From Bioeconomic Farm Models to Multi-Agent Systems: Challenges for Parameterization and Validation

Thomas Berger*

Josef G. Knoll Visiting Professorship for Development Research

University of Hohenheim, Germany (490e@uni-hohenheim.de)

Pepijn Schreinemachers

Josef G. Knoll Visiting Professorship for Development Research

University of Hohenheim, Germany (pepijn@uni-hohenheim.de)

Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference, Gold Coast, Australia, August 12-18, 2006

---

* Corresponding author

From Bioeconomic Farm Models to Multi-Agent Systems: Challenges for Parameterization and Validation

Abstract

Bioeconomic farm models have been very instrumental in capturing the technical aspects of human-nature interactions and in highlighting the economic consequences of resource use changes. They may elucidate the tradeoffs that farm households face in crop choice and farming practices, assess the profitability of various land-use options and capture the internal costs of adjusting to changes in environmental and market conditions. But they face also limitations when it comes to analyzing situations, in which heterogeneity of households and landscapes is large and increasing.

Multi-agent models building on the bioeconomic farm approach hold the promise of capturing more fully the heterogeneity and interactions of farm households. The fulfillment of this promise, however, depends on the empirical parameterization and validation of multi-agent models. Although multi-agent models have been widely applied in experimental and hypothetical settings, only few studies have tried to build empirical multi-agent models and the literature on methods of parameterization and validation is therefore limited.

This paper suggests novel methods for the empirical parameterization and validation of multi-agent models that may comply with the high standards established in bioeconomic farm modeling. The biophysical measurements (here: soil properties) are extrapolated over the landscape using multiple regressions and a Digital Elevation Model. The socioeconomic surveys are used to estimate probability functions for key characteristics of human actors, which are then assigned to the model agents with Monte-Carlo techniques. This approach generates a landscape and agent populations that are robust and statistically consistent with empirical observations.

Keywords: land use modeling, spatial and social heterogeneity, Monte-Carlo, random samples

1. INTRODUCTION

In many developing countries—for example in Uganda, the case study reported on here—, farm households are trapped in a downward spiral of interacting biophysical and socioeconomic forces (Pender et al., 2004). Poor farm households are compelled to apply unsustainable farming practices, which erode their natural resource base, reduce crop yields, and in turn, promote poverty.

Multi-agent system models (MAS) have a large potential to improve the understanding of downward spiral situations, to learn about the uncertainties related with complex agro-ecological systems and to explore new policy options (Parker et al, 2003). To fulfill their potential, MAS need to be carefully parameterized and validated with empirical data. MAS have been widely applied in hypothetical and experimental settings; only few studies have tried to build empirically based multi-agent models, and the literature on methods of empirical parameterization is therefore limited (Berger and Parker, 2002).

This paper is organized as follows. Section 2 introduces multi-agent models as building on the well-established bioeconomic farm approach. We then present in section 3 and 4 a novel approach – based on spatial interpolation and Monte-Carlo techniques – to parameterize MAS with empirical data. Taking the example of ongoing research in Uganda, we show how landscapes and agent populations can be parameterized from field measurements and farm household survey data. The last two sections discuss the validation of our approach and conclude.

2. BIOECONOMIC MODELING AND MULTI-AGENT SYSTEMS

Bioeconomic farm models have been very instrumental in capturing the technical aspects of human-nature interactions and in highlighting the economic consequences of resource use changes (Kuyvenhoven et al., 1998; Barbier, 1998; Woelcke, 2003; to name a few studies). They may elucidate the tradeoffs that farm households face in crop choice and farming practices, assess the profitability of various land-use options and capture the internal costs of adjusting to changes in environmental and market conditions (see also the more recent work of Holden and Shiferaw, 2004; Holden et al., 2004; Deybe and Barbier, 2005). But they face also limitations when households and landscapes are heterogeneous. In general, this is the case when farm households differ considerably in terms of factor endowments and decision-making processes and when resources are exchanged locally or in networks (Berger et al., 2006).

Multi-Agent Systems

Heterogeneity and interactions clearly fall into the core competence of multi-agent models (Janssen, 2002; Parker et al., 2003). MAS applied to natural resource management are generally implemented with fourth generation, object-oriented programming languages. They consist of two components; a cellular model component that represents the landscape under study and an agent-based model component that represents socio-economic decision-makers and their interactions. According to the classification proposed by Berger and Parker (2002), most of these applications are abstract or experimental; only few studies have tried to build empirical MAS. In the field of agricultural economics, MAS are often implemented as farm-based mathematical programming models (Balmann 1997; Berger 2001; Happe 2004). Their specific characteristic is that every farm households is represented by one computational agent.

Challenges to building empirical MAS

When building empirical MAS four challenges arise to the modeler:

1. *Simulating real-word decision problems*: agent decision-making needs to capture the essential features of real-world complexity and trade-offs such as – in the case of downward spiral situations – investment versus consumption, subsistence versus market production, short-term profit versus ecological sustainability.

2. *Empirical parameterization*: the cellular and the agent-based component of the MAS need to represent a real-world situation of typically heterogeneous biophysical and socioeconomic conditions, for example, various soil types, crop and vegetation growth, land holdings, social networks and human actors.
3. *Model validation*: MAS need to be validated against empirical data. Simulation results should show a sufficient goodness of fit in the baseline scenario and resemble real-world development paths.
4. *Sensitivity testing*: empirical data have measurement errors and not all model parameters might be known. Simulation results therefore need to be checked for these uncertainties.

In Schreinemachers and Berger (2006) we elaborated on the first challenge. As in Balmann (1997), Berger (2001), and Happe (2004), agent decision-making is represented by whole-farm mixed-integer linear programming problems. To be able to assess the poverty dimension of the downward spiral in Uganda, much effort was invested into better capturing the consumption and production behavior of farm households. We used micro-economic statistical approaches to estimate savings, food expenditure, and labor production functions and specified yield response functions for the biophysical modeling components.

For reasons of space, we cannot present the model equations for the Ugandan case study and refer for more details to Schreinemachers (2006). Here, we discuss in more detail the challenges related to model parameterization and validation.

3. BIOPHYSICAL PARAMETERIZATION OF MAS

This and the next section describe a novel methodology that combines *predictive soil maps* to parameterize the biophysical component of MAS, and *Monte Carlo* techniques to parameterize the agent-based component of MAS. The methodology is described on the basis of a case study for two village communities in Southeastern Uganda.

Research area

The two village communities are Magada and Buyemba; both are located in the southern part of the Iganga District, which through redistricting has recently become the Mayuge District. Climatic conditions allow the cultivation of two sequential crops in a year. Main food crops are cassava, sweet potato and beans, the main cash crop is coffee, while maize and plantain are both sold and home consumed. Farm households predominantly rely on the hand hoe in their crop management; the use of external inputs such as fertilizers, pesticides and improved seeds, is rare. Soil fertility is generally low but varies across locations (Brunner 2004; Ruecker 2005).

From soil samples to continuous soil maps

In the MAS a landscape of grid cells represents the biophysical environment that farm households manage. The landscape is organized in spatial layers with each layer containing the information about a specific property, such as soil chemical and physical properties, village boundaries and the location of farmsteads and agricultural plots. Layers are composed of grid cells of 71 x 71 meters (0.5 ha), which is the smallest amount of land cultivated by a single farm household.

Empirical information about soil properties is obtained from soil samples. The challenge here is to create continuous soil maps by interpolating soil sample values. The Uganda study used predictive soil mapping based on stepwise multiple regressions of soil properties on terrain parameters and/or other soil properties. Terrain parameters were derived from a Digital Elevation Model (DEM) and included elevation, slope, upslope area, plan curvature, profile curvature, curvature, wetness index, streampower index, and aspect (Rhew et al. 2004).

Distribution of agents into the landscape

The next challenge is to populate the landscape with agents. **Figure 1** shows the different stages in generating the spatially located agents and farm plots for the village of Magada; the same procedure was applied to Buyemba. The left upper panel (**Figure 1A**) shows the sample points within the village boundary of Magada. The figure shows that sample farm households are not evenly distributed in the landscape but are clustered around the road network.

Two different areas according to population density are therefore first demarcated: areas alongside the road network are designated as of high population density, and all other areas are of low population density (**Figure 1B**). Because the sample was random, the geographical distribution of the sample households represents the distribution of the total population. In Magada, for instance, 84 percent of the sample households live in the high-density area, which accounts for 40 percent of the total village area. Of the remaining (non-sample) households, 84 percent is thus allocated in the high-density area and 16 percent in the low-density area. Standard routines for spatial random allocation—as offered by GIS software packages—were used for this purpose.

All allocated farmsteads were then converted into grid cells, as shown in **Figure 1C**. Finally, using the estimated sample distribution from the survey, agricultural plots were allocated to the agents. A random spatial allocation was not used at this stage, as this would have produced an unrealistically scattered pattern of farm plots. The allocation was therefore done manually based on available qualitative information (**Figure 1D**).

<< Figure 1 >>

4. SOCIOECONOMIC PARAMETERIZATION OF MAS

When generating an empirically based MAS, every computational agent must represent a single real-world farm household. To increase the quality of empirical data, random samples are typically preferred to population censuses or censuses of agriculture (Carletto 1999). In the Uganda study, data were collected for about 17 percent of the farm households. The challenge hence is to extrapolate the sample population to parameterize the remaining 83 percent of farm households.

Monte Carlo approach

The Monte Carlo approach applied here is based on empirical cumulative distribution functions. **Figure 2** illustrates such a function for the distribution of goats over farm households. The figure shows that 35 percent of the farm households in the sample have no goats; the following 8 percent has one goat, etc. This function can be used to randomly distribute goats over agents, as well as all other resources in an agent population. For this, a random integer between 0 and 100 is drawn for each agent and the number of goats is then read from the y-axis. Repeating this procedure many times recreates the depicted empirical distribution function.

<< Figure 2 >>

With this procedure all resources that agents are endowed with can be allocated. By varying the random seed number, the procedure yields a differently endowed agent population each time. Yet, each resource would than be allocated independently, excluding the event of possible correlations between different resources. However, actual resource endowments typically correlate, for example, larger households have more livestock and more land. To include these correlations in the agent populations, first the resource that most strongly correlates with all other resources is identified and used to divide the survey population into a number of clusters. Empirical cumulative distribution functions are then calculated for each cluster of sample observations.

Each agent is allocated quantities of up to 80 different resources in the Monte Carlo procedure. These resources include 68 different categories of household members (34 age groups of two sexes), 4 livestock types (goats, young rams, cows, and young bulls), area under coffee plantation, female head of household, liquidity, ratio of equity and debt capital, plus innovativeness. Agents are generated sequentially, that is, agent No.1 first draws 80 random numbers in 80 different cumulative distribution functions before agent No. 2 does the same.

The Monte Carlo approach outlined here works well if correlations among agent characteristics are not too tight as in Uganda. With strong correlation some fine-tuning might be necessary; by skewing the distribution functions towards otherwise under-represented combinations of agent characteristics the random assignment may then still yield statistically consistent agent populations.

6. VALIDATION TESTS

To validate our parameterization approach, a large number of agent populations was generated by applying different random seed values. Within the scope of this paper, the full range of validation tests cannot be shown. Instead, the test results are illustrated with a few examples and snap shots from the agent populations.

Validation at population level

At the population level, it is checked whether the averages in the agent population resemble those of the survey population. For this purpose, average resource allocations for hundred generated agent populations are calculated in **Table 1**. For all resources, the average resource endowments in the agent population fall within the confidence interval of the survey average and the difference between the two averages is generally small. The random agent generator hence reproduces population averages.

<< Table 1>>

Validation at cluster level

The above has shown that the sample population is well replicated at the aggregate population level, but this might not necessarily be so at lower levels of aggregation. The following graphs and figures thus look at the cluster and agent level.

Figure 3 depicts four boxplots comparing the distribution of household size, arable land, cows and goats in the sample with an agent population with seed value 577. Each box ranges from the 25th to the 75th percentile (the inter-quartile range) with the 50th percentile, or median, also marked in it. Clusters are based on household size, which is why there is a strong correlation between these two variables in the left upper pane. The figure shows that median values do not differ much between the survey population and the agent population. In addition, most inter-quartile ranges are of comparable width, except for household size, but that is because this variable was used to define the clusters.

<< Figure 3 >>

Validation at agent level

In the Ugandan case study, assigning the spatial location of farmsteads and farm plots is not part of the Monte Carlo procedure. Because of the lack of geo-referenced data, we could not meaningfully redraw and statistically cross-check the agent location maps. The land endowment is therefore constant for each agent in each subsequent agent population. Only the non-land resources are randomly allocated to the agents based on cluster-specific distribution functions. **Figure 4** plots household size against amount of arable land per agent.

<< Figure 4 >>

One objective for generating agents randomly is to endow each agent differently in different agent populations which then allows for sensitivity analysis. **Figure 5** illustrates our success in this light; this figure is a boxplot showing the variation in resource endowments for agent No. 250 in hundred generations of different populations. Agent No. 250 has a fixed location for farmstead and plots as can be seen from the zero variance the agent's land area of 5 ha. The agent is randomly assigned to alternative clusters, though mostly it is assigned to cluster numbers 0, 1, 2 or 3, because these clusters have most agents with 5 ha of land, that is, these clusters have the highest probability that an agent with 5 ha of land is assigned to them.

<< Figure 5 >>

The reproduction of correlations is the third objective in the random agent generation. The left diagram in **Figure 6** plots the number of adults against the number of children in the survey population, while the two right panes do the same for two generated agent populations. The figures show that correlation between adults and children within the household, as observed in the survey, is well replicated in the agent populations, ensuring that the agents created are demographically consistent in this respect.

<< Figure 6 >>

Sensitivity testing

At the moment of writing this paper, sensitivity testing for the Uganda case study has not been completed. The Monte Carlo approach generates many possible and statistically consistent agent populations which may then be used for repetitions of simulation experiments. **Figure 7** shows the variation of simulation outcomes for 50 different agent populations in the baseline scenario. Variation is here measured by standard deviations expressed as percentages of the normalized mean. As can be seen, variation for most key policy indicators is low in the order of 5%; only in the case of farm assets including savings variation is in the order of 10%. This "inherent" model noise has to be considered when comparing various simulation experiments, for example on policy interventions in markets for credit and fertilizers. Again, more details on model validation and sensitivity testing can be found in Schreinemachers (2006).

<< Figure 7 >>

7. CONCLUSION

This paper showed that empirical parameterization based on digital elevation models, multiple regressions and Monte-Carlo techniques may generate statistically consistent agent populations which may be submitted to extensive validation tests. The variation of key policy indicators in the baseline scenario suggests that the model is sufficiently robust for comparing alternative policy interventions.

LITERATURE CITED

Balmann, A. *1997. Farm-based modelling of regional structural change: A cellular automata approach. European Review of Agricultural Economics **24**, 85-108.*

Barbier, B. *1998. Induced innovation and land degradation: Results from a bioeconomic model of a village in West-Africa, Agricultural Economics, **19**:15-25.*

Berger, T. *2001. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. Agricultural Economics **25** (2/3), 245-260.*

Berger, T., and D. C. Parker. *2002. Examples of specific research – introduction. In: D. C. Parker, T. Berger, and S.M. Manson, editors. Agent-based models of land use / land cover change. LUCC Report Series No. 6, Louvain-la-Neuve. [online] URL: http://www.indiana.edu/%7Eact/focus1/ABM_Report6.pdf*

Berger, T., P. Schreinemachers, and J. Woelcke. *2006. Multi-agent simulation for development of less-favored areas. In: H. Van Keulen, editor. Development Strategies for Less-Favored Areas. Agricultural Systems, Special Issue (in print).*

Brunner, A. *2004. Soil erosion in Uganda - From modelling to adapted slope use. ZEF News **16**, October 2004. [online] URL: http://www.zef.de/fileadmin/webfiles/downloads/zefnews/no16-04-2004-engl.pdf*

Carletto, C. 1999. *Constructing samples for characterising household food security and for monitoring and evaluating food security interventions: Theoretical concerns and practical guidelines. Technical Guide #8, International Food Policy Research Institute,Washington, D.C. [online] URL: http://www.ifpri.org/themes/mp18/techguid/tg08.pdf*

Deybe, D., and B. Barbier. *2005. Micro-macro linkages in bio-economic development policy, agricultural systems, special issue on less-favored areas. In H. van Keulen, editor. Development strategies for less-favored areas. Agricultural Systems, Special Issue (accepted for publication).*

Happe, K. *2004.Agricultural policies and farm structures - Agent-based modelling and application to EU-policy reform. Studies on the agricultural and food sector in central and eastern Europe, Vol. 30, IAMO. [online] URL: http://www.iamo.de/dok/sr_vol30.pdf*

Holden, S., and B. Shiferaw. *2004. Land degradation, drought and food security in a less-favored area in the Ethiopian highlands: a bio-economic model with market imperfections. Agricultural Economics, **30**(1): 31:49.*

Holden, S., B. Shiferaw, and J. Pender. *2004. Non-farm income, household welfare, and sustainable land management in a less-favored area in the Ethiopian highlands. Food Policy, **19**(4), Special Issue: 369-392.*

Janssen, M. A., editor. *2002. Complexity and ecosystem management: The theory and practice of multi-agent systems. Cheltenham, U.K., and Northampton, Mass.: Edward Elgar Publishers.*

Kuyvenhoven, A., H. Moll, and R. Ruben. *1998. Integrating agricultural research and policy analysis: Analytical framework and policy applications for bio-economic modeling. Agricultural Systems **58**(3), 331-349.*

Parker, D.C., and T. Berger. *2002. Synthesis and discussion. In D. C. Parker, T. Berger and S. M. Manson, editors. Agent-based models of land use / land cover change. LUCC Report Series No. 6, Louvain-la-Neuve. [online] URL: http://www.indiana.edu/%7Eact/focus1/ABM_Report6.pdf*

Parker, D.C., S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman. *2003. Multi-agent system models for the simulation of land-use and land-cover change: A Review. Annals of the Association of American Geographers, Vol. **93**/2.*

Pender, J., P. Jagger, E. Nkonya, and D. Sserunkuuma. *2004. Development pathways and land management in Uganda: Causes and implications. World Development,* ***32****(5): 767-792.*

Rhew, H., S. Park and G. Ruecker. *2004. Predictive soil mapping at regional scale in Iganga District, Uganda. A final report. University of Seoul, South Korea.Roeling, N. 1996. Towards an interactive agricultural science. European Journal of Agricultural Education and Extension* ***2:*** *35-48.*

Ruecker, G. R. *2005. Spatial variability of soils on national and hillslope scale in Uganda. ZEF Ecology and Development Series No.* ***24****. [online] URL:* http://www.zef.de/fileadmin/webfiles/downloads/zefc_ecology_development/ecol_dev_24_abstract.pdf

Schreinemachers, P. *2006. The (ir)relevance of the crop yield gap concept to food security in developing countries with an application of multi-agent modeling to farming systems in Uganda. Doctoral thesis submitted to the Agricultural Faculty of Bonn University.*

Schreinemachers, P., and T. Berger. *2006. Land-Use Decisions in Developing Countries and their Representation in Multi-Agent Systems. Journal of Land Use Science (in print).*

Woelcke, J. *2003. Bio-economics of sustainable land management in Uganda. Development Economics and Policy* ***37****. Peter Lang, Frankfurt am Main.*

Figure 1: Spatial generation of agent population and agricultural plots from a sample of farm households

Note: A. Survey sampling points; B. Division in areas with high and low population density; C. Location of farmsteads and conversion into grids; D. Distribution of agricultural plots.

Figure 2: Empirical cumulative distribution of goats over all households in the sample

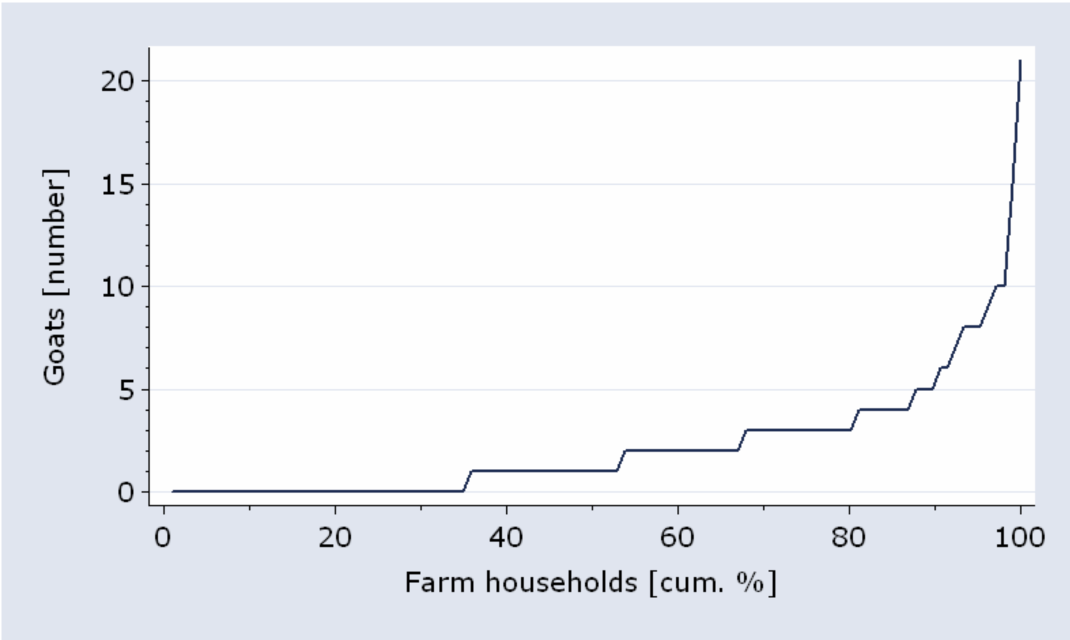


Figure 3: Boxplots for the distribution of the four major resources over clusters

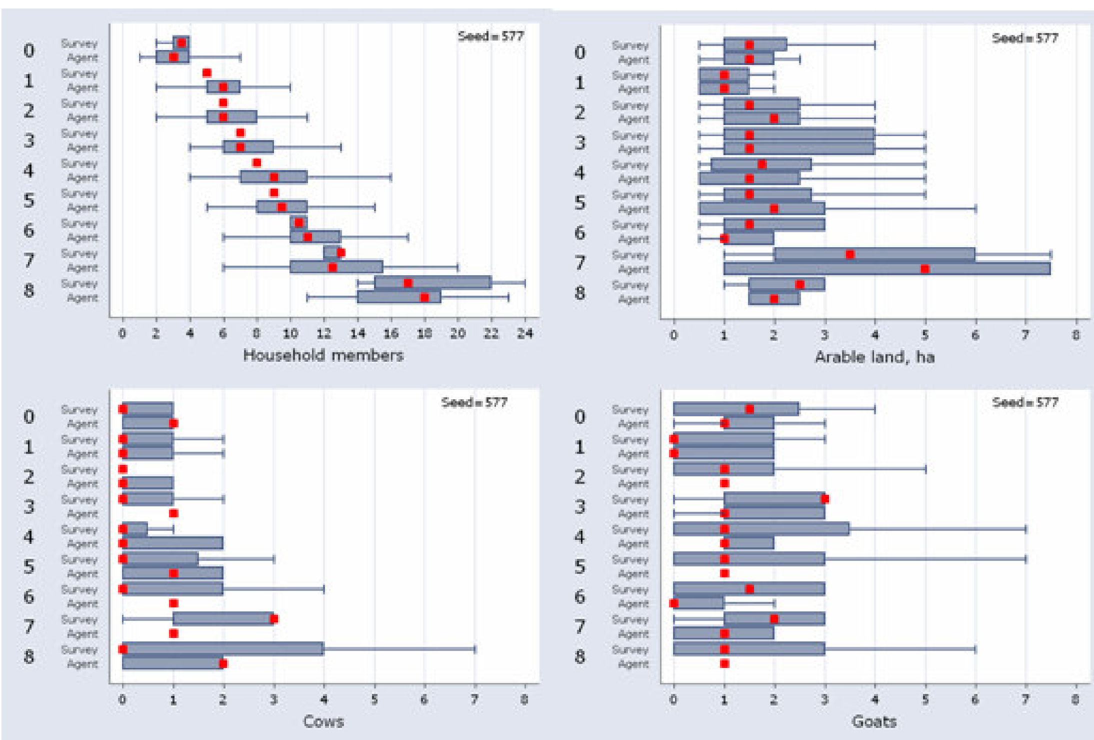


Figure 4: Correlation between household size and amount of arable land

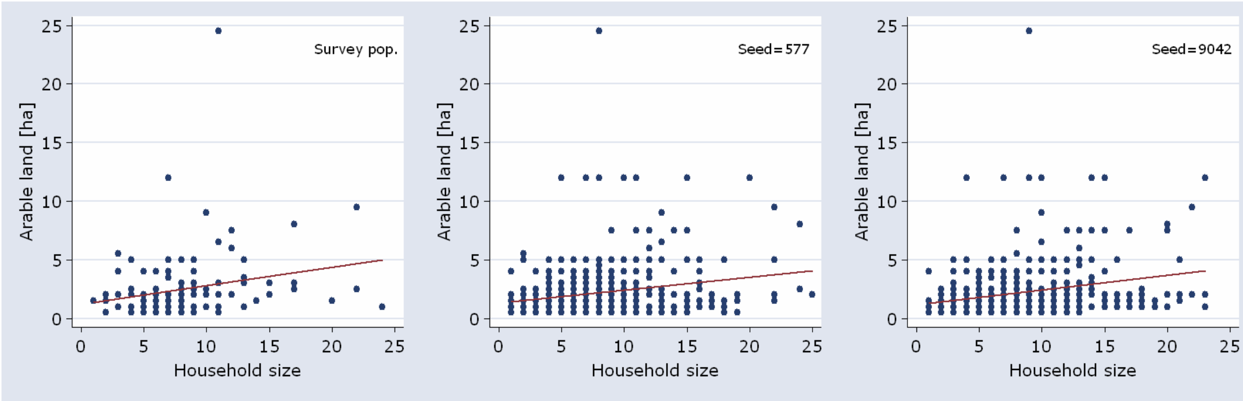


Figure 5: Boxplot illustrating the variation in agent endowments in alternative agent populations

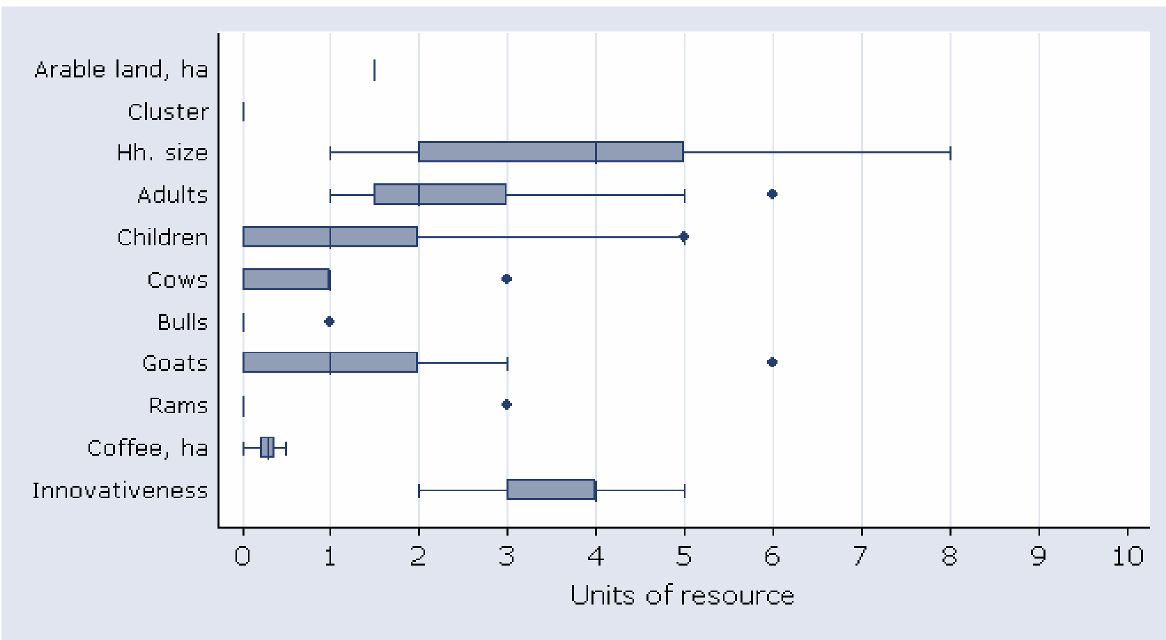


Figure 6: Scatter plots correlating the number of children and adults, with regression line

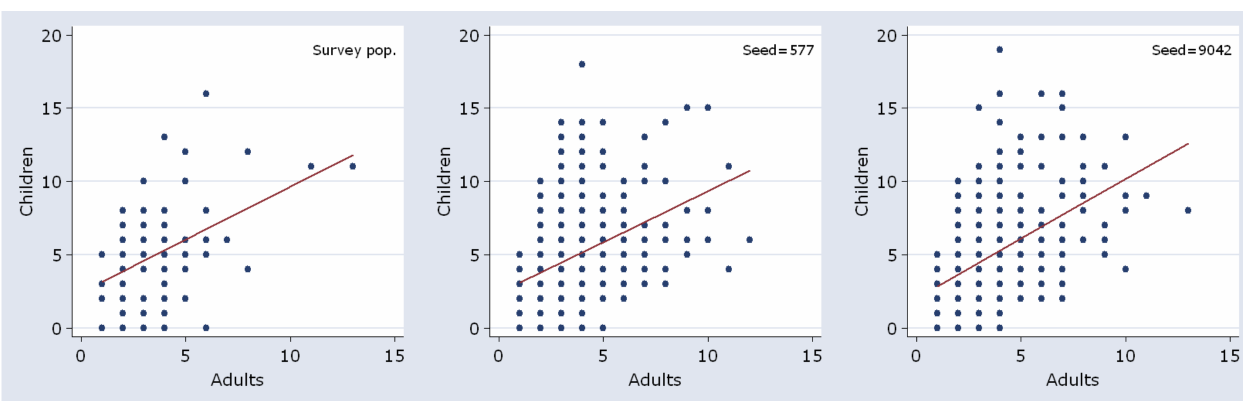
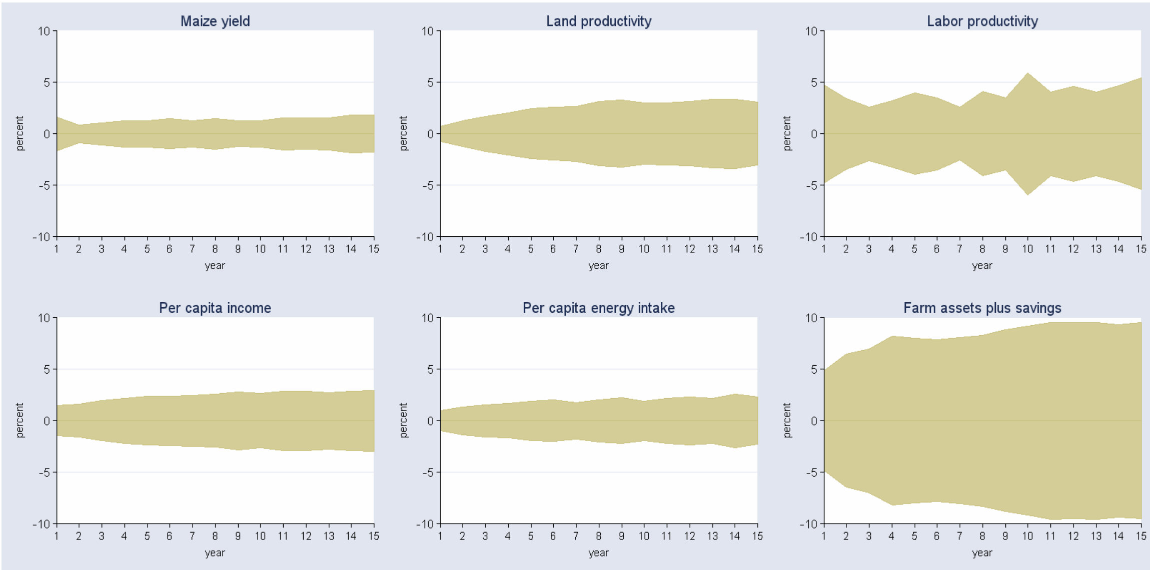


Figure 7: Variation of simulation outcomes for 50 different agent populations



Note: Variation is measured by standard deviations expressed as percentages of the normalized mean.

Table 1: Resource endowments of the survey population compared to meta-averages of the agent population

Resource	Population	Average	SE SD[1]	Confidence interval	
Household members	Survey	7.87	0.45	6.99	8.75
	Agent	7.89	0.11		
% children	Survey	55.06	2.47	50.22	59.91
	Agent	54.87	0.75		
Cows	Survey	0.81	0.18	0.45	1.17
	Agent	0.81	0.02		
Young bulls	Survey	0.08	0.04	0.01	0.16
	Agent	0.09	0.01		
Goats	Survey	1.29	0.16	0.98	1.61
	Agent	1.23	0.04		
Young rams	Survey	0.14	0.04	0.06	0.23
	Agent	0.14	0.02		
Coffee, ha	Survey	0.31	0.10	0.11	0.51
	Agent	0.31	0.02		
Plots, 0.5 ha	Survey	4.58	0.51	3.58	5.58
	Agent	4.34	0.00		
Innovativeness	Survey	3.88	0.17	2.35	3.03
	Agent	3.85	0.04		

Note: Agent population is average over 100 different agent populations.
[1] SE is Standard Error of the average referring to the average within the survey population; SD is Standard Deviation of the average referring to the average across agent populations.