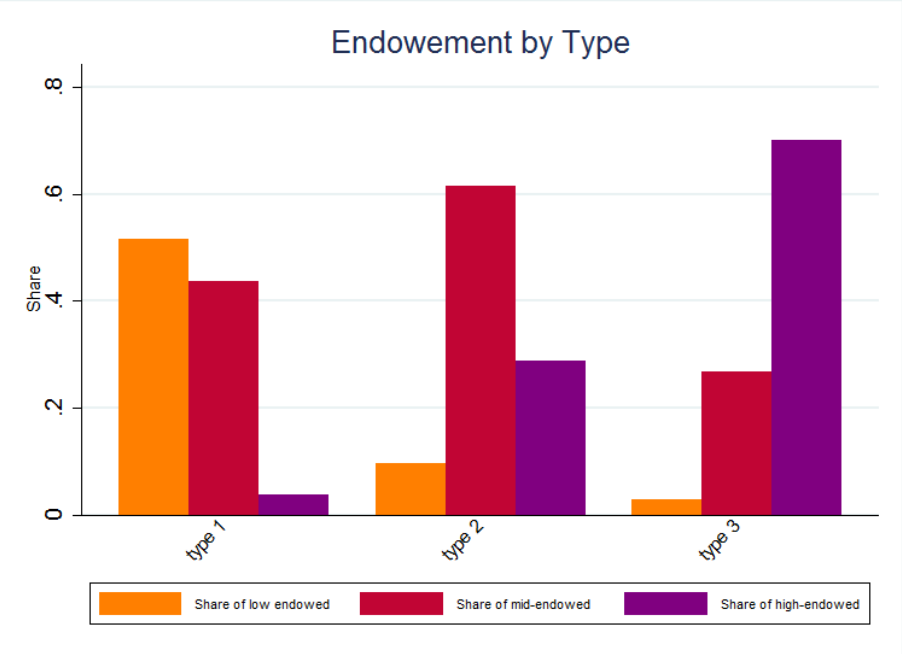
Table 6e: distribution of characteristics by type in the human domain

	Type 1	Type 2	Type 3
Human Domain			
Household size	5.35***	6.22*	7.28***
	[0.16]	[0.14]	[0.30]
Share of married heads	0.70***	0.87***	0.93**
	[0.03]	[0.02]	[0.03]
Share of female heads	0.32***	0.14***	0.09**
	[0.03]	[0.02]	[0.04]
Age of the head	44.62**	46.31	49.46**
	[0.78]	[0.80]	[1.51]
Years of education of the heads	3.4	3.4	3.51
	[0.26]	[0.24]	[0.37]
Share of literate heads	0.59**	0.66	0.73*
	[0.04]	[0.03]	[0.05]
Mean years of education in the household	4.06	4.23	4.62
	[0.18]	[0.17]	[0.23]
Max years of education in the household	6.61***	7.33	8.33***
	[0.25]	[0.23]	[0.32]
Average age of adults in the household	23.77	23.21	23.84
	[0.71]	[0.46]	[0.92]
Number of children in the household	0.76	0.79	0.85
	[0.06]	[0.06]	[0.12]
Young dependency ratio	1.01	0.98	0.87
	[0.06]	[0.05]	[0.10]
Old dependency ratio	0.08	0.06	0.07
	[0.02]	[0.01]	[0.02]
Share of HH worrying about food shortages	0.40***	0.17***	0.06***
	[0.04]	[0.03]	[0.03]
Months experiencing food shortages?	1.59***	0.64***	0.03***
	[0.14]	[0.10]	[0.03]
N. f observations	188	229	67

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

The three types differ from each other across all of the five domains, as shown in table 6. One of the striking characteristic that stands out in differentiating them is the level of endowments, as measured by a wealth index including dwelling characteristics, size of the cultivated land and ownership of agricultural and non-agricultural assets (figure 8). We defined low-endowed households as the ones in the bottom quartile of the wealth distribution, mid-endowed households as the ones in the 2nd and 3rd quartile and highly endowed households as the ones in the top quartile of the asset distribution. Figure 8 shows in which of the endowments category fall most of the households in our typologies.

Figure 8: Level of Endowments by Type



More broadly, the types can be characterized as following:

Type 1: Female headed households with low endowments

- Small female headed households with lower levels of education and high food insecurity.
- Generally low levels of gender equality.
- Small landholdings and little livestock owned.
- Little use of labor and non-labor inputs.
- Low levels of endowments and large portions of harvest going to own consumption.
- Some issues with soil erosion.

Type 2: Legumes growers with mid-levels of endowments

- Average levels of gender equality but wage gap in favor of women.
- Small landholdings but wide variety of livestock bred, often including poultry.
- Legumes growers with high legumes yields.
- Frequent employment of hired and communal labor.
- Mid-levels of endowments.
- Relatively frequent use of soil conservation practices but problems with soil erosion.

Type 3: Highly endowed households breeding large ruminants

- Large households with many children and high levels of education. High levels of food security.
- High gender equality in terms of responsibilities but severe wage gap.
- Large landholdings and ownership of big ruminants.
- High input expenditure, including in improved seeds.
- High levels of endowments and good dwelling conditions. High commercialization of the harvest.

- Relatively little use of soil conservation practices and very severe problems of soil incrustation.

Table 7 summarizes the main characteristics of every type relative to each sustainability domain, providing a simplified framework for classifying farm households into a particular type. Figure 9 shows a graphic representation of the main characteristics of each type.

Table 7: matrix of performance for each SI domain

	Productivity	Economic	Environment	Social (gender)	Human
Type 1: Female headed households with low endowments	Small land and little livestock. Low productivity.	Little inputs use. Low endowments. Harvest mostly used for own consumption.	Soil erosion issues.	Low levels of gender equality.	Female heads. Small household size. Food insecure.
Type 2: Legumes growers with mid-levels of endowments	Wide variety of livestock bred. Large production and productivity of legumes.	Frequent use of hired and communal labor. Mid-level of endowments. Half of the harvest going to own consumption.	Use of conservation practices but high soil erosion.	Average levels of gender equality and wage gap in favor of women.	Average levels of human endowments, including education and food security.
Type 3: Highly endowed households breeding large ruminants	Ownership of large ruminants. Large production and productivity, especially of cereals.	Large use of inputs. High levels of endowments. Good dwelling conditions. Commercialization of the harvest.	Low use of conservation practices and severe soil incrustation.	High levels of gender equality but severe wage gap.	Large households with children. High education. Food secure.

Figure 9: Graphic representation of types

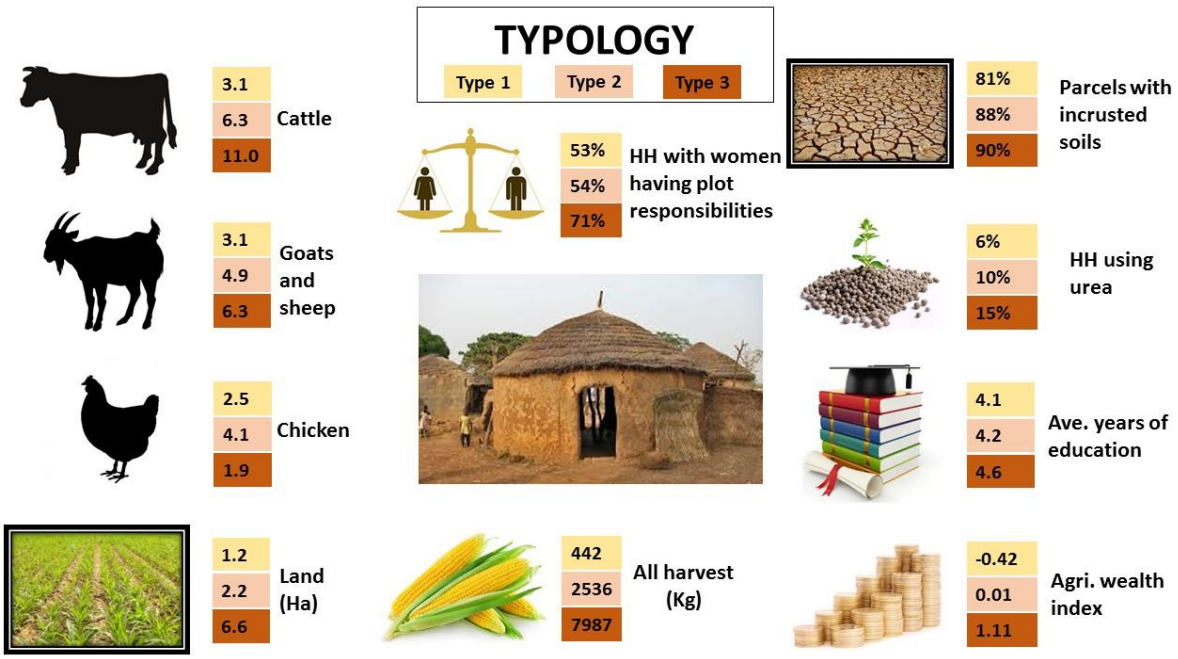
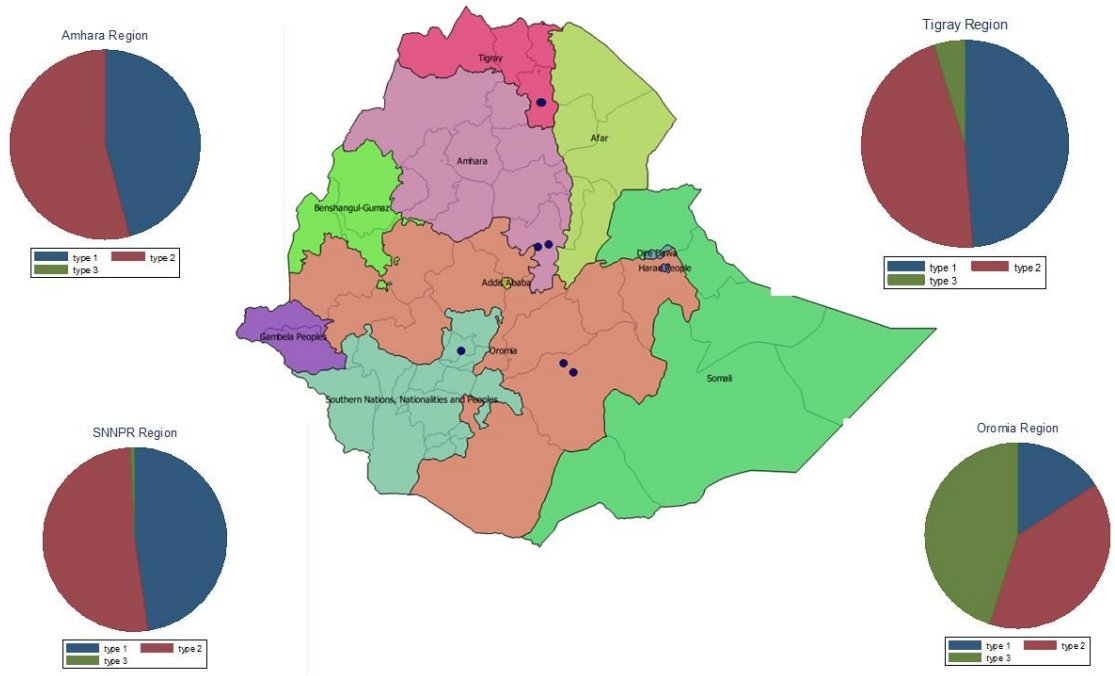


Figure 10: Distribution of Typologies by region



The typologies are heterogeneously distributed across space, as shown in figure 10. While in the SNNPR region and especially in Amhara there is a high concentration of households with low to medium levels of endowments (type 1 and 2), The Oromia region concentrates high shares of the highly endowed households (type 3). The spatial distinctions are important because they can support interventions based on the most prevalent households' typologies in the area.

The characteristics of each household type described above can be displayed clearly with a spider plot. Figure 11 summarizes the performance of each type relative by each domain as follows:

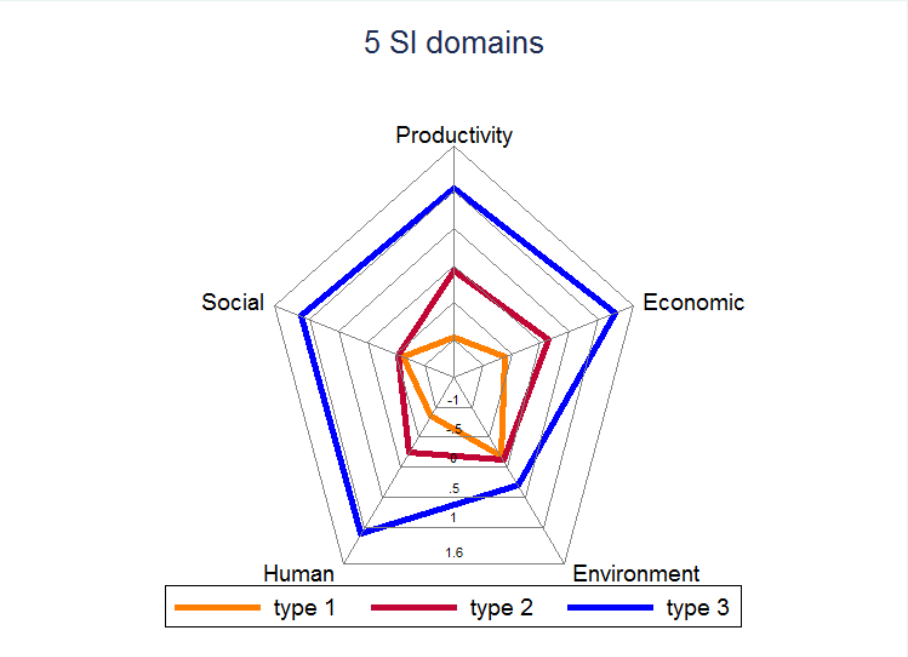
- Type 3 largely dominates in every aspect, but in the environment domain he's closer to the other two types.
- Type 1 and type 2 have similar performances in the social and environmental domains but type 2 is better endowed in terms of productive, economic and human assets.

Recommendations:

- Large efforts have to be made to improve the productive capacity and the economic resources available to group 1. This could be made through granting better access to superior agricultural technologies and trainings on how to use them to increase yields and livestock production. Better access to education can also contribute to improve their human endowments and may foster better gender inclusion.
- Africa RISING can focus on fostering the gender inclusiveness in group 2.
- Group 3 is performing very well across the five SI domains and could therefore be involved by the AR implementers to show the good example to the neighboring farmers.
- The three groups present severe problems of soil erosion and incrustation, therefore they would all benefits from interventions aiming at improving the quality of the soil.

The appendix includes additional graphs characterizing the obtained typologies.

Figure 11: Typologies performance by sustainability domain



NOTE: The following variables are used to measure each domain: cereals yield (Productivity), asset-based wealth index (Economic), soil conservation index composed of crop rotation, alternative or minimum/zero tillage, experience of soil erosion without measures for mitigating it and share of parcels with incrusted soils (Environment); gender equality index composed by female responsibility in managing certain plots (Social), and average education in the household (Human).

References

Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications* (Vol. 418). John Wiley & Sons.

Cunningham, W., & Maloney, W. F. (1999). Heterogeneity among Mexico's micro-enterprises: an application of factor and cluster analysis. *World Bank Policy Research Working Paper*, (1999).

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863-14868.

Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. CRC Press.

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago Press.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues* (Vol. 14). Sage.

McDonald, R. P. (2014). *Factor analysis and related methods*. Psychology Press.

Mulaik, S. A. (2009). *Foundations of factor analysis*. CRC press.

Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu.com.

Sethi, S. P. (1971). Comparative cluster analysis for world markets. *Journal of Marketing Research*, 348-354.

Appendix Figures

Figure A1: Typologies by domain (productivity and economic)

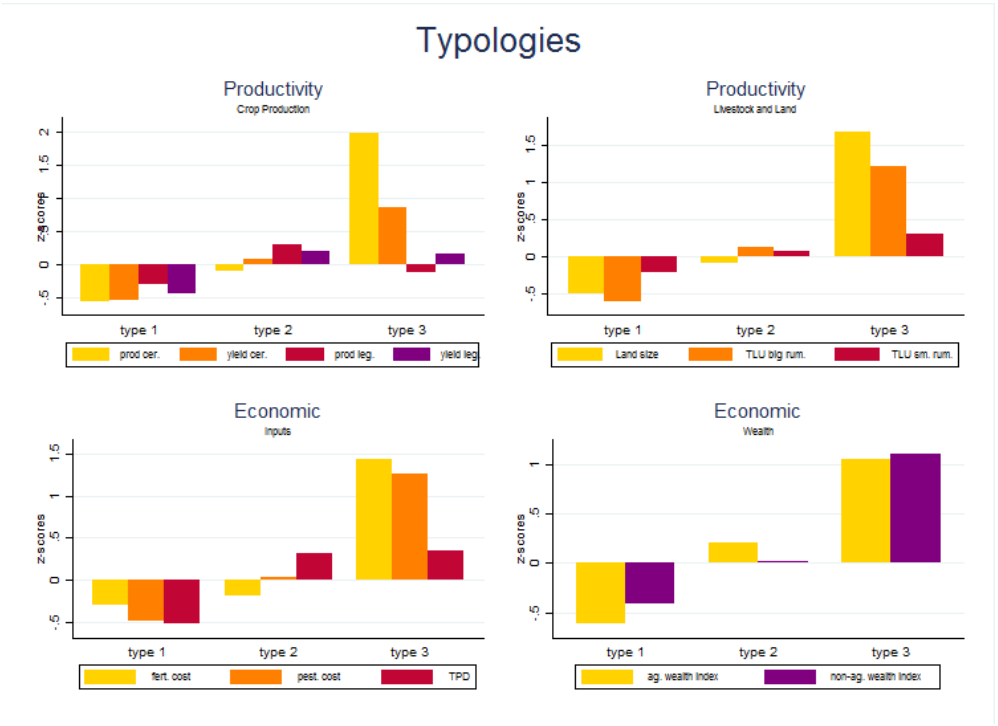


Figure A2: Typologies by domain (environment, social and human)

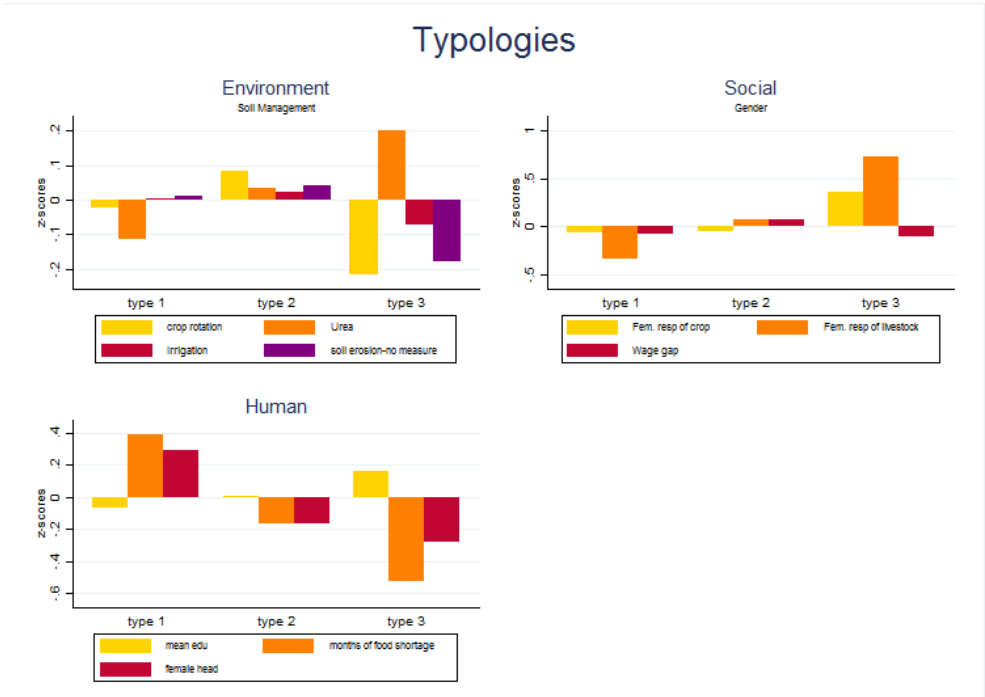


Figure A3: Radar graph – productivity (z-scores)

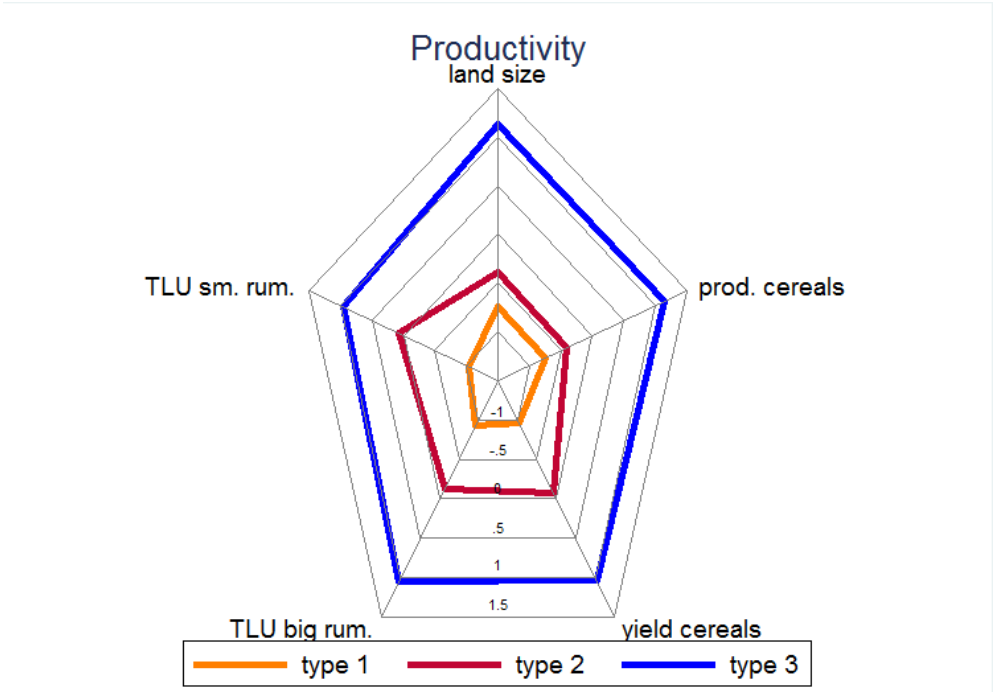


Figure A4: Radar graph – economic (z-scores)

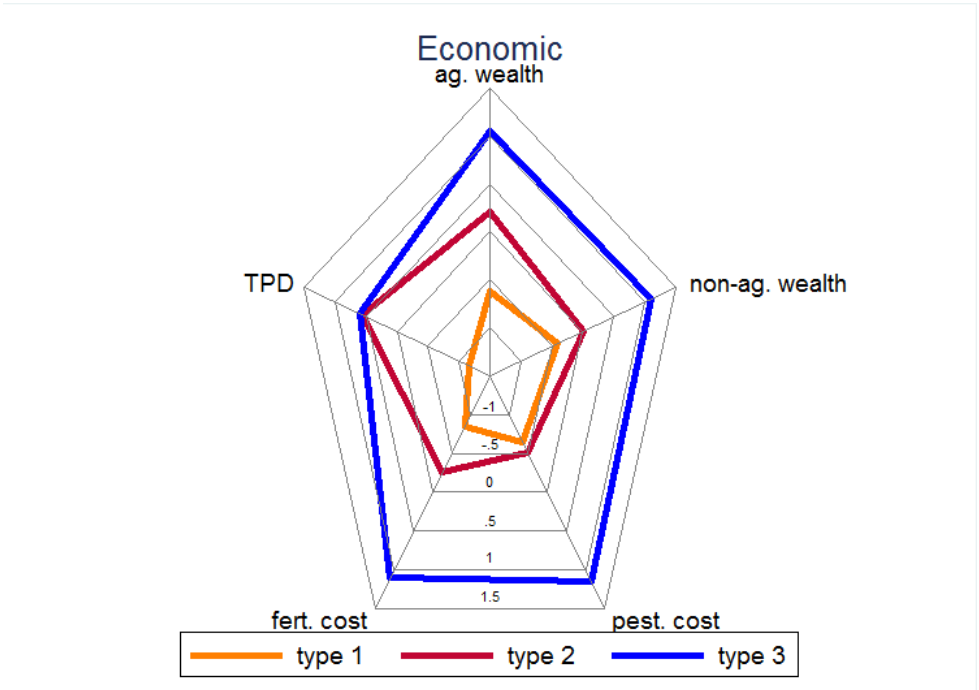


Figure A5: Radar graph – environment (z-scores)

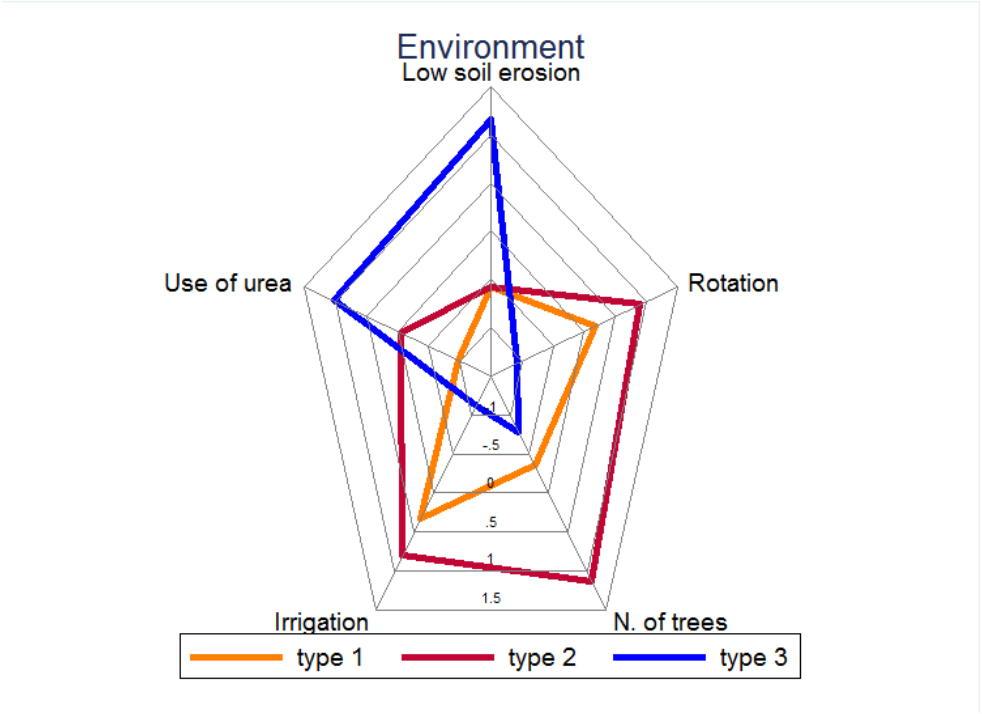


Figure A6: Radar graph – social and human (z-scores)