ISET MA Program in Economics Policy Institute

International School of Economics at Tbilisi State University

UNDP

Empowered lives. Resilient nations.

KNOWLEDGE NEEDS IN GEORGIAN AGRICULTURE:

THE CASE OF FARMING HOUSEHOLDS

This research would have been not possible without the Swiss Cooperation Office South Caucasus (SCO) and UNDP supported project "Modernization of the Vocational Education and Training and Extension Systems Related to Agriculture in Georgia" and the support from different experts, researchers, analysists, interviewers and interviewees.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Cooperation Office
South Caucasus

UNDP

Empowered lives.
Resilient nations.

We would like to emphasis our gratitude to the International School of Economics Policy Institute (ISET-PI), which has elaborated this study based on the survey results provided by Analysis and Consulting Team (ACT). Our special thanks are addressed to the authors of the study:

Dr. Florian M. Biermann, Assistant Professor at ISET and lead author
Eric Livny, President of ISET
Zurab Abramishvili, Junior Researcher at ISET
Saba Devdariani, Junior Researcher at ISET

The study was coordinated by Ruediger Heining, Project Manager at UNDP Georgia with support of Tamar Sanikidze, Tamar Kitiashvili and Tea Gulua as project team members.

Published in Georgia

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The UNDP Farmer Knowledge Project was carried out in two phases. Data on Georgian rural households[1] was collected by the polling agency *Analysis and Consulting Team* (ACT) between February and July 2015.[2] This data was analyzed with the purpose of producing policy recommendations by the ISET Policy Institute between November 2015 and July 2016.

The project pursued three overarching **goals**: (i) to understand which gaps in agricultural knowledge of Georgian farmers have the strongest impact on farmers' productivity and income, and recommend relevant agricultural extension measures; (ii) to predict structural and social changes in Georgia's agriculture under different scenarios; (iii) to suggest appropriate policy interventions to mitigate or encourage these changes.

Survey Strategy and Data

Questionnaire. The survey instrument was designed by UNDP, covering (i) household/farm and individual respondent's characteristics, (ii) self-assessed needs for different types of skills and knowledge, and (iii) inputs and outputs of the farm production process. According to survey instructions, one interview was supposed to take around 45 minutes.

To elicit training/capacity building needs, respondents were asked to assess the extent to which receiving more information related to particular knowledge areas was important for improving their farm operations. For each knowledge area, the answers were to be given on a scale ranging from 1 ("not important at all") to 5 ("very important"). Thus, higher numbers indicated greater (self-assessed) demand for knowledge. Knowledge areas were identified as: "crop production", "animal breeding", "animal fattening", "dairy production", and "farm management", and for each of these categories, between 8 and 10 specific activities were named. For each of these activities, the need for knowledge ("what to do") was identified separately from the need for skills ("how to do it").[3]

---

[1] In this study, a farming household is defined as a household that is not living in one of the urban parts of Tbilisi, Kutaisi, Batumi, Rustavi, or Poti, and in the year 2014 used land for agricultural purposes. To make the text more readable, we will use the terms "farm", "household", and "farming household" interchangeably.

[2] ACT also surveyed 100 agricultural enterprises. These are not covered in this report, as they are different from ordinary farming households in various relevant respects. Most importantly, commercial enterprises have different access to capital, e.g., because they can provide balance sheets and business plans, and, depending on the scale of their production, they may be capable of autonomously training their employees. The latter means that they do not depend on agricultural extension as much as farming households. In light of the findings of this study, it can also be assumed that they are less susceptible to the *Dunning-Kruger effect* (see Section 1.4).

[3] A detailed discussion of the difference between knowledge and skills can be found in Section 1.5.3.

Geography. A total of 3000 households have been surveyed: 100 in Tbilisi, 580 in Imereti, and 290 per each of Georgia's remaining 8 major regions (Adjara, Guria, Kakheti, Kvemo Kartli, Mtskheta-Mtinaneti, Samegrelo-Zemo Svaneti, Shida Kartli, and Samtskhe-Javakheti).

Gender. About 57% of the respondents were female, in stark contrast with the percentage of female household heads, which stood at only 25%. Women respondents tend to be more confident in their knowledge (or less interested in learning) and report on average slightly lower knowledge needs then men.

Farmers' age. The average respondent's age was almost 55 years, with the youngest being 17 and the oldest 94 years old. Interestingly, reported knowledge gaps do not significantly differ across different age cohorts.

Crops. There is a very large variation in the prevalence of different crops in the UNDP sample. 1,664, 1,609, and 1,466 households have maize, tomatoes, and cucumbers in their crop portfolios, respectively. On the other extreme, tobacco and tea are produced by only one household, each.

That said, the most common crops are grown on very small pieces of land, suggesting a very low degree of specialization and commercialization. For example, the average plot sizes under maize, tomato and cucumbers are 0.32, 0.05, and 0.04 hectares, respectively. The average wheat producer, on the other hand, grows wheat on 1.21 hectares, followed by sunflowers and dry forage, which on average are grown on 1.17 and 0.85 hectares, respectively.

Commercialization. About 40% of farmers in the UNDP sample are *subsistence farmers* who have no monetary income from selling agricultural products. About 56% are *semi-subsistence farmers* who receive monetary income from selling agricultural products, but it is less than the value of the produce they consume. Finally, only 3% in the sample can be defined as *professional farmers* who derive significant monetary income from selling agricultural products, exceeding the value of the produce they consume.[4]

Analysis and Policy Recommendations

Knowledge gaps and agricultural extension. The average reported knowledge deficits range between 3.46 (for activities related to crop production) to 2.66 (for farm management). Reported skill and knowledge deficits are highly correlated, suggesting that farmers failed to distinguish between the "what" and "how" aspects of knowledge referred to in the questionnaire.

---

[4] The definition of subsistence, semi-subsistence, and professional farmers applied here is used in other studies, too. For literature references, see Section 3.2.

Table A. Average knowledge and skill gaps by type of agricultural activity

Activities	Importance of receiving more knowledge (what)[5]	Importance of receiving more skills (How)
Crop production	3.46	3.59
Dairy	3.09	3.17
Animal fattening	3.06	3.15
Animal breeding	3.04	3.14
Farm management	2.66	2.72

Usually, one would expect knowledge gaps and output to be negatively correlated, i.e., low productivity farmers should express greater demand for relevant knowledge, and vice versa. This was indeed the case with some crops, and most notably wheat: the least productive wheat growers reported greater knowledge gaps in crop production, and vice versa. Yet, for the majority of knowledge areas a reverse relationship was often observed: lower demand for knowledge was associated with lower output.

There are two mutually non-exclusive ways to interpret such an apparently nonsensical correlation. One straightforward explanation is that people *demand* knowledge in areas which they deem important for their incomes and livelihoods. In that case, they would indicate low knowledge needs in secondary areas of their farming operations, and these will often be those in which they are not very productive. This interpretation is supported by the data, e.g., in the manner in which men and women responded to the question concerning knowledge gaps in the dairy sector. Women, who are the main dairy sector workers in Georgia, express significantly higher demand for relevant skills: 3.63 as compared with 3.17 for men. Moreover, they value dairy sector-related knowledge more than that concerned with the male-dominated animal breeding and fattening (3.13 and 3.14, respectively). Likewise, commercial skills (selling, buying) are mostly demanded by those engaged in commercial farming activities, who are almost by definition more skillful in this regard then subsistence farmers.

Another possibility, regularly encountered in self-assessments of skills and knowledge, is that people who know little about a subject are also unaware of their ignorance. Known as the *Dunning-Kruger (DK) effect,* this phenomenon would also lead to a positive correlation between reported knowledge gaps and productivity: the worse farmers are at their trade, the less interest they have in upgrading relevant skills, and vice versa.

The DK effect is not just a problem for empirical research into the link between actual (as opposed to self-reported) knowledge gaps and productivity. **To the extent that Georgian farmers are subject to the DK**

[5] Assessment ranged from 1 ("not important at all") to 5 ("very important").

effect, they will underestimate the importance of knowledge for the success of their farm operations and will lack motivation for self-improvement. Instead of admitting own mistakes and skill deficits and making efforts to learn, they will attribute low income and productivity to other factors, such a lack of government support, access to finance, or bad luck.[6]

A most interesting finding concerns the possibility of a non-linear (inverted U-shaped) relationship between the amount of land used for growing a particular crop and the demand for relevant knowledge. Sunflower is an excellent case in point. Sunflower growers exhibit the greatest thirst for knowledge at *intermediate* values of land under this crop (about 0.5-1 hectare). They are less interested in learning when they are far from achieving commercialization or are already successfully commercialized (having less than 0.5 or more than 1 hectare under sunflower). While lacking in statistical power (there are only 10 sunflower farmers in the sample), this finding is suggestive of a general trend that could be utilized to better target extension measures.

To sum up, the relationship between demand for knowledge and productivity appears to be complicated, calling for a multipronged policy response:

- First, farmers are less prone to misjudge their *relative* knowledge and skills in the production of staple crops, such as wheat or sunflower, in which it is easy to compare own productivity with that of one's neighbor. The opposite is true for small-scale tomato growers.[7] **This finding calls for an inexpensive "information" intervention, benchmarking comparable Georgian farms to each other (to motivate weak performers) and the global best practice (to motivate leaders). Farmers who are relatively more productive should receive a "menu-style" extension offer which would allow them to pick extension programs according to their preferences. At the same time, given the potential for complacency on their part, they should be informed about their standing with respect to the world's most productive farmers.[8]**

---

[6] Another challenge for agricultural extension may be the "epistemic culture" in post-Soviet agriculture. In the Soviet Union, it was decided centrally what knowledge a farmer should have (a typical "top-down approach"). In Western extension systems, on the other hand, the knowledge offering is based on the farmers' preferences (a "bottom-up approach"), which requires the farmers to evaluate and describe their knowledge needs. Farmers in the post-Soviet space may often not have the capacities to do that (for further discussions of this issue, see Part I in Hornidge et al. (2016).

[7] The same facts can be interpreted in a different way without affecting the policy recommendation. While successful wheat growers may become complacent about their own level of competency, commercially oriented tomato growers may be eager to acquire better skills because they deal with a perishable product. According to our data, tomato farmers demonstrate the highest correlation between productivity and demand for *post-harvest treatment skills (which are only relevant for commercial tomato farmers)*: one step increase in the demand for relevant knowledge is associated with a 23% increase in tomato yield per hectare.

[8] This point was suggested at the stakeholder workshop that was held at ISET on September 14, 2016.

- Second, farmers tend to overcome their ignorance as they get more specialized (in the sense of devoting a larger share of their land to commercial production of one or two crops). If we take tomato growers, for example, the DK effect is statistically significant only for the least specialized farmers. **Extension measures seeking to encourage greater specialization and commercialization are therefore also likely to increase farmers' appetite for knowledge.**
- Third, **extension measures should initially target groups that are aware of their own knowledge and skill limitations**, such as the *less productive* wheat producers, *commercial* tomato farmers, sunflower growers *on the verge of commercialization*, or *female* dairy farmers. Once extension services are successfully established for one type of farming, they could be rolled out to other types of agriculture, and other target groups.

The last point is particularly important given that a lot capacity building effort failed to target strongly motivated farmers. While perhaps subjective and biased, self-assessed knowledge needs do reflect the motivation of a farmer to learn: someone who says that additional knowledge and skills will help to "improve farm operations" (the formulation used in the questionnaire) will also be willing to make the necessary investments in terms of time and intellectual efforts to acquire new knowledge. As observe in the data, such motivation may be particularly strong for farmers working at a medium productivity level who face good prospects of becoming professional farmers and fully commercialize their agricultural activities[9].

Agricultural extension and productivity. For those crops for which we find a strong and unambiguous correlation between reported demand for knowledge and productivity, such as *wheat*, the UNDP data allow to estimate the effect of using agricultural extension to close existing knowledge deficits on farmers' incomes.[10] For example, we find that, with 99% statistical significance, a one-step reduction in the reported demand for knowledge in *soil preparation* is on average associated with an 800 kg increase in wheat yields per hectare. By plugging in data on the average size of wheat producing farms (12 hectares), and the price of wheat (55 tetri per kg), we can calculate that a one-step reduction in knowledge gaps pertaining to soil preparation would increase the income of an average wheat-producing farm by 528 lari. Similar estimates can be made for other knowledge areas.

---

[9] This point was highlighted at stakeholder workshop that was held at ISET on September 14, 2016.

[10] Such calculations rely on a rather strong assumption that causality runs from reported knowledge gaps to productivity. As a matter of fact, both farmers' productivity and demand for knowledge may be "caused" by another factor such as access to irrigation. If that's the case, extension measures alone would not fundamentally change either the farmers' appetite for knowledge or yields.

Importantly, farmers' incomes can also be increased through extension measures that target *farm management and marketing skills*. The UNDP data suggest that **gaps in entrepreneurial knowledge and marketing skills ("trading and selling") are significantly and positively (!) correlated with the yields of virtually all products**. This seemingly puzzling finding is easy to interpret: marketing skills are particularly useful for the more successful, commercial farmers who sell at least some part of their produce. Such farmers can benefit from *farm management training* in two ways: by increasing agricultural output and the *prices* at which these outputs are sold. Prices can be affected through timing decisions (when to seed, harvest and sell – assuming some degree of storability) as well as through farmers' ability to negotiate.

We estimated the correlation between selling product prices and each of the 18 farm management knowledge variables (9 subject areas each for skills and knowledge). **Statistically significant links between prices and farm management skills have been found only for those products in which farmers enjoy some degree of market power**:

- **Tomatoes**: better prices can be achieved through better post-harvest treatment, storage, packaging, transportation and choice of selling markets;
- **Tropical fruits**: given limited supply, farmers can achieve better prices through bargaining;
- **Walnuts**: prices can be very significantly affected by post-harvest treatment (drying) and storage.

We find no correlation between management skills and prices for internationally traded commodities, such as wheat. Georgian farmers are price takers as far as the latter are concerned. It goes without saying that extension measures focusing on marketing and farm management skills should first and foremost target those product categories in which they have the largest effect on prices and productivity.

Agricultural extension and structural change. Extension measures can also impact the direction and intensity of structural changes that Georgia's agricultural sector will be undergoing in the coming 5-15 years. We apply a clustering methodology to identify those households that are likely to exit the agricultural sector because they are unable to generate sufficient income and/or keep up with competition and technological change. **In our baseline scenario, which assumes legal and institutional status quo and no policy interventions, we expect close to a half (46.5%) of Georgian farmers to leave agriculture.** Of those remaining, we expect 21.13% to be *subsistence farmers* (only producing for own needs), 25.11% – *semi-subsistence farmers* (whose consumption accounts for more than half of the total value of their agricultural produce), and 7.25% – *professional farmers* (whose consumption accounts for less than half of the total value of their agricultural produce).

Further, our methodology allows to quantify farmers' transitions between different categories and out of agriculture. As may be expected, we estimate that a disproportionately large share of current subsistence farmers (almost 55%) will be among those leaving agriculture. Conversely, only 36% of the current professional farmers will exit the sector. A sizeable portion (24.13%) of today's subsistence farmers will upgrade to the semi-subsistence level, but only a tiny fraction (0.56%) of them will become professional farmers in the future. Finally, only 12% of current professional farmers will retain their status in the future. **We further estimate how structural change and internal transitions will be affected by policy interventions such as agricultural extension and capital relief**. We model extension through a counter-factual scenario in which all knowledge gaps in farm management and crop production are entirely closed. It turns out that the probabilities to leave agriculture and move from one farmer category to another are hardly affected by an extension strategy seeking to eliminate knowledge gaps of the entire farmer population. For example, the share of households that leave agriculture in the extension scenario (46,77%) is roughly equal to the baseline scenario (46,50%).

Of course, this result is very sensitive to our assumptions concerning a lack of systematic differences among farmers in terms of their learning abilities and motivation to close existing knowledge gaps. Assuming (more realistically), that more productive farmers are better placed to benefit from extension, the latter may result in greater polarization, with a relatively larger share of subsistence and semi-subsistence farmers leaving the sector, and better farmers getting even better.

Capital relief and structural change. In addition to the extension scenario, we modeled two capital relief scenarios. According to the first scenario, each household gets a capital injection equivalent to a lump sum payment of 10,000 lari (to be invested in the household's agricultural operations). According to the second scenario, each household receives capital relief that corresponds to twice its yearly income, also to be invested in farming activities. This second scenario is motivated by the fact that current agricultural income may serve as a proxy for the size of collateral available to a household. Based on data on loan interest rates, we estimated the average real return on investment in Georgian agriculture at 21%.[11]

Unlike agricultural extension, **provision of additional capital leads to considerable adjustment in farmers' transitions among different categories and out of agriculture**. The lump sum intervention reduces the number of farmers that are likely to shut down their agricultural activities from 46.50% to 42.58%, and increases the number of professional farmers from 7.25% to 12.03%. Its impact is strongest on today's professional farmers, dramatically reducing the probability of them leaving the sector or move downwards on the commercialization scale. Moreover, the percentage of today's professional farmers who maintain their status in the future more than doubles in this scenario from 11.59% to 25.67%.

[11] A detailed explanation of how we got to this percentage can be found in Section 4.4.

This result may be due to the fact that a relatively large group of commercially oriented farmers possess sufficient agricultural assets (such as land) but may be not productive enough to stay in agriculture in the medium term. By adding to their productivity and income, a capital intervention can increase their willingness and ability to remain engaged in commercial agriculture (regardless of age and education). As far as subsistence and semi-subsistent farmers are concerned, the binding constraint for most of them to stay in agriculture is apparently not income per se but meager agricultural assets.

Compared with the lump-sum scenario, the percentage of households that leave agriculture goes up from 42.58% to 44.68%, if capital relief is a function of existing income. However, it is still lower than 46.5%, the share of households that abandon agriculture in the baseline (no intervention) scenario. **To sum up, capital relief, particularly if provided in a relatively "egalitarian" way, appears to be an effective means of (a) slowing the process of structural change out of agriculture, and (b) encouraging already relatively commercial farmers to further commercialize their agricultural activities.**

Speed of transition. In addition to the sheer magnitude of structural change we concern ourselves with its speed, i.e., how fast farmers move from one category to another and/or out of agriculture. As may have been expected, we find that **the more equally a policy measure benefits different types of farmers, the slower is the process of structural change**. Using the so-called *trace index* from the theory of dynamical systems (which takes the value 1 for the most dynamic and 0 for a completely static system), we show that the speed of change is highest without any intervention (trace index of 0.80) and it is only slightly lower with extension (trace index 0.795). The process of structural change is slowed down considerably (trace index 0.77) by capital relief that is proportionate to current income, and the maximum deceleration is achieved with a lump-sum transfer of 10,000 lari (trace index 0.75). While these quantitative results are heavily dependent on the underlying assumptions, our qualitative conclusion stands firm: **a lack of targeting in the delivery of capital relief (e.g. subsidies, vouchers, etc.), agricultural extension and similar measures will inevitably slow down the speed of consolidation in the Georgian agricultural sector.** This does not necessarily mean that current government policies, which often do not discriminate between beneficiaries, are misguided. In fact, keeping people in subsistence or semi-subsistence agriculture may be a desirable policy objective as long as Georgia's urban economy fails to create enough jobs for those leaving the agricultural sector.

It is worth reiterating that in our analysis of agricultural extension we assumed that it has the same impact (closing all knowledge gaps) on all farmer categories. Even under this highly unrealistic assumption we find that agricultural extension does not have much influence on the extent and speed of transition out of agriculture, on the one hand, and farm commercialization, on the other. Yet, as has been discussed above, extension is very likely to disproportionately benefit the more successful, younger and better educated farmers, those who may be about to break into commercial farming or are already successfully producing for the market. Taking that into account, **extension may be expected to increase existing inequalities in incomes and farm profitability, pushing low-performers from the market more quickly than otherwise.**[12]

Training and structural change. The knowledge gaps elicited in the questionnaire are related to specific agricultural activities, like “soil preparation”, “business plan development”, and “animal healthcare”. Each of these knowledge gaps is connected to a well-defined set of tasks. Soil preparation, for example, comprises various types of tillage, weeding, application of chemicals, etc. These techniques and methods can be learned without understanding underlying theoretical concepts, and are for this reason the bread and butter of public and private extension efforts around the world.

Training, on the other hand, may relate to the creation of more comprehensive agricultural competencies, answering not only “what” but also “why” questions. While training typically remains below the level of higher education, it may provide the space for teaching essential theoretical foundations, allowing farmers to autonomously react to challenges posed by the environment and make the right decisions in non-standard situations.

To assess the importance of training measures for agricultural productivity and incomes, we looked at how farmers’ performance correlates with their formal educational achievements. It turns out that farmers’ performance does correlate in statistically significant way with secondary, vocational, and bachelor’s level education (but not with other education levels). This finding suggests that training measures that tackle general knowledge gaps may be an effective means of promoting productivity. Yet, while statistically significant, the impact of general knowledge on productivity is found to be rather modest in *size*. The strongest increase of the marginal annual income of a household, by 308 lari, is associated with the addition of a bachelor’s degree-holding family member. The respective values for vocational education and secondary education are 271 and 191 lari per annum.

This improvement appears to be particularly modest when compared with the potential impact of agricultural extension. For example, eliminating all 17 knowledge gaps in crop production and farm management would hypothetically entail an average increase in annual household income of 6,936 lari! As a caveat, however, one should bear in mind that even the most comprehensive extension program will

[12] This will happen, for example, because extension will increase the cost of agricultural land, incentivizing farmers who do not benefit from extension to sell their plots to the more successful neighbors.

not be able to eliminate all knowledge gaps, and that considerable efforts (and time!) would be necessary to approach this ambitious goal.

In any case, it is our view that **agricultural extension that addresses the most important knowledge gaps (e.g. in farm management and crop production) and is targeted at the most motivated learners (aspiring wheat growers, female dairy workers, etc.) is likely to be more effective** than general training measures which would be equivalent to raising the overall education level of Georgian farmers.

Finally, we looked at the extent to which extension and capital relief measures affect socially vulnerable households, which we define to be those households that fulfill two criteria: (a) their per person non-agricultural income (in *purchasing power parity*) is below the World Bank poverty threshold of $3.10 per day (which corresponds to $1.78 of nominal income), and (b) our methodology predicts that they will leave the agricultural sector within the next 5-15 years. The rationale behind this definition is that a household which cannot remain agriculturally active but does not (yet) have profitable outside options is under a high risk to end up in poverty. Currently, 22.8% of Georgian rural households are vulnerable according to this definition.

We find that **agricultural support polices, such as extension and capital relief do not greatly reduce the number of vulnerable households.** Perhaps paradoxically, their number even slightly increases with extension (22.94%). For one thing, agricultural extension does not affect people's income from non-agricultural sources (a criterion we have used to define the group of vulnerable households). Second, knowledge is apparently not the binding constraint for households that, according to our methodology, are likely to exit the agricultural sector[13]. Rather, their status has more to do with meager and unproductive agricultural assets. Also, the share of socially vulnerable households decreases from 22.8% (in the baseline scenario) to 20.9% with lump sum capital relief, however, even this result does not provide strong support for such policies as a means of reducing the share of vulnerable households in total rural population. **To be more effective, social relief measures should not be indiscriminate but target those whom we identify to be most threatened by structural change. In addition, one may try to tackle the challenge of social vulnerability by encouraging subsistence and semi-subsistence farmers to acquire professional skills that are suitable for services and manufacturing employment, and at the same time**

[13] We do not assume here (and throughout the report) that farming households systematically differ in their ability to close knowledge gaps ("learn"). If, however, learning is positively correlated with current productivity, any extension and training measures would disproportionately benefit the better off farmers, leading to greater income and social status gaps. Greater inequality would, in turn, trigger faster exit from agriculture by the weakest farmers.

foster the creation of non-agricultural jobs in rural areas. It is uncertain, however, to what extent other sectors can absorb the superfluous agricultural labor force (see Section 3.1).

Importantly, the speed of consolidation in Georgia's agricultural sector is a strategic decision to be made through the political process. Economic policies, including agricultural extension, subsidies and capital interventions should then be aligned with the strategic objective. As we have discussed above, economic assistance and social relief measures differ in their effects on the speed of consolidation. At present, some segments of Georgian agriculture can become much more productive without necessary forcing smallholders out of the sector and into (inexistent) urban jobs. Examples are export-oriented branches of agriculture such as hazelnuts[14]. However, in the slightly longer run, some consolidation is inevitable, and relief measures, including active labor market policies, would have to be enacted for the transition out of agriculture to proceed smoothly.

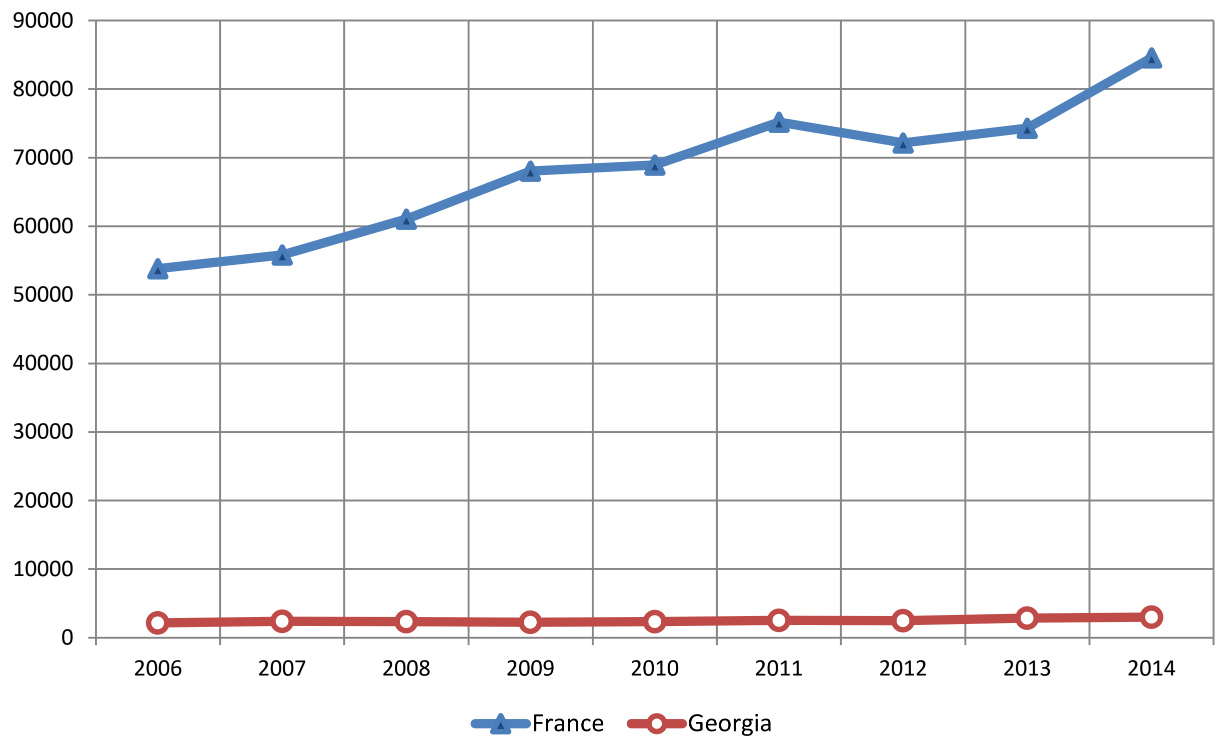
[14] This point was raised at the concluding stakeholder workshop on September 14, 2016.

CHAPTER 1. AGRICULTURAL KNOWLEDGE AND FARM PERFORMANCE

1.1. Productivity in Georgian Agriculture

In 2012, sixty people employed in Georgian agriculture generated approximately as much value as one French agricultural worker.15 Until 2014, this ratio had not substantially improved, as can be seen in Figure 1-1, which compares the value added per agricultural worker from 2006 to 2014 in Georgia and France.[16]

Figure 1-1: Value added by an agricultural worker in France and Georgia (in 2005 dollars). Source: World Bank



When comparing Georgia with France, a country with one of the most advanced agricultural sectors in the world, this result may not come as a surprise. Yet, also if one compares Georgia to other transition countries, Georgian farmers perform worse than their peers. Figure 1-2 depicts the value added per agricultural worker in Armenia, Bulgaria, Georgia, and Romania.

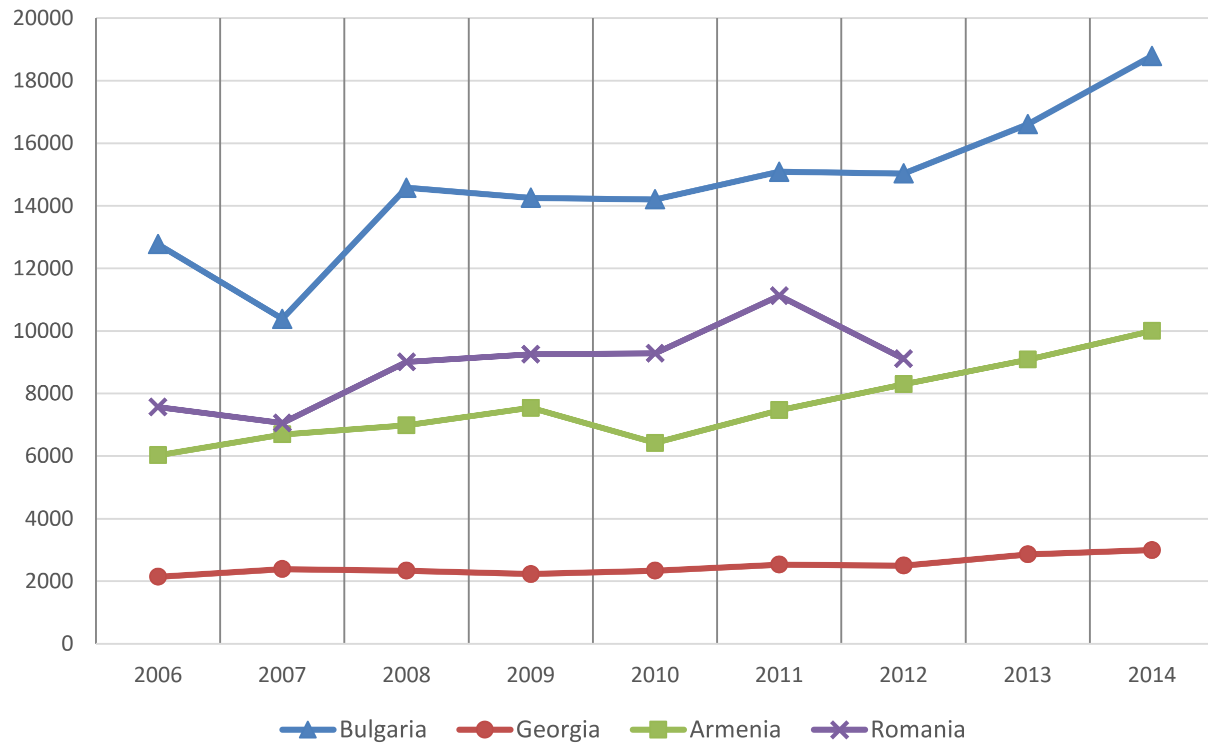
---

[15] An analysis of the strengths and weaknesses of the Georgian economy can be found in Biermann et al. (2013).

[16] World Bank data. Agricultural output per worker is defined as the revenue of the agricultural sector at market prices, reduced by the expenses for intermediate inputs which cannot be attributed to the agricultural labor force. The resulting number is divided by the number of people employed in the agricultural sector. The agricultural sector includes the cultivation of crops, livestock production, forestry, hunting, and fishing.

The productivity dynamics shown in the figure correspond to economic developments in the agricultural sectors of the respective countries. For instance, the curves of Romania and Bulgaria display the productivity leap that followed the entry into the European Union in 2007. As can be seen, in Armenia, Bulgaria, and Romania agricultural productivity evolves dynamically. In Georgia, however, the development of productivity is essentially a flat line, which implies that there were neither external events nor policy measures that changed the dreary conditions in Georgia's agricultural sector. This implies that in the last 10 years, the Government of Georgia did not give priority to agricultural development or policy measures did not lead to improvements.

Figure 1-2: Value added by an agricultural worker in Armenia, Bulgaria, Georgia, and Romania (in 2005 dollars). Source: World Bank



Consider the curve of Armenia, a country that is neighboring Georgia and shares a similar history. The productivity gap between the two countries, as far as it does not result from unreliable Armenian statistics, is often attributed to specific problems of Georgia ("civil war and many years of lawlessness", "loss of the Russian market", "land fragmentation and lack of cooperation among small farmers" etc.) and to particular strengths of the Armenian agriculture ("better utilization of Soviet era infrastructure, including irrigation", "higher concentration of agricultural machinery" etc.).[17] There is another factor, however, that was shown to play an important role in explaining productivity differences among many countries: the availability of knowledge and skills. As will be explained in the next section, its importance derives from the fact that it is inseparably entangled with all other factors that influence productivity. Besides its other objectives, this chapter attempts to take stock of the "knowledge factor" in Georgia, which could then be compared with the knowledge situation in countries like Armenia if similar studies would be carried out there. Knowledge differences may indeed explain a part of the productivity differences between the two countries.

It should be mentioned, however, that closing knowledge and skills gap is not about increasing productivity as such. Instead, the policymaker should be equally concerned about income or employment. Productivity does not always go hand in hand with these two objectives. In particular, the policymaker may want to target skill and knowledge gaps in sectors that have a future in Georgia, and sectors that are likely to generate additional employment and labor incomes, such as the relatively labor-intensive production of blueberries (as opposed to capital-intensive wheat).

To understand the exact connections between productivity, income, and employment, one has to analyze the markets for agricultural products and labor, which goes beyond the current study. In the remaining part of this chapter, we take productivity as a proxy for income and employability.

1.2. The Link between Knowledge and Agricultural Productivity

In this chapter, agricultural extension will be considered primarily as a means of increasing productivity in the agricultural sector and, ignoring specific circumstances, raise income and employability of the rural population. As shown in the last section, there is a huge potential to be realized in Georgia. At least 44% of the Georgian labor force is employed in agriculture,[18] which implies that any raise in agricultural productivity will benefit a large share of the population and in this sense be highly "inclusive".[19]

Figure 1-3 illustrates the four ways in which knowledge affects agricultural productivity:

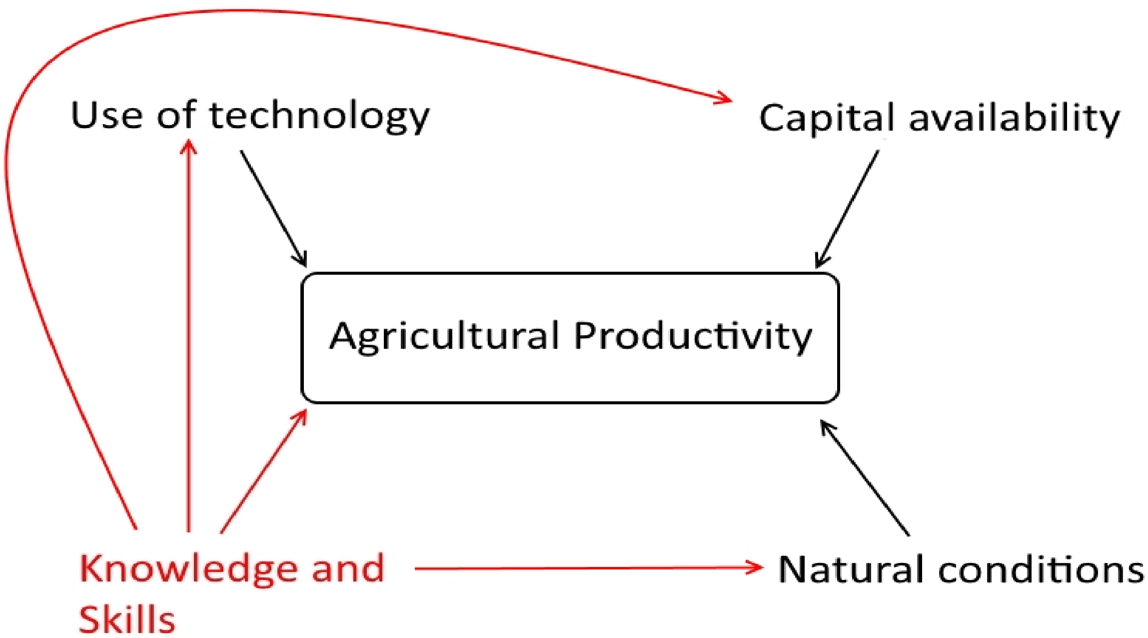
---

[17] Cf. Labadze and Livny (2012)

[18] Own calculation based on GeoStat's 2013 *Household Survey*, as the percentage of households which name agriculture as their main source of income. Other estimates are typically higher, for example, 53% in the AGRICIS TRADE Country Report Georgia 2015.

[19] At the same time, one has to consider that as agricultural productivity increases over time, the share of the labor force employed in the sector will shrink. In most advanced economies, the share of the labor force employed in agriculture is usually between one and three percent. See Bluashvili and Biermann (2013).

Figure 1-3: The interrelation between knowledge and other factors influencing productivity



Firstly, knowledge and skills are important in themselves for agricultural productivity. There is a direct link between farm output and a farmer's knowledge of when it is the best time to seed, how to prepare the soil, and how to make use of irrigation, to name just a few activities that require knowledge.

Secondly, knowledge is indispensable for the use of technology, which is in turn an important driver of productivity. Even simple agricultural machinery cannot be used and maintained without operational skills, let alone the high-tech equipment deployed in developed countries.

Thirdly, skills are also needed to acquire capital for investments. A bank will not hand out a credit in the absence of reliable figures about a farm's productivity, sound business and investment plans. Providing such information requires considerable knowledge. More generally, the perceived competence of a farmer is not only a precondition for any investors to risk their capital, but also an important determinant of the *amount* of funding that can be raised[20].

Fourthly, a farmer's ability to exploit natural conditions also depends on knowledge. For example, the quality of soil is not an unchangeable parameter. It can be improved through crop rotation and other measures. It can also be spoiled by a farmer ignorant about soil cultivation.

---

[20] The connection between knowledge and capital availability is so close that it is not unusual to treat human capital and financial/physical capital as complements in macroeconomic models. Cf., for example, Alvarez Albelo (1999) and Galor and Moav (2004).

While the role of knowledge and skills for agricultural productivity can hardly be exaggerated, it is much more difficult to measure the availability of knowledge compared with technology, capital or natural conditions. Why this is the case, and how one can tackle this problem will be explained in the next section.

1.3. Methodology

On behalf of UNDP, the Georgian polling company ACT conducted a survey among 3000 rural households and 100 agricultural companies. ACT chose a sampling strategy that included weighting and stratification, attaining representativeness in various dimensions, most notably with respect to the regions of Georgia. The details of the sampling strategy will not be discussed here (a concise description can be found in a separate document provided by ACT).[21]

The knowledge needs of rural households and agricultural companies are very different. This chapter primarily aims at the establishment of effective extension services targeting ordinary farmers in Georgia (the majority of whom are smallholders) and our analysis is, therefore, restricted to the 3000 rural households covered by the ACT survey. The much more professional agricultural companies can be assumed to have means to train their employees and acquire the knowledge necessary for their operations.

1.3.1. The questionnaire

The survey questionnaire was designed by UNDP and contains a wide range of questions on (a) household/farm and interviewee characteristics, (b) self-assessed needs for knowledge, and (c) the inputs used in agricultural production and the resulting outputs. According to the survey instructions, one interview was supposed to take around 45 minutes.

The interviews served to collect socio-economic data (e.g. age and educational achievements of the household members), data on property and ownership (e.g. size of the farm, income), and previous exposure to training and extension measures.

To elicit *knowledge needs*, the respondents were asked to assess the extent to which receiving more information about a particular activity is important in view of improving the farm operations. For each knowledge area, the answers were to be given on a scale ranging from 1 (“not important at all”) to 5 (“very important”). Potential knowledge areas included “crop production”, “animal breeding”, “animal fattening”, “dairy production”, and “farm management”. For each of these categories, between 8 and 10

[21] “Survey of Skills and Knowledge Needs in Agricultural Production: Sampling Design prepared by ACT for UNDP”, Tbilisi, March 2015. For a comprehensive overview of sampling design issues see Thompson (2012).

specific activities were named, and for each of these activities, the need for knowledge ("what to do") was elicited separately from the need for skills ("how to do it"). Figure 1-4 below shows the questionnaire excerpt for the category "crop production".

Figure 1-4: Excerpt from the survey which asks for knowledge and skills gaps in crop production

TOPICS	1. Receiving more Knowledge (What)					2. Receiving more Skills (How)					3.Most Important
K1. Crop Production											
1. Preparation of soil	1	2	3	4	5	1	2	3	4	5	1
2. Soil analysis	1	2	3	4	5	1	2	3	4	5	2
3. Selection of seeds	1	2	3	4	5	1	2	3	4	5	3
4. Seeding	1	2	3	4	5	1	2	3	4	5	4
5. Monitoring of growing process – fertilizers – agrochemicals – mechanical weed control	1	2	3	4	5	1	2	3	4	5	5
6. Harvesting technologies	1	2	3	4	5	1	2	3	4	5	6
7. Post-harvesting activities (storage, packaging, transportation)	1	2	3	4	5	1	2	3	4	5	7
8. Selling of product / Trading	1	2	3	4	5	1	2	3	4	5	8

1.3.2. The regression method

Regression analysis is a statistical method seeking to establish the covariation between the *explained variable* and a set of *explanatory variables*. In our case, the explained variable is typically an output or performance indicator, for example, the amount of wheat harvested per 100 square meters, while the explanatory variables are different self-reported knowledge gaps and household characteristics.

The most commonly applied regression technique is the so-called *ordinary least squares estimation* (OLS): the coefficient of an explanatory variable, which captures the covariation with the explained variable, is determined such that the sum of squared distances between the observed values of a variable and its estimated values, using that coefficient, is minimized. Least squares works if the underlying connection between the explained and the explanatory variables is *linear*, i.e. a change in an explanatory variable is correlated with a change in the explained variable to a degree which is independent of the value of the explanatory variable at which this change occurs. This linearity assumption may not be fulfilled when it comes to farmer's knowledge. Conveying basic knowledge to completely ignorant farmers may have a greater impact than deepening the knowledge of people who have already a certain level of expertise ("diminishing returns"). However, the exact non-linear connection is not known, and assuming that non-linearity takes a certain form, as can be done for example by estimating a logistic regression curve, presupposes a non-linear connection that, if wrong, would lead to a dramatic misinterpretation of the data. Therefore, it is common to estimate a linear regression when information about potential non-linearities is missing.

If OLS is used when the linearity assumption is violated, the estimator loses its property to be BLUE, namely to be the *Best Linear Unbiased Estimator*, which, according to the *Gauss-Markov Theorem*, it has otherwise.[22] If one applies the least-squares estimator to estimate a non-linear covariation between two variables, the estimated value can be interpreted as the *average* covariation over the whole range of values taken by the explanatory variable. A technical description of the regression method can be found in Appendix A.

1.4. The Dunning-Kruger Effect

Usually, one would expect that knowledge gaps and output to be negatively correlated, i.e. the greater the reported knowledge gaps, the lower the output, and vice versa. This is the case with some of the estimations we report. However, we observe equally often a reverse relationship, where bigger knowledge gaps are associated with *higher* output. Such an apparently nonsensical correlation is regularly encountered in studies that rely on a self-assessment of skills and knowledge, known as the *Dunning-Kruger effect*.

In one of the experiments reported in Dunning and Kruger (1999), students had to solve exercises in English grammar, and afterwards, they were supposed to estimate how well they did in comparison with the rest of the group. It turned out that those who were in the bottom quartile "grossly overestimated their abilities relative to their peers" (p. 1126), while those who were in the top quartile slightly underestimated themselves. Dunning and Kruger's explanation is that the knowledge needed to assess one's own performance is the same as the knowledge which determines the performance itself. Since the original paper's publication, the Dunning-Kruger effect has been analyzed, confirmed, questioned, and discussed in many follow-up studies, but the overall view is that the cognitive bias it represents is present whenever people are asked to self-assess their proficiency in some activity.

[22] We will forgo the explanation of what makes an estimator "unbiased". An easily accessible discussion of the least squares regression method can be found in Wooldridge (2008). A general description of the least-squares method is provided in Chapters 2 and 3, the latter also covering the Gauss-Markov Theorem.

In terms of farmer productivity, the Dunning-Kruger effect suggests that those who have little knowledge in a certain area do not recognize their deficits, as the recognition of the usefulness of additional knowledge requires that one knows already something about the respective activity. Likewise, those who are well-informed do not realize the ignorance of others and thus underestimate their knowledge advantage over others.

In regression equations reported in this study, when the estimated β-coefficient is positive, indicating a *positive* relationship between performance and a knowledge gap, there is a suspicion that the Dunning-Kruger effect plays a role, though sometimes also other explanations may be possible.

The obvious question is how to interpret the β coefficient in case it is positive. If the coefficient is negative, then there is a straightforward interpretation: closing the knowledge gap by one step increases output by β. If, on the other hand, the correlation is positive, it is clear that there is a connection between knowledge and output, yet it is unreasonable to assume that increasing knowledge will decrease output. This problem will be handled as follows:

- If the coefficient is negative, we will interpret it in the standard way.
- If the coefficient is positive, we will assume that the Dunning-Kruger effect holds, and will report a *negative* correlation between output and a knowledge gap of *unknown size*.

Note that, if the Dunning-Krueger effect holds for at least some farmers, any negative coefficient indicates a covariation that is lower than is arguably the case in reality. A negative coefficient simply implies that (statistically) the link between a knowledge gap and output is dominated by those farmers who do not fall victim to the Dunning-Kruger effect. At the same time, the true negative correlation must be stronger than what is reported. Thus, the reported coefficients represent the *lower bound* for the true size of the effect.

1.5. Results

If one is skeptical about the accuracy of the self-assessment approach pursued in this study, one would suspect that there are no statistically significant correlations between the reported knowledge needs and a farmer's productivity. It turns out, however, that this is not the case and that there is clear evidence for the relevance of self-assessed knowledge.

As described in the previous section, we have estimated a very large number of regressions of output measures on different knowledge gaps, mostly without covariates[23], and what we report as results are *systematic observations* which cannot be explained by random correlations. Because they are systematic, they allow for the development of theories about the conditions and circumstances that would generate such patterns. These theories are crucial if one wants to derive meaningful insights from the data.[24]

1.5.1. Regressing output on knowledge and skills in crop production

We find strong (statistically significant) evidence for the impact of knowledge gaps in crop production and yields, as measured by harvest (in kilos) per 100 square meters. The estimations were made without covariates, and the statistically significant results are reported in Table 1-1.

Table 1-2: Summary of regressions of yields on knowledge and skills in crop production[25]

Explained variable	Explanatory variable	Estimated effect	Significance Level
Wheat yield per sq. m.	K: Soil preparation	-8	99%
Wheat yield per sq. m.	K: Soil analysis	-8	95%
Wheat yield per sq. m.	K: Selection of seeds	-5	90%
Wheat yield per sq. m.	K: Seeding	-6	95%
Wheat yield per sq. m.	K: Growing process	-7	95%
Wheat yield per sq. m.	K: Harvesting	-5	90%
Wheat yield per sq. m.	K: Post harvesting	-6	95%
Wheat yield per sq. m.	S: Selling/trading	-6	90%
Maize yield per sq. m.	S: Soil analysis	3	90%
Maize yield per sq. m.	S: Harvesting	2	95%
Tomato yield per sq. m.	K: Post-harvesting	22	90%

[23] In econometrics terminology, a regression is conducted *from* the explained variable (in our case the performance measures) *on* the explanatory variables.

[24] In many cases, a statistically significant gender difference could be identified. If the respondent was female, the impact of knowledge gaps on production outcome was different than for males. This is an interesting aspect that may have implications for the training measures that should be taken.

[25] The yield (explained variable) is measured in kilos per 100 square meters. The explanatory variables are different reported knowledge and skill needs. If the explanatory variable is a *knowledge* need, this is indicated by the letter "K" (like "K: Post-harvesting"). If it is a *skill* need, it is indicated by the letter "S" (like "S: Seed selection"). The values in the column "Estimated effect" are the estimated β-coefficients (see Appendix A). If an estimation is significant at 99%, the probability that the reported correlation is purely coincidental is less than 1%.

Tomato yield per sq. m.	K: Sell./trad.	23	90%
Cucumber y. per sq. m.	K: Sell./tradi.	11	99%
Cucumber y. per sq. m.	S: Growing	7	95%
Cucumber y. per sq. m.	S: Sell./trad.	9	99%
Beans yield per sq. m.	K: Sell./trad.	5	95%
Beans yield per sq. m.	S: Seed Selection	5	95%
Beans yield per sq. m.	S: Seeding	3	95%
Beans yield per sq. m.	S: Harvesting	4	95%
Beans yield per sq. m.	S: Sell./tra.	7	99%
Potato yield per sq. m.	K: Seed selection	5	90%
Potato yield per sq. m.	S: Seed selection	6	90%

As already explained in Section 3, we should keep in mind that the two opposite ways in which reported knowledge gaps correlate with yields can cancel out each other. This is the case of cabbage in the survey: the knowledge needs farmers report do not appear to be correlated with cabbage yields achieved per 100 square meters. While we find no significant correlation between relevant knowledge gaps and cabbage yields for the *whole* sample, a strong correlation may be found *within* the two sub-samples of better-than-average and worse-than-average farmers.[26]

1.5.2. Where should agricultural extension start?

There is a clear link between demand for knowledge and productivity for *wheat, as shown in* Table 1-3.

Table 1-4: Estimation results for wheat production

Dependent variable	Independent variable	Estimated effect	Significance Level
Wheat yield per sq. m.	K: Soil preparation	-8	99%
Wheat yield per sq. m.	K: Soil analysis	-8	95%
Wheat yield per sq. m.	K: Selection of seeds	-5	90%
Wheat yield per sq. m.	K: Seeding	-6	95%
Wheat yield per sq. m.	K: Growing process	-7	95%
Wheat yield per sq. m.	K: Harvesting	-5	90%
Wheat yield per sq. m.	K: Post harvesting	-6	95%
Wheat yield per sq. m.	S: Selling/trading	-6	90%

The first row says that if the reported knowledge gap in soil preparation were smaller by one step, then, on average, the yield per 100 square meters would be 8 kilos higher. The significance level of this finding

[26] In principle, one could check this possibility by looking at those farmers whose yield per 100 square meters is *above* average separately from those whose yield is *below* average. If the Dunning-Kruger effect holds, one would expect that among those with high yields, the correlation between knowledge gaps and yield is negative, while for the others it would be positive. One could then look at informed farmers only, so as to get rid of the Dunning-Kruger effect, which makes the estimated coefficients difficult to interpret.

is 99%, which means that the probability that in fact there is no correlation between the two parameters, and this finding is just a statistical coincidence, is less than 1%. The other rows are to be read analogously. Given that the average wheat producing farm grows wheat on an area of 12,000 square meters, and given the reported average price 55 tetris per kg of wheat, reducing the knowledge gap by one step in soil preparation or soil analysis would increase the income of an average wheat-producing farm by **528 lari**.

It is interesting to note that wheat producers generally do not fall victim to the Dunning-Kruger effect, as all of the estimated coefficients are negative. Wheat is the only product for which this is the case. For all other outputs, the coefficients are positive (see Table 1-1).

What makes wheat producing farmers immune to the Dunning-Kruger effect? It may be the case that one can more easily compare own wheat yields and quality to those achieved - in similar or roughly comparable conditions – by one’s neighbors, injecting a greater degree of realism in people’s assessment of own knowledge gaps.

The opposite situation is characteristic of tomatoes, as shown in Table 1-3. In both post-harvesting and selling/trading areas, knowledge is strongly correlated with yields, but in the “wrong” direction (indicating the Dunning-Kruger effect). This suggests that Georgian tomato growers have a much less realistic assessment of their own skills and productivity. This could be explained by several factors. First, unlike wheat, tomato plots are typically fenced, not allowing for an easy comparison of quality and quantity. Second, tomato yields heavily depend on expensive technology such as greenhouses and drip irrigation, which may be simply out of reach for the poorer farmers (rather than admitting to knowledge gaps, they would blame their lackluster performance on a lack of cash).

Table 1-5: Estimation results for tomato production

Dependent variable	Independent variable	Estimated effect	Significance level
Tomato yield per sq. m.	K: Post-harvesting	0.22	90%
Tomato yield per sq. m.	K: Sell./trad.	0.23	90%

This finding carries important implications for practical extension strategies. First, to be receptive to training and capacity building measures, farmers have be informed about their relative strengths and weaknesses. In other words, they have to know they can do better by slightly tweaking their cultivation or post-harvest tactics (they are always convinced they could do better if armed with expensive technology, but this makes them less receptive to training measures). Second, if government or donors are willing to launch an extension initiative, they would be advised to target groups that are aware of their own knowledge and skill limitations, such as wheat growers. Once extension is successfully established for one type of farming, it can be rolled out to other branches of agriculture, including tomatoes.

1.5.3. Should agricultural extension distinguish between knowledge and skills?

The UNDP data suggest that the farmers did not distinguish between "what" and "how", as the reported skill and knowledge needs pertaining to the same farming activity are highly correlated. Somewhat exceptional are wheat producers – a particularly well-informed group of farmers. In Table 1-2, all but one significant explanatory variables refer to *knowledge (*indicated by the letter "K", as in "K: Seeding").

For an extension strategy to work, it is essential to distinguish between knowledge and skills if the former is understood as *theoretical knowledge* and the latter as *practical/applied knowledge*.[27] In the questionnaire, however, the distinction between knowledge ("what to do") and skills ("how to do it") may have been misunderstood by the respondents.

In our opinion, the distinction between "what" and "how" does only loosely coincide with the difference between theoretical and practical knowledge. Theoretical (the biological reasons crop rotation is required in potato growing) and practical knowledge (how to implement effective rotation) can indeed be kept apart. However, the distinction between "what" (what is rotation?) and "how" is much less clear cut, and less relevant for practical training measures. For example, one cannot teach the "how" of rotation without explaining what it is. Exceptions may be activities where the needed skills are very basic and do not require training at all, and activities where the skills already exist but are not applied to relevant tasks. In most cases, however, it will be necessary to deliver the skills, understood in the "what" sense, together with the knowledge. Likewise, the teaching of "how" will usually imply the teaching of "what". Also here may be exceptions, e.g. receiving instructions in calculation and math ("how") can be conveyed without applying it to specific problems ("what"). However, we think that for practical training purposes the distinction between "what" and "how" is artificial, whereas the distinction between "theoretical" and "practical" knowledge is not.

1.5.4. Transforming farmers into entrepreneurs

Besides agricultural competencies for wheat farmers, another area to be initially targeted by extension activities are management and marketing skills. Entrepreneurial knowledge and skills ("trading and selling") is the only area in which knowledge gaps are significantly (and positively!) correlated with yields of almost *all* products. All other knowledge and skill areas are correlated with yields of particular outputs (e.g. soil analysis was correlated with the yields of wheat and maize, but not of tomatoes).

---

[27] The importance of this distinction in agricultural extension was stressed in several publications of the EU funded PRO AKIS project, e.g. Labarthe et al.: "Systematic reviews of academic literature for evaluating the effectiveness of farm advisory services", PRO AKIS 2014.

The seemingly puzzling positive relationship between entrepreneurial knowledge gaps and yields is easy explain. Most Georgian smallholders do not produce for the market at all but rather for own consumption. These farmers do not report knowledge gaps in trading and selling, because they are not selling anyway. Only the more productive farmers who bring some of their output to the market report a need for additional commercial knowledge, hence a mostly positive relationship (i.e. higher output is correlated with bigger reported knowledge gaps).

This finding is highly relevant for targeting extension measures. In particular, there is no point in delivering entrepreneurial training to farmers producing only for own consumption. Further, training for farmers selling in local village or side-of-the-road markets should differ from those producing for intermediate traders, possibly on a contractual basis, not to mention those integrated into international markets. Farmers living on subsistence agriculture may benefit the most from greater specialization and production for local markets. Those who are already selling on local markets can be brought to a level where they can become partners with intermediate traders. This requires the skills to plan production and deliveries in a way that allows for meeting contractual obligations (and concluding contracts at all), which may include the skills to calculate own costs and revenues.

1.5.5. Farm management is key

Management skills benefit farmers in two ways: not only do they increase agricultural output but also the *prices* at which these outputs are sold. Management skills affect prices through timing decisions (when to seed and harvest), as well as through farmers' negotiation skills.

Interestingly, prices and farm management skills correlate not only for internationally traded commodities, such as wheat, but also for locally sold goods (tomatoes), niche products (tropical fruits), and speculative goods (walnuts).

For each of the 18 explanatory variables in farm management (9 subject areas each for skills and knowledge), one estimation was made to measure its impact on the reported price achieved for a product. While for many goods there was no impact, for tomatoes a statistically significant effect was found in 15 out of 18 regressions (at 90% significance level). For tropical fruits, a significant effect was found in all 18 regressions! Our results are summarized in Table 1-4.

Table 1-6: The impact of management skills on prices

Product	Number of significant (90%) estimations out of 18	Average coefficient
Wheat	1	0.03
Maize	0	--
Cucumber	0	--

Tomatoes	15	-0.08
Beet	1	Only 2 observations
Carrot	6	Only 16 observations
Potato	6	0.03
Cabbage	0	--
Eggplant	10	-0,03, but only 12 observations
Onion	6	0.14
Garlic	2	Only 17 observations
Beans	2	0.19
Watermelon, melon pumpkin	0	--
Radish	0	--
Soy, barley, oats, alfalfa	13	0.07
Dry forage for animals (hay, straw, stubble)	9	0.14
Dry Beans	0	--
Drupaceous fruit (cerry, plum, peach…)	5	-0.06
Apple	0	--
Pear	0	--
Other seed-produced plants (quince, medlar etc.)	0	--
Citrus	0	--
Sub-tropical fruit (persimmon, pomegranate, fig, kiwi, feijoa etc.)	18	-0.21
Table grapes	0	--
Technical /wine grapes	0	--
Berries (strawberry, rasperry etc.)	0	--
Walnut	6	0.13
Tea	0	--
Sunflower	0	--
Tobacco	0	--
Plants of vine	0	--
Young plants of vegetable	0	--
Laurel	0	--

Based on these results it seems plausible to target management skills deficits for products in which farmers exhibit a strong *negative* correlation between knowledge gaps and prices achieved, such as tomatoes and tropical fruit. For speculative goods, the price of which highly depends on own drying and storage capacity, such as walnuts, the reported effect is positive, i.e. achievers of the best prices report bigger knowledge gaps in farm management knowledge and skills. Less fortunate farmers, who sell in the glut, apparently blame a lack of access to high quality dryers and storage, rather than knowledge, and would be less receptive to management training measures. A typical instance of the Dunning-Kruger effect.

CHAPTER 2. CLUSTER ANALYSIS OF GEORGIAN RURAL HOUSEHOLDS

2.1. Introduction

The main purpose of this section is to apply a clustering methodology to the sample of Georgian farmers for various policy purposes such as to better target training and extension measures, agricultural subsidies, etc. The clustering approach allows to partition a sample into subsamples (“clusters”) such that all elements belonging into the same cluster are “similar” with regard to some criteria of interest, and therefore can be treated in a similar fashion. A survey of the purpose of clustering and different clustering methodologies is provided by Sarstedt and Mooi (2011), Chapter 9.

Section 2.2 lays out the clustering methodology we have developed and applied to the UNDP sample of Georgian farmers. Sections 3-5 describe the variables we used to cluster Georgian farmers according to their 1) chances of survival in agriculture, 2) suitability for agricultural extension and training measures, and 3) capital relief measures, such as interest subsidies, respectively.

Each section starts with a “naïve” theory motivating the choice cluster variables, and continues with a brief discussion of clustering results such as the number of data points (farms) belonging into in various cluster and each cluster’s suitability for the policy treatment at hand.

2.2. The Clustering Methodology

For instructional purposes, let us consider a sample of farms for which the following three characteristics are known: total land used for agricultural purposes, the share of rented land in total land used, and the number of cattle owned by a farm. The sample of Georgian rural households on which the UNDP project is based does indeed include these parameters. To make this discussion more interesting, let us compare the Georgian data to those of *Bavaria* in Germany. While we do not actually possess a similar database for Bavaria, we found relevant average figures in the 2014 *Agrarbericht Bayern*. Table 2-1 shows the average values of these parameters for Georgia and Bavaria.

Table 2-2: Farm characteristics in Bayern (Germany) and in Georgia

Farm characteristic	Average Bavaria	Average Georgia
Total land used for agriculture	34 ha	0.8 ha
Share of land which is rented	48.3%	2.3%
Number of cattle	62	3.6

If we are interested in studying the link between ecological damage – slurry dissemination – and the degree of farm commercialization, we may consider the number of cattle as a proxy for slurry output and the negative ecological externalities associated with it. Moreover, the problem of slurry contamination becomes more difficult to mitigate the less land is available to a farmer. Hence, the *number of cattle per ha* may be a good way to operationalize the ecological damage. The *share of rented land*, on the other hand, may be considered a proxy for farm commercialization (only farms that generate profits can afford to rent land). The baseline hypothesis we want to test is that greater commercialization leads to greater ecological damage. How can this be done with the cluster analysis methodology?

To this end, we will decompose the sample into two clusters, one with a low share of rented land, and another with a high share. It is important to notice that “high” and “low” will not be defined a priori based on some threshold value. Rather, the division into “high” and “low” (degree of commercialization) will be the *outcome* of the clustering procedure (to be discussed in more detail below). For now, just assume that Cluster 1 contains less commercial farms, according to the “rented land” criterion, and Cluster 2 contains the farms which are more commercial.[28] The criteria according to which the sample is partitioned are called the *cluster variables*, and in this example, the only cluster variable will be the share of rented land. After identifying Clusters 1 and 2, we look at them separately and compare their properties. First, we would like to know the number of farms in each cluster. If it is true that the share of rented land reflects the degree of commercialization, we should expect that in a country like Georgia, where the agricultural sector largely consists of smallholder farms, the non-commercial cluster is relatively bigger than in Bavaria (even though Bavaria has a quite fragmented and non-industrialized agriculture relative to other regions in Europe). However, it is important to understand that the range of the share of rented land of farms that are considered commercial in Georgia may be very different from the corresponding range of commercial farms in Bavaria. Being commercial in Georgia means something different than in Bavaria, and this is captured in the clustering methodology: the assignment of a farm to Cluster 1 or 2 depends on its degree of commercialization *relative to the other farms* in its respective sample, and samples are obviously very different in Georgia and in Bavaria. An attractive property of clustering is that it is not based on an exogenously set definition. What it means to be a commercial farm in Georgia and Bavaria, respectively, will be the *outcome* of the clustering procedure.

To test the theory of ecological damage through slurry dissemination, we will examine whether the less commercial farms have more land per cattle than the more commercial farms and by how much these

[28] The number of clusters does not have to be two. We could also form *three* such clusters which would contain little, medium, and highly commercialized farms or decide to have an even greater number of clusters. The maximal number of clusters that can be formed depends on the sample size, and there are well-established rules for determining this upper bound. We will not discuss this further (for details on the number of clusters that should be formed, see Sarstedt and Mooi (2011), Chapter 9), as these rule assure that given the size of 3,000 of the UNDP sample, four clusters (as we choose in our analyses) is a number which is fully permissible.

shares differ. This could be done both in Bavaria and in Georgia, so as to see whether the strength of this connection is related to the development of the agricultural sector.[29]

In the statistical analyses of the two subsamples more can be done than just considering averages. For example, one could consider the *distributions* of the cattle stock and land in the two clusters. Based on those criteria, one might identify the percentage of farms that are below some threshold of "economic survivability" in both clusters and find out that many farms in the non-commercial cluster, but not in the commercial cluster, will disappear in the next few years. This would allow for a projection of how commercialization and the related slurry dissemination is going to evolve in the future. Our analysis in Sections 3-5 of this Chapter is based on this very approach, i.e., the identification of the share of households that fulfill certain criteria in each cluster.

It is clear that many of these questions could be answered with other statistical techniques as well. For example, one could simply measure the correlation between cattle/land ratio and the share of rented land, or use multivariate regression analysis taking into account a broader range of factors affecting the environment, other than the degree of farm commercialization. Yet, cluster analysis is a particularly intuitive way to reduce complexity and test hypotheses one has about a sample because it entails the identification and comparison of meaningful subsamples, such as "commercial" and "non-commercial" farms, which most people have a more or less clear idea about. This reduction of complexity is even more valuable in case of multidimensional clustering based on multiple[30] criteria of interest, as illustrated below.

Consider a second example with the same data, but this time, we would like to test whether the ecological balance of a farm is driven by an underlying, unobservable factor called "environmental compliance". Let us assume that compliance becomes more important the more commercialized a farm is (e.g., because inspectors will focus on larger commercial farms). Another assumption is that environmental compliance leads farms to use more land and less animals.

To divide the sample into compliant and non-compliant farms we shall identify two benchmark farms are: one with much land and few animals (the "compliant" farm) and another with little land and many animals

---

[29] Note that neither availability of land nor the number of cattle are cluster variables in this example, as they were not used to define the clusters. The only cluster variable is the degree of commercialization, i.e., the share of rented land.

[30] Note that the number of clusters is independent of the number of cluster variables. In principle, one can define several clusters based on one variable, and alternatively one could have many cluster variables which are used to define only, say, two clusters.

(the "non-compliant" farm)[31]. These benchmark farms will now be used to define the clusters. Note that in this case we are employing as cluster variables the number of cattle and land used, and not the share of rented land as in our previous example. The clustering procedure assigns each data point to one of the two benchmark farms, namely to the one which is *closer*[32] (and, therefore, more similar[33]) to that data point. Note that the sample may include many farms that are quite dissimilar to *both* benchmark farms, i.e., farms which have both *many* animals and *much* land and farms which have both *few* animals and *little* land. Also these cases will be assigned to one of the two clusters, as a farm that appears to be very different from both benchmark farms will still be mathematically closer to one of them[34].

Now, having implemented the clustering procedure, we can finally test our theory that compliance considerations cause farms to increase the ratio of land to animals. Since compliance is more important the more commercial a farm is, one should expect that the share of rented land (as before considered a proxy of commercialization) is higher among farms in the cluster of compliant farms. If that were not the case, our theory would be (statistically) refuted.

A detailed technical description of the clustering methodology can be found in Appendix B.

2.3. Clustering by Prospects of Survival in Agriculture

2.3.1. Motivation

The UNDP survey is based on a sample of 3,000 rural households which were surveyed in 2015. Assume that one would visit the very same households five years later, i.e. in the year 2020. How many of those who engage in agriculture today would still be farmers then? Moreover, what can be said about the characteristics of those people who stay in agriculture and those who leave?

The motivation for these questions is straightforward. UNDP is about to devise an agricultural extension system for Georgia, and to maximize its impact, the efforts should target specific groups of households where agricultural extension is expected to make the greatest difference. The likelihood of a household to stay in agriculture in the longer run is an important criterion in this respect.

---

[31] The benchmark farms may be actual data points that exist in the sample, or they may be "fictitious farms", created to define the clusters. In the latter case, they are not counted when the subsamples are analyzed.

[32] For the notion of closeness, several mathematical measures (so-called "metrics") are available that can be applied in n-dimensional space, the most common of which is the so-called *Euclidean distance,* discussed in the next section.

[33] Mathematically, the notion of similarity will be captured by "closeness", as described further down.

[34] The case that two data points are equally far away from the benchmarks is extremely unlikely with real-world data if at least one of the cluster variables is continuous and generated by some random or noisy process.

Similarly, the government of Georgia may accompany the upcoming structural changes in the agricultural sector by policies that alleviate social distress and support economically weak groups. These policies may focus measures to improve agricultural productivity (investment in infrastructure such irrigation, collection centers, and storage), structural relief (strengthening industries and sectors that provide alternative employment) and/or social assistance (early retirement programs, transfer payments, etc.). For designing these policies in ways that maximizes impact, it is important to know which rural inhabitants will leave agriculture in the middle run.

Both aspects are examined in detail in separate chapters on sectoral migration out of agriculture and training/extension targeting. Here we will set the framework of analysis by clustering the farm sample according to a naïve theory about sectoral migration out of agriculture.

2.3.2. The cluster variables

We postulate that the decision whether or not to give up agriculture is driven by two factors:

1. The economic value of the farm.
2. The *professional potential* of the people working on the farm.

The economic value of the farm can be assessed in terms of its assets and profitability. If the economic value of a farm is high, its owner can be assumed to be less likely to give up agriculture in the future.

Professional potential, on the other hand, refers to properties of the *farmer* which influence the likelihood of him/her to stay in agriculture.

We approximate economic value by (a) the income that is generated through agricultural activities and (b) the land that is used in the agricultural production process. We assume that the higher these parameters, the more likely is the farm to survive.

The professional potential of a farmer is operationalized through (a) the age and (b) qualification of the household head. The older a farmer, the less likely s/he is to be working in agriculture in five years from now; the lower his/her qualification, the more difficult it will be for him/her to keep up with modern technology and mechanization. The farmer's human capital also plays a central role in commercial agricultural activities which require knowledge in farm management, ranging from production planning to bookkeeping to skills to fulfill bureaucratic requirements. The existing human capital of a farmer indicates the capacity to acquire, update, and upgrade such skills.[35]

[35] A more detailed discussion of the role of knowledge and skills in agriculture can be found in Chapter 1.

Education (years of schooling) is typically a key parameter of human capital. In the following section we discuss the ways in which education could be measured in a way which suits the above considerations.

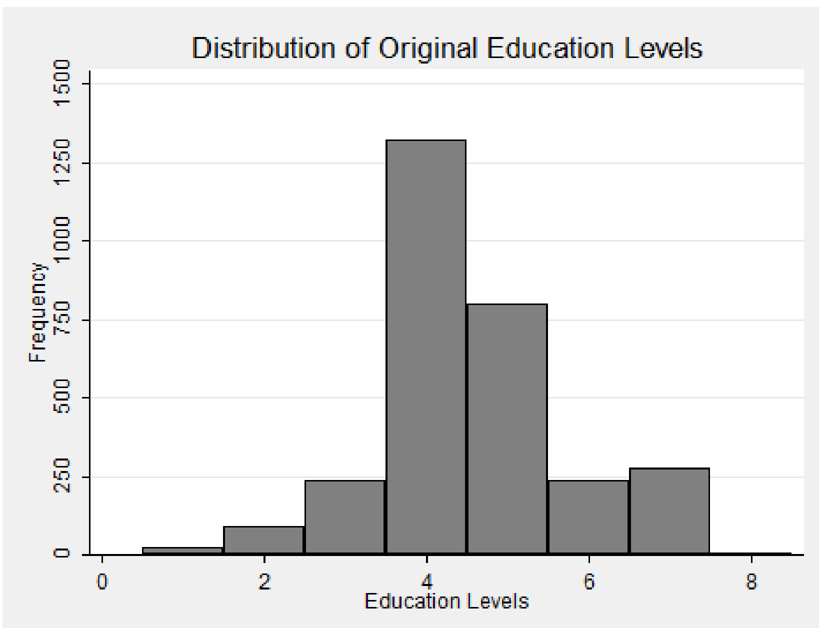
2.3.3. Measuring education

Respondents were asked to state household members' qualifications on a scale of 1 to 8, as follows:

Table 2-3: Educational achievements according to the questionnaire

Level	Qualification
1	No formal education
2	Basic education (grades 1-4)
3	Incomplete secondary education (grades 7-9)
4	Complete secondary education (grades 10-12)
5	Vocational education
6	Bachelor's degree or equivalent
7	Master's degree of equivalent
8	Ph.D. or equivalent

The distribution of answers among these 8 values is shown in the following diagram:



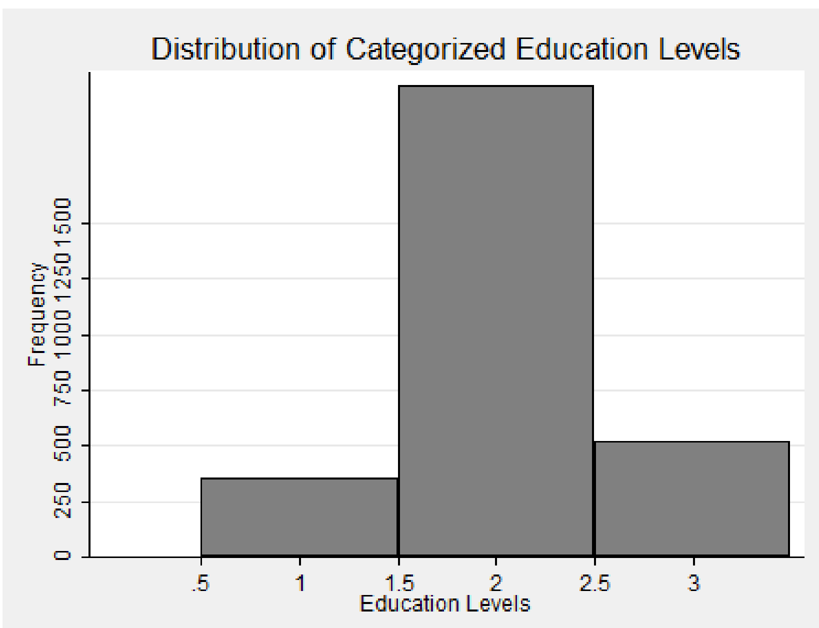
As can be seen, most respondents have incomplete secondary education, and the second biggest group is comprised of those who have vocational training (though many farmers who received a vocational degree did not complete their secondary education, which is higher on the scale).

The problem with the categorization as it can be found in the questionnaire is that for agricultural productivity, it is unclear whether complete secondary education (Level 4) or vocational education (Level 5) is more relevant. While in Georgia, secondary education arguably enjoys higher reputation (and therefore may have been reported by the respondent as the "highest achievement" even if the person also learned in a VET college), for specific agricultural expertise and skills vocational education is often much more relevant. At least, this is the case in countries where the vocational education system is well-developed and effective. In Georgia, however, the vocational education colleges are highly heterogeneous in terms of facilities, employability of graduates, competencies of teachers, and other relevant aspects (cf. the evaluation of colleges in four Georgian cities in Biermann et al. (2014)). The lack of homogeneity in the Georgian VET system makes it difficult to assess the proficiency of somebody who graduated from a vocational training college, so that it is unclear whether vocational education or completed secondary education are more relevant with respect to the potential of a households to develop its farming activities. In light of these circumstances, it would be most appropriate if the categorization would identify only three kinds of farmers:

Table 2-4: Categorization of qualifications according to relevance for farming operations

Group	Qualification	Level according to the questionnaire (see Table 2-2 above)
1	No vocational education, no secondary education	1,2,3
2	Secondary education *or* vocational education	4,5
3	Higher education	6,7,8

How the respondents are distributed according to this scheme is shown in the following diagram:



This is a categorization that will be referred to in other chapters. However, for clustering a dataset it is generally advantageous to use cardinally scaled variables instead of categorical[36] variables whenever this

[36] To remind the reader, a variable is *categorical* if it can assume only a finite number of values that do not represent different quantities or degrees of something. Rather, the value of the variable just stands for a category, not for an

is possible.[37] That this is advisable can be seen by the following consideration: if we would use the original categorical variable, which can assume 8 values, the different possible combinations of the categorical variable with the other cluster variables would call for a lot of distinct clusters. If, on the other hand, we can order the different possible values according to some cardinal criterion, the variable becomes one-dimensional (“high” vs. “low”). In each of the clusters, the variable will have either a high or a low variable instead of assuming one of the many possible categorical values. In addition, if the cluster variables are cardinally scaled, it allows for more intuitive interpretations of the resulting clusters (“highly educated” vs. “lowly educated” households).

In principle, one could also order the values of the categorical variable “education”, both as given in Table 2-2 or in the more aggregate form of Table 2-3, according to the criterion “level of education”. While it would not be categorical anymore, the distances between different levels of education could hardly be reasonably interpreted (it would not make sense to say that vocational education is equally far “above” basic education as a bachelor’s degree is “above” incomplete secondary education, though the distances between the respective values were the same). In this case, the variable would just be *ordinally* scaled, i.e., the different values of the categorical variables were put in some order according to “higher” and “lower”, yet the size of the differences in the variable’s values would not correspond to differences in education levels.

Ordinal scaling would only be a second-best solution if cardinal scaling is possible, as it is the case here. To this end, we measure educational achievements by the cardinally-scaled variable “years of schooling”, as shown in Table 2-4.

Table 2-5: Years of schooling associated with different qualification levels

Level	Qualification	Years of schooling
1	No formal education	0
2	Basic education (grades 1-4)	4
3	Incomplete secondary education (grades 7-9)	9
4	Complete secondary education (grades 10-12)	12
5	Vocational education	10.5
6	Bachelor’s degree or equivalent	16
7	Master’s degree of equivalent	18
8	Ph.D. or equivalent	22

amount. Examples are variables that denote whether the respondent has children, whether sh/e is married or not, the religion or ethnicity, province of residence, etc. In contrast, a variable is scaled *cardinally* if the difference between two values the variable assumes can be interpreted in a meaningful way.

[37] Sarstedt and Mooi (2012) even go so far to say that cluster variables should never be categorical.

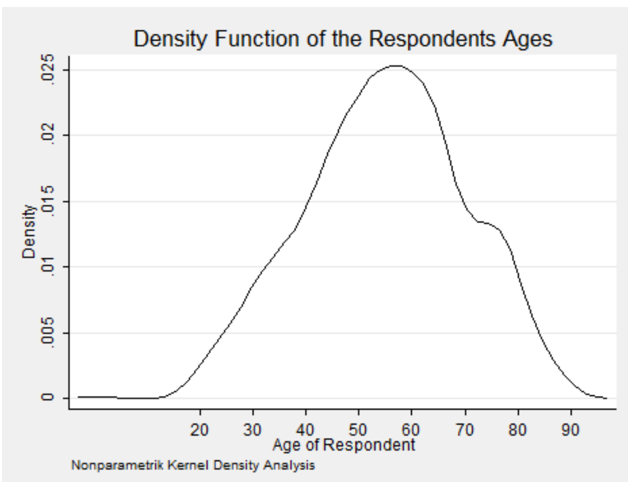
While for other types of analysis, education will also be captured by the categorization given in Table 2-3, for clustering purposes we will use “years of schooling” according to Table 2-4.

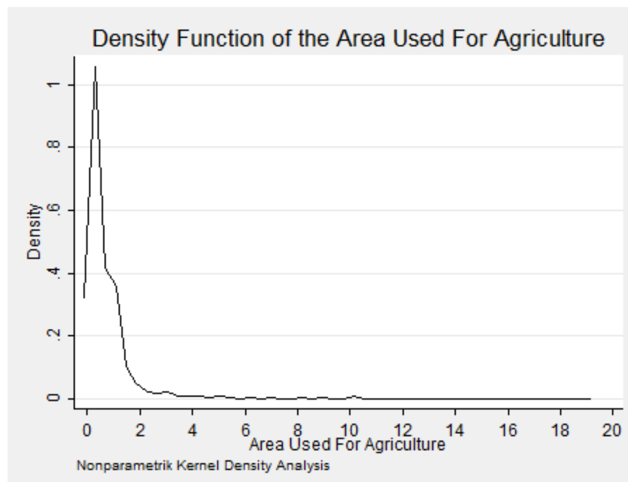
2.3.4. Clustering variables and results

Threshold households can be identified based on the means and standard deviations of relevant cluster variables:

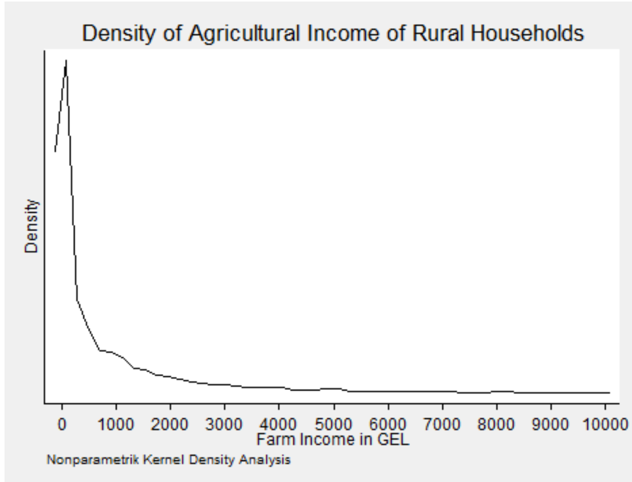
Variable	Mean	Standard Deviation
Years of schooling of household head	11.88 years[38]	2.98 years
Age of respondent	54.87 years	15.14 years
Area used for agriculture	0.8 ha	2.17 ha
Income from Agriculture	1,262 lari	4,112 lari

The following three figures display the density curves of three continuous cluster variables: *age of respondent*, *area used for agriculture*, and *income from agriculture*:





[38] 11.88 years of schooling correspond to a qualification between vocational education (level 5 in the questionnaire) and complete secondary education (level 4 in the questionnaire).



For all the threshold farmers of our clusters, we define the “high value” of a cluster variable to be about one standard deviation above the mean.[39] To define the “low value”, we often cannot use one standard deviation distance, as this could lead to a negative value (which would not make sense when it comes to income, land ownership, etc.). In such cases, we use a value which is about one quarter of the distance between 0 and the mean (the exact values chosen for the benchmark households are not important).

By these rules, we define four benchmark households as follows:

Cluster 1 benchmark household (low economic value, low potential)	
Years of schooling of household head	10.3 years
Age of respondent	64.7 years
Area used for agriculture	0.4 ha
Income from agriculture	433.6 lari

Cluster 2 benchmark household (low economic value, high potential)	
Years of schooling of household head	13.5 years
Age of respondent	45.9 years
Area used for agriculture	0.5 ha
Income from agriculture	495.3 lari

Cluster 3 benchmark household (high economic value, low potential)	
Years of schooling of household head	10.6 years
Age of respondent	62 years
Area used for agriculture	3.3 ha
Income from agriculture	6,229.2 lari

[39] If a cluster variable is Gauss-distributed, approximately 15% of farmers would have values higher than the chosen “high value”. Though none of the variables in the data set are Gauss-distributed, we chose to keep to the convention of selecting threshold values according to standard deviations so as to avoid arbitrariness.

Cluster 4 benchmark household (high economic value, high potential)	
Years of schooling of household head	13 years
Age of respondent	44.3 years
Area used for agriculture	3.5 ha
Income from agriculture	8,567.1 lari

After assigning all data points to the closest benchmark households[40] we obtained the following clusters:

Cluster	Number of observations	Frequency
#1 Low value, low potential	1,251	43.56%
#2 Low value, high potential	1,278	44.50%
#3 High value, low potential	146	5.08%
#4 High value, high potential	197	6.86%
Total[41]	**2,872**	**100%**

Importantly, training and extension measures can influence the probability of staying in agriculture for members of different clusters. This will be analyzed in chapters 3-5 of the current chapter.

2.4. Clustering for Targeted Extension and Training Measures

2.4.1. Motivation

Given that resources available for agricultural extension and training are scarce, it is important that these services are targeted at farmers who are likely to stay in agriculture and who may benefit the most in terms of productivity and profitability. One way of achieving efficient targeting is to cluster farmers according to criteria that are relevant for extension measures to achieve expected outcomes. In this section we set the framework for this type of analysis.

2.4.2. Clustering variables and results

Our naïve theory is that effectiveness of agricultural extension and training depends on two factors:

1. Farmers' *motivation.* We propose to operationalize this variable through the *average reported knowledge/skill gap*. The questionnaire asks how much the upgrading of knowledge or skill in a particular area would, in the respondent's view, improve the operation of the farm. Hence, a farmer's answer to this question can be said to reflect his/her subjective recognition of, and willingness to close, a knowledge gap.

[40] As described in Appendix B, the data were first *standardized*, so as to get rid of unit effects.
[41] The total number of observations does not add up to 3,000 because some respondents did not answer a question related to a cluster variable. These data points were removed from the sample.

2. Farmers' *aptitude* (ability or potential). We capture this variable, as before, by using information on the respondents' age and education, which reflect (however imperfectly) their ability to adopt new practices and techniques.

The following table reports the means and standard deviations of those cluster variables that capture motivation and aptitude:

Variable	**Mean**	**Standard Deviation**
Years of schooling of household head	11.88	2.98 years
Age of respondent	54,87 years	15.14 years
Average reported knowledge gap	3.21	1.25

We identify clusters for targeted extension and training measures by selecting four benchmark households:

Cluster 1 benchmark household (low motivation, low aptitude)	
Years of schooling of household head	10.1 years
Age of respondent	65.5 years
Average reported knowledge gap	2.0

Cluster 2 benchmark household (low motivation, high aptitude)	
Years of schooling of household head	13.5 years
Age of respondent	45.5 years
Average reported knowledge gap	2.1

Cluster 3 benchmark household (high motivation, low aptitude)	
Years of schooling of household head	10.5 years
Age of respondent	63.7 years
Average reported knowledge gap	4.2

Cluster 4 benchmark household (high motivation, high aptitude)	
Years of schooling of household head	13.3 years
Age of respondent	45.7 years
Average reported knowledge gap	4.2

With these clusters, we get the following frequencies:

Cluster	**Number of observations**	**Frequency**
#1, low motivation, low aptitude	682	22.74%
#2 low motivation, high aptitude	672	22.41%
#3 high motivation, low aptitude	786	26.21%
#4 high motivation, high aptitude	859	28.64%
Total	**2,999**	**100%**

One can improve the targeting of extension and training measures by focusing on farmers who simultaneously belong to two (or more) clusters. In our case, the government or donors could, for instance, decide to work with farmers at the intersection of Cluster 2 according to their survival prospects (i.e. high potential farmers whose farms are (yet) of low economic value) and Cluster 4 according to their aptitude and motivation for training measures (high motivation, high aptitude).

2.5. Clustering for Targeted Capital Relief Measures

2.5.1. Motivation

It is widely acknowledged that a lack of capital is a bottleneck for the productivity of Georgian agriculture (cf., for example, Pellillo et al. (2014)).[42] The availability of agricultural credit for Georgian farmers, however, can be influenced by policy measures. The most direct approach is to subsidize agricultural loans, as is done in Georgia, for example, by the German development bank KfW, which is re-financing agricultural loans granted by banks and micro-finance institutes (MFIs) at favorable conditions.[43] According to Pellillo et al. (2014), 13% of Georgian farmers who successfully applied for credit, reported being granted a subsidized loan.

For policies that aim to alleviate capital scarcity in Georgian agriculture, it is important to know *which* farmers are most constrained by a shortage of capital. This is essential for subsidy schemes and other relief measures to have the greatest impact. In this section we set the framework for this type of analysis.

2.5.2. Clustering variables and results

We postulate that the likelihood to obtain a loan from a bank or a Microfinance Institution depends on:

1. Size of collateral, as approximated by *total income* and *land owned* variables.
2. Creditworthiness, as reflected by personal properties of the farmer, namely, as before, age and education.

Note that the variables we used to cluster farms according to their survival chances (Section 2 above) are different from those used here to operationalize creditworthiness. For the former, we considered only income from agricultural activities; what is relevant for access to credit, is *total* income of a household.

---

[42] In Chapter 1, we discussed how different factors, including availability of capital, determine agricultural productivity.

[43] There are various other programs that subsidize agricultural investments, some of which offer loans under preferential conditions. An overview is available on the website of the *Agricultural Projects' Management Agency* www.apma.ge.

Likewise, for a farm to continue its operations, we consider the amount of land used in the production process, which in many cases does not coincide with land owned by the household. For creditworthiness, however, only the land that is actually owned is relevant, because other (e.g. rented) land cannot be used as collateral.[44]

Besides collateral and income, also the age and education of the loan applicant affect their creditworthiness. Loan applications are often refused if the applicant exceeds a certain age (if the expected remaining years of active work do not suffice to cover a realistic repayment scheme). Education also comes into play because financial literacy and communication skills provide loan applicants with the ability to better explain and justify their investment plans.

The following table summarizes the means and standard deviations of relevant cluster variables:

Variable	**Mean**	**Standard Deviation**
Years of schooling of household head	11.88	2.98 years
Age of respondent	54.87 years	15.14 years
Area of land owned	0.74 ha	1.66 ha
Total income	7,578 lari	8,586 lari

We identify clusters for targeted capital relief measures by selecting 4 benchmark households as follows:

Cluster 1 benchmark household (low creditworthiness, small collateral)	
Years of schooling of household head	10.3 years
Age of respondent	65 years
Land owned	0.5
Total income	4,649.2 laris

Cluster 2 benchmark household (high creditworthiness, small collateral)	
Years of schooling of household head	13.4 years
Age of respondent	45.5 years
Land owned	0.5 ha
Total income	4,979.1 laris

Cluster 3 benchmark household (low creditworthiness, large collateral)	
Years of schooling of household head	10.4 years
Age of respondent	62.3 years
Land owned	1.6 ha
Total income	16,220.5 laris

Cluster 4 benchmark household (high creditworthiness, large collateral)

[44] Land used as collateral must be officially registered, which is often not the case in Georgia. However, the ownership of land, even if not registered, yields *the potential* to provide collateral after registration.

Years of schooling of household head	13.7 years
Age of respondent	46.2 years
Land owned	1.6 ha
Total income	19,810.8 laris

The sizes of the clusters are given in the following table:

Cluster	**Number of observations**	**Frequency**
#1 low creditworthiness, small collateral	1,149	40.05%
#2 high creditworthiness, small collateral	1,101	38.38%
#3 low creditworthiness, large collateral	245	8.54%
#4 high creditworthiness, large collateral	374	13.04%
Total	**2,865**	**100%**

The most interesting fact about this table is that farmers who lack creditworthiness and/or a sizeable collateral (belonging in Clusters 1, 2 and 3) make up almost 87% of Georgian farmers. This finding is very much in line with previous discussion about general capital scarcity and credit constraints as major impediments for productivity improvement in Georgia's agriculture.

If access to agricultural lending is primarily constrained by farmers' advanced age, there is little one can do about it other than provide social assistance and facilitate rural dwellers' retirement from agriculture. If, on the other hand, there is a problem is with education and skills, some relief could be offered through training and extension measures.

CHAPTER 3. STRUCTURAL CHANGES IN GEORGIAN AGRICULTURE: BASELINE SCENARIO

3.1. Introduction

In this chapter we seek to forecast the structural changes in the Georgian agricultural sector that are likely to occur within the next 5-15 years assuming a baseline scenario in which the legal and regulatory frame remains as is, existing policy interventions are maintained, and no new policies are commenced. Our focus will be on migration out of agriculture, which we consider to be an inevitable consequence of Georgia's current development trajectory.[45] The primary goal of our analysis will be to characterize those farmers who are likely to terminate their agricultural activities and estimate their share in the total farmer population. In addition, we attempt to project the changes in the degree of commercialization of those farms that will stay in business. As discussed in Chapter 1, commercialization is a key characteristic of Georgian farmers in terms of farmers' productivity, income, and the extent to which they can benefit from future extension measures[46]

Among farmers who will exit the agricultural sector, we distinguish between two categories:

1. *People who leave agriculture due to old age.* Clearly, for this group there is no need to provide new employment in other sectors of the economy. That said, many of the older farmers have been engaged for many years in subsistence or semi-subsistence farming and did not save enough for retirement.[47] As they grow older, such farmers may not be able to support themselves, requiring social support exceeding the Georgian standard pension, which currently stands at 150 lari per month.
2. *People who leave agriculture before pension age.* For this group, the policy challenges are considerably more complex. The decision to leave agriculture in almost all such cases will be triggered by insufficient farming income, and there is a great danger that those selling land and moving to Georgia's urban agglomerations may not find suitable employment. In many African and Asian countries, the movement out of agriculture has led to rural depopulation and to urban

[45] As mentioned in Chapter 1, around 50% of the Georgian workforce is still employed in agriculture, while in developed economies this figure usually stands at 1-3%. Chapter 1 contains a more comprehensive discussion of the trasformation process.
[46] This finding of Chapter 1 coincides with much of the existing literature. References are provided in Chapter 2.
[47] It follows directly from the definition of subsistence farming, as presented in Chapter 2, that the accumulation of savings is not possible.

> degradation into slums (see, for example, Abumere (1981)). As elsewhere, the absorption of former subsistence farmers in other sectors of Georgian economy would require professional and personal qualities that rural dwellers do not normally possess, such as work ethics, service culture and language skills. Relatively large numbers of people who are lacking in human capital could be absorbed manufacturing (and, to a lesser degree, construction and tourism). For extra manufacturing jobs to be created, however, this sector would have to grow at rates exceeding gains in labor productivity due to increased automation.[48] In Georgia, like in other countries, this is not likely to happen. As a second-best but more feasible solution, the Georgian policymakers may want to slow down the process migration out of agriculture. This can be done, as will be discussed and quantified in the subsequence chapter, through agricultural extension measures and capital relief, lifting the agricultural incomes of rural households and reducing incentives to close down farm operations. Those deciding to leave may be targeted by social assistance measures.

An important question concerning structural change is whether and how the *recirculation* of a farm's resources takes place when it shuts down. Land is the most important factor in this respect. Raising productivity necessitates the exploitation of economies of scale which forces those farms which continue operating to utilize a greater amount of land. Yet, if smallholders who leave agriculture resist such a consolidation process by not selling their farm's resources, or if the small-scale farm operations are continued in the same unproductive fashion by successors (usually children or other relatives), this may slow down the transformation of the whole agricultural sector.

As a methodological basis of the analyses in this chapter, we develop a taxonomy of Georgian rural households according to their degree of commercialization, classifying each sampled household as a subsistence farmer, semi-subsistence farmer, or professional farmer.[49] This taxonomy is described in section 2. A projection of future mobility between these three groups is discussed in section 3.

---

[48] If manufacturing activities grow below the rate of productivity increases, the sector will not create additional demand for labor and will not be able to absorb excess rural population, as has been the case during England's Industrial Revolution in the 18th and 19th centuries. In recent years, a century-old discussion has revived, questioning the importance of labor as a factor of production in future, high-tech production processes. While the hypothesis that labor loses significance was for a long time only held by a minority of unorthodox economists, recently also mainstream economists have argued along these lines (e.g., McAfee and Brynjolfsson (2012)). If true, such a development might make the absorption of superfluous agricultural workers even more difficult, as the Georgian economy might leap directly to a technology-intensive production mode (with little demand for labor) instead of going through all the traditional stages of economic development.

[49] All households in the survey engage at least in some farming activity, as this was a condition to be included in the sample. Hence, we do not have data on rural households not connected to agriculture.

3.2. A Taxonomy of Georgia's Rural Households

While the concept of *subsistence farming* is not uniquely defined in the literature, most definitions focus on the *degree of commercialization* of agricultural activities. As Abele and Frohberg (2003) write based on a survey of scholarly literature,[50] "The preferred definition of subsistence agriculture relates it to the share of marketed produce. The lower this share, the higher is the degree of subsistence orientation."
In line with Abele and Frohberg conclusion, the definition of subsistence farming we use in this study (see Table 3-1) is based solely on the ratio of monetary income from selling agricultural produce in total income. We abstract from such context-specific parameters as the absolute value of income or any other variable (which would typically be different across countries and even regions).[51]

Table 3-2: A taxonomy of Georgian rural households according to their degree of commercialization and their distribution in the sample

Category	Definition	Frequency	Percentage
Subsistence farmer	The household has no monetary income from selling agricultural products.	1,212	40.4
Semi-subsistence farmer	The household has monetary income from selling agricultural products, but it is less than the value of the produce that is self-consumed.	1,701	56.7
Professional farmer	The household has monetary income from selling agricultural products, and it is higher than the value of the produce that is self-consumed.	87	2.9
Total		3,000	100.0

3.3. Sectoral Transformations in Georgia's Agricultural Sector: Baseline Scenario

This section revolves around the following question: which households will remain involved in agricultural activities 5-15 years from now, and if so, in what form? The methodology applied here does not warrant a more precise specification of time given the absence of time-series data on the dynamics of Georgian farmer characteristics. By projecting past patterns into the future, one could more accurately forecast the probability of exit from agriculture for different types of farmers. As such data is not available, we simply count the number of households that fulfill certain criteria which we believe are relevant for the decision to stay in agriculture and to move from one farmer category to another.

---

[50] A comprehensive survey (as of 1986) of how the term subsistence farming is used in different contexts is given in Sharif (1986). Already in 1986, there was a multitude of definitions.
[51] The taxonomy used here is similar to Davidova et al. (2013).

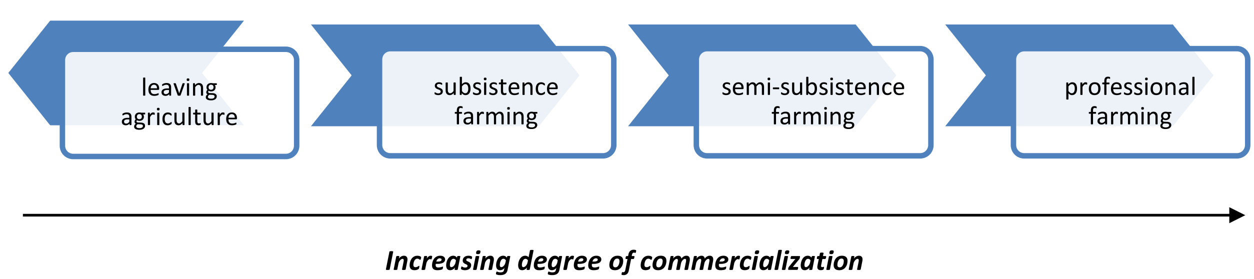
The basis for the analysis will be the first clustering conducted in Chapter 2. To remind the reader, the four clusters were defined by four benchmark households as summarized in Table 3-3:

Table 3-4 Clustering of Georgian farmers according to professional potential and economic value

Cluster #	Benchmark household characteristics (clustering variables)		Interpretation	Percentage in the sample
Cluster 1	**Potential**: Age of respondent: Education of household head: **Economic value**: Income from agriculture: Land used for agriculture:	 high low low little	Low economic value, low potential	43.56%
Cluster 2	**Potential**: Age of respondent: Education of household head: **Economic value**: Income from agriculture: Land used for agriculture:	 low high low little	Low economic value, high potential	44.50%
Cluster 3	**Potential**: Age of respondent: Education of household head: **Economic value**: Income from agriculture: Land used for agriculture:	 high low high much	High economic value, low potential	5.08%
Cluster 4	**Potential**: Age of respondent: Education of household head: **Economic value**: Income from agriculture: Land used for agriculture:	 low high high much	High economic value, high potential	6.86%

What factors are relevant with regard to the decision to leave agriculture or move from one farmer category to another? We argue that farmer's *professional potential* and *economic value*, as defined by the clustering variables in Table 3-3, can be seen as forces that push a household up or down on the "scale of commercialization", as presented in Figure 3-1:

Figure 3-1: The "scale of commercialization" according to the taxonomy presented in Chapter 2



Relatively young and better educated (*high potential*) farmers are more likely to move their farms towards higher commercialization. Likewise, *high value* farms that possess greater productive assets (land) and generate more income for their owners are more suitable for future commercialization.
Based on these considerations, the following connections between the taxonomy of Chapter 2 and the clustering of the last chapter form the basis of our estimations:

(1) **Households that are likely to leave agriculture**. We assume that low-potential-and-low-value farms (Cluster 1) are likely to exit agriculture within the next 5-15 years. By definition, farmers belonging to this cluster lack the qualities that are necessary to adopt modern farming practices. At the same time, their current farming practices and assets do not provide for acceptable income generation. The likely response of people who find themselves in such circumstances is to close down their agricultural activities and either enter retirement or search for alternative employment.[52]

(2) **Households that will engage in subsistence and semi-subsistence farming**. Farmers belonging in Clusters 2 and 3, lacking in either professional potential or economic value (but not both), are assumed to become (or remain) subsistence or semi-subsistence farmers. To achieve a higher degree of commercialization and take farm operations from subsistence to the semi-subsistence stage, farmers must be *willing* to make use of professional abilities and resources that are available to them in order to expand.[53] We infer farmers' motivation from responses to the following question "In thinking about the future, do you have a wish/desire to expand your agricultural activities/expand your farm?" In the absence of appropriate motivation, we assume that available potential will not be utilized, leaving relevant households in subsistence agriculture.

(3) **Professional Farmers.** Finally, we assume that households who have both the professional potential and high value farming assets (Cluster 4) will likely engage in professional farming activities.

Table 3-5: Projection of structural changes/sectoral movements among Georgian farmers

Today	**In 5-15 years**				
	left agriculture	is subsistence farmer	is semi-subsistence farmer	is professional farmer	Total
Subsistence farmer	**54.91%** (share of Cluster 1 among subsistence farmers)	**20.39%** (share of Clusters 2 and 3 among subsistence	**24.13%** (share of Clusters 2 and 3 among subsistence	**0.56%** (share of Cluster 4 among subsistence farmers)	100%

[52] Cf. the distinction made in Chapter 1. Trivially, old farmers will tend to enter retirement while young farmers do not have this option.
[53] However, they are unlikely become professional farmers. We believe that for this to happen, both potential and economic value would have to be present.

		farmers not willing to expand)	farmers willing to expand)		
Semi-subsistence farmer	**41.16%** (share of Cluster 1 among semi-subsistence farmers)	**21.68%** (share of Clusters 2 and 3 among semi-subsistence farmers not willing to expand)	**25.48%** (share of Clusters 2 and 3 among semi-subsistence farmers willing to expand)	**11.68%** (share of Cluster 4 among semi-subsistence farmers)	100%
Professional farmer	**36.23%** (share of Cluster 1 among professional farmers)	**21.13%** (share of Clusters 2 and 3 among professional farmers not willing to expand)	**31.88%** (share of Clusters 2 and 3 among professional farmers willing to expand)	**11.59%** (share of Cluster 4 among professional farmers)	100%
Total	**46.50%**	**21.13%**	**25.11%**	**7.25%**	**100%**

The cells of Table 3-6 represent the intersections of two partitions of the sample, one given by the clustering according to survival parameters, and another being the farmer taxonomy developed in this section. Additionally, w*illingness to expand* enters in the assignment of households to the subsistence and semi-subsistence categories.

To summarize our results (see last row in Table 3-7), we expect almost a half of Georgian farmers (46.5%) to have left agriculture in 5-15 years; 21.13% are expected to be subsistence farmers; 25.11% – semi-subsistence farmers, and 7.25% – professional farmers.

Table 3-8 provides for a side-by-side comparison of the current and future distribution of farmers across different categories, ignoring those will have left the agricultural sector.

Table 3-9: Current and future distributions according to the farmer taxonomy

Category	Today	In 5-15 years from now
Subsistence farmer	40.4%	39.5%
Semi-subsistence farmer	56.7%	46.9%
Professional farmer	2.9%	13.6%

As can be seen, the groups of subsistence and semi-subsistence farmers are expected to shrink in percentage terms – some will become professional farmers, but the majority will exit the agricultural sector; the professional farmer group will increase its relative share by a factor of five.

Some other results are also worth highlighting:

- Almost 55% of the current subsistence farmers will leave agriculture, while this number stands at only 36% for the current professional farmers.
- Only 12% of professional farmers today will also be professional farmers in the future.
- Just a marginal fraction of 0.56% of today’s subsistence farmers will become professional farmers in the future, but 24.13% will upgrade to the semi-subsistence level.

When assessing our methodology, one might question our assumption that households carrying out farming activities today ("professional farmers") need professional potential and high value agricultural assets to *retain* their status. In our view, operating in an increasingly competitive environment, the productivity threshold that has to be met by farmers in order to stay in agriculture increases over time. This requires "professional" farmers of today (who derive most of their income through commercial selling of agricultural produce) to continuously improve their skills and develop their businesses. The slogan "grow or disappear", frequently used to describe the choice faced by family farms in Western countries (e.g., Nuppenau (2009)), equally applies in a developing country that is subject to forceful economic transformations. Georgian farmers will only be able to retain their status if they continuously realize their development potential.

For the same reason, semi-subsistence farmers of today, who do have the right farm assets or professional potential to maintain their business but lack the motivation to expand, will move downwards in the degree of commercialization and will become subsistence farmers in the future. Our methodology, therefore, captures the well-established fact that in a growing and consolidating environment those who don't grow fall behind.

CHAPTER 4. THE IMPACT OF EXTENSION AND CAPITAL RELIEF MEASURES

4.1. The Two Objectives of Policy Interventions

Considered from a strategic perspective, agricultural policies are pursued to achieve structural (economic) and social outcomes. Any specific policy program one may think of, such as agricultural extension, insurance, subsidies, export facilitation or assistance with the introduction of new products and technologies, plays out in these two dimensions. On the one hand, the policymaker may want to influence structural change in the agricultural sector, e.g., promote productivity gains and speed up (or slow down) the farm consolidation process. On the other, they may seek to reduce poverty and cushion the economic consequences of agricultural modernization.

In this Chapter, we will study the expected social and economic outcomes of policy interventions applied to Georgia's agricultural sector. Our focus will be on agricultural extension and capital relief measures, both of which will be modeled in highly stylized ways, reflecting a *maximum*, or the *upper bound*, of what could be achieved through the respective actions. To capture the maximal effect of agricultural extension, we will look at a hypothetical situation where the knowledge gaps of all households in crop production and farm management were closed.[54] For capital relief, we will look, for example, at a scenario where each household receives subsidies or other financial support equivalent to 10,000 lari, which will be invested in farm activities. If such measures are implemented, farmers will see their incomes increased, resulting in social and structural (economic) outcomes that are very different from the status quo (no intervention) scenario discussed in the previous chapter.

To assess *structural changes*, we will consider how different policy interventions affect the likelihood to exit agriculture or move from one farmer category to another[55]. Using indices that capture the speed of convergence in dynamic systems, we will also estimate the time for these structural changes to be consummated under the different scenarios.

As for the *social consequences* of modeled policy interventions, we will project the share of *socially vulnerable* households, defined as those who are likely to leave agriculture but do not have sufficient non-agricultural income, under each scenario.

---

[54] Why we choose these two knowledge areas will be explained in Chapter 3 of this report.

[55] See Chapter 3 for a detailed farmer taxonomy.

Note that when assessing the social and structural consequences of different policy interventions, we strictly focus on the rural population and disregard any potential repercussions for the rest of society, which, for instance, may come in the form of increased taxes, changes in food security and prices.

In Section 2, we will discuss the concept of *smart interventions*, namely such policies which aim at utilizing essential information that is available to farmers instead of merely giving them to-do instruction. We will focus on how training, extension, and capital relief measures can be implemented in a smart way.

In Sections 3 and 4, we will discuss the impact of agricultural extension and capital relief interventions, respectively. In Section 5, we use analytical tools that are usually applied to the study of dynamic processes to analyze the *speed* of consolidation in Georgia's agricultural sector. Finally, in Section 6 we assess the impact of various policy interventions on the number of socially vulnerable households.

4.2. Smart Interventions

In the international context, the most frequently applied policy interventions are training, extension, and capital relief measures. The comprehensive study by Davidova et al. (2013) discusses in a lot of detail how these policies are adjusted to the specific needs of small and semi-subsistence farms (SSFs) in the EU.

Also in this report, we will focus on these types of support for farms. Policy measures that do not belong to this class of interventions are those which affect the legal environment in which farms operate (e.g., animal welfare legislation, rules regarding the use of fertilizers and pesticides, labor standards, etc.) and tax regulations. Taxes, however, can be seen as the reverse of subsidies or "negative capital relief". They are therefore implicitly covered in our analysis since we assess how the availability of capital affects structural and social developments in the agricultural sector.

The legal environment, however, is left out in our discussion. The legal environment is comprised of regulations which define the limits within which farm operations can take place. Importantly, regulations do not give farmers any freedom of choice. They force certain types of behaviors upon all farms alike, independent of their characteristics.[56] This is a flaw, as each farmer knows much better than anybody else, and in particular the regulator (government), the specifics of their farm operations. If the government gives economic agents some degree of *autonomy* in the way they adjust to a policy intervention, the informational advantage people have over the regulator can be utilized to achieve better

[56] This is not to say that the legal framework is irrelevant or that it would be possible to fully substitute the legal environment by "incentive interventions".

results.[57] For this reason, setting economic incentives and giving those who are targeted the freedom how to respond to these incentives may represent a very effective way of influencing behavior.[58]

Whether training, extension, and capital relief measures make use of the information available to farmers depends on how they are implemented. These interventions must come as *offers* made to the farmers, which they are allowed to reject, and it is important that accepting these measures is costly for the farmers. These costs must not take the form of "fees" – for extension, the farmers' costs may simply be the investment of time and intellectual efforts; for capital relief, farmers may be expected to pay interest, and/or provide necessary collateral.

The farmers' decisions whether to accept extension services or capital relief will incorporate their private information, and so these interventions can be considered "smart" in the above sense. Farmers will only opt for a loan or extension service if they sincerely believe that it will benefit them.[59]

To sum up, if interventions on which we focus come as costly offers, the private information of the farmers is tapped through what economists call self-selection: farmers will sort themselves into the right assistance categories, and in this way increase the effectiveness and efficiency of relevant interventions.

4.3. Agricultural Extension

This Section focuses on the question how extension measures will influence farmers' incomes, and through them, sectoral movements within and out of agriculture.

To answer this question, we repeat the analysis of the baseline scenario in Chapter 3 but replace actual household incomes by fictitious incomes that *could* be achieved if certain knowledge gaps were closed through extension activities. Here, we restrict ourselves to the 17 knowledge areas in the questionnaire that were related to crop production and farm management, as these were the only fields where the Dunning-Kruger Effect did not make the regression results uninterpretable.[60] The fact that crop

---

[57] A prime example of such interventions are so-called *steering taxes* or *incentive taxes*, which do explicitly not aim at generating revenue for the government. Their goal is to influence the behavior of the economic subjects, and the introduction may be budgetary neutralized through the simultaneous reduction of other taxes. For a discussion of steering taxes, see Rodi and Ashiabor (2012).

[58] The fact that quotas and prohibitions do not allow economic agents to optimally adjust their behavior is at the heart of the famous argument about the benefit of private negotiations made by Nobel Prize Laureate Ronald Coase (see Coase (1960)).

[59] The availability of capital may also be made conditional on preliminary training, as it is practiced, for example, in the loan program of the *Farm Service Agency* (FSA) of the United States (cf. Parsons et al. (2000)).

[60] As discussed in Chapter 1, the Dunning-Kruger-Effect does not make it impossible to identify knowledge gaps, yet it prevents the importance of knowledge gaps from being accurately quantified. Given that the questionnaire assessed the *importance* (rather than *existence*) of knowledge gaps, a positive correlation between a knowledge gap and farmers' performance implies that farmers who lack expertise and achieve poor results in, say, animal

production and farm management knowledge areas are not subject to the Dunning-Kruger effect suggests that the majority of farmers are willing to improve their capacity in these two areas, and the more so the weaker is their performance. This is precisely the reason we recommend (see Chapter 1) that extension activities target knowledge gaps in those particular areas.

For each household, we calculate the *maximal additional income* (MAI), expressed in Georgian lari, which could be obtained by closing all knowledge gaps in crop production and farm management. MAI is given by the following formula:

$$MAI = 17 \cdot (ARKG - 1) \cdot HUA \cdot 220$$

Here, *ARKG* is the average reported knowledge and skills gap of that farmer in the mentioned fields, and *HUA* is the number of hectares used for agriculture by that farm. There are 17 potential knowledge/skills gaps that can be closed in crop production and farm management. Hence, we multiply 17 by the average knowledge gap of that farmer (in crop production and farm management), and we deduce 1, because a farmer who reports 1 can be interpreted as having no knowledge gap (the interpretation of the answer "1", provided in the questionnaire, is "not important at all").

The term $HUA \cdot 220$ is the additional income that we assume can be gained by particular household if a knowledge gap is closed by one "step". As discussed in Chapter 1, the maximum income improvement per hectare that can be achieved by a one-step reduction in a knowledge gap is 440 GEL. We assume the average income gain per hectare to be equal to 50% of this value, i.e. 220 lari[61]. Based on these assumptions, we computed the average MAI achieved by eliminating all knowledge gaps in crop production and farm management to be 6,936 lari.

Once we have calculated MAI, we derive for each household the *potential income with extension* (PIE) as

$$PIE = EAI + MAI.$$

husbandry, prefer to specialize in other types of agricultural activities and do not attach much importance to closing knowledge gaps in animal husbandry (and vice versa).

[61] The average income gain from a one-step reduction in a knowledge gap has been calculated based on a heuristic but statistically sound solution, namely the principle of indifference for continuous variables (cf. Keynes (1921), Chapter 4). Plainly speaking, the principle of indifference says that if there is no information given about the distribution of a random variable, one should assume that it is uniformly distributed. If we apply this principle to the problem at hand, then the estimate should be the middle point between the greatest income impact that was measured for any combination of a knowledge gap and a product (440 GEL per hectare), and the lowest, (0 GEL per hectare). The alternative would be to work with the average impact over all knowledge gaps. Yet, just in the field of crop production, which has 8 knowledge gaps, there are 36 outputs with different prices. Computing the average over all $8 \cdot 36 = 288$ combinations would not be reasonable, because many households grow only few of the 36 products, and reported prices are highly volatile and unreliable for many products.

[61] To derive the table, we have re-calculated the clustering of Report 2 with the same seeds of the k-means procedure (of which we ran only one iteration) but using the fictitious income PIE.

Here, *EAI* is the existing agricultural income of that household.

PIE, the total income resulting from eliminating all knowledge gaps in crop production and farm management was computed to be equal to 8,239 lari on average. Substituting PIE for actual income, we get the following transformation (see Table 4-1[62])

*Table 4-1: Projection of structural changes/sectoral movements among Georgian farmers assuming all knowledge gaps are eliminated through agricultural **extension** measures*

Today	**In 5-15 years (extension / baseline scenario)**				
	left agriculture:	is subsistence farmer	is semi-subsistence farmer	is professional farmer	Total
Subsistence	54.11% / 54.91%	20.92% / 20.39%	23.85% / 24.13%	1.12% / 0.56%	100%
Semi-subsistence	43.96% / 41.16%	19.94% / 21.68%	27.63% / 25.48%	8.48% / 11.68%	100%
Farmer	39.36% / 36.23%	22.17% / 21.13%	25.50% / 31.88%	12.97% / 11.59%	100%
Total	**46.77% / 46.50%**	**21.11% / 21.13%**	**25.30% / 25.11%**	**6.82% / 7.25%**	**100%**

Interestingly, while the estimated additional income that can be achieved by closing the knowledge gaps is substantial, agricultural extension appears to be only marginally affecting sectoral movements within and out of agriculture. This happens because our methodology assumes (not very realistically) that farmers do not systematically differ in their a) learning abilities and b) motivation to close existing knowledge gaps. Under these assumptions, when applied to the entire farmer population, agricultural extension only marginally affects the *relative* standing of farmers. As a result, the distribution of farmers across different farmer types, as well as movement out of agriculture, are hardly affected by extension measures, failing to support our initial hypothesis that extension could serve as a means of preventing or delaying people's exit from agriculture.

If, however, today's better farmers also happen to be better *learners and technology adopters*[63], agricultural extension will disproportionately benefit the more productive farmers, generating polarization: exit from agriculture by the weaker (subsistence and semi-subsistence) farmers, on the one hand, and greater professionalization among the most productive farmers of today, on the other.

[62] To derive this table, we re-calculated Table 3-3 in Chapter 3 using fictitious income PIE.

[63] Unfortunately, the UNDP data contain no information on farmers' innate talent and learning capacity. That said, the polarization outcome (exit by the weakest farmers, greater professionalization of the stronger ones) could be generated by plugging in data on the importance of knowledge gaps pertaining to *all* knowledge areas (and not only crop production and farm management). Given that on the whole less productive farmers tend to underestimate the importance of knowledge gaps, they will be less affected by the extension treatment relative to stronger farmers who are better aware of existing knowledge and skill gaps (and/or are more motivated to learn new tricks).

4.4. Capital Relief

In this Section, we implement the same analysis looking at sectoral transitions and movement out of agriculture that could be achieved through two different capital relief measures:

1. Every household gets a lump sum payment of 10,000 lari. This money is invested in the household's farming activities.
2. Every household gets capital relief equal to two times its yearly income[64], to be invested in farming activities.

4.4.1. Scenario 1: every household receives a lump sum payment of 10,000 lari

Here, we compute the addition of income from investment (IAC) of 10,000 lari as follows:

$$IAC = 10{,}000 \cdot r,$$

where r is the average interest rate reported by farmers who received loans (21%[65]). Assuming, conservatively, a return rate of 21% and 10,000 lari of invested capital, IAC is equal to 2,100 lari.

We now compute for each household the potential income with 10,000 lari additional capital (PIC) as:

PIC = existing agricultural income + IAC.

On average, PIC, the total income resulting from the injection of additional capital, was computed to be equal to 3,360 lari. Table 4-2 provides our estimates of the expected changes/sectoral movements among Georgian farmers when substituting PIC for actual income.

Table 4-3: Projection of structural changes/sectoral movements among Georgian farmers assuming each household invests additional ***10,000 lari*** *in its farming activities*

Today	**5-15 years from now (10,000 GEL investment / baseline scenario)**				**Total**
	left agriculture	subsistence farmer	Semi-subsistence farmer	professional farmer	
Subsistence farmer	54.54% / 54.91%	20.74% / 20.39%	23.94% / 24.13%	0.78% / 0.56%	100%
Semi-subsistence farmer	39.15% / 41.16%	21.07% / 21.68%	26.57% / 25.48%	13.21% / 11.68%	100%
Farmer	29.63% / 36.23%	22.20% / 21.13%	22.53% / 31.88%	25.67% / 11.59%	100%

[64] This scenario takes into account the (realistic) possibility that lending to households may be a function of available income because the latter can serve as a proxy of collateral when applying for loans.

[65] This is a conservative estimate of the return on capital, as the average return on investment should be (well) above the average interest rate. If this were not the case, investment would not yield sufficient return to cover the cost of capital and demand for loans would shrink, causing interest rates to go down. A return rate of 21% is also unusual for agriculture in emerging economies, where returns of 30-50% for established food crop cultivation were estimated (see Udry and Anagol (2006)). As the range of different return rates found in the literature is rather large, going down to estimates of 2-4% for the United States (cf. Erickson et al. (2004)), we also report results for a return rate of 12%.

Total	**42.58% / 46.50%**	**21.31% / 21.13%**	**24.09% / 25.11%**	**12.03% / 7.25%**	**100%**

Unlike in the extension intervention, provision of additional capital leads to considerable changes in the sectoral movements. Most strikingly, it greatly reduces downward mobility among today's professional farmers while only marginally affecting the future status of subsistence and semi-subsistence farmers. This result follows from the fact that a relatively large group of commercially oriented farmers (defined as those who generate more than a half of their income from selling agricultural products) possess sufficient agricultural assets (such as land) but are not productive enough to stay in agriculture (according to the clustering procedure we implemented). By adding to their productivity and income, the capital intervention is increasing their willingness and ability to remain engaged in commercial agriculture (regardless of age and education). As far as subsistence and semi-subsistent farmers are concerned, the binding constraint for most of them to stay in agriculture is apparently not income per se but meager agricultural assets.

To see the extent to which the results are driven by our determination of the return on capital investment, we conducted identical analysis for a return rate of 12%. In this case, IAC is 1,200 lari and PIC is 2,460 lari. As could be expected, with a lower return on capital, the positive impact of the capital intervention on professional farmers is not as pronounced, however it remains significantly stronger than on other farmer categories.

While quite interesting, these results should be taken with a grain of salt given their sensitivity to the assumptions we make about the nature of the capital intervention (increasing income, but not agricultural assets) and the clustering methodology we apply.

4.4.2. Scenario 2: every household receives two times their yearly income

According to our second capital relief scenario, farmers are provided with financial support twice larger than their agricultural income. The return on capital is assumed to be 21%. Table 4-4 provides the results:

Table 4-5: Structural changes/sectoral movements among Georgian farmers assuming each household invests an amount equal to ***twice their annual income*** *in farming activities at a return rate of* ***21%***

Today	**5-15 years from now (investment equal to twice annual income/ baseline scenario)**				**Total**
	left agriculture	is subsistence farmer	is semi-subsistence farmer	is professional farmer	
Subsistence farmer	53.93%/54.91%	20.66%/20.39%	23.77%/24.13%	1.64%/0.56%	100%
Semi-subsistence farmer	41.30%/41.16%	20.57%/21.68%	27.69%/25.48%	10.44%/11.68%	100%
Professional farmer	35.15%/36.23%	21.47%/21.13%	22.80%/31.88%	20.58%/11.59%	100%
Total	**44.68%/46.50%**	**20.91%/21.13%**	**24.37%/25.11%**	**10.04%/7.25%**	**100%**

With the amount of financial relief depending on current household income (which, as one may realistically assume, could serve as a collateral when applying for loans), the result is somewhat greater polarization compared to the lump sum scenario: those who are doing better get more funding than those who are doing worse, accelerating the process of consolidation in Georgia's agriculture. For example, the number of households closing down their agricultural activities in this case (44.68%) is lower than in the baseline (46.50%) but higher than in the lump sum scenario (42.58%). As before, the positive impact is strongest on today's commercial farmers. Our results were not significantly affected when recalculated for a lower return on capital (12%), as shown in Table 4-6.

*Table 4-7: Structural changes/sectoral movements among Georgian farmers assuming each household invests an amount equal to twice their annual income in farming activities at a return rate of **12%***

Today	**5-15 years from now**				**Total**
	left agriculture	is subsistence farmer	Is semi-subsistence farmer	is professional farmer	
Subsistence farmer	54.41%	20.59%	23.53%	1.47%	100%
Semi-subsistence farmer	42.27%	20.19%	27.44%	10.09%	100%
Professional farmer	35.56%	21.33%	23.22%	19.89%	100%
Total	45.24%	20.74%	24.35%	9.67%	100%

4.5. The Speed of Consolidation

The tables about structural change computed in the preceding Section and in Chapter 3 can be interpreted as transition matrices of a stochastic dynamical system.[66] A dynamical system describes how a farm may change its "state" (or status, to use more conventional terminology) over time. The technical details will be provided in Appendix C.

To illustrate the point in non-technical language, let us focus on the development of one particular household over time in the intervention scenario given by Table 4-8 above (farms receive capital equal to twice their annual income and invest it at a return rate of 12%).

Let t denote the *time index*, which starts at $t = 0$ and assumes the subsequent integer values $1,2,3$ ... In any $t \neq 0$, the household may be in one of four states, namely *outside agriculture*, *subsistence farmer*, *semi-subsistence farmer*, or *professional farmer*. At $t = 0$, which we consider to be *today*, the household cannot be outside agriculture, because then it would not be in the UNDP sample (it was a sampling criterion that a household engages in some agricultural activities). However, the analysis can start at any point t, also one that is in the future, when the household has already left agriculture.

[66] For being proper transition matrices, one row has to be added, as will be described in Appendix C.

Given we know the state in which the household is in t, the question we ask is in which state the household will be in $t+1$. If in t the household has left agriculture, then also in $t+1$ it will be outside agriculture (we assume that households, once they have left agriculture, will not return). Yet, what about the case that the household is in t a subsistence farmer? Then, the entries in the first row in Table 4-3 can be interpreted as the probabilities of the categories to which the farm may move:

	has left agriculture	is subsistence farmer	Is semi-subsistence farmer	is professional farmer
Probabilities for the state in $t+1$	54.41%	20.59%	23.53%	1.47%

Similarly, if the household is in t in the state *semi-subsistence farmer*, then the probabilities in which it will be in the state $t+1$ are given by the second row of Table 4-9:

	has left agriculture	is subsistence farmer	Is semi-subsistence farmer	is professional farmer
Probabilities for the state in $t+1$	42.27%	20.19%	27.44%	10.09%

Finally, if the household is in t in the state *professional farmer*, then the probabilities of its state in $t+1$ are given by the third row of Table 4-10:

	has left agriculture	is subsistence farmer	Is semi-subsistence farmer	is professional farmer
Probabilities for the state in $t+1$	35.56%	21.33%	23.22%	19.89%

For $t+1$ and subsequent points of time the same analysis can be made, and in this way the development of is modeled as what is known in mathematics as a stochastic dynamical system (technical details and literature references can be found in Appendix C).

For dynamical systems, indices were developed that capture *how* dynamic the system is, i.e., whether the state of the farm changes more frequently, which means that the system is more dynamic, or rather rarely, which means that the system is rather static. The most simple of these is the *trace index* (Shorrocks (1978)) (formal definition is stated in Appendix C). If there is zero mobility between the categories, the trace index has the value 0, and if the probability to enter any of the other states is the same (25% in our case), a situation that could be considered as maximal mobility, the index takes the value 1. A higher value of the trace index therefore implies a faster dynamic process.

Table 4-11 presents the trace indices of the transition matrices corresponding to the different scenarios:

Table 4-12: Trace index scenario comparisons

Scenarios	Trace index
No intervention	0.808
Extension	0.795
2 x yearly income, 21% return	0.770
2 x yearly income, 12% return	0.757
10,000 lari, 21% return	0.756

Table 4-5 shows that *any* intervention slows down the consolidation process in Georgian agriculture (the non-intervention scenario has the highest trace index). The strongest deceleration could be achieved by providing households with capital equivalent to 10,000 lari, assuming a return rate of 21%.

Extension reduces the speed of consolidation rather marginally, which is not surprising. While the availability of knowledge increases the resilience of farms, it helps not everyone equally, e.g., the older a farmer is, the less likely it is that they will benefit from extension. Extension *increases* inequality in incomes and farm profitability, so that low-performers will be ousted from the market more quickly than without this intervention.[67] This is different with the unconditional (and less realistic) capital supply of 10,000 lari, which would benefit everyone equally.

Generally, one can see that the deceleration effect is smaller for measures which increase the inequality of farmers. This is the case whenever the support is not provided equally to everyone but with a greater likelihood to those who are already in favorable situations. Capital relief that depends on the current income and agricultural extension have this property.

The inequality-increasing effect of government interventions was observed before. With a view on the USA, Baltensperger (1987) writes: “Larger farms are better able to take advantage of government programs and policies, including tax laws, agricultural programs, and research and extension activities”. If it is a policy objective to slow down the speed of consolidation, measures which reduce inequality (in terms of farm performance and profitability) should be given priority.

[67] If farmers with potential receive training, the land prices may increase, because land will become a more profitable asset for the more productive farmers. As a result, those who did not receive extension or for whom it was not productivity-enhancing, now face stronger incentives to sell their land and leave agriculture, accelerating the consolidation process.

4.6. Socially Vulnerable Households

We define a household to be socially vulnerable if (a) the household income from outside agriculture divided by number of household members is below the *World Bank poverty threshold* of $3.10 per day (PPP), which corresponds to $1.78 daily nominal income, and (b) the household is expected to terminate agriculture in the next 5 years.

The rationale for this definition is straightforward. A household that cannot remain agriculturally active but does not have profitable outside options (yet) is under a high risk to end up in poverty.

For the baseline scenario as well as the two intervention scenarios, we have computed the expected shares of socially vulnerable rural households in Georgia. The results are summarized in Table 4-6:

Table 4-13: Share of socially vulnerable households under different intervention scenarios

	With no intervention	With extension	With additional 10,000 lari per household
Percentage of vulnerable households	22.80%	22.94%	20.90%

As shown in Table 4-6, the share of socially vulnerable households cannot be changed substantially through the two policy interventions discussed here. Indeed, it seems paradoxical that the extension intervention is predicted to *increase* the number of socially vulnerable households. The probable reason is, however, that the extension leads to increased inequality and, more importantly, to a slightly higher share of households that will leave agriculture.[68]

[68] As shown in the last chapter, the speed of the structural change is lower in the extension than in the baseline scenario. However, the number of households that leave agriculture is marginally higher in the extension scenario (46,77% with extension vs. 46,50% in the baseline scenario). This is no contradiction, as the speed of the dynamics that we measured in the last chapter not only captures how quickly farms leave agriculture, but also the movements between the three farmer categories, which are more sluggish when there is extension.

CHAPTER 5. THE IMPACT OF TRAINING

5.1. Introduction

The analyses in this Chapter focus on training, which we consider to be different from agricultural extension. Extension is directed at closing specific knowledge gaps, while training provides more comprehensive knowledge that goes beyond the mere application of techniques and methods (for a detailed discussion, including literature references, see the next Section).

There are multiple policy implications that can be derived from our analysis in this chapter. We shed light on the question who could profit from training (as opposed to extension, which was discussed previously), and what difference training could make in terms of income. The learning outcomes that could be achieved through training interventions are similar to those achieved through formal education programs (general secondary, vocational and bachelor-level education). Intellectually, training and formal education programs pose similar demands on its subject and might be achievable for considerable parts of Georgia's rural population. Very importantly, the UNDP data lend strong indirect evidence to the potential benefit of training measures: the presence of household members who have completed secondary education, vocational training, and bachelor degree programs has a statistically significant impact on farm incomes (this is not the case for other qualification levels).

As previously, we do not consider the efforts that might be required to implement policy measures of a certain scope (e.g., how many trainers or extension officers are needed etc.). Hence, we cannot compare the resulting income *gains* with the *costs* of a policy intervention. What we estimate is the *maximum* impact on household income to be achieved through training measures that upgrade the qualification of *all* household members to a certain level.

Our findings are relevant for targeting training interventions because they show that certain qualification levels do play a role in Georgian agriculture, while others do not. In this way, the analysis in this report may help design effective training programs by targeting relevant *learning outcomes* (that are equally demanding to those that are shown to affect farmers' incomes) and *people* who are capable of achieving them (e.g., able to complete secondary education, vocational training, or bachelor degree programs).

Interestingly, our analysis suggest that vocational education is associated with better farmer income, lending support to policies implemented by the Georgian government in recent years seeking to give priority to this branch of the educational system.

5.2. Training vs. Extension

The first chapters in this volume were concerned with the identification of knowledge gaps that are common among Georgian farmers and estimating the potential for income gains that could be achieved by closing them. The knowledge gaps elicited in the questionnaire are related to specific agricultural activities, like “soil preparation”, “business plan development”, and “animal healthcare”. Each of these knowledge gaps is connected to a well-defined set of tasks. Knowledge of soil preparation, for example, covers various types of tillage, weeding, and the application of organic materials and chemicals. Relevant techniques can be learned in abstraction from underlying theoretical concepts, and have therefore been in the focus of agricultural extension efforts, which typically use an applied and hands-on approach (cf. Jones and Garforth (1997)). Therefore, it is not surprising that suppliers of fertilizers, pesticides, and agricultural equipment frequently and successfully engage in extension activities by offering instructions very closely related to the application of their respective products.

Training, on the other hand, relates to the creation of more comprehensive agricultural competencies, answering not only “what” but also “why” questions. While training typically remains below the level of higher education, it may cover the teaching of essential theoretical foundations, allowing the farmer to autonomously react to challenges, diagnose a problem, and make the right decisions in non-standard situations. As Prevost (1996) writes, training “has to address the complex nature of the relationship between farming and the environment together with the objectives of professional training”. Training related to soil preparation, for example, will typically cover the beneficial role of microorganisms and earth fungi for soil fertility, explain why organic material is important, and explain which soil characteristics are conducive to the biological quality of the soil (cf. Eyhorn (2005)). This goes far beyond a hands-on “cookbook approach” to address specific knowledge gaps.[69]

In this chapter, we considers the role of more comprehensive training programs of the kind described above. We will estimate the potential for income increases resulting from more fundamental and theoretical competencies that can be conveyed in training programs. As we cannot measure the impact of training directly (too few farmers have undergone training programs so far), we look at the significance of formal qualifications in the agricultural production process in Georgia. We base our methodology on the assumption that the move from, say, completed secondary education to higher education comes

[69] While training and extension efforts have different focus, there is no clear dividing line between the two. Obviously, the distinction is rather artificial, as training can be hands-on and focusing on specific tasks, making it similar to extension, while extension can also cover some theoretical background subjects.

along with improved theoretical understanding of the same kind that could be achieved through training measures. In other words, we use the impact of formal educational degrees on farming success as a proxy for the importance of more fundamental “why” knowledge that could also be achieved through training.

5.3. Which Qualification Matters?

Our goal in this Section is to understand which qualification levels are significant for agricultural success and in this way help design training measures that make a difference in farm performance.[70]

Let us stress that before carrying out the econometric analysis, we had no preconceptions about which qualification levels might be relevant in Georgian agriculture. We would not have been surprised to find out that – unlike the closing of practical knowledge gaps through extension – the more comprehensive knowledge that can be acquired through formal education would turn out to be of little significance for farming success. In the end, the patchy quality of Georgian formal educational institutions does not warrant that obtaining formal degrees would make any difference.[71] However, it turns out that three educational levels *are* in fact relevant for agricultural achievement.

We regressed total agriculture income of households (dependent variable) on the area of land owned and variables indicating the number of household members who have achieved each of the eight different levels of education. [72] The results of our estimations are summarized in Table 5-1.

Table 5-2: Agricultural income (dependent variable) regressed on land and the number of household members with different qualifications

	Explanatory variable	Coefficient (in lari)	Standard Deviation (in lari)	Significance level (approximate values)
Land used for agriculture (in hectares)		567	41	> 99%
Number of household	**Level 1**: No formal education	25	408	4%
	Level 2: Basic education (grades 1-4)	-131	265	30%
	Level 3: Incomplete secondary education (grades 7-9)	82	150	40%

[70] Training will in most cases not result in a formal degree (though it may result in the conferment of certificates). It will therefore not be possible to formally achieve a relevant qualification level through training measures, but it may be possible to acquire *equivalent* knowledge, so that a farmer may *de facto* reach the same competence level that is associated with a certain formal degree.

[71] The severe flaws of the Georgian educational system, in particular on the schooling level, have been frequently addressed in the policy debate (see, for example, Biermann and Mzhavanadze (2016) and Kelbakiani and Livny (2015), as well as many other relevant contributions on the ISET Economist Blog, and the empirical evidence cited in those articles).

[72] For deliberations on the relevance of this categorization, the way the different educational levels should be ranked according to their relevance for farmers, as well as the translation of educational achievements in “years of schooling”, see Section 2.3 in Report 2 of this project. The analysis offered here yields additional insights for that discussion.

	Level 4: Complete secondary education (grades 10-12)	219	68	> 99%
	Level 5: Vocational education	300	93	> 99%
	Level 6: Bachelor's degree or equivalent	332	107	> 99%
	Level 7: Master's degree of equivalent	134	118	70%
	Level 8: Ph.D. or equivalent	191	1199	12%

Let us briefly remind the reader how to interpret the regression results. The numbers in the second column of Table 5-1, under the header "Coefficient (in lari)", stand for change in agricultural income resulting from a one-unit increase of the respective explanatory variable. For example, having one additional hectare of agricultural land is estimated to increase agricultural income by 567 lari, while having an additional household member of educational level 3 (incomplete secondary education) increases the annual household income from agriculture by only 82 lari.

The numbers reported in the third column, under the header "Standard Deviation (in lari)", can be used to estimate an interval around the reported coefficients in which the true coefficient will be with a certain probability[73]. The true coefficient is to be found with a probability of about 68% in a one-standard deviation neighborhood around the estimator. Thus, if we take land as an example, we can be about 68% sure that the impact of one additional hectare on the agricultural income is in the interval of 526 to 608 lari (567 ±41). If we extend the interval to two standard deviations, we will capture approximately 95% of the probability. Hence, we can be about 95% certain that the true impact of one additional hectare of agricultural land is in the interval of 485 to 649 lari (567 ±82).

The significance level of an estimator gives the probability that its real impact is not 0 (and the reported correlation is just spurious). As a regression measures a stochastic connection between to variables, it can never be fully ruled out that the reported correlation is just a random outcome. In our particular case, the observed correlations appears to be genuine at the general accepted significance level of 95% for only four of the explanatory variables. For example, the likelihood that adding one household member with level one educational achievement (no formal education) will help increase agricultural income is as little as 4%. In other words, the coefficient associated with level one educational achievement (25 lari) is absolutely devoid of statistical significance.

[73] The true coefficient is with a probability of about 68% in a one-standard deviation neighborhood around the estimator.

It is common to remove statistically insignificant explanatory variables from a regression since their inclusion reduces the reliability of all other estimates. In line with this practice, we have regressed agricultural income on only those explanatory variables that are statistically significant.[74] Our estimation results are reported in Table 5-3:

Table 5-4: The results of the regression of agricultural income (dependent variable) on land and the number of household members with different qualifications. Only significant variables.

	Explanatory variable	Coefficient (in lari)	Standard Deviation (in lari)	Significance level (approximate values)
Land used for agriculture (in hectares)		570	41	> 99%
Number of household members with different qualification levels	**Level 4**: Complete secondary education (grades 10-12)	191	62	> 99%
	Level 5: Vocational education	271	88	> 99%
	Level 6: Bachelor's degree or equivalent	308	104	> 99%

The fact that the educational levels 4, 5 and 6, namely secondary education, vocational education, and Bachelor's level, are significant for income generation in Georgian agriculture (while others are not), lends indirect support to training initiatives that aim to achieve similar learning outcomes. Our recommendation is therefore that training efforts targeting Georgian farmers be not focused on very basic abilities like reading and elementary calculations skills (levels 1 to 3). Neither should they aim to achieve learning outcomes normally associated with Master's and Ph.D. programs (levels 7 and 8).

At the same time, our estimation results suggest that while statistically significant, the economic impact of formal education (and, by implication, of training interventions targeting similar learning outcomes) turns out to be rather modest, as can be inferred from the *size* of relevant coefficients. Achieving the Bachelor's qualification has the strongest impact on the marginal income of a household. Adding a household member with this level of education increases the annual household income by 308 lari on average. The respective values for vocational and secondary education are 271 and 191 lari.

[74] The lack of statistical significance of some explanatory variables could be caused by the low numbers of people reporting relevant educational qualification. Statistical significance is very much determined by the quality and composition of the sample, and if only very few household members report e.g. Ph.D.-level qualification, it will be impossible to derive statistically significant results about their average impact on agricultural income.

Among others, these numbers illustrate that structural change in Georgian agriculture has not yet fully set in. According to economic theory, employees are paid by their marginal productivity, which, according to our estimation, is 308 lari for people with bachelor's degree[75]. The value of 308 lari is very low and indicates that many people with higher education are "stuck" in agriculture – they are lacking the ability to move to sectors that pay higher prices for their labor. With the inevitable acceleration of consolidation and structural change processes, many of the better educated people, whose qualifications *do* make a difference for the income of farms, will leave the agricultural sector and sell their labor elsewhere. Training could be conceived as a means of countering this development, making up for the loss of qualified labor in Georgian agriculture.

It should be mentioned that in this regression, land and educational achievements of household members together explain only about 7% of the variation in the agriculture income of households.[76] This number could be increased by introducing other relevant variables such as geographic location, climate, soil quality, access to capital or machinery, among many other factors. Their (purposeful) exclusion, however, does not detract from the validity of our estimation results.

It is interesting to compare the potential of income increases through training with the possible effects of agricultural extension. In Chapter 4, we computed that by closing all 17 knowledge gaps in crop production and farm management, we could bring about an average increase in farmers' annual income of 6,936 lari. While perhaps plausible, tailored extension measures that would be required to completely eliminate all knowledge gaps are by definition much more costly to deliver than general training. Quite possibly, requisite extension measures have to be very comprehensive, and their delivery would have to spread over a long time. The number of 6,953 lari is therefore to be taken as a (very high) upper bound of what can be achieved with agricultural extension in the Georgian reality.

If one offers training instead, we can consider a scenario where all household members who have completed secondary education would be lifted to bachelor's level. While practically not realistic, this exercise allows us to derive an upper bound of what could be achieved with a training aiming at such a goal. The average household in the UNDP sample has three members with completed secondary education. Upgrading their knowledge to the Bachelor's qualification would increase the average household's annual income from agriculture by $(308 - 191) \times 3 = 351$ lari.

---

[75] As these people are often owners or co-owners of their farm and not pure employees, their overall income may be higher than their marginal product, yet the marginal product is the "market value" of their labor in agriculture

[76] The R^2 of the regression is 0.073.

Of course, the analysis offered in this chapter is quite limited in the sense of only looking at the potential maximum benefits of training and extension interventions. For a more meaningful comparison of both policy alternatives, one would have to know more about the costs and administrative efforts associated with each specific policy measure at hand.

APPENDIX. METHODOLOGICAL TECHNICALITIES

A. The Regression Method

All equations of the linear regression models used in this study take the form

$$y = \beta x + \gamma_1 z_1 + \ldots + \gamma_n z_n + \alpha.$$

- In this equation, variable y is to be explained as a linear function of the explanatory variables x and z_1 to z_n. The goal of our estimation is to estimate the coefficients β and γ_1 to γ_n.
- The explained variable y will typically be some performance or output measure of the household, e.g. yield per 100 square meters, income, or selling market prices achieved.
- The explanatory variable whose impact on the explained variable y we are most interested in is x. Its coefficient β can be interpreted as the change in y if x changes by one unit. For instance, if x is "knowledge needs in soil preparation" and y is "wheat yield per 100 square meter, measured in kilos", a value of beta of -20 means that if the knowledge gap is closed by one step, the wheat yield per square meter goes up by 20 kilos.
- Variables z_1 to z_n are explanatory variables, such as individual characteristics of the respondent (age, gender, etc.), that may also account for some of the variation of y but do not correspond to the primary relation of interest.
- α is the *intercept* of the linear regression function, i.e. the estimated value of y in case all explanatory variables take the value 0.

Note that throughout this project we regress various performance and output indicators y on only *one* explanatory variable of interest together with some *covariates* (see below) rather than regressing them on all possible explanatory variables (knowledge gaps) at once. For example, we do not include skills in soil preparation and skills in soil analysis as explanatory variables in the *same* equation. The reason is that, from an econometric perspective, including several explanatory variables of interest in the same estimation equation bears the risk of collinearity (which is a very substantial problem when the explanatory variables are self-assessed knowledge gaps with a large degree of correlation among them).
Many of the reported regressions will not include any covariates. Doing so is equivalent to estimating the average effect of a particular knowledge gap, i.e. its average effect on men and women, people of different ages, educational background etc.

The number of regression equations estimated as part of the project is very high. Just for crop production, we estimated the impact of gaps in eight knowledge and eight skill areas on six explained variables, namely yields per 100 square meters for wheat, maize, tomato, cucumber, beans, and potato. This gives $16 \times 6 = 96$ regressions.

In a regression with only one explanatory variable x and an explained variable y, there is a linear relation between the estimated coefficient β and the so-called *correlation coefficient* ρ of Pearson, which is defined as

$$\rho = \frac{cov(x, y)}{\sqrt{var(x)}\sqrt{var(y)}}.$$

Here, $cov(x, y)$ is the covariance between the variables x and y, $var(\cdot)$ is the variance, and $\sqrt{var(\cdot)}$, the square root of the variance, is the standard deviation.

The correlation coefficient ρ takes values between -1 and 1 and measures the degree of linear correlation between the two variables x and y. If $\rho = 1$, then the relationship between x and y is described by a non-stochastic linear function of the form

$$y = \alpha + \beta x$$

with $\beta > 0$ and α being parameters. If $\rho = -1$, then the relation between the two variables becomes

$$y = \alpha - \beta x$$

again with with $\beta > 0$ and α being parameters. Hence, ρ is a normalized measure for the degree to which the relation between the two variables can be described by a linear function. It can be shown that in a OLS regression model with one explanatory variable, the coefficient β of that variable is given as

$$\beta = \frac{cov(x, y)}{var(x)}.$$

In view of the definition of ρ, we can write β as

$$\beta = \rho \frac{\sqrt{var(x)}\sqrt{var(y)}}{var(x)},$$

which shows that the estimated regression coefficient β is a linear function of the correlation coefficient ρ with a positive slope (both the standard deviation and the variance of a variable are positive values). This means that in a regression with one explanatory variable, the estimated β essentially measures the correlation between the two variables x and y.[77]

[77] It is important to keep in mind that correlation does not say anything about causality between two variables.

B. Clustering

Each data point in the sample can be considered a vector in n-dimensional space, where each dimension corresponds to one farm characteristic, such as the amount of land, the number of cattle, a particular reported knowledge gap, etc.

There are different possibilities to measure the *distance* between two vectors / data points. The most commonly used metric is the so-called *Euclidean distance*. If we have two vectors

$$v = (v_1, \dots, v_n)$$

and

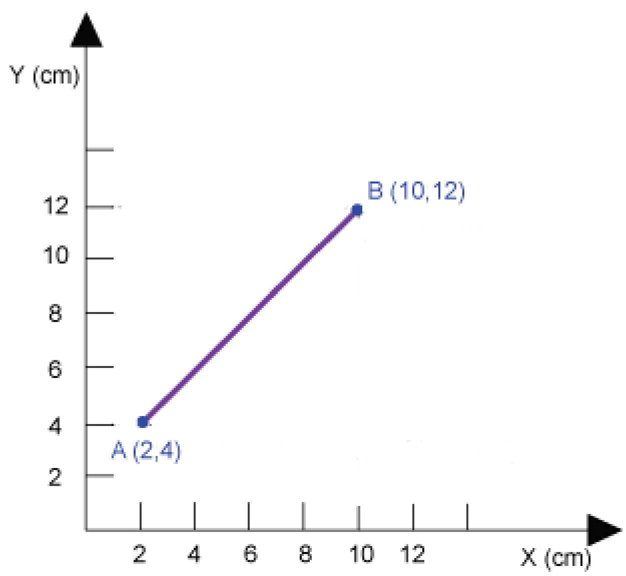
$$w = (w_1, \dots, w_n)$$

then the Euclidean distance $d(v, w)$ is defined as

$$d(v,w) = \sum_{i=1}^{n} \sqrt{(v_i - w_i)^2}.$$

In two and three dimensional space, when vectors can be represented as points in the plane or space, respectively, the Euclidean distance coincides with the intuitive notion of distance, i.e., the shortest distance between the two points, if measured with a measuring tape. Figure 1 illustrates this fact. There are two vectors $A = (2{,}4)$ and $B = (10{,}12)$, which might represent, for example, two farms where the first entry stands for the hectares of agricultural land and the second for the number of cattle.

Figure 1: An illustration of Euclidean distance

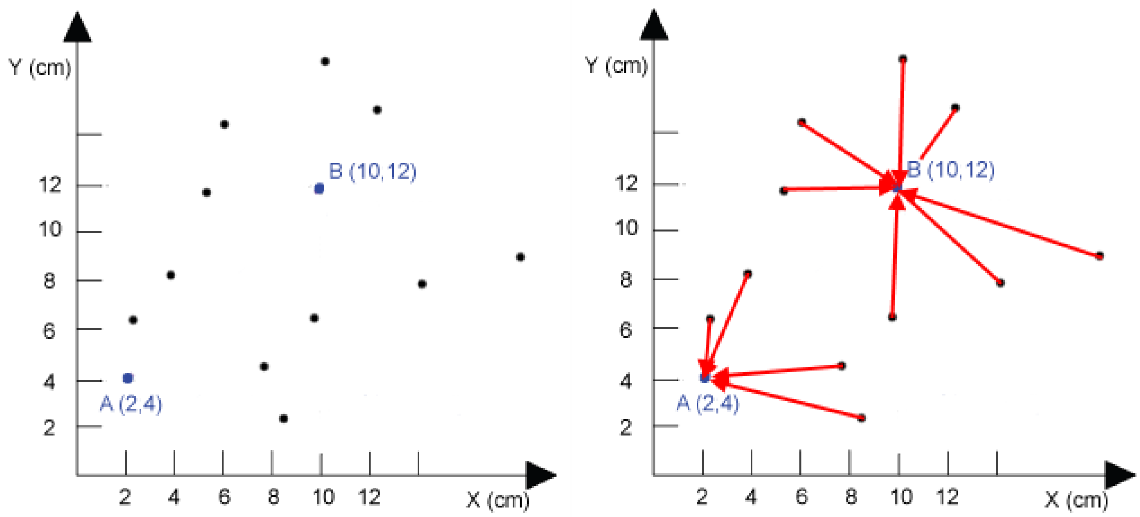


The lengths of the purple line is given by the Euclidean distance

$$d(A, B) = \sqrt{(2-10)^2 + (4-12)^2} = 11.31$$

Assume that points A and B are the values of the benchmark farms, as described in the second example of the preceding section. How will the benchmark farms be used to partition the sample? On the left hand side of Figure 2, additional data points represent other farms in the sample. On the right side, they are partitioned into clusters by assigning them to the closest benchmark farm.

Figure 2: Assigning data points to clusters



Each farm is assigned to that benchmark farm to which the red arrow points. The first cluster, defined by the farm A, contains 4 additional data points, and the second cluster, defined by the farm B, contains 6 additional data points.

This procedure can be generalized to more than two benchmark farms. If there are more than two benchmark farms, each defining one cluster, a data point would still be assigned to the cluster of the benchmark farm to which it is closest.

In this procedure, a technical problem has to be solved. In the definition of Euclidean distance, all numbers enter with their respective units of measurement. Take the example in which the two cluster variables are agricultural land and number of animals, and let us now assume that land is not measured in hectares but in square meters. In this case, the distance between two points would be primarily determined by land, which has numbers in the range of several thousands, while the number of cattle, which may be between 10 and 30, would be largely irrelevant. To see this, consider farm A with 10 cattle and 30,000 square meters of land; farm B with 30 cattle and 30,100 square meters of land, and farm C with 11 cattle and 31,500 square meters of land. Note that in terms of land, all three farms do not differ much in relative terms, yet farm B has three times more cattle than farm A, while farm C has only 10% more cattle than farm A. Nevertheless, Euclidean distance between farm A and farm C is greater than the distance between farm A and farm B. Moreover, if one would change the unit of measurement and express the land in, say, hectares, this result would change. This influence of the units of measurement is obviously an unattractive feature of defining clusters based on Euclidean distance.

To solve the problem, data can be *standardized*. In this transformation, we (a) divide each component in the vector by the standard deviation, and (b) subtract the mean of that dimension[78] Both standard deviation and mean come from all values in the sample. As a result, one obtains data which have a mean of 0 in all dimensions, and the standard deviation replaces the original units as the unit of measurement. When performing the clustering procedure in this paper, we standardized all data, regardless of the original units of measurement, as recommended in the literature (see Sarstedt and Mooi (2011), p. 247).

It should be noted that the methodology used here is a simplified version of the more commonly used *k-means* clustering method. In the k-means method, certain (fictitious or real) initial data points, termed *seeds* or *initial centroids*, are chosen as the starting points of an iterative process.[79] The seeds correspond to the *benchmark farms* in the examples above.

In the first stage of the k-means process, each data point is assigned to the seed which is closest according to Euclidean distance or some other metric. Next, in each cluster, the location of the seeds is readjusted so as to become the *geometric median* of the cluster, i.e. the point at which the average distance to the elements in that cluster is minimal. Then, a re-assignment of data points is done, based on the new location of the centroids (each data point assigned to the centroid that is closest), followed by a new re-adjustment of the centroids and so on. It can be shown that the k-means procedure converges after a finite (but possibly very large) number of steps (cf. Bottou and Bengio (1995)). However, an important caveat is that the k-means procedure does *not* converge to a unique clustering. Rather, the finally obtained clusters depend on the initial choice of seeds (cf. von Luxburg (2010)). Hence, it is common to run the k-means procedure multiple times, choosing the seeds randomly, and then take that final clustering where the average distance of the data points to their respective centroids is minimal.

Obviously, the method used in this project, as it was described above, coincides with the first stage of the k-means process. After the data points were assigned to the benchmark farms, no further adjustments were made. There are advantages and disadvantages of the method used here compared to standard k-means. The latter has two desirable properties:

1. Each data point is assigned to the centroid to which it is closest.

---

[78] Both standard deviation and mean are functions of *all* values in the sample.

[79] The term centroid is indeed misleading, as the points that are chosen in the procedure are in fact the *geometric medians*, i.e., the points where the average distance to the other points in the cluster is minimal. The geometric median does not generally coincide with the centroid.

2. Each centroids is located within a cluster at that point where the average distance to the data points of that cluster is minimal (the geometric median).

By applying only the first stage of the procedure, the first property is retained but the second is lost. However, classical k-means also has a disadvantage that is more relevant in a social science context than in machine learning and marketing, where k-means is applied most frequently. Namely, the final clustering generated through the k-means procedure may be difficult to interpret with regard to the scientific question one wants to address through the clustering. This is indeed what was experienced when running standard k-means with the sample of Georgian rural households (with varying initial seeds). In this study, we chose to have clusters that are economically and socially interpretable rather than have centroids that correspond to the geometric medians. Thus, an alteration of the standard methodology was called for.

C. Dynamical Systems

The tables about structural change computed in the Chapters 3 to 5 can be interpreted as transition matrices of a stochastic dynamical system.[80] A dynamical system describes how a farm may change its "state" (or status, to use more conventional terminology) over time. In mathematical language, such a system may be presented as a stochastic function

$$f(t,s): N \times S \to S$$

which maps the time index t, which is a natural number, and the current state of the system, s, which is an element of the finite state space S, into the state space S. Applied to the dynamics of structural change in agriculture, the state space S has the following elements:

$s_1 =$ "outside agriculture",

$s_2 =$ "subsistence farmer",

$s_3 =$ "semi-subsistence farmer", and

$s_4 =$ "professional farmer"

The value of the function $f(t,s)$ stands for the state of a farm in the next period $t+1$ if the current state of that farm is s.

As an example, consider the scenario given by Table 4-3 in Chapter 4 (farms receive capital equal to twice their annual income and invest it at a return rate of 12%). At the start, in $t=1$, assume that a farm is in the subsistence state s_2. Then, the probabilities for the values of $f(1,s_2)$, where $f(1,s_2)$ is to be

[80] For being proper transition matrices, one row has to be added, as will be described below. For an introduction to dynamical systems of the kind used here, which are also called *Markov Chains*, see Florescu (2014), Chapter 12.

interpreted as the farm's state in $t = 2$, i.e., the state into which this farm moves, are given in the first row of Table 4-4:

	has left agriculture	is subsistence farmer	Is semi-subsistence farmer	is professional farmer
Probabilities for the state in $t = 2$	54.41%	20.59%	23.53%	1.47%

In $t = 2$, the farm may have become, for example, a semi-subsistence farmer. Then, the probabilities of $f(2, s_3)$ are given by the second row in Table 4-4, which is:

	has left agriculture	is subsistence farmer	Is semi-subsistence farmer	is professional farmer
Probabilities for the state at $t = 3$	42.27%	20.19%	27.44%	10.09%

This may go on ad nauseam, with the sequence of values of the function f being the trajectory the farm takes through the state space. The matrix in Table 4-4 captures the transition probabilities of the farm for each point of time and current state of the farm. In this dynamical system, the state $s_1 =$ "outside agriculture" is what is called a *trap* (cf. Brin and Stuck (2002), p. 25), i.e., a state that, once it is assumed, will not change anymore (because a farmer that has left agriculture will, by assumption, never return).

For dynamical systems, indices were developed that capture *how* dynamic the system is, i.e., whether the state of the function f changes more frequently, which means that the system is more dynamic, or rather rarely, which means that the system is rather static.

The most simple of these, the *trace index* (Shorrocks (1978)), is defined as

$$m_T = \frac{K - trace(P)}{K - 1}$$

where K is the number of elements in the state space, which in our case equals four (the number of categories in which a farm can be), and $trace(P)$ is the sum of the probabilities for staying in the same category in which the farm is (it corresponds to the *trace* of the transition matrix, i.e., the sum of diagonal elements).

If there is zero mobility between the categories, we have $m_T = 0$, because all diagonal elements are ones (no farm leaves the category in which it is initially), and hence $trace(P) = 4$. Maximal mobility, on the other hand, is achieved when the probability to enter any of the available states is the same. In this case, $m_T = 1$, because the probability to remain in a given category would be 0.25 and $trace(P) = 1$. A higher value of the trace index therefore implies a faster dynamic process.

BIBLIOGRAPHY

Abele, Klaus and Frohberg, Steffen (2003): Introduction, in: Abele and Frohberg (editors), Subsistence Agriculture in Central and Eastern Europe: How to Break the Vicious Circle?, *Studies on the Agricultural and Food Sectorin Central and Eastern Europe*, IAMO.

Abumere, Sylvester I. (1981): Population Distribution Policies and Measures in Africa South of the Sahara: A Review, *Population and Development Review* 7, pp. 421-433

Alvarez Albelo, Carmen D. (1999): Complementarity between physical and human capital, and speed of convergence, *Economics Letters* 64, 357-361

Baltensperger, Bradley H. (1987): Farm Consolidation in the Northern and Central States of the Great Plains, *Great Plains Quarterly* 7, pp. 256-265

Biermann, Florian, Nino Doghonadze, Robizon Khubulashvili, Lasha Labadze and Eric Livny (2014): Vocational training for job seekers in Georgia: Technical assistance for the Ministry of Labor, Health, and Social Affairs, ISET-PI/World Bank

Biermann, Florian, Lasha Labadze and Giorgi Mekerishvili (2013): Trade Relations of Georgia: a report for the Business Association of Georgia, ISET Policy Institute

Biermann, Florian and Giorgi Mzhavanadze: Don't Talk about Georgia's Future!, *The ISET Economist Blog*

Bluashvili, Aleqsandre, and Florian Biermann (2013): Rural Unemployment through Productivity Gains, *The ISET Economist Blog*

Brin, Michael and Garrett Stuck (2003): Introduction to Dynamical Systems, Cambridge University Press

Bottou, Leon and Yoshua Bengio (1995): Convergence Properties of the K-means Algorithms, in: Tesauro and Touretzky, editors, *Advances in Neural Information Processing Systems* 7, pp. 585-592. MIT Press

Coase, Ronald H. (1960): The Problem of Social Cost, *Journal of Law and Economics* 3, pp. 1–44

Davidova, Sophia, Alastair Bailey, Janet Dwyer, Emil Erjavec, Matthew Gorton and Kenneth Thomson (2013): Semi-Subsistence Farming – Value and Directions of Development, *Study for the European Parliament Policy Department B: Structural and Cohesion Policies*

Erickson, Kenneth W., Charles B. Mossa and Ashok K. Mishra (2004): Rates of Return in the Farm and Nonfarm Sectors: How Do They Compare?, *Journal of Agricultural and Applied Economics* 36, pp. 789-795

Eyhorn, Frank (2005): Soil Fertility Training Manual, Forschungsinstitut für Biologischen Landbau

Florescu, Ionut (2014): Probability and Stochastic Processes, John Wiley and Sons

Galor, Oded, and Omer Moav (2004): From Physical to Human Capital Accumulation: Inequality and the Process of Development, *Review of Economic Studies* 71, pp. 1001-1026.

Garibaldi, Pietro (2006): Personnel Economics in Imperfect Labor Markets, Oxford University Press

Hornidge, Anna-Katharina, Anastasiya Shtaltovna and Conrad Schetter (2016) (editors): Agricultural Knowledge and Knowledge Systems in Post-Soviet Societies, Peter Lang Publishers

Jones, Gwyn E. and Chris Garforth (1997): The history, development, and future of agricultural extension, in: Swanson, Burton E. (editor), Improving Agricultural Extension: A Reference Manual, 3rd edition, FAO

Kelbakiani, Giorgi, and Eric Livny (2015): If You Are So Smart, Why Are You Stuck in Kutaisi?, *The ISET Economist Blog*

Keynes, John M. (1921): A Treatise on Probability, Macmillan

Kruger, Justin and David Dunning (1999): Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments, *Journal of Personality and Social Psychology* 77, pp. 1121–34

Labadze, Lasha and Eric Livny (2012): Agricultural Productivity in Georgia and Armenia, a Sequel, *The ISET Economist Blog*

McAfee, Andrew and Erik Brynjolfsson (2012): Race Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy, Digital Frontier Press

Nuppenau, Ernst-August (2009): Economies of scale, energy use and enterprise size in agriculture: modeling of policies to reduce carbon dioxide and greenhouse gas emissions, *Agricultural Sciences (Žemės ūkio mokslai)* 16, pp. 203-210

Parsons, Robert L. and Gregory D. Hanson, Wesley N. Musser, Roland Freund, and Lehan Power: A Financial Training Program for USDA/FSA Borrowers: Evolution and Impacts, *Agricultural and Resource Economics Review* 29, pp. 240-250

Pellillo, Adam, Irakli Kochlamazashvili and Nino Kakulia (2014): Agriculture and Rural Development in Western Georgia: A Baseline Assessment, ISET Policy Institute

Prevost, Philippe (1996): Environment, complexity and professional training in agriculture 'turning local learning into global knowledge', *European Journal of Agricultural Education and Extension* 2, pp. 25-33

Rodi, Michael and Hope Ashiabor (2012): Legal Authority to Enact Environmental Taxes, in: Handbook of Research on Environmental Taxation, by Janet E. Milne and Mikael Skou Andersen (editors), Edward Elgar Publishing

Sarstedt, Marko and Erik Mooi (2011): A Concise Guide to Market Research, Springer 2011

Sharif, Mohammed (1986): The concept and measurement of subsistence: a survey of the literature, *World Development* 14, 555–577

Shorrocks, Anthony F. (1978): The Measurement of Mobility, *Econometrica* 46, pp. 1013-1024

Steger, Thomas M. (2000): Economic growth with subsistence consumption, *Journal of Development Economics* 62, 343–361

Thompson, Steven K. (2012): Sampling, John Wiley & Sons

Udry, Christopher and and Santosh Anagol (2006): The Return to Capital in Ghana, *American Economic Review* 96, pp. 388-393

von Luxburg, Ulrike (2010): Clustering Stability: An Overview, *Foundations and Trends in Machine Learning* 2, pp. 235-274

Wooldridge, Jeffrey M. (2008): Introductory Econometrics: A Modern Approach, 4th edition, South Western College Publishers.