Typology Characterization of farmers in Malawi

April 5th 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.

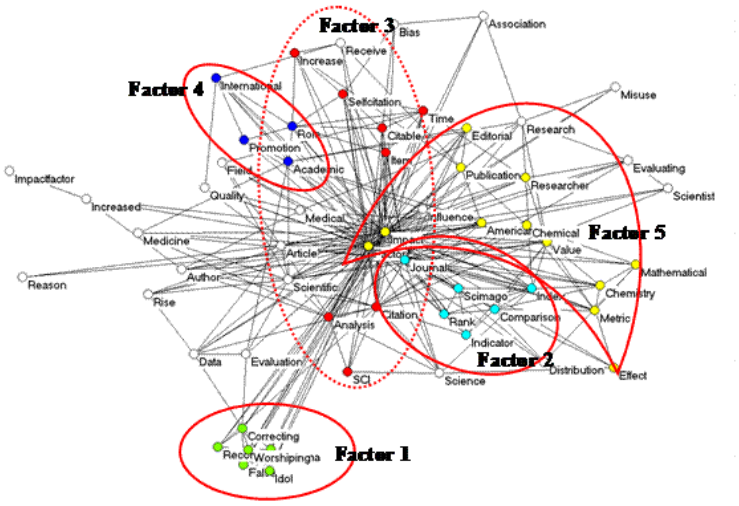
This document presents a summary of a typology study done using quantitative statistical methods (discussed below) applied to micro data from the Malawi Africa RISING Baseline Evaluation Survey (MARBES) (conducted in 2013) and secondary data on environmental/biophysical variables from various source. 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 Malawi).

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 TARBES 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 MARBES 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 two factors (represented by the first two dots at the top of the line graph) are highly relevant but that the 3rd factor starts to be less important in explaining the variation (smaller vertical jump).

Figure 2: Scree plot of productivity variables

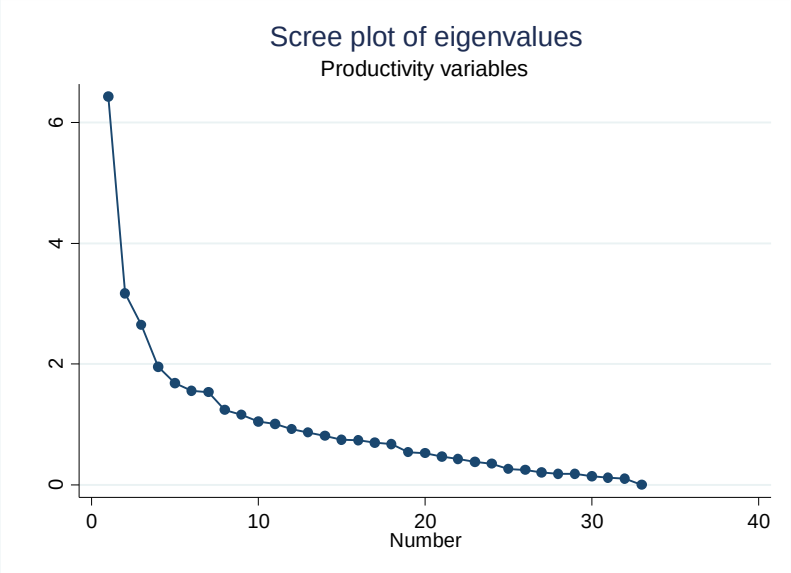


Table 1 shows the rotated matrix of factor loads for the two factors we have chosen, with the relevant variables highlighted (>0.5 or <-0.5). Factor 1 captures elements related to vegetables cultivation while Factor 2 is related to total land size and cereal cultivation. The final selection of variables for the cluster analysis include share of households cultivating legumes and total production of legumes for factor 1, and total land size and area cultivated with cereals for factor 2.

Table 1: Factor loads of productivity variables

Variable	Factor1	Factor2
Land size (Ha)	0.0471	0.822
N. parcels	0.114	0.4388
Min distance plot	-0.0566	-0.1516
Max distance plot	0.0567	0.0964
N. trees	0.0954	0.4959
N. crops	0.3576	-0.059
N. plots	0.1347	0.2996
HH does intercropping	0.1158	-0.0055
HH does intercropping with legumes	-0.0573	0.0542
N. of intercropped plots	0.2608	0.1194
Size intercropped land (Ha)	0.0033	0.0577
Size legumes-intercropped land (Ha)	-0.0187	-0.0015
Ownership mixed livestock	0.0221	0.1067
N. livestock types owned	0.05	0.1253
Maize only crop	-0.0556	-0.0472
Mixed crops	0.0556	0.0472
Cultivation of cereals	-0.0175	0.1577
Cultivation of vegetables	0.8187	-0.0794
Cultivation of legumes	-0.0957	0.1967
Area cultivated with cereals (Ha)	0.007	0.851
Area cultivated with vegetables (Ha)	0.7713	0.1502
Area cultivated with legumes (Ha)	-0.0853	0.4032
Production cereals (Kg)	0.0627	0.5149
Production vegetables (Kg)	0.8236	0.1033
Production legumes (Kg)	-0.0521	0.2476
Yield cereals (Kg/Ha)	0.0942	-0.2255
Yield vegetables (Kg/Ha)	0.7589	-0.0552
Yield legumes (Kg/Ha)	0.0061	-0.0818
TLU small ruminants	-0.0184	0.206
TLU big ruminants	-0.004	0.0818
TLU poultry	0.0066	0.0553
Fertilizer used (Kg)	-0.0322	0.1379
HH does irrigation	-0.0463	0.0183

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 two (Figure 3). Table 2 shows that factor 1 captures total harvest and different harvest use while factor 2 captures wealth and dwelling conditions. The final list of variables considered includes total Kg of grains harvested and Kg of harvest used for own consumption for factor 1, and non-agricultural wealth coupled with quality of floor material for factor 2.

Figure 3: Scree plot of economic variables

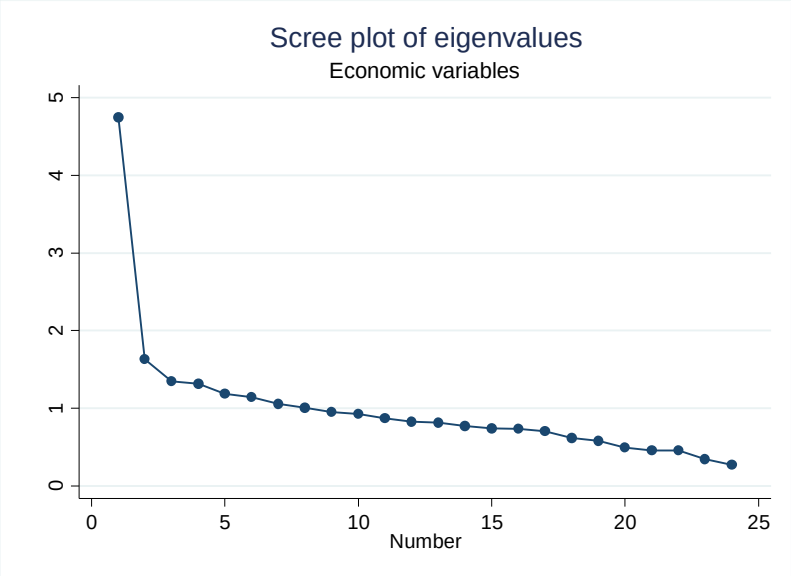


Table 2: Factor loads of economic variables

Variable	Factor1	Factor2
Fertilizer cost	0.5193	0.3294
Traditional seeds cost	0.2112	-0.0335
Improved seeds cost	0.3644	0.1511
Pesticide cost	-0.0002	0.2266
Other non-labor cost	0.0646	0.045
Animal feed cost	0.0475	0.1701
Agricultural wage	0.0597	0.0825
HH uses community labor	0.0456	-0.0632
HH uses hired labor	0.2534	0.252
Total PD used for crops	0.3714	0.0205
Total harvest of grains (Kg)	0.8028	0.1636
Total harvest of stover (Kg)	0.2498	0.0484
Total harvest used for animal feed (Kg)	0.1908	0.0042
Total harvest used for crop residual (Kg)	0.3521	0.0521
Total harvest used for seeds (Kg)	0.6192	0.0631
Total harvest used for gifts (Kg)	0.5346	0.2621
Total harvest used for own consumption (Kg)	0.6258	-0.0509
Total harvest used for other reasons (Kg)	0.0517	0.1186
Total harvest sold (Kg)	0.5721	-0.0023
Agri wealth index	0.129	0.5666
Non-agri wealth index	0.1942	0.8258
Good floor material in dwelling	-0.0264	0.7507
Good source of drinking water	0.2565	0.1921
Good toilet facility	-0.0656	0.0377

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 includes the trees owned by the household on the land, the third concerns soil

erosion issues and the fourth captures the use of manure. Our final selection of variables includes the share of parcels with clay/loam soils and incrusted soils (Factor 1), the number of leguminous and fruit trees (Factor 2), the share of farmers experiencing soil erosion as well as the ones not taking any preventive measure (Factor 3), and the share of households using manure (Factor 4).

Figure 4: Scree plot of environment variables

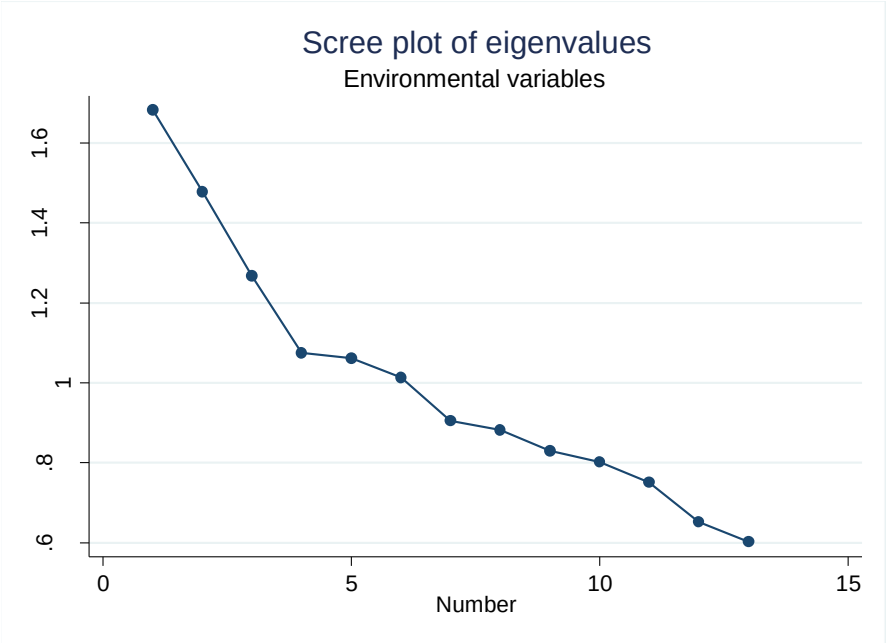


Table 3: Factor loads of environment variables

Variable	Factor1	Factor2	Factor3	Factor4
HH uses irrigation	0.0081	0.3354	0.089	0.4041
HH uses crop rotation	-0.0431	0.4036	0.0457	0.4706
HH uses fallowing	0.0487	0.349	0.0061	-0.2752
HH uses alternative tillage	-0.0729	-0.2923	0.0701	0.3041
HH uses manure	0.0651	0.0134	0.0207	0.7751
HH uses urea	0.0643	0.084	-0.0044	0.0872
HH experiences soil erosion	0.0669	-0.0004	0.7968	0.1209
HH experiences soil erosion and does not takes any preventive measure	-0.0684	0.0334	0.807	-0.0708
Share of parcels with clay or loam soil	0.7919	0.0217	-0.0382	0.0414
Share of parcels with brown or black soil	0.5333	-0.023	-0.0623	0.0667
Share of parcels with incrusted soil	0.7282	-0.0171	0.068	-0.0147
N. of leguminous trees	-0.0165	0.7222	-0.0592	0.1157
N. of fruit trees	0.021	0.6443	0.1394	-0.0401

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 and females-only managed livestock as the main variables of interest.

Figure 5: Scree plot of social variables

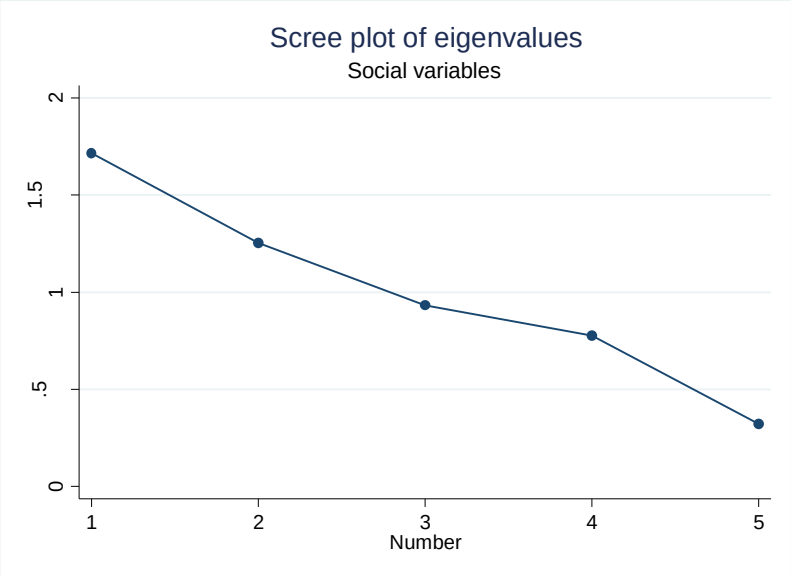


Table 4: Factor loads of social variables

Variable	Factor1
Females also responsible for plots	0.1622
Females only responsible for plots	0.0782
Females also responsible for livestock	0.8475
Females only responsible for livestock	0.7884
*Wage gap (wage women/wage men*100)*	0.4306

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, including the prevalence of older age groups (factor 1) and younger age groups (factor 2),the level of education of household members (factor 3), and the basic characteristics of the household head (factor 4). Experiencing food shortages in the 12 months preceding interview date do not appear to play a key role in differentiating the sample. We finally select mean age and mean adult age in the household (factor 1), young and total dependency ratio (factor 2), mean level of education in the household and years of education of the household head (factor 3) and indicators of whether the household head is married and whether is both female and married (factor 4).

Figure 6: Scree plot of human variables

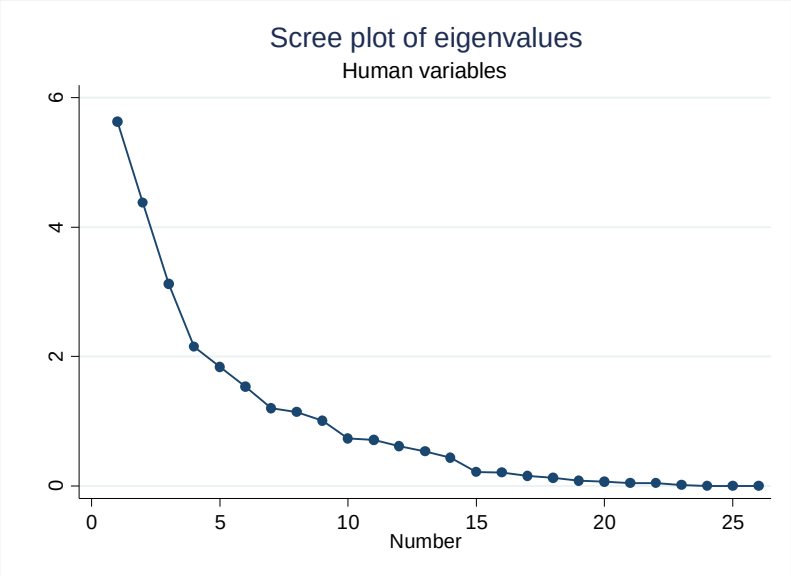


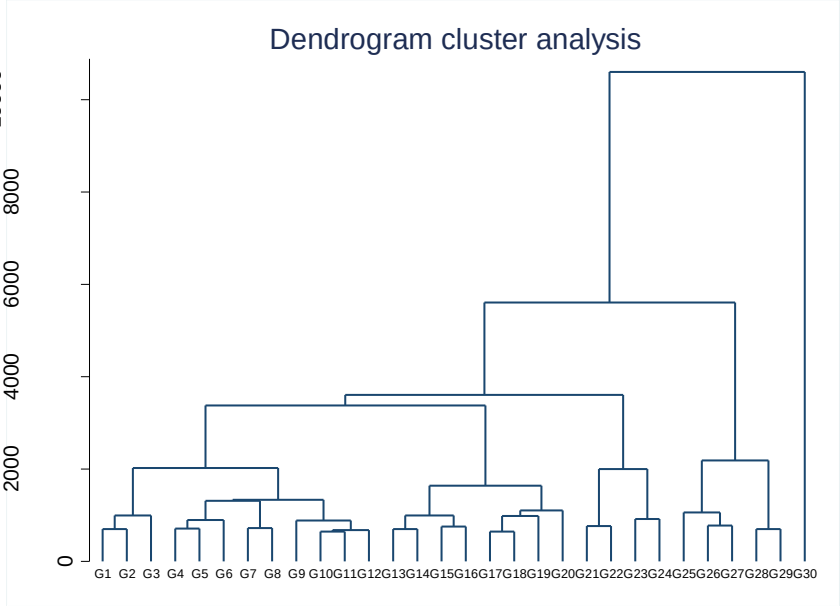
Table 5: Factor loads of human variables

Variable	Factor1	Factor2	Factor3	Factor4
HH size	-0.1708	0.3467	0.0235	-0.1479
Head is married	-0.0958	0.0153	0.0777	-0.5248
Head is widow	0.1609	-0.0561	-0.076	-0.2345
Head is single	-0.0346	0.0276	-0.0193	0.9575
Head is female	0.084	0.0509	-0.1053	0.4822
Head is female and single	-0.0472	0.0754	-0.0292	0.9666
Head is male and single	0.0249	-0.1106	0.0209	0.1283
Head's age	0.8195	-0.1408	-0.1199	-0.0536
Head's years of educ	-0.1361	0.0316	0.8853	-0.0556
Head is literate	-0.0582	0.0237	0.7195	-0.0417
Mean years of edu.	-0.1688	-0.1096	0.9144	0.0161
Highest years of edu.	-0.0657	-0.2176	0.8394	-0.0483
Mean age	0.8696	-0.4188	-0.0764	-0.0068
Mean adult age	0.9198	0.1621	-0.1471	-0.0293
N. of males adults	0.0265	-0.37	0.0904	-0.2534
N. of females adults	0.1477	-0.2154	0.0774	0.107
children	-0.4555	0.5456	0	-0.1876
Young dep. Ratio	-0.1826	0.9413	-0.0754	0.0794
Old dep. Ratio	0.691	0.1712	-0.0046	-0.0718
Total dep. ratio	0.0548	0.9627	-0.0741	0.0524
Share of 0-14 y.o.	-0.3926	0.8373	-0.0603	0.0253
Share of 15-29 y.o.	-0.3305	-0.5927	0.1532	-0.017
Share of 30-44 y.o.	-0.2125	0.072	0.0623	-0.0392
Share of > 45 y.o.	0.8627	-0.279	-0.1381	0.0205
HH worries for food shortages	-0.0265	0.1086	-0.1308	0.0361
Months experienced food shortages	-0.1129	0.0098	-0.1157	0.0508

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 4 productivity variables, 4 economic variables, 7 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 four final groups, or "types" of farmers. Tables 6a to 6e illustrate 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 304 of the farmers in the sample, type two is the biggest and defines 330 farmers, type 3 accounts for 288 farmers and finally type 4 is the smallest, with 137 farmers.

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

	Type 1	Type 2	Type 3	Type 4
Productivity Domain				
Total land size (Ha)	0.58***	0.90***	1.24***	1.68***
	[0.03]	[0.03]	[0.04]	[0.06]
Share of households doing intercropping	0.75***	0.86	0.90***	0.92***
	[0.02]	[0.02]	[0.02]	[0.02]
Share of households doing intercropping with legumes	0.61***	0.78	0.84***	0.85***
	[0.03]	[0.02]	[0.02]	[0.03]
Area of intercropped plots	0.56**	0.98	1.48**	1.54
	[0.12]	[0.21]	[0.35]	[0.32]
Area of plots intercropped with legumes	0.21*	0.54	0.57	0.67
	[0.05]	[0.17]	[0.25]	[0.15]
Share of households owning mixed livestock	0.22***	0.45	0.55***	0.77***
	[0.02]	[0.03]	[0.03]	[0.04]
N. of different livestock types owned	0.83***	1.43	1.73***	2.25***
	[0.05]	[0.06]	[0.06]	[0.09]
Share of households cultivating maize only	0.13***	0.02***	0.01***	0.00***
	[0.02]	[0.01]	[0.00]	[0.00]
Share of households growing cereals	0.98**	0.99	1	1
	[0.01]	[0.00]	[0.00]	[0.00]
Share of households growing vegetables	0.22**	0.26	0.29	0.31
	[0.02]	[0.02]	[0.03]	[0.04]
Share of households growing legumes	0.70***	0.91**	0.98***	0.99***
	[0.03]	[0.02]	[0.01]	[0.01]
Area of cereals(ha)	0.32***	0.47***	0.61***	0.86***
	[0.02]	[0.02]	[0.02]	[0.04]
Area of vegetables(ha)	0.01***	0.02	0.03**	0.05***
	[0.00]	[0.00]	[0.00]	[0.01]
Area of legumes(ha)	0.22***	0.38	0.44	0.72***
	[0.03]	[0.04]	[0.03]	[0.06]
Production of cereals(kg)	254.87***	607.69***	1089.49***	1870.64***
	[15.90]	[19.86]	[43.97]	[84.05]
Production of vegetables(kg)	4.42***	14.79	21.89	48.14***
	[0.70]	[2.70]	[4.13]	[10.83]
Production of legumes(kg)	71.60***	170.86***	309.02***	464.91***
	[10.88]	[10.10]	[15.33]	[30.31]
Yield of cereals(kg/ha)	1095.04***	1685.23	2046.35***	2329.37***
	[56.05]	[63.87]	[69.19]	[89.74]
Yield of vegetables(kg/ha)	512.18***	959.88	792.94	1273.00***
	[74.61]	[153.75]	[113.34]	[198.76]
Yield of legumes(kg/ha)	473.22***	644.18**	881.55***	824.76**
	[40.77]	[34.53]	[37.36]	[53.81]
TLU small ruminants	0.06***	0.11***	0.19***	0.31***
	[0.01]	[0.01]	[0.01]	[0.03]
TLU big ruminants	0.00*	0	0.01	0.03***
	[0.00]	[0.00]	[0.01]	[0.02]
TLU poultry	0.02***	0.04	0.06***	0.07***
	[0.00]	[0.00]	[0.00]	[0.01]
Kg fertilizer used	149.04***	380.12	568.87***	755.62***
	[15.56]	[29.59]	[39.48]	[67.52]
N. of observations	304	330	288	137

Standard errors of means in brackets

* 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	Type 4
Economic Domain				
Value of fertilizer used (GHC)	3250.23***	10021.80***	20641.26***	37064.52***
	[455.04]	[861.40]	[1541.15]	[2588.10]
Value of traditional seeds purchased (GHC)	551.97	536.67**	745.62	987.02***
	[63.83]	[59.15]	[86.41]	[155.43]
Value of improved seed purchased (GHC)	577.80***	1067.04***	1907.58***	2907.44***
	[73.31]	[121.58]	[187.30]	[350.71]
Value of pesticides used (GHC)	77.34***	336.27	537.09*	876.19***
	[35.61]	[68.47]	[129.25]	[243.86]
Share of households using communal labor	0.24***	0.32	0.39***	0.39
	[0.02]	[0.03]	[0.03]	[0.04]
Share of households using hired labor	0.18***	0.33***	0.55***	0.72***
	[0.02]	[0.03]	[0.03]	[0.04]
Total person-days used, male & female	146.70***	241.49**	332.32***	440.78***
	[8.62]	[10.37]	[16.20]	[25.71]
Total Kg of grains harvested	230.97***	638.04***	1222.70***	2561.75***
	[6.26]	[7.78]	[13.65]	[85.75]
Total Kg harvest used for own consumption	154.26***	308.34**	466.16***	523.86***
	[4.60]	[6.69]	[19.65]	[23.88]
Total Kg harvest sold	33.11***	112.40***	219.83*	567.84***
	[13.56]	[29.30]	[13.46]	[71.47]
Agricultural wealth index	-0.15***	-0.08*	0.08	0.41***
	[0.10]	[0.02]	[0.02]	[0.04]
Non-agricultural wealth index	-0.31***	-0.15***	0.16***	0.75***
	[0.04]	[0.03]	[0.06]	[0.14]
Share of households with good floor in dwelling	0.05***	0.11	0.13	0.18***
	[0.01]	[0.02]	[0.02]	[0.03]
Share of households with good source of drinking water	0.07***	0.1	0.15*	0.21***
	[0.01]	[0.02]	[0.02]	[0.04]
Share of households with good toilet facility	0.02	0.01	0.01	0.01
	[0.01]	[0.01]	[0.01]	[0.01]
N. of observations	304	330	288	137

Standard errors of means in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

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

	Type 1	Type 2	Type 3	Type 4
Environmental Domain				
Share of households practicing irrigation	0.03***	0.09	0.15***	0.23***
	[0.01]	[0.02]	[0.02]	[0.04]
Share of households practicing rotation	0.49***	0.73	0.84***	0.86***
	[0.03]	[0.02]	[0.02]	[0.03]
Share of households practicing fallowing	0.05	0.07	0.07	0.05
	[0.01]	[0.01]	[0.02]	[0.02]
Share of households practicing alternative tillage	0.01	0.01	0.01	0.03*
	[0.00]	[0.01]	[0.01]	[0.01]
Share of households using manure on (any) plot in either season	0.38***	0.56	0.61***	0.64***
	[0.03]	[0.03]	[0.03]	[0.04]
Share of households using urea on (any) plot in either season	0.14	0.13	0.16	0.1
	[0.02]	[0.02]	[0.02]	[0.03]
Share of households affected by soil erosion	0.62	0.6	0.65	0.71**
	[0.03]	[0.03]	[0.03]	[0.04]
Share of households with soil erosion but no erosion control measure	0.14	0.11**	0.15	0.18
	[0.02]	[0.02]	[0.02]	[0.03]
Average share of parcels with clay or loam soil	0.49	0.49	0.51	0.55
	[0.03]	[0.02]	[0.02]	[0.03]
Average share of parcels with black or brown soil	0.41	0.43	0.41	0.43
	[0.03]	[0.02]	[0.02]	[0.03]
Average share of parcels with incrusted soil	0.70**	0.64	0.62**	0.68
	[0.02]	[0.02]	[0.02]	[0.03]
N. of leguminous trees owned	1.94***	3.48	4.63***	6.34***
	[0.16]	[0.24]	[0.34]	[0.56]
N. of fruit trees owned	2.55***	5.71	7.10***	9.88***
	[0.23]	[0.71]	[0.56]	[1.25]
N. of observations	304	330	288	137

Standard errors of means in brackets

* 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	Type 4
Social Domain				
Share of HH with female having shared plot responsibility	0.84	0.85	0.84	0.83
	[0.02]	[0.02]	[0.02]	[0.03]
Share of HH with female having exclusive plot responsibility	0.49***	0.36	0.31**	0.25***
	[0.03]	[0.03]	[0.03]	[0.03]
Share of HH with female having shared livestock responsibility	0.03***	0.05	0.06***	0.07***
	[0.00]	[0.00]	[0.00]	[0.00]
Share of HH with female having exclusive livestock responsibility	0.02***	0.02	0.03*	0.03**
	[0.00]	[0.00]	[0.00]	[0.00]
N. of observations	304	330	288	137

Standard errors of means in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

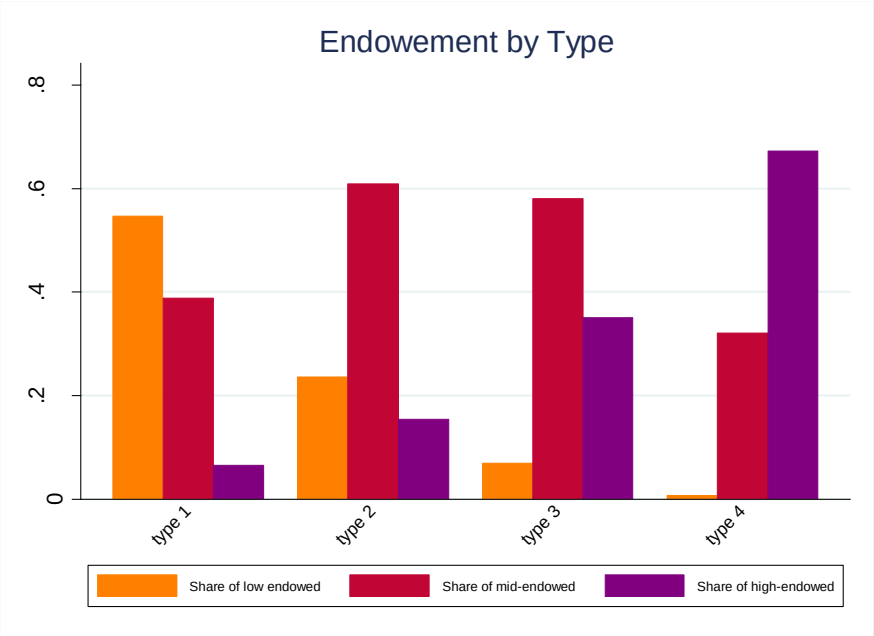
Table 6e: distribution of characteristics by type in the human domain

	Type 1	Type 2	Type 3	Type 4
Human Domain				
Household size	4.35***	4.87	5.14***	5.47***
	[0.11]	[0.10]	[0.11]	[0.16]
Share of married heads	0.60***	0.73	0.76**	0.85***
	[0.03]	[0.02]	[0.03]	[0.03]
Share of female heads	0.41***	0.3	0.25**	0.18***
	[0.03]	[0.03]	[0.03]	[0.03]
Age of the head	41.45***	44.5	44.14	47.43***
	[0.86]	[0.83]	[0.75]	[1.19]
Years of education of the heads	3.89***	4.84	5.50***	6.00***
	[0.18]	[0.19]	[0.21]	[0.32]
Share of literate heads	0.57***	0.72	0.77***	0.83***
	[0.03]	[0.02]	[0.02]	[0.03]
Mean years of education in the household	4.01***	4.94	5.52***	6.13***
	[0.14]	[0.15]	[0.16]	[0.23]
Max years of education in the household	5.62***	6.92	7.72***	8.58***
	[0.18]	[0.17]	[0.18]	[0.26]
Average age of adults in the household	22.75*	23.71	23.64	25.28*
	[0.65]	[0.58]	[0.65]	[0.88]
Number of children in the household	0.84	0.84	0.82	0.66**
	[0.05]	[0.05]	[0.05]	[0.06]
Young dependency ratio	1.21***	1.09	1.01	0.93**
	[0.06]	[0.05]	[0.04]	[0.07]
Old dependency ratio	0.09	0.14***	0.07*	0.1
	[0.02]	[0.02]	[0.02]	[0.03]
Share of HH worrying about food shortages	0.76***	0.51	0.38***	0.18***
	[0.02]	[0.03]	[0.03]	[0.03]
Months experiencing food shortages?	2.77***	1.55	1.01***	0.54***
	[0.15]	[0.09]	[0.09]	[0.09]
N. of observations	304	330	288	137

Standard errors of means in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 8: Level of Endowments by Type



The four 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. More broadly, the types can be characterized as following:

Type 1: Female-headed, low educated households with low levels of endowments

- High number of female headed households, with heads less likely to be married and with low education attainments and literacy rates. High young dependency rates.
- High proportion of women with exclusive plot responsibilities but low proportion of women with livestock responsibilities. Wage gap very favorable to women.
- Very high food insecurity.
- Little asset ownership (land below 0.6 Ha, very little livestock, low agricultural and non-agricultural wealth).
- Low production and productivity of all major crops, also due to low input use (both in terms of labor inputs, which are mainly composed by family labor, and non-labor inputs). Much less frequent cultivation of legumes with respect to other groups.
- More than half of crop harvest devoted to own consumption, almost no crop sales.
- Low levels of soil conservation practices and problems of incrusted soils.

Type 2: Old households with medium-low levels of endowments

- Relatively uneducated households with high old dependency rates and high food insecurity.
- Low productivity and input use, even though better than type 1, and mid-levels of endowments. Small land size (below 1 Ha).
- Half of harvest is devoted to own consumption, little sales.
- Good soil quality.

Type 3: Medium-high endowed households with high levels of productivity

- Households with a large active population and fairly high levels of educational attainment.
- Low levels of gender equality with respect to other groups, especially in terms of wage gap.
- Medium levels of crop production but high productivity (especially for legumes) and frequent intercropping practices. Very high share of households growing legumes (98%).
- Frequent use of communal labor.
- Medium levels of endowments, with average land size around 1.2 Ha.
- Low levels of incrusted soils and frequent use of soil conservation practices.

Type 4: Highly endowed households breeding small ruminants

- Large male headed households with high levels of educational attainments. High percentage of active population.
- High percentage of women with some livestock responsibilities.
- High levels of food security.
- Extremely high asset ownership (large land above 1.6 Ha, high number of livestock types and units – especially small ruminants -, high agriculture and non-agriculture index).
- High production and productivity of crops with high input use (especially irrigation). Very high share of households growing legumes (99%) and comparatively high share growing vegetables (31%).
- Frequent use of soil conservation practices but, despite that, high levels of soil incrustation and soil erosion.

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, low educated households with low levels of endowments	Low crop production and productivity. Little legume cultivation. Little livestock owned.	Low wealth (agri and non-agri), land size below 0.6 Ha, low input expenditure, most harvest going to own consumption and no crop sales.	Low levels of soil conservation practices and problems of incrusted soils.	High frequency of female responsibility for crops but opposite for livestock. Very favorable wage gap.	Single female heads with low levels of literacy and education. High young dependency ratio. Very low food security.
Type 2: Old households with medium-low levels of endowments	Low crop production and productivity. Little livestock owned.	Low-medium wealth (agri and non-agri), land size below 1 Ha, low input expenditure, half of harvest going to own consumption and little crop sales.	Good soil quality.	Average gender equality.	Small households with high old dependency ratio. Relatively low food security.
Type 3: Medium-high endowed households with high levels of productivity	High crop production and productivity, especially for legumes. Frequent intercropping.	Medium-high wealth (agri and non-agri), high input use (especially communal labor).	Low levels of incrusted soils and frequent use of soil conservation practices.	Below average levels of gender equality.	Households with large active population and mid-high levels of education.
Type 4: Highly endowed households breeding small ruminants	Very high crop production and productivity. High livestock ownership, especially small ruminants. Frequent intercropping and vegetables cultivation.	Very high wealth (agri and non-agri), high input use (especially irrigation and hired labor). The harvest going to sales is the same amount as the one going to own consumption.	High frequency of soil conservation practices but severe problems of soil erosion and incrustation.	High frequency of female responsibility for livestock but opposite for crops.	Very large households with married male heads and high levels of education. High food security.

Figure 9: Graphic representation of types

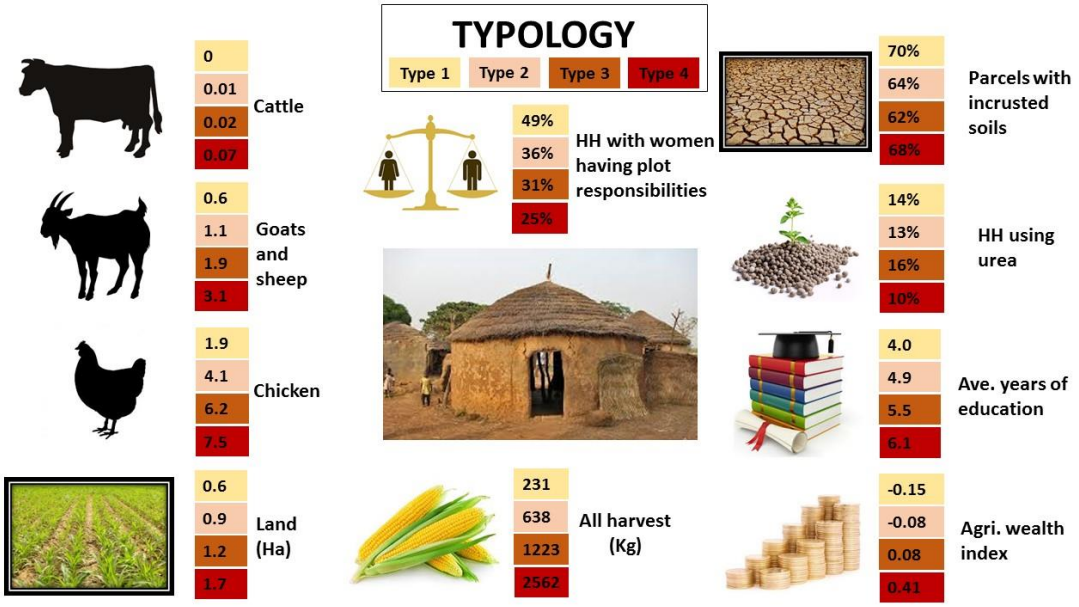
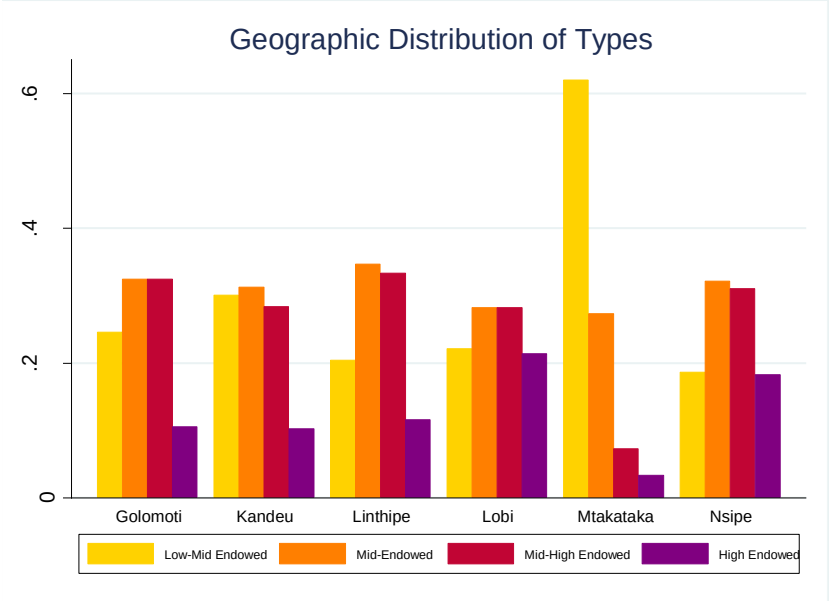
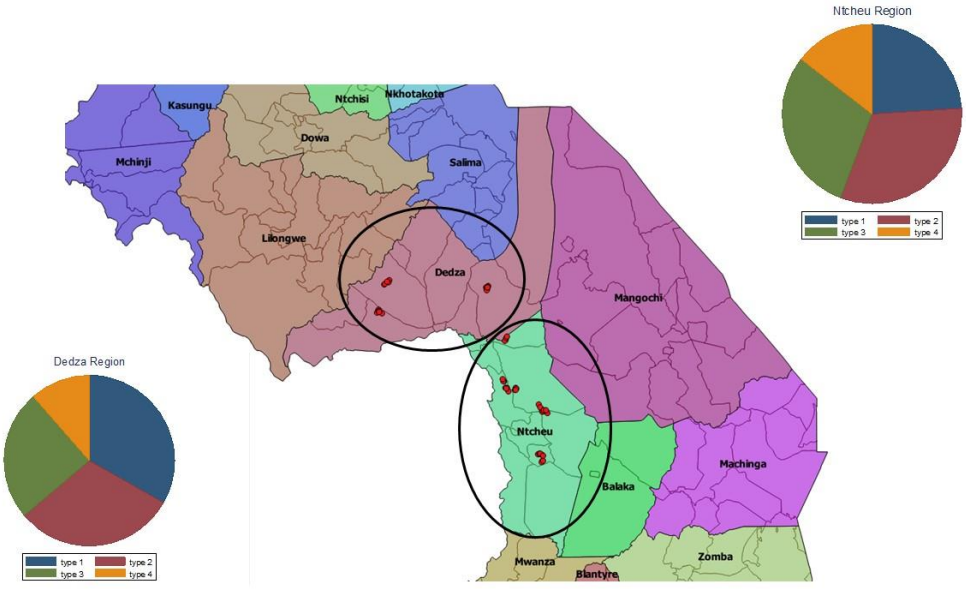


Figure 10: Distribution of Typologies by Districts



The differences in climatic conditions between groups are an indication of heterogeneity of typology distribution across space. Figure 9 shows the typology composition of each district in the sample. While in Mtakataka there is a very high concentration of female-headed, low educated households with low levels of endowments (type 1), Nsipe and espacially Lobi concentrate high shares of Mid-endowed and high endowed households (type 3 and 4). However, if we look at the typology distribution by region (figure 11) the differences are much less pronounced. The spatial distinctions are important because they can support interventions based on the most prevalent households' typologies in the area.

Figure 11: Distribution of Typologies by Regions



The characteristics of each household type described above can be displayed clearly with a spider plot. Figure 12 summarizes the performance of each type relative by each domain as follows:

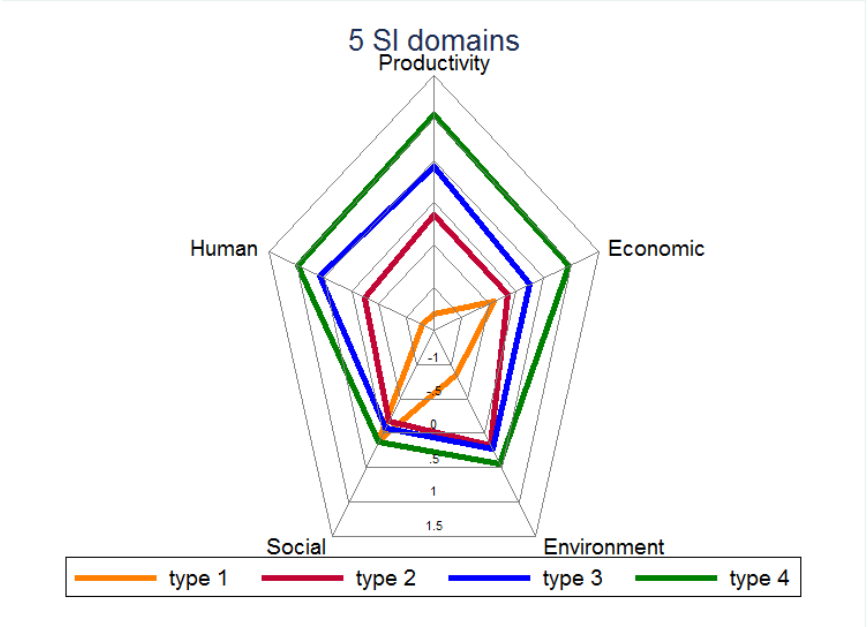
- The largest differences are observed in the productivity and human domains, with group 3 and especially group 4 presenting much higher levels with respect to the other two groups.
- In terms of economic endowments, type 4 differentiate itself with a very strong performance, while the other groups are fairly close to each other at a lower level.
- Endowments in the social aspect, here measured by gender equality, are rather equally distributed across groups and are on average fairly high, especially in terms of wage gaps.
- Finally, type 1 lags behind in terms of soil conservation practices while type 4 is the group that performs the best. Nevertheless, the groups with the least problems of soil quality are group 2 and 3.

Recommendations:

- Farmers in type 1 need an integrated intervention from AR supporting a raise in their productive capacity, an improvement of their endowment level and a training about soil conservation practices.
- Farmers in type 2 need a targeted intervention from AR improving their productive capacity.
- Farmers in type 3 and 4 are already performing quite well across all the aspects and can be involved by the project to facilitate adoption. AR can also support farmers in type 4 in mitigating their soil degradation problems.

The appendix includes additional graphs characterizing the obtained typologies.

Figure 12: 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 and livestock (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)

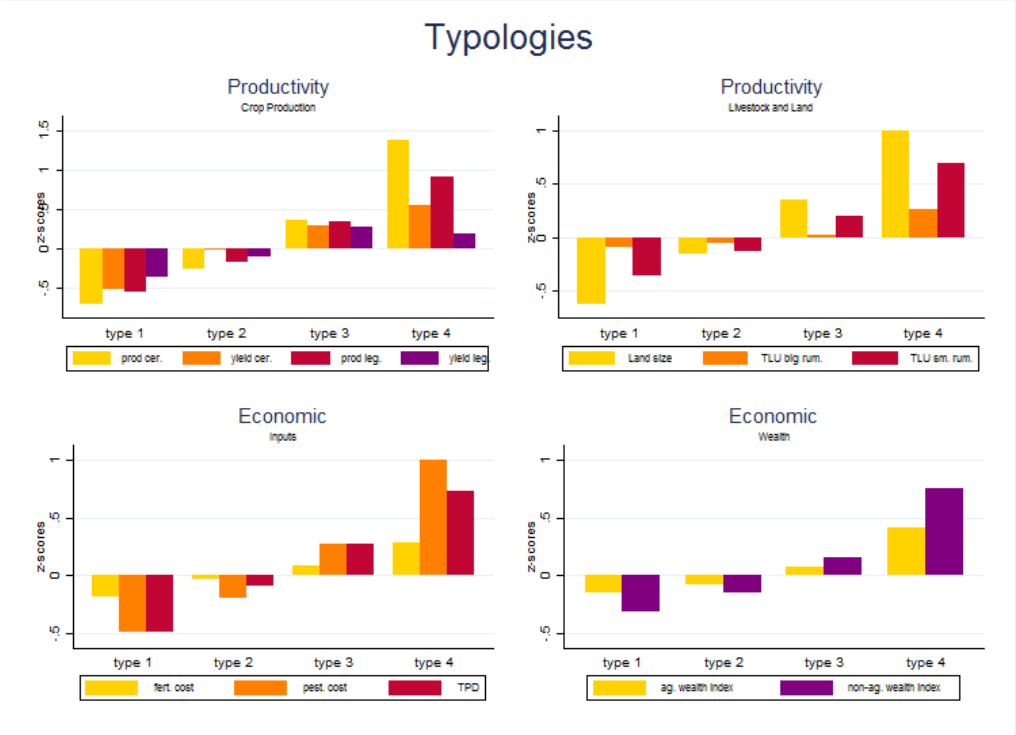


Figure A2: Typologies by domain (environment, social and human)

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

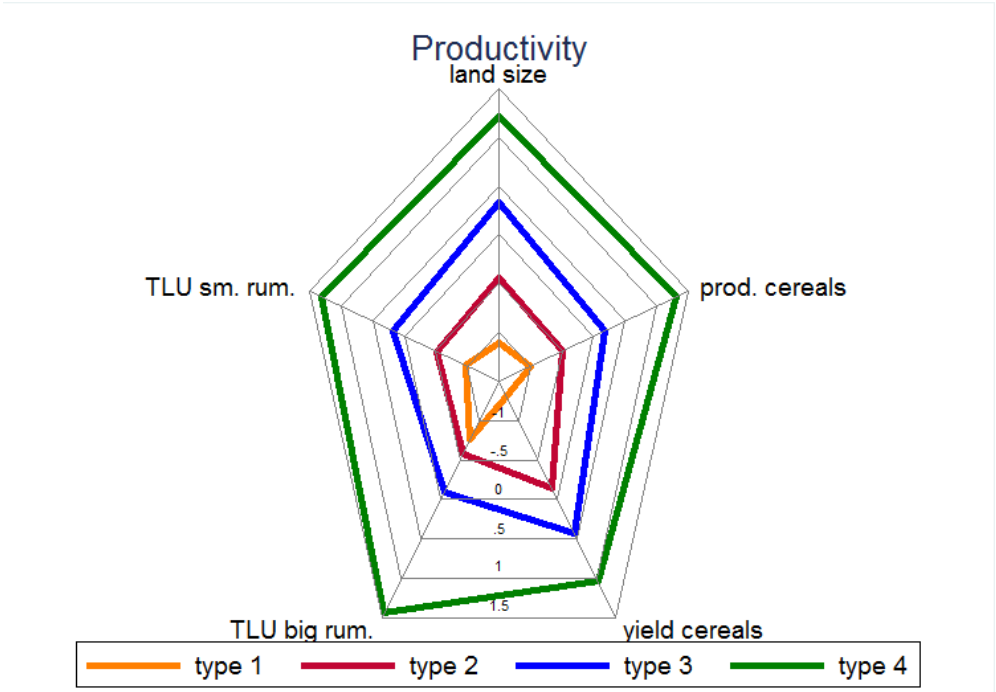


Figure A4: Radar graph – economic (z-scores)

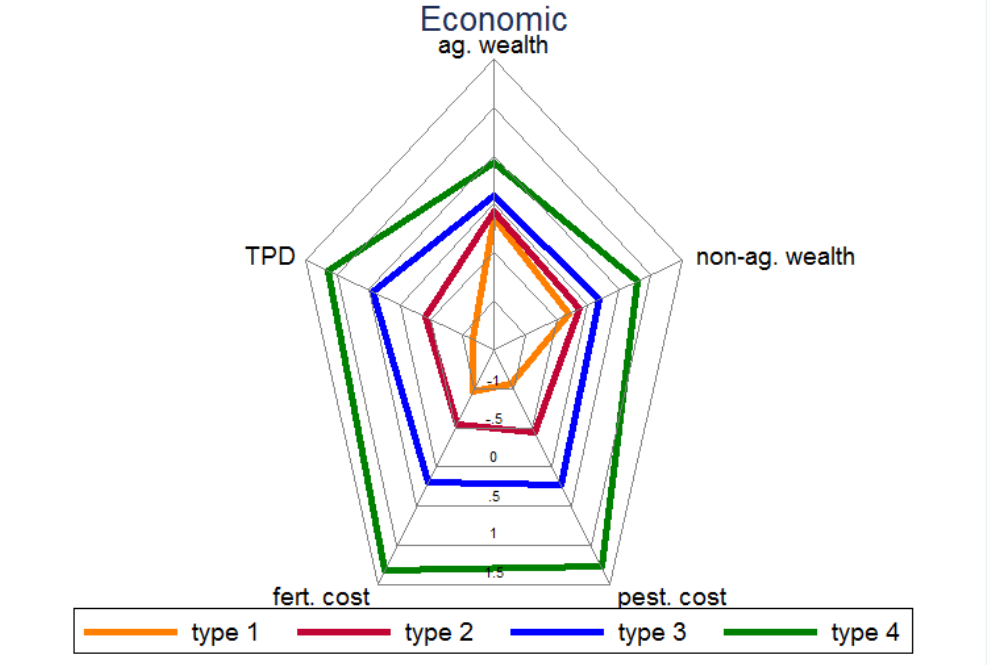


Figure A5: Radar graph – environment (z-scores)

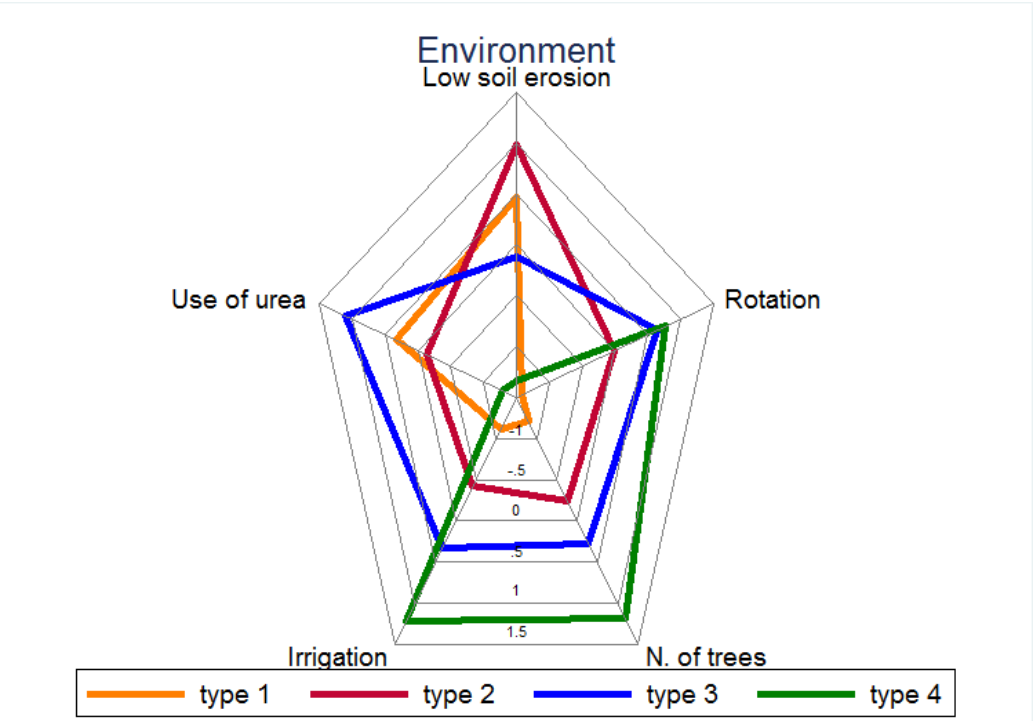


Figure A6: Radar graph – social and human (z-scores)