Bio-based and Applied Economics 8(1): 75-99, 2019
DOI: 10.13128/bae-8147

Full Research Article

Estimating a Dual Value Function as a Meta-Model of a Detailed Dynamic Mathematical Programming Model

Claudia Seidel, Wolfgang Britz

Institute for Food and Resource Economics, University of Bonn, Germany

Abstract. Mathematical programming (MP) is a widespread approach to depict production and investment decisions of agents in agent-based models (ABM) related to agriculture. However, introducing dynamics and indivisibilities in MP models renders their solution computing time intensive. We present a meta-modeling approach as an alternative to directly integrating MP in an ABM. Specifically, we estimate a dual symmetric normalized quadratic (SNQ) value function from a set of MP solutions. The approach allows us to depict relationships between key attributes, like the farm endowment with (quasi-) fixed factors and discounted farm household incomes, without modeling the technology in detail. The estimated functions are integrated in the ABM to derive agents' decisions. The meta-modeling approach relaxes computational restrictions such that spatial interactions in large regions can be simulated improving our understanding of structural change in agriculture. It can also be used to extrapolate to farming populations where data availability might be restricted.

Keywords. Mathematical programming, Mixed Integer Program, meta-model, duality, symmetric normalized quadratic value function, agent-based modeling.

JEL codes. Q15, C61, C63.

1. Introduction

Agent-based modeling is a popular approach to simulate phenomena depending on spatial interactions between farmers (Berger 2001; Britz 2013a; Huber *et al.* 2018). It allows integrating behavioral rules that differ from standard micro-economic assumptions such as full information and full rationality (Bonabeau 2002; Nolan 2009). These features render agent-based models (ABMs) particularly suited to investigate the complex dynamic processes underlying structural change in agriculture (Zimmermann *et al.* 2009).

ABMs rapidly become complex and computing intensive if behavior of agents is modeled in detail (Zimmermann *et al.* 2009). This is especially true for ABMs that, first, use mathematical programming (MP) to derive agents' behavior in the ABM and that, second, explicitly model land markets as key driver of structural change in agriculture (Bal-

Corresponding author: wolfgang.britz@ilr.uni-bonn.de

ISSN 2280-6180 (print)
ISSN 2280-6172 (online)

mann 1999). In this type of ABMs, decision making of farmers regarding production and investment quantities or willingness to pay for (quasi-) fixed resources is typically (partly) derived from (discounted) profits or household incomes[1], as well as from marginal values of (quasi-) fixed resources simulated by MP (Schreinemachers and Berger 2011; Happe *et al.* 2006). Accordingly, a MP model has to be solved at least once for each agent and in each time step of the ABM. To represent realistically the decision space of agents, especially when considering investments, MP models require a large set of constraints as well as binary and integer variables (Mixed Integer Program, MIP). This can result in model set-ups with several ten thousand equations and variables, of which several hundreds are binary or integer variables (Britz *et al.* 2016). As a consequence, running an ABM on large farming populations is very resource demanding if each agent's behavior is derived from solving a large MIP.

MP solutions serve as inputs for different elements of ABMs focusing on agricultural structural change. Farm household income drives exit decision of farmers; and marginal returns to land (or other factors distributed by auctions) determine the bids of agents in simulated markets. As a consequence, income and marginal returns to land determine which farms grow, shrink or exit in the ABM and are, therefore, key drivers of dynamic processes in ABMs (Balmann 1999). In particular, simulation of land auctions is computationally challenging. Bids have to be calculated for each farmer and each plot of land that is available for rent at each time step, which can require solving a MP model for each combination of plot and farmer. Such ABM applications require efficient sampling schemes and sufficient computing power, especially if a whole agricultural region with many agents and a long time horizon should be investigated (Troost and Berger 2016). Even with increasing computational power such as using computing cluster and efficient MIP algorithms solving in parallel (e.g. Britz 2013b; Troost and Berger 2016) direct implementation of large MP models in an ABM results in high computing intensity.

The tension between computing needs and increased detail and coverage is a long-standing problem in the scientific and engineering simulation domain despite the tremendous increase in computer power and algorithmic progress. It persists since increased data availability, using sensitivity analysis and growing model sizes, by e.g. integrating more interactions between the agents, drive up computing needs. Indeed, the higher computing power itself invites researchers to increase model size and complexity to overcome shortcomings in previous set-ups. This can also be observed for agricultural ABMs using MP models (e.g. Arsenault *et al.* 2012; Brown *et al.* 2016; Kellermann *et al.* 2008; Lobianco and Esposti 2010; Polhill *et al.* 2007; Schreinemachers and Berger 2011; Zimmermann *et al.* 2015). They are now solved with far more agents, integrate different types of agents and/or different types of market interactions, or they use MP models which are harder to solve. Furthermore, large-scale sensitivity to address model uncertainty has become widespread. Thus, to keep computing time at bay, agricultural ABMs using MP models still face restrictions with regard to the number of agents and/or to the design of the MP

[1] In the following, we only refer to "income" for the sake of readability. Dependent on the model set-up, a MP model derives profit rather than household income if off-farm labor or other non-agricultural activities are not included. In the MP model we use, farmers also generate income from non-agricultural sources. The term "discounted" applies to dynamic settings where MP models are solved for several years. In this case, yearly profits or incomes are discounted.

approach. Reflecting that MIP problems are NP-hard to solve[2], MP models are set-up without dynamics, with no or a decreased number of binaries or integers, or with an overall reduced number of constraints and variables.

To overcome the need to sacrifice detail in a simulation model in favor of speed, meta-modeling strategies that require less computational power have been developed (Meckesheimer *et al.* 2002). They provide a simple mathematical approximation of the input/output relations in the underlying simulation model (Kleijnen 2018; Kleijnen and Sargent 2000; Meckesheimer *et al.* 2002; Pierreval 1996), drawing on statistical approaches such as (polynomial) regression models, splines or neural networks (Kleijnen and Sargent 2000). Their aim can be threefold: first, to improve the understanding of the behavior of the simulation model and the problem entity; second, to optimize the model with respect to the determination of the input set; and third, to make predictions of the model's simulation behavior (Bouzaher *et al.* 1993; Kleijnen 1979; Kleijnen 2005; Kleijnen and Sargent 2000). In the latter case, the meta-model is run instead of the simulation model itself, mainly to reduce computing needs (Kleijnen and Sargent 2000; Meckesheimer *et al.* 2001; Meckesheimer *et al.* 2002).

The objective of this paper is to develop a dual value function as a meta-model of a complex MP model motivated by the opportunity to reduce computational limitations of simulations with ABMs. The meta-model also allows to extrapolate to larger farming populations where the necessary detailed information on endowments such as structures and machines, as well as costs, labor and investment needs to set-up a MP model, is not available for each agent. Furthermore, setting-up and calibrating MP models to yield realistic solution behavior is a time consuming process which can be only partially automatized (see Troost and Berger 2016). The two aims of reducing computing needs and covering a larger population are therefore interrelated.

Our approach integrates a meta-model of a MP model in an ABM to compute optimal production and investment quantities, related discounted income and marginal values of (quasi-) fixed production factors. We, first, solve a suitable number of farm optimization problems with the MP model to obtain a farming population with individual production plans. We, next, estimate a meta-model from the MP results and subsequently integrate the estimates in the ABM. Specifically, we estimate a dual symmetric normalized quadratic (SNQ) value function which simulates discounted farm household income, input and output (i.e. netput) quantities and marginal values. As in the underlying MP model, the meta-model reflects production and investment decisions under maximization of discounted income at given prices and endowments, providing results under full rationality. However, simulated values can also be used to depict agent behavior that deviates from this assumption. To give an example, instead of using the marginal values of land, the average discounted income per ha of land can be used to define the marginal willingness to pay for an additional plot of land.

The structure of the paper is as follows. First, we present the general methodology to develop a meta-model of a MP model to be integrated in an ABM (chapter 2). In chapter 3, we apply the proposed method to a specific setting in order to precisely describe

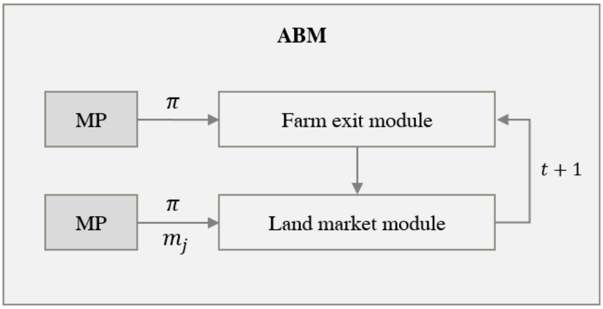
[2] "NP" stands for non-deterministic polynominal time algorithm. NP-hard means that so far, no algorithm has been found which could solve MIP problems in polynominal time. Clearly, the actual solution time depends on problem size and structure, the solver and hardware used.

our methodological approach. Chapters 4 and 5 present and discuss the estimation results before we briefly conclude in chapter 6.

2. General methodology

Many ABMs focusing on structural change in agriculture directly integrate a MP model in the model set-up (Fig 1)[3]. The MP model, solved for each farmer and at each time step, delivers income, input quantities bought and output quantities sold, as well as marginal values to (quasi-) fixed factors such as land and labor. Often, simulated incomes drive farm exit decisions in ABMs. In land market auctions, marginal returns to land can be used to determine the agents' willingness to pay for an additional plot. Aggregated quantities of inputs and outputs over the farming population might be used to define price feedbacks in the ABM, such that aggregated macro-level phenomena have an impact on agents' behavior on the micro-level (Chen and Liao 2005).

Figure 1. Classical set-up of an ABM integrating MP.



Note: π = (discounted) profit or income, m_j = marginal returns to (quasi-) fixed factors, t = yearly time steps of ABM.

Computing time restrictions for solving an instance of the MP model limit the complexity of the MP approach and/or the number of agents. Therefore, we develop a meta-model of the MP model which we integrate in the ABM. The meta-model delivers the same information as the MP model, but much faster by approximating the behavior of the MP model based on an estimated dual value function. The estimated dual value function mimics the simulation behavior of the MP model. Using the same inputs as the MP model, i.e. prices and endowments of each farm in the population, it generates outputs (income, netput quantities, marginal returns to (quasi-) fixed factors) that are very close to those simulated by the MP model. The MP model has to be solved only once for the whole farming population. Production quantities, income and marginal values of each

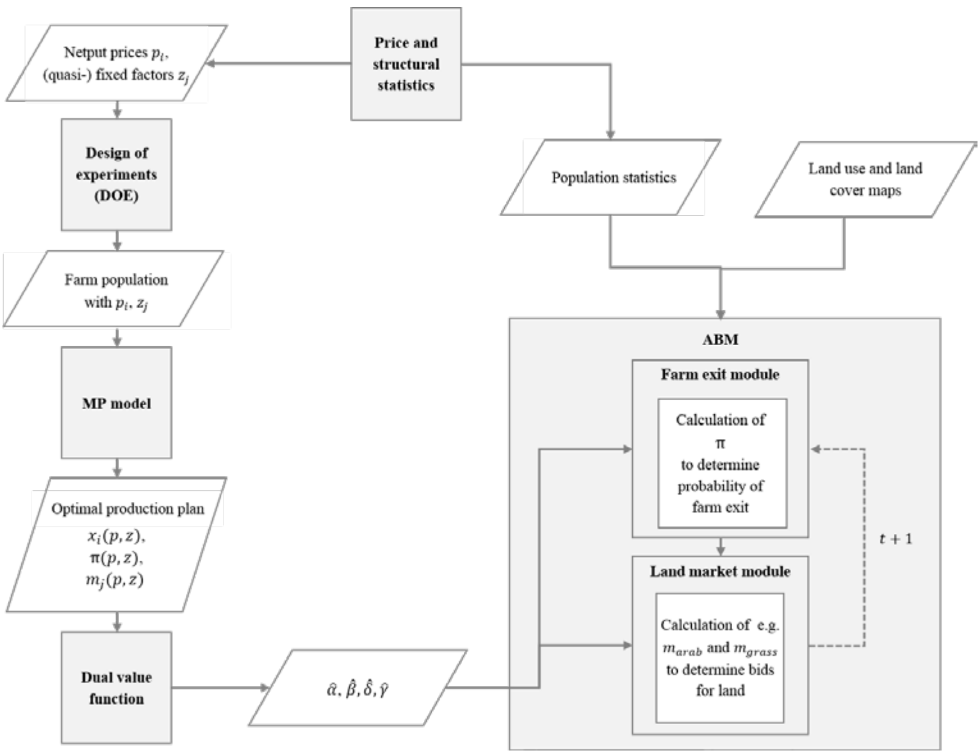
[3] In this and subsequent figures of ABMs in this paper, only the two modules of farm exit and land market are presented. Obviously, an ABM can include other and/or further modules to which our approach can also be applied if agent's decision making is based on MP.

farmer are updated in the ABM using the estimates of the dual value function. In the classical approach, the MP model has to be solved in each simulation period to derive production quantities, income and marginal values.

The advantage of estimating a dual value function over independent regressions of variables of interest is that the value function represents income maximizing behavior just like the MP. According to duality theory, the dual function depicts the optimal frontier, i.e. income maximizing netputs at given prices and limiting production factors. Thus, it maintains microeconomic consistency and indirectly comprises the information on the production feasibility set (Diewert 1971; Sidhu and Baanante 1981; Thijssen 1992). While in the classical approach, the technology of agricultural production is directly integrated in the ABM through the MP model; in the meta-modeling approach, it is represented in its dual form by the estimates of the value function.

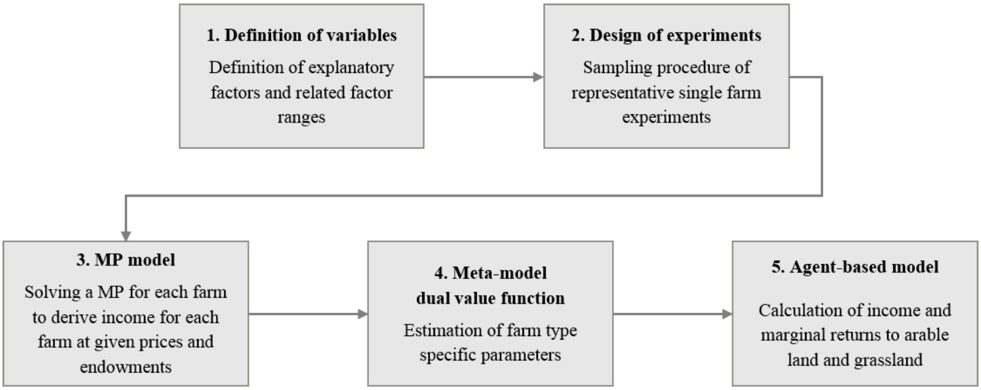
The overall modeling approach is depicted in Figure 2; the steps to take in the modeling approach are presented in Figure 3.

Figure 2. Meta-modeling approach.



Note: p_i = prices of inputs and outputs, z_j = farm's endowments with (quasi-) fixed factors, x_i = quantities of inputs bought and outputs sold, π = (discounted) profit or income, m_j = marginal returns to (quasi-) fixed factors, m_{arab} = marginal returns to arable land, m_{grass} = marginal returns to grassland, $\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\gamma}$ = estimated coefficients of value function, t = yearly time steps of ABM.

Figure 3. Steps in the meta-modeling approach.



First, the most important explanatory factors, which differentiate the farmers in the region to be investigated, and their factor ranges need to be defined based on price and structural statistics, e.g. the number of farms of different farm types in a region, the distribution of farm sizes in a region etc. Second, a suitable observation sample has to be defined using design of experiments (DOE) to cover the farming population. Third, in comparison to the classical approach (Figure 1), a MP model designed to capture important interactions at farm level is solved for each observation of the farming population once outside of the ABM before the start of the ABM-simulation. This yields the optimal production plan for each farmer, i.e. a dataset of optimal investment and production quantities, related farm household income and marginal returns to (quasi-) fixed factors. Fourth, these solutions are used to estimate a dual value function that becomes the meta-model of the MP model. Fifth, the function along with estimated coefficients are integrated into the ABM and used for calculating income and marginal returns to (quasi-) fixed factors for each agent at each time step.

3. Application

We apply the suggested meta-modeling approach to the MP model FARMDYN, the ABM ABMSim and the German region of North Rhine-Westphalia (NRW). NRW encompasses 3.4 million hectares of which 1.4 million hectares are agricultural land managed by 33,700 farms. The agricultural structure is dominated by livestock farming with 67 % of the farms holding cattle and/or pigs. NRW is characterized by high agricultural productivity and strong economic pressure on the land market with a rental share above 50% (IT.NRW 2019). Therefore, we assume that agents act in a (bounded) rational way, i.e. they optimize under limited information. Furthermore, we represent land rental markets in the ABM as auctions.

To investigate structural change in NRW in an ABM, the around 34,000 farms need to be depicted as agents. In an approach integrating a MP model in the ABM, the MP model must be solved at least once in any year of the simulation horizon for each farmer. In our setting, the MIP optimization problem for a single farm comprises roughly 20,000 equations with 30,000 variables including 3,000 binary or integer decision variables. Solving 34,000 of these MIPs takes several days. Even if the MP model would be less complex, the computing time would be still too high to allow investigations of structural change of the whole region over several years. With the meta-modeling approach, however, the several ten thousand optimization problems can be solved within a couple of minutes which allows us to solve an ABM over 10 years for the whole region of NRW within 15 minutes.

In the following, we present the meta-modeling approach along the five steps as depicted in Figure 3.

3.1 Generating the farm sample (Steps 1 and 2)

Observation samples for farms of different specializations (arable cropping, dairy, pig fattening, cattle fattening, mixed) are generated considering variations in (1) input and output prices, (2) endowment with (quasi-) fixed factors and (3), where appropriate, factors describing the technology such as the milk yield per cow. The factor ranges are chosen to capture the possible minimum and maximum values found in the farming population according to statistical data from the Association for Technology and Structures in Agriculture (KTBL) (KTBL 2016) and regional data of NRW (IT.NRW 2019).

Design of experiments (DOE) generates for each farm specialization a sample of farms that differ in initial conditions and other attributes. Initial conditions are, among others, available family labor, capital stock (stables, machinery and storage facilities), arable land and grassland owned by the farm. Other attributes are input and output prices that describe the farm's market environment as well as yield potentials and household expenditures (Britz *et al.* 2016).

To make the solution procedure more efficient, we make sure that only plausible combinations of factor ranges are generated. Unrealistic set-ups such as a farm with 250 cows, 10 ha and 0.25 labor units are likely to either lead to infeasibilities, i.e. to a loss of observations, or to unrealistically high or low marginal returns. Therefore, instead of drawing independent factor values from absolute factor ranges, we define for each farm branch one key attribute, e.g. total farm size in hectares for arable farms or dairy herd size for dairy farms. Factor ranges of further attributes are defined relative to the key attribute and from there mapped into absolute values. As an example, for the farm branch dairy, the farm's endowment with arable land and grassland is defined by its individual amount of hectares of arable and grassland per number of cows. The absolute amount of arable and grassland is then defined by multiplying sampled number of cows with sampled hectares of arable and grassland per number of cows.

In order to consider many factors and decrease computing time, we construct our sample based on Latin-hypercube sampling (LHS) as an efficient quasi-random sampling procedure. LHS is a space filling random sampling design that distributes the randomized factor level combinations smoothly over the range of factor level permutations (Iman and Conover 1980; McKay *et al.* 1979). Specifically, we apply the LHS package of R by Carnell (2016), assuming a uniform distribution over each considered factor.

After steps 1 and 2, we have a farming population with individual endowments with e.g. arable land, grassland and labor units, as well as prices of netputs that farmers face, reflecting the farming population and prices in the region under investigation.

3.2 Description of the MP model

The MP model FARMDYN (Britz *et al.* 2016), that we use for our application, simulates economic optimal production and investment decisions, assuming a fully informed, fully rational and profit maximizing farmer[4]. It considers farm profits including subsidies plus potential earnings from off-farm work, given constraints such as a detailed depiction of the production feasibility set of the farm, the maximum willingness to work on the farm or off-farm, liquidity or restrictions relating to the Common Agricultural Policy and German environmental laws. FARMDYN can be run in either comparative-static or dynamic mode with a finite planning horizon.

Decisions of investments and labor supply are modeled as integer variables to consider indivisibilities and to reflect returns to scale, for instance relating to stable sizes or labor needs for the management of farm branches. As an example, the MIP assumes that the farm can work at higher wages for 20 or 40 hours a week and/or to supply a low amount of off-farm labor at the legal minimum wage. Different farming systems can be simulated (arable, dairy, beef, pig fattening and biogas plants) and combined to depict diversified farms.

FARMDYN currently reflects German conditions drawing on technological and economic data from KTBL (Britz *et al.* 2016; KTBL 2016). Originally developed to derive marginal abatement cost functions in German dairy farming under differently detailed emission accounting schemes (Lengers *et al.* 2013), it was subsequently extended by a detailed description of pig farms (Garbert 2013), arable farming (Remble *et al.* 2013) and biogas plants (Schäfer 2014; Schäfer *et al.* 2017). FARMDYN is a bottom-up model. It is evaluated by means of its gross margins. Gross margins as simulation results of typical farms of a particular region (e.g. taken from structural data of North Rhine-Westphalia, IT.NRW 2019) are compared with data provided by KTBL (KTBL 2016).

FARMDYN is realized in GAMS and solved by the industry MIP solver CPLEX 12.6 (Britz *et al.* 2016), in our application on a 44 core computing server profiting from parallel processing in CPLEX. Furthermore, efficient solution strategies are implemented in FARMDYN such as parallel computing on multiple cores and the reduction of the solution space of the MIP by, first, solving a relaxed MIP (RMIP). A Graphical User Interface based on GGIG (GAMS Graphical Interface Generator, Britz 2014) allows to steer model runs and to exploit results (Britz *et al.* 2016). A detailed description of the model can be found in the FARMDYN model documentation (see Britz *et al.* 2016).

3.2.1 MP model run (Step 3)

As third step in our meta-modeling approach, we use FARMDYN to derive optimal farm household income, netput quantities and marginal returns to (quasi-) fixed factors for each farmer in the farming population sampled in the previous step.

We run FARMDYN in dynamic mode. In the dynamic set-up, optimal decisions are simultaneously determined at each point in time based on the current state of the system, reflecting the Principle of Optimality by Bellman (Bellman 1954). A value function dis-

[4] A dynamic-stochastic variant of the model is also available which can capture risk behavior based on different approaches to which the approach could also be applied.

counts the incomes that are simulated at each point in time (Bellman 1954). Thus, FARM-DYN delivers discounted farm household income at given netput prices and endowments. The corresponding average quantities of outputs sold, inputs bought, investment made and off-farm work supplied reflect yearly average activities of the optimal production plan over the planning horizon.

As marginal values of a MIP are conditioned on the current integer solution and do not consider that integers might change if a (quasi-) fixed factor increases, they might not reflect the actual shadow prices. That is why we derive the marginal returns to arable land and grassland from solving the model with increased endowments of arable land and grassland by one hectare and report the change in discounted income. Therefore, the marginal values of arable land and grassland consider how farming activities would change in the next ten years if an additional hectare of arable land or grassland could be used for agricultural production for ten years (the usual duration time of rental contracts in German agriculture is between eight and twelve years, Albersmeier *et al.* 2010), also including possible investments in a new stable if additional land becomes available.

In Step 3, we obtain optimal netput quantities, income and corresponding marginal returns to (quasi-) fixed factors for each farm of the farming population.

3.3 Determination of the meta-model (Step 4)

As a meta-model, we estimate a dual value function from the solutions of the MP model provided by step 3. Since the meta-model comprises the information about the technology of farms, a meta-model has to be estimated for each farm type by specialization separately. The possibility of farmers to switch from one agricultural production to another could be implemented by generating and estimating a sample of mixed farms that use various technologies.

3.3.1 Choosing the variables to be included in the meta-model

The inputs and outputs that are used as explanatory variables determine the farm household's costs and revenues from agricultural production and off-farm work. The endowments with (quasi-) fixed factors and prices of netputs are used as independent variables to explain the dependent variables discounted income, netput quantities as well as marginal returns to land. In opposite to estimating from real-world data, we control the data generation process by solving the MP model which allows us to also generate observations on marginal values.

As an example, Table 1 presents the lists of netputs simulated for dairy farms. The inputs include feed concentrates bought, variable costs of crops that are produced for feeding (such as maize silage, incl. fertilizer, plant protection products, electricity etc.), as well as investments made. Outputs are the amount of milk produced (other revenues such as from slaughtered cows or solved calves are reflected in the milk price), hours worked off-farm and exported manure. In regions with high livestock density, a farmer who exports manure makes a payment to an importing farmer. Therefore, exporting manure means a cost and is considered as a negatively valued output in the estimation of the value function (Kuhn *et al.* 2019; Schäfer and Britz 2017). The (quasi-) fixed factors characterize a farm household as an agent depicted in the ABM. For a dairy farm, we consider the number of hectares of arable land and grassland, the amount of labor available, and the construction year of the existing stable, as well as the milk yield per cow as an indicator of productivity. As described above, the farm endowment with land and labor is derived from the initial number of cows since this is the key attribute in the DOE for the farm branch dairy. We define a range of 40 to 150 milk cows per farm.

The construction year of the stable is included as (quasi-) fixed factor because a stable can only be used for 30 years. Once the stable reaches an age of 30, the farmer has to invest in a new stable in order to continue milk production. If the stable reaches the maximum age of 30 years in the optimization horizon of ten years, the farmer decides to reinvest in a stable or to quit milk production. This way, we can consider the age of a stable in the ABM and also include the possibility of farm exit due to a necessary large investment in a new stable. This is achieved by increasing the age of the stable with each time step of the ABM until a maximum age of 30 years and if a farmer invests in a new stable at a certain time step of the ABM, setting the age of the stable back to zero. Whether an agent has recently invested or will have to invest in a new stable within the next ten years is reflected in its discounted income. This way, it is possible to also include sunk costs related to a recent investment in a new stable and path dependencies which play a crucial role in agricultural production decisions (Huber *et al.* 2018). As the discounted income can be used to derive bids for land plots, the age of the current stable will have an influence on the willingness to pay for an additional plot of land.

Table 1. Variables included in the meta-model for the farm branch dairy.

Variable	Description	Factor range		
		Min	Max	Unit
Outputs		Netput price ranges		
Milk produced	Amount of milk produced [t]	310.00	360.00	€/t
Off-farm labor	Hours that a farmer works off-farm [h]	8.00	15.00	€/h
Manure exported	Amount of manure exported from the farm [m^3]	1.00	20.00	€/m^3
Inputs				
Feed concentrates	Sum of feed concentrates bought [kg]	0.80	1.20	€/kg
Crops	Sum of crops produced for feeding [kg]	0.80	1.20	€/kg
Investments	Sum of investments [1]	0.80	1.20	€
(Quasi-) fixed factors		Factor level ranges		
Arable land	Number of hectares of arable land	0.38	0.42	ha/cows
Grassland	Number of hectares of grassland	0.31	0.35	ha/cows
Milk yield	Milk yield per cow in 100 kg	80.00	86.00	*100 kg/cow
Labor units (lu)	Amount of farm labor available	28.00	38.00	cows/lu
Stable year	Construction year of the existing stable	1985	2010	

Note: Prices of feed, crops and investments are based on price indices composed of mean prices. Produced crops are solely fed to the animals and not sold on the market. Therefore, the crop price reflects the costs of crop production. Please also note that the export of manure means costs to the exporting farm. In the estimation of the value function, its price will, therefore, have a negative sign.

3.3.2 Definition of functional form

To approximate the behavior of the highly detailed MP model as good as possible, flexibility of the functional form of the value function is important. It must depict a multiple input, multiple output production function. Its derivatives will define input demand and output supply functions as well as marginal returns to limiting factors (Lopez 1982).

Inter alia, Diewert (1971, 1974), Christensen *et al.* (1973), Lau (1976) and Sidhu and Baanante (1981) propose flexible functional forms applicable to duality theory. To our knowledge, the choice of functional forms is quite limited if multiple inputs and multiple outputs are to be considered and convexity can be imposed to guarantee regularity, which is important for latter simulations. Based on the work of Diewert and Wales (1987, 1988) and Diewert and Ostensoe (1988), Kohli (1993) developed the symmetric normalized quadratic (SNQ) profit function (in our case value function as the MP delivers discounted income) that allows imposing global convexity, stays flexible and treats all inputs and outputs identically:

$$\pi\left(p,z\right)=\sum_{i=1}^{n}\alpha_i\ p_i+\frac{1}{2}\omega^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}p_ip_j+\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}p_iz_j+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{m}\gamma_{ijk}p_iz_jz_k \tag{I}$$

where π is the profit (in our case discounted income), $\alpha_i,\beta_{ij},\delta_{ij},\gamma_{ij}$ are the parameters to be estimated, p_i are the input and output prices, z_i are the (quasi-) fixed factors, $\omega=\sum_{i=1}^{n}\theta_i p_i$ is the price index for normalizing the prices and θ_i is the weights of prices for normalization (Henningsen 2014).

The estimation equations encompass the output supply and input demand equations x_i, derived by taking partial derivatives of the value function with respect to price p_i, and marginal returns m_j derived as partial derivatives towards the factor quantities z_j, according to the envelope theorem (Henningsen 2014; McKay *et al.* 1983).

$$x_i=\frac{\partial\pi(p,z)}{\partial p_i}=\ \alpha_i+\omega^{-1}\sum_{j=1}^{n}\beta_{ij}p_j-\frac{1}{2}\theta_i\omega^{-2}\sum_{j=1}^{n}\sum_{k=1}^{n}\beta_{jk}p_jp_k+\sum_{j=1}^{m}\delta_{ij}z_j+\frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m}\gamma_{ijk}z_jz_k \tag{II}$$

$$m_j=\frac{\partial\pi(p,z)}{\partial z_j}=\sum_{i=1}^{n}\delta_{ij}p_i+\sum_{i=1}^{n}\sum_{k=1}^{m}\gamma_{ijk}p_iz_k \tag{III}$$

As shown by the shadow price equation (equation III), marginal returns to fixed factors are determined by price effects and price-fixed factor effects. Accordingly, marginal returns to land vary not only due to different netput prices that the agents face but also due to joint effects of netput prices and endowments of farms with land, working units and other (quasi-) fixed factors. Therefore, the value function meta-modeling approach allows maintaining heterogeneity among farms of the same specialization with different farming structures and/or facing different prices. Only farmers with a similar farming structure and prices are assumed to derive the same optimal netput quantities and marginal returns to (quasi-) fixed factors. This observation points out that the value function is a dual representation of the technology. The value function does not fully depict the behavior of the agents in the ABM. Additional factors determining agents' decision making such as irrational or social behavior can be explicitly modeled in the ABM resulting in further heterogeneity among agents.

3.3.3 Estimation of the value function

Since we are particularly interested in the derivation of the marginal values to land, resp. the shadow prices of land, we estimate the netput equations (equation II) and shadow price equations (equation III) simultaneously, to inform the estimator on the marginal returns to land (McKay *et al.* 1983) that are also provided by FARMDYN. To our knowledge, that is a rather novel approach which reflects that other data sources such as farm samples used to estimate dual value functions are not providing observations on marginal values.

Corner solutions, resulting from in- or output quantities simulated as zero, frequently occur in our generated dataset and represent a particular challenge for the meta-modeling approach. For example, a farm may not supply off-farm labor as the returns to labor in the farm exceed the reserve wage; or the reserve wage becomes so high that the farm does not produce agricultural output and family members only work off-farm. If no off-farm labor is supplied, we can conclude that the internal return to labor is at or above the reserve wage, but we cannot assume that it is exactly at the reserve wage as required for a consistent estimation of the value function with a standard estimator. In real-world observed samples in which the data generation process is not controlled, such zero observations are potentially subject to self-selection bias such that two-stage procedures like limited information maximum likelihood (LIML) drawing on Heckman (1979) may become necessary. In our case, we can exclude all observations with any zero input or output from the estimation since we know that the underlying technology is identical for all farms by definition as defined by the structure and parameterization of the MP model. Still, corner solutions remain in the solution space of the MP model because of its integer variables. This is a particular challenge for the meta-modeling approach and discussed in chapter 4.

The SNQ value function is estimated as a seemingly unrelated regression (SUR) using the R package micEconSNQP (Henningsen 2014). Convexity on prices, which is an assumption of duality theory (Diewert 1973; Lau 1976; Lau 1986; Thijssen 1992), is imposed post-estimation where necessary based on Koebel *et al.* (2003). We slightly modified the R code to include equations for marginal returns to land.

At step 4, we obtain the estimates of the dual value function which represent the optimal production decisions of farmers originally simulated in the MP model. The estimated dual value function is now the meta-model of the MP model and can be integrated in the ABM.

3.4 Description of the ABM

ABMSim, the ABM we use for our modeling approach, was constructed to analyze structural change in farming in a spatial explicit setting. The landscape is generated using

CORINE (coordination of information on the environment) land cover data (European Topic Centre on Terrestrial Environment 2000) and differentiates between arable land, grassland, forest, housing, other urban fabrics, water bodies and other land types. The farming population is disaggregated in groups by specialization such as dairy, arable, pig fattening or mixed farming types. For each county in NRW, based on data from IT.NRW (2019), it generates the observed number of farms by specialization and size class (<5 ha, 5-10 ha, 20-50 ha, 50-100 ha, 100-200 ha, >200 ha). Initialization takes place by distributing the generated farms on available spots in the landscape, making sure that the farming structure at commune and county level is reflected (Schäfer *et al.* 2019). Once the farmsteads are allocated, the algorithm generates the agricultural plots with a random plot size from 1 pixel (= 1 ha) up to a chosen maximal plot size (Britz 2013a).

ABMSim consists of five modules in which the estimated coefficients of the meta-model are used to calculate incomes or marginal returns to (quasi-) fixed factors to depict decision making of agents in the ABM: land use change module, farm exit module, land market module, nutrient auction module (Schäfer *et al.* 2019) and milk delivery module. The modules of ABMSim are solved iteratively over distinct time steps of one year. In each year, economic drivers like exogenous prices or policies can be updated (Britz 2013a).

All markets included in ABMSim are represented as auctions and depict the interaction space of agents where they compete for e.g. land, manure disposal or milk delivery contracts. The discounted household income derived from the MP, resp. from the dual value function, can be used in the ABM to represent economic optimal production and investment decisions. In order to mimic real-world behavior of agents, a variety of behavioral rules can be applied. Bounded rational behavior of agents is included by e.g. the possibility to derive bids from observations in the agent's neighborhood or from an agent's average discounted income per ha. These behavioral rules can be applied only for a part of the agents. As a consequence, the population can differ in a way that some agents behave according to full economic rationality while others take bounded rational decisions (Britz 2013a).

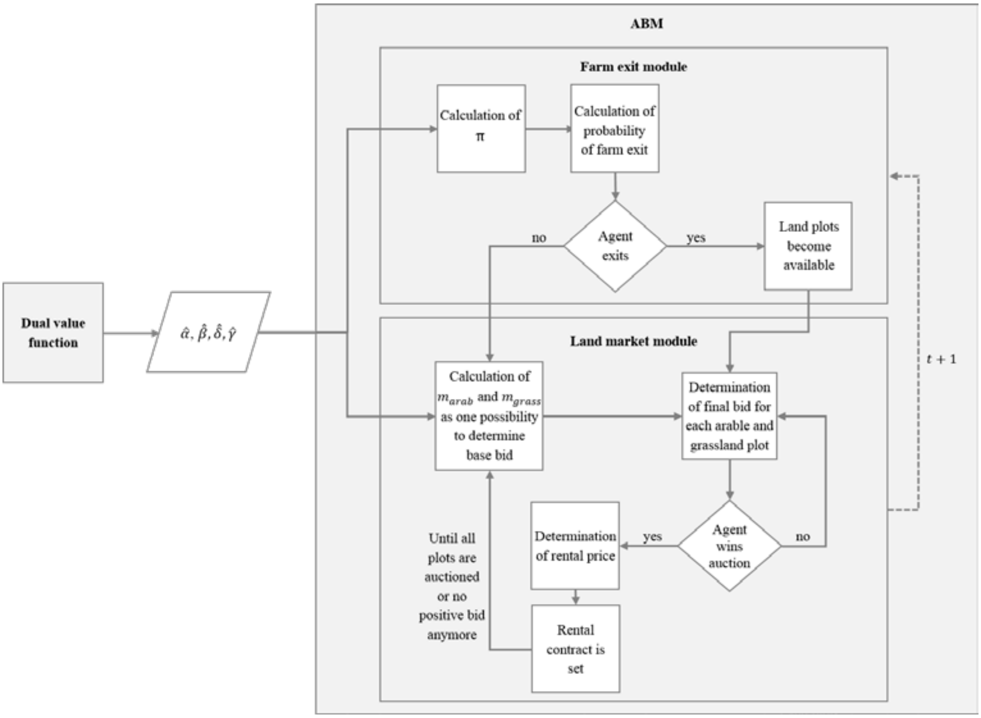
In the following, the farm exit and the land market modules of ABMSim will be briefly described in order to present the integration of the SNQ value function estimation in the ABM. These two modules are of particular importance for the simulation of structural change since they depict actions and interactions of agents which result in exit decisions and farm growth – the typical indicators of structural change in agriculture. A full description of ABMSim can be found in its model documentation (see Britz 2013a).

3.4.1 Integration of the meta-model in farm exit and land market modules (Step 5)

The estimated coefficients of the SNQ value function are used for the identification of agents that exit agricultural production based on discounted income calculations, and for the derivation of the bids of the agents based on the marginal returns to arable land and grassland. Figure 4 presents how the SNQ value function estimates are used in the two modules.

In the farm exit module, the probability of a farm exit in each period depends inter alia on each agent's current discounted income from farming (net of off-farm income). Combined with other information such as the agent's age and the probability to be employed outside agriculture, the calculated discounted income from farming (using

Figure 4. Connection of dual value function and ABM.



Note: Besides determining the base bid by calculating marginal returns to arable land (m_{arab}) and grassland (m_{grass}), it can also be defined based on average returns to land or average rents in neighborhood.

equation I) drives the probability of a farm exit[5]. If an agent exits, its current renting contracts end and the land owned (with the exemption of the farmstead) will be rented out. These plots are handed over to the land market module. Agents that do not exit agricultural production become potential bidders on plots in the land market.

The land market in ABMSim is a pure rental market and represented by a spatially explicit auction mechanism[6]. Agents who want to rent an additional plot of land put bids on the plots they are interested in. Free plots are plots where the rental contract ended or where the recent user exited the market. The agricultural plots are heterogeneous in location, size and type (arable land, grassland). The bidding behavior of agents is based on a base bid. One way to define it, is to use the marginal returns to land calculated from current prices, farm endowment and the estimated coefficients of the SNQ value function, as

[5] The derivation of the probability of farm exit can be found in the appendix.

[6] The auction mechanism is modeled as generic as possible in order to be applied to other market implementations, such as a market for milk delivery contracts or for manure disposal rights. A more detailed description of the auction algorithm can be found in the model documentation of ABMSim (Britz 2013a).

presented in formula III, and assuming full economic rationality. As the estimates of the value function are based on MP solutions of a ten year optimization, the calculated base bid includes information on the optimal production plan for the next ten years at current price expectations (in our case constant prices), also including potential large investments in the future, as presented in chapter 3.3.1. However, the base bide can also be defined according to the simulated discounted income per unit of land; or, as another possibility, an agent might use the average rent paid for rental contracts in the neighborhood. The last two options depict bounded rational behavior. The base bid is, first, reduced by transport costs to the plot depending on the distance to the farmstead and, second, increased by a markup for plots larger than one ha. The markup is used to reflect cost saving opportunities due to a large plot size. The resulting bid is restricted to not be larger than the base bid. As grassland and arable land are separate fixed inputs in the SNQ value function, agents place different bids on plots of arable land and grassland.

A rental contract of 10 years is set at a specific rental price between land owner and farmer winning the auction which depends on the chosen rules on auction order and price determination. After this land transaction, all bids for the remaining plots are recalculated for the winner of the auction because the willingness to pay for another plot of land has changed due to changed land endowment. The new marginal returns to land can easily be calculated by means of the SNQ shadow price equations (equation III) taking into account the increased land endowment. Due to the binding nature of rental contracts, bids may turn out to become unfortunate in the future because changes in prices or non-renewed rental contracts might change marginal returns to land and cause sub-optimal rental prices of current rental contracts.

4. Results

The dual value function is supposed to provide a good approximation for simulated netputs, discounted income and marginal returns to land to reduce the additional uncertainty introduced in the overall framework due to the replacement of the MP by a meta-model (Meckesheimer *et al.* 2002). Therefore, we focus on the fit of the meta-model in the result section, using a dataset of 1,002 dairy farms simulated by FARMDYN. These observations were kept from a sample of 5,000 farms after removing zero observations. We run the MP model as a dynamic programming model over the period from 2015 to 2025. Netput quantities refer to averages of 2015 to 2025. A descriptive summary of the simulated and estimated netput quantities, simulated incomes and marginal returns to land is presented in Table 2.

The values are discounted household incomes comprising not only returns from the farm operation, but also from working off-farm and from returns to accumulated cash. The model comprises optimal financing decisions based on different types of loans which differ in length and rates. The discount rate hence captures the time preference of a farmer and differs from the market based one. The farms simulated with FARMDYN are medium to large farms with a herd size between 40 and 150 cows and a land endowment of 32 to 110 ha in total, representing well the dairy farming structure in NRW (IT.NRW 2019).

Figure 5 presents scatterplots of the MP-simulated and fitted quantities from the SNQ value function of milk and off-farm labor supplied as outputs, and concentrates bought

Table 2. Key descriptives of simulated and estimated variables.

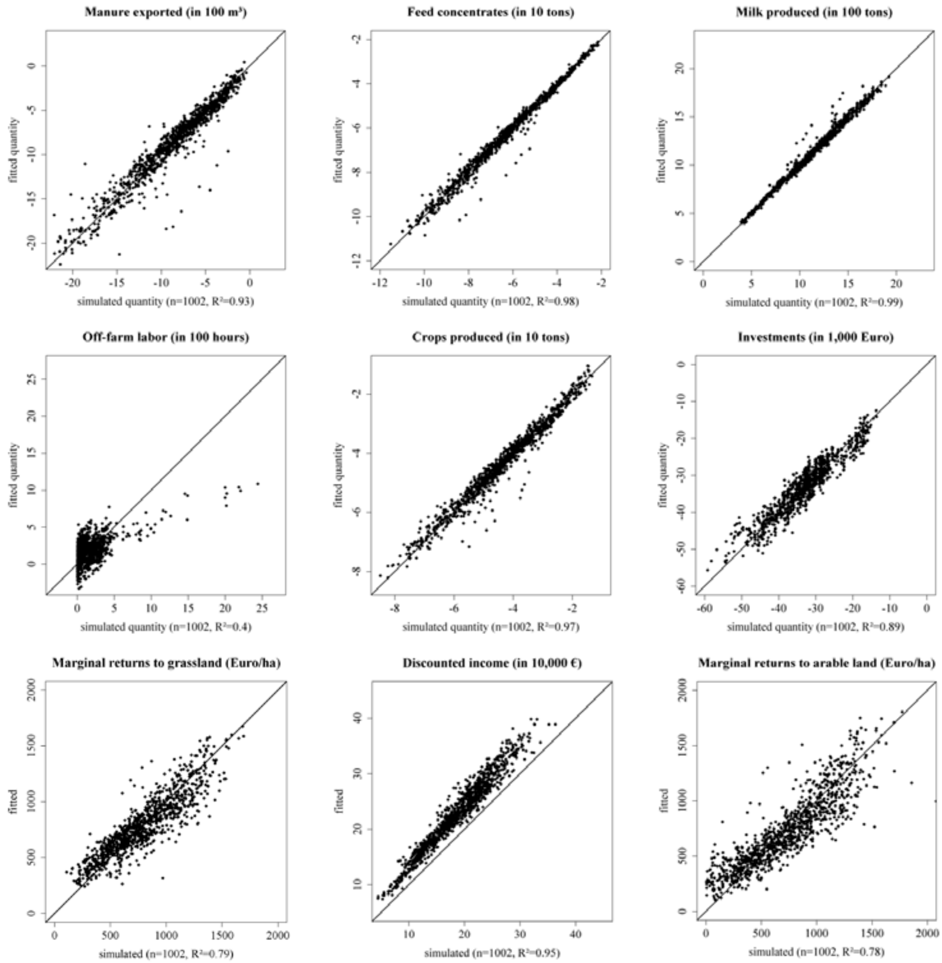
	Milk		Off-farm labor		Feed concentrates	
	MP-simulated	SNQ-fitted	MP-simulated	SNQ-fitted	MP-simulated	SNQ-fitted
Min	391.0	398.0	5.0	-299.6	-113,287.0	-96,854.0
Median	1,142.6	1,141,6	105.9	167.3	-62,849.0	-62,712.0
Max	1,922.3	1,913,3	2,442.0	793.5	-21,396.0	-19,769.0
R^2	0.99		0.35		0.95	
	Crops produced		Investments		Manure exported	
	MP-simulated	SNQ-fitted	MP-simulated	SNQ-fitted	MP-simulated	SNQ-fitted
Min	-84,903.0	-81,915.0	-59,192.0	-55,646.0	-2,593.3	-2,338.2
Median	-43,145.0	-42,838.0	-32,733.0	-32,996.0	-794.0	-785.7
Max	-13,354.0	-10,313.0	-13,587.0	-12,510.0	-40.6	40.0
R^2	0.97		0.89		0.93	
	Discounted income		Marginal returns to arable land		Marginal returns to grassland	
	MP-simulated	SNQ-fitted	MP-simulated	SNQ-fitted	MP-simulated	SNQ-fitted
Min	44,666.0	74,335.0	3.3	82.0	107.9	232.1
Median	189,989.0	233,354.0	716.8	673.6	771.6	754.2
Max	363,546.0	407,976.0	2,079.9	1,798.9	1,693.0	1,691.1
R^2	0.95		0.79		0.80	

Note: MP-simulated values represent the values that are provided by FARMDYN; SNQ-fitted values are values that are based on the estimation of the SNQ value function.

and crops produced for feeding, investments made and manure exported as inputs, as well as of the discounted incomes and marginal returns to land. The adjusted R^2 are very high (>93%) for the netputs milk, feed concentrates, crops and exported manure. The slightly lower R^2 for investments (89%) results from the assumption made in the MP that stables have to be bought in pre-determined sizes to reflect returns-to-scale. The integer character of variables make their estimation more difficult (see also discussion section in chapter 5). This can be especially seen in the moderate fit of the variable off-farm labor (40%). The dual value function with its continuous derivatives fails to fit the step-function that results from the integer character of the variable.

In opposite to that, the fit of average annual discounted income is with 95% very high, with a slight tendency to overestimate at high levels. The fit of the marginal returns to arable land and grassland is high (about 80%). The slightly lower fit compared to the estimated netput quantities is due to the complex interactions between the limiting production factors land and labor. The binary character of labor results in hard to predict changes in discounted incomes, if land endowment changes. These interactions are not fully captured by the shadow price equation derived from the SNQ value function.

Figure 5. Scatterplots of the dynamic dataset.



Note: Figures were created using R.

5. Discussion

To our knowledge, although a vast amount of literature can be found that investigates meta-modeling approaches for simulation models (e.g. Friedman and Pressman 1988; Jalal *et al.* 2013; Kleijnen 1979; Madu and Kuei 1994), inter alia simple Linear Programming (LP) models (e.g. Bailey *et al.* 1999; Johnson *et al.* 1996; Thangata *et al.* 2004), there is a lack of research that explicitly presents a meta-modeling approach of a complex MP model. Consequently, there is no evidence about the general performance of a linear meta-model of complex MIPs.

Our results suggest that a dual value function is able to provide, on average, a high fit for netput quantities and discounted income for the MP model FARMDYN analyzed in here. This might come as a surprise since MIPs provide corner solutions (due to the presence of integers) and are prone to overspecialization. The high fit found in here suggests that the large set of constraints of FARMDYN dampens that tendency and leads to a plausible and robust simulation behavior in the sense that changes in netput prices and factor endowments lead to a, on average, smooth response. This might imply that, if MP is used in an ABM to depict farming decisions, a certain degree of complexity is needed if a jumpy and hard to predict behavior has to be avoided. However, computing needs would be driven up – that is the starting point of using a meta-model instead.

Although replacing the integers by continuous variables would reduce computing time, a "normal" MP would eliminate returns-to-scale in investments and labor use which are now endogenously captured by the integers. We consider capturing returns-to-scale as important for the investigation of structural change. Note here that while the dual value function imposes convexity in netput prices, both convexity and concavity of income in fixed factors can be depicted.

As expected, the fit of the meta-model is lower if quantities are depicted by integer variables. Integers violate the assumed continuous relation between prices and netput quantities underlying the dual value function. However, this can also be considered as an advantage in some cases. The linear world of a MIP requires that investments come in pre-defined sizes if returns-to-scale are to be captured, whereas in reality, especially for building and structures, sizes can be rather flexibly chosen by the investing farmer. FARMDYN tries to overcome that problem partially by offering fine-grained stable sizes, but the basic problem remains. Compared to buildings and farming structure, for off-farm labor, the restrictive assumption is made in the MP that only 20 and 40 hour contracts are possible, besides a minimum wage job with only a few off-farm working hours a week. This explains the low fit of the netput off-farm work. In reality, however, family members might have some more flexibility to work part-time such that the smoothing effect of the meta-model might actually lead to a more realistic behavior. As such, the meta-model can also be understood as a way to interpolate over distinct points of the technology and the resulting solution space.

Furthermore, actions of agents in the presented ABM are derived from discounted income and marginal returns to (quasi-) fixed factors which are very accurately represented by the dual value function. Therefore, the more moderate fit for the variable off-farm labor should not invalidate the overall approach.

The main advantage of using a meta-model based on duality is that it provides a coherent framework to derive simultaneously netput quantities, (discounted) incomes and marginal returns. That is especially relevant if all these variables are needed in the ABM. If, for instance, only marginal returns to land are required, a simpler estimation approach not requiring a system estimation focusing on a high fit might be sufficient and more promising. Even if results for several variables are needed and a relatively high fit is obtained for all of them, the missing consistency might not be a concern. That would especially be true if the estimators are able to improve the fit for cases such as off-farm work where the dual approach cannot perform well by definition. Thus, approaches for instance from machine learning could be used instead of a theory consistent system estimation. As such, our results with the more restrictive dual approach define a kind of lower bound on the potential fit of a meta-model using more flexible fitting approaches.

Furthermore, in order to differentiate decision making of agents regarding planning horizons (e.g. milk delivery quantities in the current year as, to a certain extent, short-term decisions, medium-term decisions with regard to rental contracts, and long term decisions related to farm survival) more precisely, different value functions differentiated by planning horizons could be estimated and integrated in the model.. This way, differences in decision making of farmers could be depicted more precisely already at the estimation stage of the modeling approach. The modeling approach as a whole would yet become more complex, partly offsetting its advantages. As the paper showed, the dual value function is able to explain a complex MP model to a certain degree and is, therefore, suited to be implemented in an ABM to derive agent's decision making.

6. Conclusion

We present an approach to meta-model netput quantities, discounted farm-household incomes and respectively marginal returns to (quasi-) fixed factors from a highly detailed Mathematical Programming (MP) model with a dual symmetric normalized quadratic (SNQ) value function. The objective of the modeling approach is to set up a meta-model that represents the MP model in an agent-based model (ABM) to derive agent's decision making from it.

A set of parameters characterizes farm types by specialization, e.g. dairy farms, datasets are generated using MP for each and a dual SNQ value function is estimated. Based on duality theory, the estimation results shall be integrated in an ABM. This approach represents a less computing intensive and technically easier set-up of an ABM compared to the direct integration of a MP model into an ABM. This reflects that solving the MP model for each farm is computing time intensive and coding efforts to integrate the MP model into the ABM are higher compared to coding some few assignments necessary for the dual value function.

As presented in the paper, the estimation of a SNQ value function is able to fit the netput quantities, discounted incomes and marginal values of the MP model FARMDYN very well. Slightly lower fits can be explained by the integer character of some netputs which is more difficult to capture by the continuous character of the dual value function.

The value function meta-modeling approach maintains micro-economic consistency and derives mutually consistent simulation results at single farm-scale for discounted incomes, netputs and marginal returns to land. Even though the meta-model reflects micro-economic optimal behavior, the modeling approach still allows to introduce deviations from fully rational behavior. The reader shall be reminded that bidding behavior can be derived from both marginal and average costs which can be further modified in the ABM by behavioral rules. Furthermore, similar to MP, the dual value function provides a behavioral benchmark that can be compared to outcomes underlying alternative behavioral assumptions. By replacing the complex MP model, the dual value function relaxes computational restrictions while maintaining at the same time complex agent's behavior. It allows to easily depict large number of agents. Given that we obtained a high fit for most variables in our system estimation, more flexible approaches, e.g. from machine learning, might be interesting alternatives which could overcome inherent restrictions of a duality based approach such as continuous derivatives.

7. References

Albersmeier, F., Plumeyer, C.-H., Schulze, B. and Theuvsen, L. (2010). Zur Stabilität von Geschäftsbeziehungen auf dem Landpachtmarkt: Eine Kausalanalyse zur Wechselbereitschaft von Verpächtern. Presentation at the 50th conference of GEWISOLA, Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaus e.V., Braunschweig, Germany.

Arsenault, A., Nolan, J., Schoney, R. and Gilchrist, D. (2012). Outstanding in the Field: Evaluating Auction Markets for Farmland Using Multi-Agent Simulation. *Journal of Artificial Societies and Social Simulation* 15(1): 1–11.

Bailey, T.G., Jensen, P. and Morton, D.P. (1999). Response Surface Analysis of Two-Stage Stochastic Linear Programming With Recourse. *Naval Research Logistics* 46: 753–778.

Balmann, A. (1995). *Pfadabhängigkeiten in Agrarstrukturentwicklungen. Begriff, Ursachen und Konsequenzen*. Berlin, Germany: Volkswirtschaftliche Schriften.

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

Balmann, A. (1999). Path dependence and the structural development of family farm dominated regions. In: European Congress of Agricultural Economists, Organized Session Papers, Warsaw, Poland, 263–284.

Balmann, A., Odening, M., Weikard, H.-P. and Brandes, W. (1996). Path-dependence without increasing returns to scale and network externalities. *Journal of Economic Behavior and Organization* 29(1): 159–172.

Baumol, W. (1987). Indivisibilities. In: Eatwell, J., Milgate, M. and Newman, P. (eds.). The New Palgrave: A Dictionary of Economics, Vol. II. London/Basingstoke, UK: Macmillan.

Bellman, R. (1954). The Theory of Dynamic Programming. *Bulletin of the American Mathematical Society* 60: 503–515.

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): 145–260.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems'. *Proceedings of the National Academy of Science* 99(3): 7280–7287.

Bouzaher, A., Cabe, R., Carriquiry, A.L., Gassman, P.W., Lakshminarayan, P.G. and Shogren, J.F. (1993). Metamodels and nonpoint pollution policy in agriculture. *Water Resources Research* 29: 1579–1587.

Britz, W. (2013a). ABMSim – A flexible framework for Agent Based Models to simulate spatially explicit structural change in agriculture, Methodological and technical documentation, Version 1.0. University of Bonn, Germany. http://www.ilr.uni-bonn.de/em/rsrch/abmsim/abmsim_version1.pdf. Accessed 19 November 2018.

Britz, W. (2013b). The Graphical User Interface for CAPRI version 2010. University of Bonn, Germany. https://www.capri-model.org/docs/GUI2013.pdf. Accessed 19 November 2018.

Britz, W. (2014). A New Graphical User Interface Generator for Economic Models and its Comparison to Existing Approaches. *German Journal of Agricultural Economics* 63(4): 271-285.

Britz, W., Lengers, B., Kuhn, T. and Schäfer, D. (2016). A highly detailed template model for dynamic optimization of farms – FARMDYN, Model Documentation - Matching SVN version rev813. University of Bonn, Germany. http://www.ilr.uni-bonn.de/em/rsrch/farmdyn/farmdyn_docu.pdf. Accessed 19 November 2018.

Brown, C., Holzhauer, S., Metzger, M.J., Paterson, J.S. and Rounsevell, M. (2016). Land managers' behaviours modulate pathways to visions of future land systems. *Regional Environmental Change* 18(3): 831–845.

Cabral, L. (1995). Sunk Costs, Firm Size and Firm Growth. *The Journal of Industrial Economics* 43(2): 161–172.

Carnell, R. (2016). lhs – Latin Hypercube Samples, R-Package, Version 0.13. https://cran.r-project.org/web/packages/lhs/index.html. Accessed 19 November 2018.

Chen, S.-H., Liao, C.-C. (2005). Agent-based computational modeling of the stock price–volume relation. *Information Sciences* 170(1): 75–100.

Christensen, L.R., Jorgensen, D.W. and Lau, L.J. (1973). Transcendental Logarithmic Production Frontiers. *Review of Economics and Statistics* 55(1): 28–45.

Diewert, E.W. (1971). An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function. *Journal of Political Economy* 79(3): 481–507.

Diewert, E.W. (1973). Functional Forms for Profit and Transformation Functions. *Journal of Economic Theory* 6: 284–316.

Diewert, E.W. (1974). Applications of Duality Theory. In: Intriligator, M. D. and Kendrick, D. A. (eds.). Frontiers of Quantitative Economics, Volume II. New York, USA: North-Holland Pub. Co.

Diewert, W.E. and Ostensoe, L.A. (1988). Flexible Functional Forms for Profit Functions and Global Curvature Conditions. In: Barnett, W. A., Berndt, E. R. and White, H. (eds.). Dynamic Econometric Modeling. Cambridge, UK: Cambridge University Press.

Diewert, W.E. and Wales, T.J. (1987). Flexible Functional Forms and Global Curvature Conditions. *Econometrica* 55: 43–68.

Diewert, W.E. and Wales, T.J. (1988). Normalized Quadratic Systems of Consumer Demand Functions. *Journal of Business and Economic Statistics* 6: 303–312.

European Topic Centre on Terrestrial Environment (2000). CORINE land cover database (Version 12/2000). https://land.copernicus.eu/pan-european/corine-land-cover. Accessed 19 November 2018.

Friedman, L. W. and Pressman, I. (1988) The Metamodel in Simulation Analysis: Can It Be Trusted? *Journal of the Operational Research Society* 39(10): 939–948.

Garbert, J. (2013). Ökonomische Auswirkungen von Politiken zur Umsetzung der Wasserrahmenrichtlinie auf die Schweinehaltung im Münsterland. Dissertation, University of Bonn, Germany. http://hss.ulb.uni-bonn.de/2013/3389/3389.htm. Accessed 19 November 2018.

Happe, K., Kellermann, K., Balmann, A. (2006). Agent-Based Analysis of Agricultural Policies: an Illustration of the Agricultural Policy Simulator AgriPoliS, its Adaptation and Behavior. *Ecology and Society* 11(1): 49–74.

Heckman, J.T. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47(1): 153–161.

Henningsen, A. (2014). 'micEconSNQP: Symmetric Normalized Quadratic Profit Function. R-Package, Version 0.6–6. https://cran.r-project.org/web/packages/micEconSNQP/index.html. Accessed 19 November 2018.

Huber, R., Bakker, M., Balmann, A., Berger, T., Bithell, M., Brown, C., Grêt-Regamey, A., Xiong, H., Le, Q.B., Mack, G., Meyfroidt, P., Millington, J., Müller, B., Polhill, J.G., Sun, Z., Seidl, R., Troost, C., Finger, R. (2018). Representation of decision-making in European agricultural agent-based models. *Agricultural Systems* 167: 143–160.

Iman, R.L. and Conover, W.J. (1980). Small sample sensitivity analysis techniques for computer models, with an application to risk assessment (with discussion). *Communication in Statistics* 9: 1749–1842.

IT.NRW (Information und Technik Nordrhein-Westfalen) (2019). Landesdatenbank NRW. https://www.landesdatenbank.nrw.de/ldbnrw/online/data;jsessionid=2F3E5E0A8D51B5A8F42279DAD7B61782.ldb1?operation=sprachwechsel&language=en. Accessed 04 July 2019.

Jalal, H., Dowd, B. and Sainfort, F. (2013). Linear Regression Metamodeling as a Tool to Summarize and Present Simulation Model Results. *Medical Decision Making* 33(7): 880–890.

Johnson, K.E., Bauer, K.W., Moore, J.T. and Grant, M. (1996). Metamodelling Techniques in Multidimensional Optimality Analysis for Linear Programming. *Mathematical and Computer Modelling* 23(5): 45–60.

Kellermann, K., Sahrbacher, C. and Balmann, A. (2008). Land Markets in Agent Based Models of Structural Change. Paper prepared for presentation at the 107th EAAE Seminar "Modeling of Agricultural and Rural Development Policies", January 29th – February 1st, 2008, Sevilla, Spain.

Kleijnen, J.P.C. (1979). Regression Metamodels for Generalizing Simulation Results. *IEEE Transactions on Systems, Man, and Cybernatics* 9(2): 93–96.

Kleijnen, J.P.C. (2005). An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research* 164(2): 287–300.

Kleijnen, J.P.C. (2018). Design and Analysis of Simulation Experiments. In: Pilz J., Rasch D., Melas V., Moder K. (eds). Statistics and Simulation. IWS 2015. *Springer Proceedings in Mathematics & Statistics* 231: 3–22.

Kleijnen, J.P.C. and Sargent, R.G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research* 120(1): 14–29.

Koebel, B., Falk, M. and Laisney, F. (2003). Imposing and Testing Curvature Conditions on a Box–Cox Cost Function. *Journal of Business and Economic Statistics* 21(2): 319–335.

Kohli, U. (1993). A Symmetric Normalized Quadratic GNP Function and the U.S. Demand for Imports and Supply of Exports. *International Economic Review* 34(1): 243–255.

KTBL (Association for Technology and Structures in Agriculture) (2016). Betriebsplanung Landwirtschaft 2016/2017. Darmstadt, Germany: KTBL Datensammlung.

Kuhn, T., Schäfer, D., Holm-Müller, K. and Britz, W. (2019). On-farm compliance costs with the EU-Nitrates Directive: A modelling approach for specialized livestock production in northwest Germany. *Agricultural Systems* 173: 233–243.

Lau, L.J. (1974). Applications of Duality Theory: A Comment. In: Intriligator, M.D. and Kendrick, D. A. (eds.). Frontiers of Quantitative Economics, Volume II. New York, USA: North-Holland Pub. Co., 176–180.

Lau, L.J. (1976). A Characterization of the Normalized Restricted Profit Function. *Journal of Economic Theory* 12: 131–163.

Lau, L.J. (1986). Functional Forms in Econometric Model Building. In: Griliches, Z. and Intriligator, M. D. (eds.). Handbook of Econometrics, Volume III. New York, USA: North-Holland Pub. Co.

Lengers, B., Britz, W. and Holm-Müller, K. (2013). Comparison of GHG-emission indicators for dairy farms with respect to induced abatement costs, accuracy and feasibility. *Applied Economic Perspectives and Policy* 35(3): 451–475.

Lengers, B., Britz, W. and Holm-Müller, K. (2014). What Drives Marginal Abatement Costs of Greenhouse Gases on Dairy Farms? A Meta-modelling Approach. *Journal of Agricultural Economics* 65(3): 579–599.

Lobianco, A. and Esposti, R. (2010). The Regional Multi-Agent Simulator (RegMAS): An open-source spatially explicit model to assess the impact of agricultural policies. *Computers and Electronics in Agriculture* 72(1): 14–26.

Lopez, R.E. (1982). Applications of Duality Theory. *Western Journal of Agricultural Economics* 7(2): 353–365.

Madu, C.N. and Kuei, C.-H. (1994). Regression metamodeling in computer simulation – The state of the art. *Simulation Practice and Theory* 2(1): 27–41.

McKay, M.D., Conover, W.J. and Beckman, R.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from computer code. *Technometrics* 21(2): 239–245.

McKay, L., Lawrence, D. and Vlastuin, C. (1983). Profit, Output Supply, and Input Demand Functions for Multiproduct Firms: The Case of Australian Agriculture. *International Economic Review* 24(2): 323–339.

Meckesheimer, M., Barton, R.R., Simpson, T., Limayem, F. and Yannou, B. (2001). Metamodeling of Combined Discrete/Continuous Responses. *American Institute of Aeronautics and Astronautics* 39(10): 1950–1959.

Meckesheimer, M., Booker, A.J., Barton, R.R. and Simpson T.W. (2002). Computationally Inexpensive Metamodel Assessment Strategies. *American Institute of Aeronautics and Astronautics* 40(10): 2053–2060.

Nolan, P., Parker, D., van Kooten, G.C. and Berger, T. (2009). An Overview of Computational Modeling in Agricultural and Resource Economics. *Canadian Journal of Agricultural Economics* 57(4): 417–429.

Pierreval, H. (1996). A Metamodeling Approach Based on Neural Networks. *International Journal of Computer Simulation* 6(3): 365–378.

Polhill, J.G., Parker, D.C. and Gotts, N. (2008). Effects of land markets on competition between innovators and imitators in land use: results from FEARLUS-ELMM. In Edmonds, B., Hernandez, C., Troitzch, K. G. (eds), Social Simulation Technologies: Advances and New Discoveries, Information Science Reference, Hershey, USA: IGI Global, 81–97.

Remble, A., Britz, W., Keeney, R. (2013). Farm Level Tradeoffs in the Regulation of Greenhouse Gas Emissions. In: Agricultural and Applied Economics Association, Annual Meeting, August 4-6, 2013, Washington, D. C., USA.

Schäfer, D. (2014). Effects of the Renewable Energy Act Amendments Biogas Plant Investment and Farm Management. Master Thesis, University of Bonn, Germany.

Schäfer, D. and Britz, W. (2017). Estimating impacts of the revised German fertilizer ordinance on manure transport flows within North Rhine-Westphalia, Poster presented at the EAAE XV Congress, Parma (Italy).

Schäfer, D., Britz, W. and Kuhn, T. (2017). Flexible Load of Existing Biogas Plants: A Viable Option to Reduce Environmental Externalities and to Provide Demand-driven Electricity? *German Journal of Agricultural Economics* 66(2): 109–123.

Schäfer, D., Grauer, Z. and Britz, W. (2019). Regionale Modellierung von agrarstrukturellen Änderungen in einem agentenbasierten Ansatz unter Berücksichtigung des überbetrieblichen Austausches von organischem Dünger. Agricultural Faculty of the University of Bonn, Germany: Schriftenreihe des Lehr- und Forschungsschwerpunktes USL, 1–51, forthcoming.

Schreinemachers, P. and Berger, T. (2011). An agent-based simulation model of human-environment interactions in agricultural systems. *Environmental Modeling and Software* 26(7): 845–859.

Sidhu, S.S. and Baanante, C.A. (1981). Estimating Farm-Level Input Demand and Wheat Supply in the Indian Punjab Using a Translog Profit Function. *American Journal of Agricultural Economics* 63(2): 237–246.

Thangata, P.H., Alavalapati, J.R.R. and Hildebrand, P.E. (2004). Metamodeling Agroforestry Adoption. In Alavalapati, J. R. R., Mercer, D. E. (eds), Valuing Agroforestry Systems, Dordrecht, Netherlands: Springer, 219–236.

Thijssen, G. (1992). A comparison of production technology using primal and dual approaches: The case of Dutch dairy farms. *European Review of Agricultural Economics* 19(1): 49–65.

Troost, C. and Berger, T. (2016). Simulating structural change in agriculture: Modelling farming households and farm succession. International Environmental Modelling and Software Society (iEMSs), 8th International Congress on Environmental Modelling and Software, Toulouse, France. International Congress on Environmental Modelling and Software, 1.

Weaver, R.D. and Lass, D.A. (1989). Corner Solutions in Duality Models: A Cross-Section Analysis of Dairy Production Decisions. *American Journal of Agricultural Economics* 71(4): 1025–1040.

Zimmermann, A., Heckelei, T. and Domínguez, I.P. (2009). Modelling farm structural change for integrated ex-ante assessment: review of methods and determinants. *Environmental Science and Policy* 12: 601–618.

Zimmermann, A., Möhring, A., Mack, G., Ferjani, A. and Mann, S. (2015). Pathways to Truth: Comparing Different Upscaling Options for an Agent-Based Sector Model. *Journal of Artificial Societies and Social Simulation* 18(4): 1–11.

Appendix
I. Calculation of probability of farm exit

The probability of farm exit is calculated from two elements:

1. The square root of the relation between the farmer's current yearly profit π_f and the maximum of (1) a pre-determined quantile of the profits in the farming population $\pi_{pop}^{quantile}$, and (2) the expected yearly net wage in the industrial sector minus commuting costs,
2. A normally distributed random number which accounts for not controlled determinants of exit decisions.

By settings parameters to zero, the effect of the different elements can be switched off.

Algebraically, the probability can be expressed as:

$$p_f = \sqrt{\pi_f / \left(\max\left(\pi_{exo}, \pi_{pop}^{quantile}\right)\right)} + N\left(\mu, \sigma^2\right)$$

If the stochastic variable p is below 0.5, the farm will exit. In this case, the agent's current renting contracts will end, while the land owned (with the exemption of the farm stead) will be rented out.

The expected yearly net wage of the farm in the industrial sector is determined from land cover data and the agent's age. First, the share of the industrial land cover *indShare* in a search radius around the farm stead is determined. This search radius is equal to the maximum commuting distance an agent is willing to accept. Thus, in rural regions with little urbanized cover characterized as non-residential, off-farm working opportunities are low. The probability to find work in the industrial sector is determined by the square root of the share of industrial land cover, multiplied by a factor f_{ind} expressing the relation between industrial land cover and open positions, and corrected for a term f_{age} that depends on the agent's age. The expected wage is then determined as the product of the wage in the industrial sector *wage* and the probability shown in the bracket:

$$E\left[wage\right] = wage * \left(f_{ind}\sqrt{indShare} - f_{age}\left[age_{cur} - age_{min}\right]\right)$$

Expected commuting costs *CommCost* are defined from the share of industrial land cover *indShare* times the maximal commuting distance *maxCommDistance,* and the commuting costs per km *commCostPerKm*:

$$E\left[commCost\right] = indShare * maxCommDistance * commCostPerKm$$

The exogenous alternative profit π_{exo} from working off-farm is finally defined as:

$$\pi_{exo} = E\left[wage\right] - E\left[commCost\right]$$

(Britz 2013a).