Agricultural, Food and Environmental Policy Analysis (AFEPA)

Master's Thesis

10 January 2022

Bridging Positive Mathematical Programming and Econometrics for Agricultural Policy Analysis

by

Josef Lubentius Baumert

1st supervisor:

Prof. Bruno Henry de Frahan
UC Louvain / Earth and Life Institute

2nd supervisor:

Prof. Paolo Sckokai
UCSC Piacenza / Department of Agricultural and Food Economics

Additional supervisor and referee:

Prof. Frédéric Gaspart
UC Louvain / Earth and Life Institute

Dear reader,

this work is the fruit of the many hours that I have spent studying with and learning from extremely kind, dedicated, and inspiring fellow students, teachers, and professors. Their commitment for intercultural friendship, their scientific curiosity to understand the world, their passion for agricultural economics and their appreciation for commonly enjoyed food and beverages are the four columns of what is affectionately known as "our AFEPA family".

Almost one year ago I started to immerse myself into the world of positive and econometric mathematical programming. Today I am proudly looking back at what I have learned, all thanks to Professor Bruno Henry de Frahan, who guided and supported me along this journey with invaluable commitment, knowledge, patience, and advice. I owe him my deepest respect and gratitude and truly enjoyed working with him.
I would also like to thank Professor Paolo Sckokai and Professor Frédéric Gaspart for their constant and reliable support and advice, not only related to this thesis, but throughout the entire master's programme.
The genuine interest in the wellbeing and the success of their students can barely be appreciated enough.

Josef Baumert

January 2022

Abstract

It has long been recognized that Positive Mathematical Programming (PMP) and econometrics can be combined to estimate consistent cost function parameters for agricultural supply analysis. Yet, the literature on Econometric Mathematical Programming (EMP) and its practical applicability is comparably scarce. This thesis methodologically refines a recently proposed EMP modelling framework and then applies it to simulate the introduction of a greenhouse gas emission cap for three major agricultural regions in France. The first-order conditions of a non-linear total net revenue maximization problem are directly employed to estimate the cost function parameters of a model that calibrates to reference values. Error terms are specified as the deviation between the observed and the true values of activity levels, fixed-input levels and fixed-input prices. Then, the developed approach is applied to estimate the cost functions of French dairy, cattle, and crop farms, using panel data from the EU's Farm Accountancy Data Network (FADN) for the years 2005-2012. The estimated cost functions are finally used to simulate the introduction of a greenhouse gas (GHG) emission cap at regional level and retrieve marginal abatement costs (MAC) as the dual value of the emission constraint, which can be interpreted as the market price of an emission cap-and-trade system. A comparison with MAC estimated by other researchers confirms that the MAC simulated in this thesis are in a reasonable range. While the levels of almost all farm activities are reduced if an emission cap is imposed, the magnitude of the reduction depends on the emissions per unit value and the level of an activity produced by a farm in the reference scenario. The simulations with GHG emission caps reveal that reductions in activity levels and GHG emissions fall proportionally more on farms with smaller activity levels than on farms with larger activity levels.

Table of contents

List of Abbreviations

AWU	annual work unit
CAPRI Modelling System	Common Agricultural Policy Regionalised Impact Modelling System
CSC	complementary slackness conditions
EC	European Commission
EEA	European Environment Agency
EMP	Econometric Mathematical Programming
EU ETS	EU Emission Trading System
FADN	Farm Accountancy Data Network
GHG	greenhouse gas
GME	Generalized Maximum Entropy
GR	gross revenue
IPCC	Intergovernmental Panel on Climate Change
KKT conditions	Karush-Kuhn-Tucker conditions
LU	livestock unit
LS	least squares
MAC	marginal (emission) abatement costs
MC	marginal costs
MP	mathematical programming
MR	marginal revenue
OLS	ordinary least squares
PMP	Positive Mathematical Programming
RMSE	Root Mean Square Error
TC	total costs
TNR	total net revenue
TVC	total variable costs
UAA	utilized agricultural area

1. Introduction

Agricultural policies are globally transitioning from price or income support schemes towards measures to reduce environmental degradation and incentivize the provision of ecosystem services. This shift fuels the demand for tools for agri-environmental policy analysis that allow to spatially disaggregate resource allocation and output production and link decisions made on farms with environmental outcomes at regional or national level (Mérel and Howitt (2014); Pe'er *et al*. (2019)). A frequently used framework to investigate the impact of an agricultural or an environmental policy on farmers' choices *ex ante* is provided by Mathematical Programming (MP) models, which represent farmers' objectives as constrained optimization problems (Carpentier *et al*., 2015). A milestone in the development of MP models was the emergence of Positive Mathematical Programming (PMP) and its formalization by Howitt (1995): By introducing calibrated non-linear terms in the objective function, PMP allows to construct models that exactly replicate observed reference values of produced activities. In opposite to linear programming models this is possible even if the number of activities exceeds the number of justifiable constraints. At least equally important is the much more realistic simulation behaviour obtained with PMP, as the non-linearity obviates abrupt discontinuities in the modelled supply response. These features have led to a renewed interest in MP in the early 2000s, especially as PMP makes it possible to handle decision problems even if data (for example, at the subregional or farm level) is too scarce to perform econometric procedures (Henry de Frahan *et al*., 2007).
However, it soon became apparent that the response behaviour of standard PMP models would always exhibit a certain arbitrariness if the number of parameters to be specified exceeds the number of observations (Heckelei and Britz, 2000). Additionally, Heckelei and Wolff (2003) demonstrated that PMP yields inconsistent parameter estimates and that it is therefore not well suited for the exploitation of additional information, in opposite to econometrics. They propose a general alternative which directly employs the optimality conditions without the need to first set-up a linear model with a calibration constraint. Their attempt to combine econometrics and MP initiated a field of research to develop models belonging to a class that was later termed Econometric Mathematical Programming (EMP) (Buysse *et al*., 2007). In spite of its theoretical soundness and its usefulness for agri-environmental policy simulations, empirical works applying EMP are still scarce or contain potentially harmful vestiges of PMP, such as first-phase linear models that may introduce bias into the final model used for simulations. (Britz and Arata (2019); Heckelei *et al*. (2012)).

Contributing to the methodological literature on EMP, Henry de Frahan (2019) proposes a framework to simultaneously estimate and calibrate model parameters using the Least Squares criterion. This thesis builds on the suggested approach and follows three objectives:

First, we assess the applicability of the procedure suggested by Henry de Frahan (2019) for agricultural policy analysis. Second, we develop a methodological refinement to relieve the rigid relationship between fixed-inputs and outputs and, thereby, also acknowledge the impact of fixed-input levels on total variable costs. Third, we use the estimated model to simulate the introduction of a hypothetical greenhouse gas (GHG) emission cap-and-trade system for specialized crop, dairy, and cattle farms in the three main agricultural regions in France.

The next chapter sets the scene by outlining the theoretical framework of PMP, starting with a description of the original procedure and a discussion of its shortcomings. Then, the increasing incorporation of econometrics into PMP is described from early applications of the Generalized Maximum Entropy and Bayesian approaches to more recent examples that use the Least Squares criterion and sometimes also account for uncertainty and risk aversion. Finally, the method proposed by Henry de Frahan (2019) is delineated.

Chapter 3 describes and then critically reviews the methodology which the tested cost function estimation approach is based upon. Using a Monte Carlo simulation with artificial data we show that the estimated cost function parameters indeed converge to their true values with increasing sample sizes. However, we also reveal a decisive flaw that is rooted in the Leontief fixed-input-output coefficient. A modification of the model specification proposed by Henry de Frahan (2019) is developed to circumvent this limitation.

Chapter 4 shows the implementation of the cost function estimation using cross-sectional FADN data. After describing the data and the necessary preparation processes, the calibration to observed reference values is validated. Then, output price changes are simulated.

Using a model specification identified as promising in chapter 4, chapter 5 simulates the regional introduction of a GHG emission cap-and-trade system. After providing a brief theoretical background on emission trading systems, an emission constraint is added to the model. With average values for GHG emissions per output level, the simulation is performed for multiple emission reduction scenarios and marginal emission abatement costs are estimated.

Finally, chapter 6 concludes by discussing the suitability of the modelling framework for applied agricultural and environmental policy analysis, drawing from the methodological discussion in chapter 3, but most importantly from the empirical analyses in chapter 4 and 5. Several pathways for further developments of the modelling framework are provided.

The simulation of the emission trading system helps to evaluate in how far the model can be used to answer highly relevant applied research questions. However, it is important to note that the objective of this thesis is *not* to provide exact information on emission mitigation costs. At best, we contribute to a better understanding of emission trading systems and how they might affect different farm types. Most importantly, this thesis aspires to illustrate the potential of EMP models and address some of the challenges that have thus far confined the applicability of EMP.

2. Theoretical Framework

This chapter describes the theoretical foundations behind Positive Mathematical Programming (PMP). The weaknesses of the original approach and the alternatives that led to the development of Econometric Mathematical Programming (EMP) are illustrated. Finally, this chapter introduces the theoretical framework suggested by Henry de Frahan (2019), which provides a novel cost function estimation method that can be assigned to the realm of Econometric Mathematical Programming.

2.1. Evolution of Positive Mathematical Programming

2.1.1. Original Positive Mathematical Programming Framework

Positive Mathematical Programming is an approach for cropping decision modelling that allows for exact calibration of decision variables to observed levels, while responding smoothly to changes in prices or other external variables. The method was used by several agricultural economists already in the late 1980s and early 1990s, but first formalized by Howitt in 1995 (Henry de Frahan *et al.*, 2007). Even though several shortcomings of PMP were noted and addressed by researchers in subsequent years, the paper by Howitt (1995) initiated the widespread application and discussion of calibrated optimization models in agricultural economics. It is continuously considered a seminal contribution to the field by reviewers and is cited by modellers (see for example Henry de Frahan *et al.* (2007), Heckelei *et al.* (2012), Mérel and Howitt (2014), Paris (2017), or Britz and Arata (2019)). The success of PMP originates in the advantages that the framework provides compared to its alternatives available in the late 20th century:

Before the development of the PMP approach, attempts to model farmers' production decisions often relied on normative mathematical programming approaches based on concepts from operations research. A substantial problem of these models is the often-imposed linearity assumption for the objective function and the constraints, which at least partly was provoked by lack of computational power. Consequently, the number of activities produced by a farm in the optimal solution was determined by the number of binding constraints, which in many cases impeded calibration to observed activity levels. While various ad hoc approaches were developed to improve calibration, a fundamental shortcoming of linear programming approaches remained unsolved, namely the jumpy supply behaviour induced by the changes of prices or other external variables. This significantly impaired their aptitude for policy simulations. The PMP approach formalized by Howitt (1995) finally provided a framework to generate more realistic simulation results without the need to impose weakly justified constraints on the model (Heckelei and Britz (2005) and Mérel and Howitt (2014)).

The PMP approach consists of a three-step procedure at the end of which a non-linear total net revenue maximization function for a single economic agent (e.g., a farm) is received with the observed activity levels as the optimal solution.
In the first step a linear total net revenue (TNR) maximization function of the following form is specified (Howitt, 1995):

$$max_x TNR = \mathbf{p}'\mathbf{x} - \mathbf{d}'\mathbf{x} \tag{2.1}$$

subject to

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \qquad [\mathbf{y}] \tag{2.2}$$

$$\mathbf{x} \leq \mathbf{x}^{\mathbf{o}} + \boldsymbol{\varepsilon} \qquad [\boldsymbol{\lambda}] \tag{2.3}$$

$$\mathbf{x} \geq \mathbf{0} \tag{2.4}$$

With the vector $\mathbf{x}$ being the vector of decision variables (a vector of output or activity levels) and $\mathbf{x}^{\mathbf{o}}$ being a vector of the *I* observed and realized output or activity levels. Vector $\mathbf{p}$ is a vector of market output prices and $\mathbf{d}$ is a vector of variable costs per output unit, both with dimension $I \times 1$. Vector $\mathbf{b}$ is referring to the *J* fixed-input constraints. $\mathbf{A}$ represents a fixed-input-output matrix of the dimension $J \times I$ and indicates the usage of fixed-resource J per unit of output or activity I. Finally, $\boldsymbol{\varepsilon}$ is a vector of small perturbations that must be positive and that are determined by the modeller. It guarantees that the calibration constraint (2.3) is de-coupled from the resource constraint (2.2) and therefore the dual value $\mathbf{y}$ of each of the J resource constraints is larger than zero (Paris, 2017).
The intention of the first step is to retrieve the $I \times 1$ vector of dual values $\boldsymbol{\lambda}$ of the calibration constraint that can be interpreted as the differential marginal costs for each activity. Thus the model (2.1)-(2.4) allows to obtain a measure of the total marginal cost of producing the output or activity vector $\mathbf{x}$, which is the sum of the accounting cost vector $\mathbf{d}$ and the differential marginal cost vector $\boldsymbol{\lambda}$ faced by the farmer (Paris and Howitt, 1998). The utilization of all available information on observed decisions of an economic agent to infer essential economic behavioural determinants like the marginal costs is the principal tenet of the PMP framework. It renders the approach "positive", in contrast to "normative" mathematical programming that is used to optimize the production plan of a firm (Paris, 2010).

In the subsequent steps a non-linear total net revenue maximization function is set up, for example with a quadratic functional form[1]:

primal:

$$max_x \, TNR \; = \; \mathbf{p}'\mathbf{x} - \mathbf{c}'\mathbf{x} - \tfrac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} \tag{2.5}$$

subject to

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \qquad [\mathbf{y}] \tag{2.6}$$

with the corresponding dual problem:

dual:

$$min_y \, TC = \mathbf{b}'\mathbf{y} + \tfrac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} \tag{2.7}$$

subject to

$$\mathbf{A}'\mathbf{y} + \mathbf{c} + \mathbf{Q}\mathbf{x} \geq \mathbf{p} \qquad [\mathbf{x}] \tag{2.8}$$

[1] Howitt (1995) defines the derivation of the calibrating parameters as the second step and the specification of model (2.5) - (2.6) as the third step.

with non-negativity imposed on output or activity levels $\mathbf{x}$ for both the primal and the dual problem. To retrieve the linear vector $\mathbf{c}$ $(I \times 1)$ and the quadratic and symmetric matrix $\mathbf{Q}$ $(I \times I)$, the only terms in the model that are unknown thus far, the PMP approach makes use of the values for the total marginal costs received in the first step:

$$\mathbf{MC} = \mathbf{c} + \mathbf{Qx} = \mathbf{d} + \boldsymbol{\lambda} \qquad (2.9)$$

Consequently, by setting the marginal costs derived from the total net revenue maximization model (2.5) – (2.6) equal to the sum of the accounting costs $\mathbf{d}$ per unit and the differential marginal costs $\boldsymbol{\lambda}$, the cost function parameters $\mathbf{c}$ and $\mathbf{Q}$ can be retrieved, completing the quadratic programming model.

The total net revenue maximizing activity levels obtained by solving model (2.5) – (2.6) are identical to the optimal activity levels derived by solving the linear model (2.1) – (2.4), indicating that the PMP model calibrates perfectly to observed activity levels without having to impose additional constraints (Howitt, 1995). While the non-linearity of model (2.5) – (2.6) therefore is an agreeable feature from a pragmatic angle, Howitt (1995) also provided an agronomic justification: Heterogeneous land quality caused the decreasing gross margin, together with declining marginal yields as the share of an activity increases within a region. This interpretation also illustrates why the non-linear cost function parameter must be convex, requiring that matrix Q be positive (semi)definite.

The model (2.5) – (2.6) also allows to quantify the impact of a change in output or activity prices ($\mathbf{p}$) or fixed-input quantities ($\mathbf{b}$) on optimal output or activity levels ($\mathbf{x}$) or shadow prices ($\mathbf{y}$), a property that makes the PMP approach so interesting for policy analysis (Paris, 2010).

Heckelei and Britz (2005) explain the popularity of PMP-type modelling at the end of the 1990s and at the beginning of the 2000s also partly by a general shift in the paradigm of (European) agricultural policy away from price support: They argue that policy instruments such as quotas or set-aside obligations could in many cases not be sufficiently well simulated by standard econometric approaches. Programming models, on the other hand, allow the linkage between economic and bio-physical aspects (such as natural resource requirements) explicitly, a feature that largely prevents implausible results. Mérel and Howitt (2014) add that the modest data requirements of PMP models facilitate their specification at very fine geographical resolution (e.g., farm level), making them attractive for the simulation of agri-environmental policies.

2.1.2. Shortcomings of Positive Mathematical Programming and Development of Alternatives

The attention that PMP received in the modelling community evoked a critical discussion about both the flaws and the merits of the methodology, culminating in an array of papers suggesting modifications to the original approach. Heckelei and Britz (2005), Henry de Frahan *et al.* (2007) and Mérel and Howitt (2014) review the early developments extensively.

This section focusses on three aspects: first, on the justification for a non-linear, convex cost function specification, which deserves attention as it illustrates the conceptual link between the rather theoretical model formulation and the agronomic reality. Second, on the second-order behaviour of under-determined PMP models and third, on the estimation bias caused by inconsistency between the models received in the different PMP phases. The remedies that were developed in the early 2000s to address the under-determinacy and the estimation bias constitute the foundation for econometric mathematical programming approaches, such as the methodology developed and evaluated in this thesis.

2.1.2.1. Justification for the Global Convexity of the Cost Function

As mentioned in the previous section, Howitt (1995) justified the nonlinearity in the second and third step of PMP with declining marginal crop yields, mainly due to the heterogeneity of land. Besides, increasing marginal costs are caused by the fixed nature of many farm inputs (Howitt, 1995). A first objection against land heterogeneity as the reason for decreasing marginal yields was expressed by Heckelei (2002): If an increase in activity levels of a crop led to continuously decreasing yields of this crop, one must assume that the additional unit of land is the most productive unit of all available land that is not yet allocated to the respective activity. However, as in reality land is allocated to multiple activities, it is hard to justify why the expansion of one crop by one unit should always lead to marginally lower average yields. Mérel and Howitt (2014) later shared these reservations, remarking that land heterogeneity could nevertheless lead to decreasing yields if one assumes that the land characteristics that make land favourable vary between activities.

Mérel and Howitt (2014) also discuss the justification that fixed endowment led to increasing marginal costs. They argue that, for example, crop-specific machinery or labour that cannot be adjusted in the short run could decrease average gross margins if an activity is expanded. However, they admit that this argument does no longer hold if the fixed-input is not activity-specific.

The different justifications for nonlinearity also imply different assumptions about which part of the net revenue function causes the convexity, as pointed out in the review by Mérel and Howitt (2014): either, the marginal yield is decreasing or, as it is much more frequently assumed in literature, the marginal cost is increasing in activity levels. The few papers that applied the former conceptualization usually implemented it using a yield function (like the original paper by Howitt (1995)), a Cobb-Douglas function or a Constant-Elasticity-of-Substitution framework. On the other hand, increasing marginal costs are mostly implemented in literature using a simple quadratic functional form or, in some cases, the generalised Leontief cost function (Henry de Frahan, 2019).

In the review of Mérel and Howitt (2014), the authors express their moderate preference for the decreasing marginal yield approach, arguing that the marginal cost approach implied the rather unrealistic assumption of constant yields. On the other hand, Heckelei and Britz (2005) remark that it was unlikely that either only the yield changed but input application remained constant or vice versa.

Given these equivocal argumentations, Mérel and Howitt (2014) summarize that most PMP studies have taken the rather pragmatic stance that the net revenue maximization problem must be concave to ensure that a solver may find a global optimum, regardless of whether this is reached through decreasing marginal yields or increasing marginal costs.

2.1.2.2. Under-Determination Problem

The original PMP model as formalized by Howitt (1995) was developed as an approach to represent the total net revenue maximization problem of *one* individual economic agent, typically a farm manager. This reflects the initial objective to develop a modelling framework that can be used if data is too scarce to apply standard econometric approaches based on behavioural functions. However, the consideration of only one observation led to a problem that was soon understood to be a general shortcoming of the PMP approach: the under-determination of the system of equation (2.9), which consists of I equations used to derive I linear and $I(I+1)/2$ quadratic cost function parameters. The implication of this is that equation (2.9) can be fulfilled by an infinite number of parameters, all of which calibrate the model but imply different response behaviour of the economic agents. Heckelei and Britz (2005) elaborate this point and show that the change of the optimal level of each activity for a change in output prices depends on the *entire* $\mathbf{Q}$-matrix, i.e., not only the cost function parameters for the respective activity (see Annex I):

$$\frac{\partial \mathrm{x}}{\partial \mathrm{p}} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}')^{-1}\mathbf{A}\mathbf{Q}^{-1} \tag{2.10}$$

Several ad-hoc approaches were developed to address the under-determination problem, as summarized by Henry de Frahan *et al.* (2007). They mainly concerned the assumptions made for the cost function parameters $\mathbf{c}$ and $\mathbf{Q}$: By assuming that the off-diagonal elements $q_{i,j}$ with $i \neq j$ of matrix $\mathbf{Q}$ are all equal to 0, the number of parameters to be estimated can be reduced to $I + I$. As the resulting model would still be ill-posed, additional assumptions have to be made for the linear cost function parameter $\mathbf{c}$:

By setting the parameter $\mathbf{c}$ equal to 0, the diagonal elements of matrix $\mathbf{Q}$ would reflect the accounting costs *and* the differential marginal costs (divided by the activity level) of each activity i in I:

$$q_{ii} = \frac{(d_i+\lambda_i)}{x_i} \quad , \quad c_i = 0 \qquad\qquad \forall i = 1, \dots, I$$

Or, by assuming that the vector $\mathbf{c}$ is equal to the accounting cost vector $\mathbf{d}$, the quadratic cost parameter $\mathbf{Q}$ would be equal to the differential marginal costs $\boldsymbol{\lambda}$ divided by the activity level $\mathbf{x}$:

$$q_{ii} = \frac{\lambda_i}{x_i} \;,\quad c_i = d_i \qquad\qquad \forall i = 1, \dots, I$$

The average cost approach provides yet another alternative to calculate the parameters $\mathbf{c}$ and $\mathbf{Q}$, by presuming that the accounting cost vector $\mathbf{d}$ for each activity unit is equal to the average total variable costs. This assumption can then be used, together with the first order condition (equation 2.9), to derive the parameters $\mathbf{c}$ and $\mathbf{Q}$ as follows:

$$average\, TVC_i = c_i - \frac{1}{2} x_i q_{ii} = d_i \qquad \forall i = 1, \dots, I$$

$$q_{ii} = 2\frac{\lambda_i}{x_i} \;,\quad c_i = d_i - \lambda_i \qquad\qquad \forall i = 1, \dots, I$$

It is noteworthy that this average cost approach requires that variable inputs can be attributed to each of the activities in I, an assumption that renders the off-diagonal elements of $\mathbf{Q}$ to be zero by definition, but that is also inconsistent with the joint technology of a multi-product cost function (Heckelei and Britz, 2005).

All of these approaches facilitate calibration, but they have some significant drawbacks, which become apparent if one brings to mind the meaning of the elements of matrix $\mathbf{Q}$. They are the second derivatives of the variable cost function with respect to activity levels $\mathbf{x}$ and, consequently, indicate how the marginal costs change when activity levels change. If only one observation for the dual values of the calibration constraint (2.3) is considered it is simply impossible to infer information on the *change* of marginal costs, meaning that the second-order behaviour of the respective model becomes arbitrary and potentially unsatisfactory (Heckelei and Britz, 2000).

An option to derive the parameters of the quadratic cost function in a less arbitrary way is the incorporation of exogenous elasticities as prior information. Helming *et al.* (2001), who implemented this method first, consider only own price elasticities (leaving aside the off-diagonal elements of $\mathbf{Q}$) and ignore the marginal effect of price changes on shadow prices (Heckelei, 2002). This leads to a myopic calibration with model elasticities being lower than the exogenous values. Mérel and Bucaram (2010) conclude that the utilization of exogenous elasticities for better second-order behaviour may under certain circumstances allow perfect calibration. However, they also note that exogenous cross-price elasticities are not often available. Heckelei and Britz (2005) argue that the incorporation of exogenous information on supply changes nevertheless presented the only convincing use of PMP if just one observation is considered.

Yet another approach was introduced and tested by Paris and Howitt (1998), who recovered a total variable cost function by combining PMP with elements from the maximum entropy approach. By using the Generalized Maximum Entropy (GME) estimator (see section 2.1.3.1) as an econometric criterion, this “path breaking” contribution also prepared the ground for the *estimation* rather than calibration of the cost function parameters (Heckelei and Britz (2005, p. 59)). Besides, even though Paris and Howitt (1998) continued to use just one observation for the derivation of the cost function, their approach generally made the incorporation of multiple observations possible. A first application of PMP for cross-sectional datasets was then provided by Heckelei and Britz (2000). By including multiple observations, they could derive changes in marginal costs based on observed differences between regions.

2.1.2.3. Inconsistent Parameter Estimation

The incorporation of multiple observations and the utilization of the Generalized Maximum Entropy approach addressed the under-determination problem of PMP, but they largely built on the original framework with its three-step procedure. Heckelei and Wolff (2003) scrutinized the PMP approach more generally. They concluded that it does not allow for consistent parameter estimation and therefore is not well suited for the exploitation of additional data information. The reason for this is that the shadow prices obtained in the first step are different from the ones obtained in the second step, indicating that equation (2.9) cannot be seen as unbiased. To elaborate this argument, equation (2.11) displays the first order condition of the quadratic model (2.5) – (2.6), with rearranged terms, assuming strictly positive activity levels and binding resource constraints:

$$\mathbf{y} = \left(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\mathbf{Q}^{-1}(\mathbf{p}-\mathbf{c})-\mathbf{b}\right) \tag{2.11}$$

However, the vector $\mathbf{y}$ of shadow prices of the resource constraints is different in stage one of the PMP procedure, when being obtained with the linear model (2.1) – (2.4). To see this, it is necessary to introduce the reader to the notion of 'preferable' and 'marginal' activities, that follows from the linear model defined in the first PMP step: The vector $\boldsymbol{\varepsilon}$ of perturbations in the calibration constraint (2.3) allows the optimal activity levels $\mathbf{x}$ to be $\boldsymbol{\varepsilon}$ units larger than the observed levels $\mathbf{x}^{\mathbf{o}}$. However, to comply with the resource constraint (2.2), the elements of $\mathbf{x}$ cannot be larger than the corresponding elements of $\mathbf{x}^{\mathbf{o}}$ for all activities. Thus, the calibration constraint (2.3) is binding only for those activities with the highest marginal net revenue $\mathbf{p_i} - \mathbf{d_i}$. Those activities are usually named 'preferable' activities. The remaining 'marginal' activities are confined by the resource constraint (2.2). Consequently, the dual value $\boldsymbol{\lambda}$ of the calibration constraint is zero for all marginal activities and strictly positive for all preferable activities, whereas the dual value $\mathbf{y}$ of the resource constraint takes positive values only for marginal activities and is zero for all preferable activities. This property allows to partition the activity level vector $\mathbf{x}$ into an $((I-M)\times 1)$ vector of preferable activities $\mathbf{x}^{\mathbf{p}}$ with the respective parameters $\mathbf{A}^{\mathbf{p}}$, $\mathbf{p}^{\mathbf{p}}$ and $\mathbf{d}^{\mathbf{p}}$, and an $(M\times 1)$ vector of marginal activities $\mathbf{x}^{\mathbf{m}}$ with the respective parameters $\mathbf{A}^{\mathbf{m}}$, $\mathbf{p}^{\mathbf{m}}$ and $\mathbf{d}^{\mathbf{m}}$, with M being the number of marginal activities. Therefore, the fixed-input shadow price derived from the linear model (2.1) – (2.4) in step one is

$$\mathbf{y} = \left(\mathbf{A}^{\mathbf{m}\prime}\right)^{-1}\left(\mathbf{p}^{\mathbf{m}}-\mathbf{d}^{\mathbf{m}}\right) \tag{2.12}$$

This shows that in the linear model in the first PMP step the dual values of the resource constraints are determined by different parameters than the dual values of the resource constraints in the quadratic model (equation 2.11). This also has implications for the dual values of the calibration constraints, $\boldsymbol{\lambda}^{\mathbf{p}}$, which can be derived for preferable activities from the first order condition of the linear model:

$$\boldsymbol{\lambda}^{\mathbf{p}} = \mathbf{p}^{\mathbf{p}} - \mathbf{d}^{\mathbf{p}} - \mathbf{A}^{\mathbf{p}\prime}\mathbf{y} \qquad (2.13)$$

As the dual values $\boldsymbol{\lambda}^{\mathbf{p}}$ depend on the fixed-input shadow prices $\mathbf{y}$ determined in the first step, and since the other values $\boldsymbol{\lambda}$ are used to specify the marginal costs of the quadratic model, Heckelei and Wolff (2003) argue that this leads to inconsistency between the linear and the non-linear models. This inconsistency can also be grasped by recalling the reason for why the calibration constraint was introduced at all, in the first PMP step as formulated by Howitt (1995): It represents the unobserved factors that limit the expansion of the 'preferable', thus the most profitable activities. In the quadratic model, however, calibration is achieved by ensuring that marginal returns decrease in activity levels. Consequently, whether an activity is among the most profitable activities depends on its level. This also implies that the dual value of the fixed-input constraint is not necessarily determined only by those activities that are among the 'marginal' activities in the first step. Given this erroneous calculation of the dual values, the shadow prices $\mathbf{y}$ do not converge to the true shadow prices $\mathbf{y}$ when additional observations are added (Henry de Frahan *et al.*, 2007).

2.1.2.4. A General Alternative to Positive Mathematical Programming

From the perspective of an econometrician, the selected model specification should represent what is assumed to be the true data generating process. Heckelei and Wolff (2003) elaborated an approach to *estimate* a model that is consistent with the modeler's assumption about the 'true' model. The important feature to ensure consistency is the simultaneous estimation of dual values and parameters in one step. In the case of a concave, quadratic total net revenue maximization function, the objective function can be expressed as follows:

$$max_{\mathbf{x_f}} TNR_f = \mathbf{gm_f'x_f} - \mathbf{c'x_f} - \tfrac{1}{2}\mathbf{x_f'Qx_f} \qquad (2.14)$$

subject to $\quad \mathbf{A_f x_f} \leq \mathbf{b_f} \qquad [\mathbf{y_f}] \qquad (2.15)$

where $\mathbf{gm}$ is a $I \times 1$ vector of (observed) gross margins. Besides, non-negativity for the activity levels $\mathbf{x}$ is imposed. A striking difference between model (2.14) – (2.15) and the non-linear model set up in the second and third step of the PMP process (equation (2.5) – (2.6)) is the introduction of indices for every observation (e.g., farm) $\mathrm{f} = 1, \dots, \mathrm{F}$. This reflects the econometric stance of considering a subsample of a population, in contrast to the original PMP approach that was explicitly developed for cases with only one observation. Therefore, $\mathbf{gm_f}$,

$\mathbf{x_f}$, $\mathbf{b_f}$ and $\mathbf{y_f}$ are the vectors of gross margins, activity levels, fixed-input constraints and fixed-input shadow prices, respectively, for each farm f. However, apart from the fact that Heckelei and Wolff (2003) replace the vector of prices $\mathbf{p}$ by the vector of gross margins $\mathbf{gm}$, the objective function at the farm level is equivalent to the non-linear model specified by Howitt (1995). The first-order conditions of model (2.14) – (2.15) can be written as

$$\mathbf{gm_f} - \mathbf{y_f'A_f} - \mathbf{c} - \mathbf{Q}(\mathbf{x_f^o} + \mathbf{h_f}) \leq \mathbf{0} \tag{2.16}$$

$$\mathbf{A_f}(\mathbf{x_f^o} + \mathbf{h_f}) \leq \mathbf{b_f} \tag{2.17}$$

To ensure that curvature conditions are fulfilled (such as positive semi-definiteness of matrix Q), second order conditions like the Cholesky factorisation must be imposed as well[2]. The vector $\mathbf{h_f}$ represents an $I \times 1$ vector of stochastic errors with mean zero and standard deviations σ_i. Consequently, and in opposite to the vector of perturbations $\boldsymbol{\varepsilon}$ introduced in the first step of the traditional PMP approach, the elements of the vector $\mathbf{h_f}$ may take negative values. Heckelei and Wolff (2003) provide several interpretations of this error specification: It may represent an error incurred by the analyst (e.g., a measurement error) or an optimization error by the farmer. In any case does it imply that the optimal activity level for each observation is not necessarily equal to the observed activity level, as an observed activity level $\mathrm{x_{fi}^o}$ varies from the optimal activity level $\mathrm{x_{fi}}$ by this idiosyncratic error $\mathrm{h_{fi}}$.

Making use of the Karush-Kuhn-Tucker conditions, it is possible to transform equations (2.16) – (2.17) into equalities:

$$(\mathbf{x_f^o} + \mathbf{h_f}) \odot (\mathbf{gm_f^o} - \mathbf{y_f'A_f} - \mathbf{c} - \mathbf{Q}(\mathbf{x_f^o} + \mathbf{h_f})) = \mathbf{0} \tag{2.16'}$$

$$\mathbf{y_f} \odot (\mathbf{b_f^o} - \mathbf{A_f}(\mathbf{x_f^o} + \mathbf{h_f})) = \mathbf{0} \tag{2.17'}$$

Where the symbol $\odot$ represents the elementwise (or Hadamard) product of two matrices. As it will be elaborated in section 3.1.2, the representation of these conditions as an equality is particularly useful if constraints may not be binding for every observation.

Heckelei and Wolff (2003) chose a generalized maximum entropy specification as the framework for the parametrization of the error vectors, which they used to set up an objective function with equations (2.16) and (2.17) as the constraints. This allowed them to simultaneously estimate the unknown dual values $\mathbf{y}$ and the cost function parameters $\mathbf{c}$ and $\mathbf{Q}$ and to analyse the impact of additional prior information (see also Heckelei and Britz (2005)). They remarked that other estimation techniques such as the Least Squares or Generalized Methods of Moments could have been applied as well, given that their estimation problem is "well-posed".

[2] The Cholesky factorisation is not explicitly written as a constraint here and for the following non-linear models, as it just constitutes an easily implementable means of ensuring convexity, without being part of the generic model. For a more detailed description of the implementation, see subchapter 3.1.1.

Heckelei and Wolff (2003) pointed out that the direct use of optimality conditions for estimation purposes had already been used before, even though mainly in the context of investment models, but not yet as an alternative to the standard PMP framework. Heckelei and Britz (2005) discussed the reasons for this comparably late adoption for programming model estimation. They concluded that, apparently, mathematical programmers did not see the necessity for the estimation of programming models with multiple observations, while, on the other hand, econometricians seem to have largely ignored the mathematical programming literature.

2.1.3. Evolution Towards Econometric Mathematical Programming

The road for a more systematic incorporation of econometrics into the positive mathematical programming framework was paved by Paris and Howitt (1998), who introduced an econometric criterion (the GME estimator) and then by Heckelei and Wolff (2003), who provided a more consistent alternative to PMP. Besides, the inconsistency between the linear and the nonlinear model of the traditional PMP approach unveiled by Heckelei and Wolff (2003) prompted most researchers after 2003 to estimate first- and second-order conditions directly, without estimating a linear model first (Henry de Frahan, 2019). These developments largely describe the advent of a framework that Buysse *et al*. (2007) termed “Econometric Mathematical Programming” (EMP). The following section concisely summarizes the evolution towards EMP: First, the GME principle and the Bayesian approach are briefly outlined. Then, more recent developments are depicted with a focus on those papers that provide the conceptual foundations for the methodology developed and implemented in this thesis.

2.1.3.1. The Generalized Maximum Entropy and Bayesian Approaches

The possibility to systematically extract information provided by more than one observation is a fundamental precondition for cost function parameter *estimation*, rather than calibration. Heckelei and Wolff (2003) argue that the incorporation of information from multiple observations could in principle be accomplished with classical estimation techniques, such as Least Squares. One may therefore wonder why so many modellers have employed the much less classical (Generalized) Maximum Entropy procedure for estimation, such as Paris and Howitt (1998), Heckelei and Britz (2000), Paris (2001), Arfini *et al*. (2008), Petsakos and Rozakis (2011) or Graveline and Mérel (2014). One reason for the popularity of GME in early papers, mainly Paris and Howitt (1998), certainly is that it provides a solution to the problem of underdeterminacy, even if only one observation is available. To comprehend this property, it is important to understand the basic functionality of the Maximum Entropy approach.

The concept of entropy in statistics was first formalized by Shannon (1948), who measures it as follows, for discrete probabilities:

$$-\sum_{i=1}^{n} p_i ln(p_i) \tag{2.18}$$

where p_i stands for the probability of the ith among n support points. If entropy is maximized, this usually implies that n probabilities p_i are calculated so that equation (2.18) is maximized, subject to the sum of all probabilities being equal to 1. One can easily verify that this is the case if $p_{i=1} = \frac{1}{n} = p_{i=2} = \cdots = p_{i=n}$, i.e., if all support points are equiprobable, in concurrence with Laplace's rule. The intuition behind maximum entropy is to find the probability distribution that best represents available information about a system while imposing as little assumptions as possible about unknown aspects of this probability distribution. In a context of econometrics the approach experienced a breakthrough in the late 1990s and early 2000s, following the publication of a book by Golan, Judge and Miller (1996). They present a generalized maximum entropy framework, consisting of a set of techniques that make it possible to use prior information on estimation results. A feature that made GME particularly relevant for the early PMP literature is that, given its strictly concave function of probabilities, it allows to find a unique solution also for ill-posed problems with comparably little prior assumptions (Paris and Howitt (1998)).

Heckelei and Wollf (2003) employ GME as a framework for the cost function parameter estimation of a "well-posed" problem. They do this by reparametrizing the vector of error terms $\mathbf{h_f}$ of model (2.16) – (2.17) as *expected* errors, with a discrete probability distribution. They set a support point at five standard deviations above and below the observed activity level $\mathbf{x_f^o}$, respectively. Striving to assign each support point the same probability, the entropy is maximized for the vector $\mathbf{c}$ and matrix $\mathbf{Q}$ that allow the vector of error terms $\mathbf{h_f}$ to be as close as possible to 0. Consequently, the GME approach can be used as an error term minimization framework, comparable to the Least Squares method. In fact, Preckel (2001) shows that for support points that are centred around zero and spanning a wide range, Least Squares and GME produce essentially identical coefficient estimates.

GME also provides a useful framework for the incorporation of exogenous prior information, such as elasticities. Again, Heckelei and Wolff (2003), provide an instructive example: First, they write the vector of land allocation elasticities $\mathbf{e}$ with respect to own gross margins as a function of the cost function parameters, the fixed-input-output coefficient, the sample mean of the gross margin and the sample mean of observed land allocations, for each activity. Then, they define an upper and a lower support point for each own-price elasticity, based on exogenous information. Entropy maximization pulls the elements of vector $\mathbf{e}$ towards the mean of the two elasticity support points, while at the same time the other entropy criterion continues to minimize the vector $\mathbf{h_f}$ of error terms. The authors point out that smaller support ranges for the exogenous information lead to higher penalties for deviation from the support centre, as this would require unequal probabilities for each support point and consequently lower the entropy. The perspective on GME as a penalty function that punishes deviations

from prior information with an intensity that depends on the strength of this prior information is outlined by Preckel (2001) in more detail.
Also several more recent studies used GME for the incorporation of exogenous elasticities, for example Severini and Cortignani (2012) or Petsakos and Rozakis (2015).

While GME has evidently served as a useful procedure in a PMP context and facilitated the evolution from standard PMP towards EMP, it also entails some pitfalls that were outlined for example by Heckelei *et al*. (2008). An important shortcoming is the lack of interpretability of the interaction between the choice of support points, the choice of reference prior probabilities and data information. Besides, the additional variables (such as probabilities) and equations that are introduced to incorporate GME in the estimation process render the model computationally difficult. Therefore, the authors suggest a Bayesian alternative, which also allows to find unique solutions to underdetermined systems of equations by identifying Bayesian highest posterior density solutions. They argue that this approach successfully addressed the above-defined flaws of GME, while being able to reproduce the behaviour of GME models perfectly, if desired. One example for the practical applicability of the Bayesian approach was provided by Jansson and Heckelei (2011), who use it to estimate behavioural parameters of supply models for 219 regions in the EU, within the framework of the CAPRI model. Henry de Frahan (2019) provides a more extensive summary of the functionality and the various applications of the Bayesian approach in recent literature.

2.1.3.2. Standard Econometric Approaches to Cost Parameter Estimation

2.1.3.2.1. Application of Least Squares Estimators

An important advantage of GME or the Bayesian approach for estimation purposes certainly is that it allows to find unique solutions even for underdetermined problems. However, with a sufficiently large number of degrees of freedom more classical approaches, such as Least Squares or the Generalized Methods of Moments, can be applied for parameter estimation (Heckelei and Wolff (2003)). In the EU, the Farm Accountancy Data Network (FADN) provides a rich source of farm data and is frequently used in mathematical programming models (Heckelei *et al*. (2012)). It is conceivable that its continuous extension in the early 2000s facilitated the rise of modelling approaches with comparably high data requirements. The impact assessment of the “Health Check” of the EU’s Common Agricultural Policy by Arfini and Donati (2011) presents an example for the incorporation of the Least Squares estimation method in a PMP setting. In a first step, the authors set up an estimation model with the minimization of the squares of error terms $\mathbf{u}$ as the objective function:

$$min_{\mathbf{u_f},\mathbf{d_f},\boldsymbol{\lambda}_\mathbf{f},\mathbf{Q}} LS = \frac{1}{2}\mathbf{u}_\mathbf{f}'\mathbf{u_f} \quad (2.19)$$

subject to

$$\mathbf{d_f} + \boldsymbol{\lambda}_\mathbf{f} \leq \mathbf{Q}\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{u_f} \quad (2.20)$$

$$\mathbf{d}_\mathbf{f}'\mathbf{x}_\mathbf{f}^\mathbf{o} \leq TVC_f \quad (2.21)$$

$$\mathbf{u}_\mathbf{f}'\mathbf{x}_\mathbf{f}^\mathbf{o} + \frac{1}{2}\mathbf{x}_\mathbf{f}^{\mathbf{o}\prime}\mathbf{Q}\mathbf{x}_\mathbf{f}^\mathbf{o} \geq TVC_f \tag{2.22}$$

$$\mathbf{d}_\mathbf{f} + \boldsymbol{\lambda}_\mathbf{f} + \mathbf{A}_\mathbf{f}'\mathbf{y}_\mathbf{f} \geq \mathbf{p}_\mathbf{f} \tag{2.23}$$

$$\mathbf{b}_\mathbf{f}'\mathbf{y}_\mathbf{f} + \boldsymbol{\lambda}_\mathbf{f}'\mathbf{x}_\mathbf{f}^\mathbf{o} = \mathbf{p}_\mathbf{f}'\mathbf{x}_\mathbf{f}^\mathbf{o} - \mathbf{d}_\mathbf{f}'\mathbf{x}_\mathbf{f}^\mathbf{o} \tag{2.24}$$

$$\textstyle\sum_{f=1}^{F} \mathbf{u}_\mathbf{f} = 0 \tag{2.25}$$

To ensure convexity of the quadratic cost matrix $\mathbf{Q}$ the Cholesky factorisation is implemented, too. Using the same notation as, e.g., in equations (2.1) – (2.8), the vectors $\mathbf{d}$ represent the accounting variable costs and the vectors $\boldsymbol{\lambda}$ the differential marginal costs. The scalars TVC are the total farm variable costs retrieved from the FADN. Thus, they are observed parameters. It is noteworthy that, in opposite to the original PMP approach, the accounting cost vectors $\mathbf{d}$ are endogenous. Constraint (2.21) ensures that the accounting costs cannot be larger than the observed TVC. The left-hand side of constraint (2.22) becomes the new cost function and must be at least as large as the observed TVC. Constraint (2.23) is the first order condition of a TNR maximization function, equivalent to equation (2.8) and (2.16) of the previously defined nonlinear models. Equation (2.24) states that the optimal solution for the dual problem must equate the optimal value for the primal problem. The estimated quadratic cost function parameter $\widehat{\mathbf{Q}}$ and the vector of error terms $\widehat{\mathbf{u}}$ can then be employed in a TNR maximization problem that can be used to simulate farm responses:

$$max_{\mathbf{x}_\mathbf{f}} TNR_f = \mathbf{p}_\mathbf{f}'\mathbf{x}_\mathbf{f} - \widehat{\mathbf{u}}_\mathbf{f}'\mathbf{x}_\mathbf{f} - \frac{1}{2}\mathbf{x}_\mathbf{f}'\widehat{\mathbf{Q}}\mathbf{x}_\mathbf{f} \tag{2.26}$$

subject to
$$\mathbf{A}_\mathbf{f}\mathbf{x}_\mathbf{f} \leq \mathbf{b}_\mathbf{f} \qquad [\mathbf{y}_\mathbf{f}] \tag{2.27}$$

with non-negative activity levels $\mathbf{x}_\mathbf{f}$. Here, the meaning of the error terms $\mathbf{u}$ becomes evident: They are the linear terms of the quadratic cost function that we denoted with $\mathbf{c}$ in the abovementioned models. First, this implies that the interpretation of the vector $\mathbf{u}$ of error terms is different from the interpretation of the vector $\boldsymbol{\varepsilon}$ of error terms in Howitt (1995) or $\mathbf{h}$ in Heckelei and Wolff (2003). The latter defined the elements of $\mathbf{h}$ as the deviation of the observed activity level from the "true" unobserved activity level, a perspective that does not apply for the elements of $\mathbf{u}$. Second, by minimizing the elements of $\mathbf{u}$ Arfini and Donati (2011) must assume that, with increasing sample size, the linear term of the cost function converges to zero.

Henry de Frahan (2019) remarked that the zero-profit condition imposed by constraint (2.24) was a questionable assumption, as it implies that all farms observed in the sample must have already reached their long-run equilibrium.

Another example for the application of Least Squares estimation in a PMP framework is Paris (2017). As several aspects of his paper provide the foundation for the model implemented and analysed in this thesis, we will outline his approach in some detail. Different from Arfini and Donati (2011) but similar to Heckelei and Wolff (2003), Paris (2017) distinguishes between economically optimal activity levels $\mathbf{x}$ and observed activity levels $\mathbf{x^o}$ which deviate from $\mathbf{x}$ by an unrestricted error term $\mathbf{h}$, so that $\mathbf{x} = \mathbf{x^o} + \mathbf{h}$. Another important feature is the explicit incorporation of observed shadow prices $\mathbf{y^o}$ in the cost function estimation model. The idea to make use of potentially available information on the prices of fixed-resources certainly is not new in itself. Early approaches to include observed land rents into a PMP model specification to obtain more reliable results were developed for example by Gohin and Chantreuil (1999), Cypris (2000) and Röhm and Dabbert (2003). Heckelei and Wolff (2003) then describe how exogenous information on the mean value of shadow prices of the fixed factors may be incorporated as prior information in their GME framework. Based on their Monte Carlo simulations they conclude, that especially for small sample sizes this additional piece of information may actually significantly reduce the Mean Square Errors of estimated parameters. Paris (2017) incorporates land prices observed at farm level in his model, arguing that "if the information on land price and other important limiting inputs is available, it should be used (…) in order to avoid violating the principal tenet of the methodology: all the available information should be used" (p. 21). By assuming that observed land prices $\mathbf{y^o}$ deviate from the "true" shadow prices $\mathbf{y}$ by an unrestricted error $\mathbf{u}$, he establishes the relation $\mathbf{y} = \mathbf{y^o} + \mathbf{u}$. The vector of error terms $\mathbf{u}$ makes it possible to include information on fixed-resource prices, even if it is only available at a higher level of aggregation, for example at regional level. However, as the vector $\mathbf{u}$ is finally minimized together with the vector $\mathbf{h}$ in a Least Squares setting, a consistent parameter estimation hinges on the assumption that the expected value of the vector $\mathbf{u}$ is 0. By alluding to the original PMP approach developed by Howitt (1995), Paris (2017) starts by setting up a linear model in a first phase, starting with the primal problem:

$$max_{\mathbf{x_f},\mathbf{h_f}} AUX_f = \mathbf{p}_\mathbf{f}'\mathbf{x_f} - \mathbf{d}_\mathbf{f}'\mathbf{x_f} - \tfrac{1}{2}\mathbf{h}_\mathbf{f}'\mathbf{W_f}\mathbf{h_f} \tag{2.28}$$

subject to

$$\mathbf{A_f}\mathbf{x_f} \leq \mathbf{b_f} \qquad [\mathbf{y_f}] \tag{2.29}$$

$$\mathbf{x_f} = \mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{h_f} \qquad [\boldsymbol{\lambda}_\mathbf{f}] \tag{2.30}$$

with $\mathbf{x_f} \geq 0$. Equation (2.28) is an auxiliary function that is equal to a TNR maximization function subtracted by a weighted square of the error terms, $\frac{1}{2}\mathbf{h_f}'\mathbf{W_f}\mathbf{h_f}$. The matrix $\mathbf{W}$ is a weight matrix with the only purpose to ensure consistent units of measurement[3]. The vectors $\mathbf{d}$ represent the unit accounting costs of each farm that are known to the modeller. From equations (2.28) – (2.30) the following Lagrangian function with its first-order conditions can be derived:

$$LMAX_f(\mathbf{x_f}, \mathbf{h_f}, \mathbf{y_f}, \boldsymbol{\lambda}_\mathbf{f}) = \mathbf{p}_\mathbf{f}'\mathbf{x_f} - \mathbf{d}_\mathbf{f}'\mathbf{x_f} - \tfrac{1}{2}\mathbf{h}_\mathbf{f}'\mathbf{W_f}\mathbf{h_f} + \mathbf{y}_\mathbf{f}'[\mathbf{b_f} - \mathbf{A_f}\mathbf{x_f}] + \boldsymbol{\lambda}_\mathbf{f}'[\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{h_f} - \mathbf{x_f}] \tag{2.31}$$

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{x_f}} = \mathbf{p_f} - \mathbf{d_f} - \mathbf{A}_\mathbf{f}'\mathbf{y_f} - \boldsymbol{\lambda}_\mathbf{f} \leq \mathbf{0} \tag{2.32}$$

[3] See section 2.2 for more details.

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{h_f}} = -\mathbf{W_f h_f} + \boldsymbol{\lambda}_\mathbf{f} = \mathbf{0} \tag{2.33}$$

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{y_f}} = \mathbf{b_f} - \mathbf{A_f x_f} \geq 0 \tag{2.34}$$

$$\frac{\partial \mathbf{LMAX_f}}{\partial \boldsymbol{\lambda}_\mathbf{f}} = \mathbf{x_f^o} + \mathbf{h_f} - \mathbf{x_f} = 0 \tag{2.35}$$

substituting (2.33) in (2.32) we can rewrite equation (2.32) as:

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{x_f}} = \mathbf{A_f' y_f} + \mathbf{W_f h_f} + \mathbf{d_f} \geq \mathbf{p_f} \tag{2.36}$$

The respective dual problem looks as follows:

$$min_{\mathbf{y_f},\mathbf{u_f}} AUX_f = \mathbf{b_f' y_f} + \frac{1}{2}\mathbf{u_f' V_f u_f} \tag{2.37}$$

subject to

$$\mathbf{A_f' y_f} + \mathbf{d_f} \geq \mathbf{p_f} \qquad [\mathbf{x_f}] \tag{2.38}$$

$$\mathbf{y_f} = \mathbf{y_f^o} + \mathbf{u_f} \qquad [\boldsymbol{\psi}_\mathbf{f}] \tag{2.39}$$

with $\mathbf{y_f} \geq 0$. Analogous to the primal problem, the squared and weighted error terms that indicate the difference between the optimal and the observed value of the shadow prices are appended to the objective function to form an auxiliary function. Paris (2017) names the weight matrix ensuring consistency of measurement units in the dual problem $\mathbf{V}$. The respective Lagrangian function and its first order conditions are as follows:

$$LMIN_f(\mathbf{y_f}, \mathbf{u_f}, \mathbf{x_f}, \boldsymbol{\psi}_\mathbf{f}) = \mathbf{b_f' y_f} + \frac{1}{2}\mathbf{u_f' V_f u_f} + \mathbf{x_f'}[\mathbf{A_f' y_f} + \mathbf{d_f} - \mathbf{p_f}] + \boldsymbol{\psi}_\mathbf{f}'[\mathbf{y_f} - \mathbf{y_f^o} - \mathbf{u_f}] \tag{2.40}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \mathbf{y_f}} = \mathbf{b_f} - \mathbf{A_f x_f} + \boldsymbol{\psi}_\mathbf{f} \geq 0 \tag{2.41}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \mathbf{u_f}} = \mathbf{V_f u_f} - \boldsymbol{\psi}_\mathbf{f} = 0 \tag{2.42}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \mathbf{x_f}} = \mathbf{A_f' y_f} + \mathbf{d_f} - \mathbf{p_f} \geq 0 \tag{2.43}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \boldsymbol{\psi}_\mathbf{f}} = \mathbf{y_f} - \mathbf{y_f^o} - \mathbf{u_f} = 0 \tag{2.44}$$

and again, by substituting (2.42) in (2.41) we can rewrite equation (2.41) as:

$$\frac{\partial \mathbf{LMIN_f}}{\partial \mathbf{y_f}} = \mathbf{A_f x_f} \leq \mathbf{b_f} + \mathbf{V_f u_f} \tag{2.45}$$

It is possible to combine the first-order conditions (2.36) and (2.45) to receive a system of the following equations:

$$\mathbf{A_f x_f} \leq \mathbf{b_f} + \mathbf{V_f u_f} \qquad [\mathbf{y_f}] \tag{2.46}$$

$$\mathbf{A_f' y_f} + \mathbf{W_f h_f} + \mathbf{d_f} \geq \mathbf{p_f} \quad [\mathbf{x_f}] \tag{2.47}$$

$$\mathbf{x_f} = \mathbf{x_f^o} + \mathbf{h_f} \qquad [\mathbf{W_f h_f}] \tag{2.48}$$

$$\mathbf{y_f} = \mathbf{y_f^o} + \mathbf{u_f} \qquad [\mathbf{V_f u_f}] \tag{2.49}$$

Paris (2017) shows that the Least Squares solution for equations (2.46) – (2.49) is unique. According to the Karush-Kuhn-Tucker conditions, condition (2.46) is an equality if the resource constraint is binding, and condition (2.47) is an equality if the respective activity level is positive. The objective of this first step is to generate estimates of the marginal cost levels $\mathbf{A}'_\mathbf{f}\hat{\mathbf{y}}_\mathbf{f} + \mathbf{W}_\mathbf{f}\hat{\mathbf{h}}_\mathbf{f} + \mathbf{d}_\mathbf{f}$ and of the input demand levels $\mathbf{A}_\mathbf{f}\hat{\mathbf{x}}_\mathbf{f}$. In a second phase Paris (2017) uses these estimated levels to estimate a nonlinear cost function, more specifically an extended Leontief specification of the following form:

$$C_f(\mathbf{x}_\mathbf{f}, \mathbf{y}_\mathbf{f}) = (\mathbf{g}'_\mathbf{f}\mathbf{y}_\mathbf{f})(\mathbf{l}'_\mathbf{f}\mathbf{x}_\mathbf{f}) + \tfrac{1}{2}(\mathbf{g}'_\mathbf{f}\mathbf{y}_\mathbf{f})\mathbf{x}'_\mathbf{f}\mathbf{Q}\mathbf{x}_\mathbf{f} + (\mathbf{l}'_\mathbf{f}\mathbf{x}_\mathbf{f})\left[\left(\mathbf{y}_\mathbf{f}^{\mathbf{1/2}}\right)'\mathbf{G}\mathbf{y}_\mathbf{f}^{\mathbf{1/2}}\right] \tag{2.50}$$

$\mathbf{l}_\mathbf{f}$ and $\mathbf{g}_\mathbf{f}$ may take on any value, but $\mathbf{l}'_\mathbf{f}\mathbf{x}_\mathbf{f}$ and $\mathbf{g}'_\mathbf{f}\mathbf{y}_\mathbf{f}$ must both be strictly positive. As in all non-linear models described above, the matrix $\mathbf{Q}$ is a symmetric, positive definite matrix of the dimension $I \times I$. The matrix $\mathbf{G}$ is of dimension $J \times J$ with the only restriction that all its off-diagonal elements be non-negative. Paris (2017) connects this nonlinear model with the estimates from the first phase as follows: first (equation 2.51), he derives the marginal costs of the Leontief specification (equation 2.50) and equates it to the estimated marginal costs of the linear model (equation 2.47). Then (equation 2.52), he calculates the input demand function by applying the Shephard lemma as the derivative of equation (2.50) with respect to the fixed-input price $\mathbf{y}_\mathbf{f}$ and equates it to the estimated fixed-input demand from the linear model (equation 2.46)[4]:

$$(\mathbf{g}'_\mathbf{f}\hat{\mathbf{y}}_\mathbf{f})\mathbf{l}_\mathbf{f} + (\mathbf{g}'_\mathbf{f}\hat{\mathbf{y}}_\mathbf{f})\mathbf{Q}\hat{\mathbf{x}}_\mathbf{f} + \mathbf{l}_\mathbf{f}\left[\left(\hat{\mathbf{y}}_\mathbf{f}^{\mathbf{1/2}}\right)\mathbf{G}\hat{\mathbf{y}}_\mathbf{f}^{\mathbf{1/2}}\right] = \mathbf{A}'_\mathbf{f}\hat{\mathbf{y}}_\mathbf{f} + \mathbf{W}_\mathbf{f}\hat{\mathbf{h}}_\mathbf{f} + \mathbf{d}_\mathbf{f} \tag{2.51}$$

$$(\mathbf{l}'_\mathbf{f}\hat{\mathbf{x}}_\mathbf{f})\mathbf{g}_\mathbf{f} + \tfrac{1}{2}\mathbf{g}_\mathbf{f}(\hat{\mathbf{x}}'_\mathbf{f}\mathbf{Q}\hat{\mathbf{x}}_\mathbf{f}) + (\mathbf{l}'_\mathbf{f}\hat{\mathbf{x}}_\mathbf{f})\Delta\left(\hat{\mathbf{y}}_\mathbf{f}^{-\mathbf{1/2}}\right)'\mathbf{G}\hat{\mathbf{y}}_\mathbf{f}^{\mathbf{1/2}} = \mathbf{A}_\mathbf{f}\hat{\mathbf{x}}_\mathbf{f} \tag{2.52}$$

Where $\Delta\left(\hat{\mathbf{y}}_\mathbf{f}^{-\mathbf{1/2}}\right)$ is a diagonal matrix with terms $\hat{\mathbf{y}}_\mathbf{f}^{-\mathbf{1/2}}$ on the main diagonal. To ensure convexity, the Cholesky factorization is applied on matrix $\mathbf{Q}$. The objective of equations (2.51) and (2.52) is to estimate parameters $\mathbf{l}_\mathbf{f}$, $\mathbf{g}_\mathbf{f}$, $\mathbf{Q}$ and $\mathbf{G}$, thanks to the estimates $\hat{\mathbf{x}}_\mathbf{f}$, $\hat{\mathbf{y}}_\mathbf{f}$ and $\hat{\mathbf{h}}_\mathbf{f}$ derived in the first phase[5]. In addition to the two equations depicted here Paris (2017) appends his model by four equations that allow for the incorporation of exogenous elasticities of supply in the model. Besides, he notes that matrices $\mathbf{Q}$ and $\mathbf{G}$ could be specified at farm, regional or even national level. Finally, he formulates the following calibrating equilibrium model:

$$min_{\mathbf{z}_{\mathbf{pf}},\mathbf{z}_{\mathbf{df}},\mathbf{x}_\mathbf{f},\mathbf{y}_\mathbf{f}} CSC_f = \boldsymbol{\rho}'_{\mathbf{pf}}\mathbf{y}_\mathbf{f} + \boldsymbol{\rho}'_{\mathbf{df}}\mathbf{x}_\mathbf{f} \tag{2.53}$$

subject to

$$(\hat{\mathbf{l}}'_\mathbf{f}\mathbf{x}_\mathbf{f})\hat{\mathbf{g}}_\mathbf{f} + \tfrac{1}{2}\hat{\mathbf{g}}_\mathbf{f}(\mathbf{x}'_\mathbf{f}\hat{\mathbf{Q}}\mathbf{x}_\mathbf{f}) + (\hat{\mathbf{l}}'_\mathbf{f}\mathbf{x}_\mathbf{f})\Delta\left(\mathbf{y}_\mathbf{f}^{-1/2}\right)'\hat{\mathbf{G}}\mathbf{y}_\mathbf{f}^{1/2} + \boldsymbol{\rho}_{\mathbf{pf}} = \mathbf{b}_\mathbf{f} + \mathbf{V}_\mathbf{f}\hat{\mathbf{u}}_\mathbf{f} \tag{2.54}$$

$$(\hat{\mathbf{g}}'_\mathbf{f}\mathbf{y}_\mathbf{f})\hat{\mathbf{l}}_\mathbf{f} + (\hat{\mathbf{g}}'_\mathbf{f}\mathbf{y}_\mathbf{f})\hat{\mathbf{Q}}\mathbf{x}_\mathbf{f} + \hat{\mathbf{l}}_\mathbf{f}\left[\left(\mathbf{y}_\mathbf{f}^{1/2}\right)\hat{\mathbf{G}}\mathbf{y}_\mathbf{f}^{1/2}\right] = \mathbf{p}_\mathbf{f} + \boldsymbol{\rho}_{\mathbf{df}} \tag{2.55}$$

[4] Note that the Shephard lemma is generally applied with respect to the prices of variable inputs and not fixed-inputs. However, Paris (2017, p. 24) calculates the vector of demand functions for fixed-inputs by deriving the cost function with respect to fixed-input shadow prices $\mathbf{y}$.

[5] Paris (2017) estimates the parameters of the nonlinear cost function by minimizing the sum of the square of two auxiliary slack variables. For reasons of brevity, this is not explicitly stated in the depicted model. See Paris (2017, p. 27) for details.

where slack-surplus variables for the primal constraints ($\boldsymbol{\rho}_{\mathbf{pf}}$) and for the dual constraints ($\boldsymbol{\rho}_{\mathbf{df}}$) are added. Note that this model now uses the estimates from the second phase (model (2.51) – (2.52)) to simulate activity levels and fixed-input prices, $\hat{\mathbf{x}}_{\mathbf{f}}$ and $\hat{\mathbf{y}}_{\mathbf{f}}$, respectively. Paris (2017) remarks that model (2.53) – (2.55) does no longer include matrix $\mathbf{A}$, whose presence rendered the relation between fixed-inputs and outputs fixed. He shows that the objective value of equation (2.53) reaches zero and the optimal values received for $\hat{\mathbf{x}}_{\mathbf{f}}$ and $\hat{\mathbf{y}}_{\mathbf{f}}$ are identical to the solutions of model (2.46) – (2.49), indicating that the model calibrates perfectly to observed values adjusted by the error terms $\hat{\mathbf{h}}_{\mathbf{f}}$ and $\hat{\mathbf{u}}_{\mathbf{f}}$. Model (2.51) – (2.53) may then be used for policy simulation purposes.

2.1.3.2.2. Incorporation of Uncertainty

The models developed and analysed in the later chapters of this thesis ignore how uncertainty (e.g., about prices) might affect farmers' decision making. In spite of the neglect of this important aspect in this work, the reader should be aware that various papers have been published in recent years showing how risk can be incorporated in a PMP or EMP framework (for example, Petsakos and Rozakis (2011), Severini and Cortignani (2012), Paris (2015), Arata *et al*. (2017), Britz and Arata (2019)).

These approaches typically replace the vector of prices $\mathbf{p}$ (or gross margins $\mathbf{gm}$, as for example in Severini and Cortignani (2012)) by a vector of *expected* prices $\boldsymbol{E}(\tilde{\mathbf{p}}_{\mathbf{f}})$ (or gross margins $\boldsymbol{E}(\widetilde{\mathbf{gm}}_{\mathbf{f}})$, respectively). Then, risk is introduced via a matrix of (co-)variances that determines the uncertainty about prices or gross margins. A setup that is frequently used to link these two elements is the so called "mean-variance" (or "E-V") approach that was mainly developed by Markowitz (1952) in the context of investment portfolio selection. The basic idea in a PMP context is to estimate cost function parameters as well as farmers' risk aversion by maximizing the total net revenue minus a risk component. The model developed by Arata *et al*. (2017) integrates risk into the setup conceptualized by Arfini and Donati (2011) that was already presented above (model (2.19) – (2.25)) and may therefore serve as a useful illustration. They first specify a minimization function as follows:

$$min_{\mathbf{u_f},\mathbf{y_f},\boldsymbol{\lambda}_\mathbf{f},\boldsymbol{\alpha}_\mathbf{f}} = \frac{1}{2}\mathbf{u}_\mathbf{f}'\mathbf{u}_\mathbf{f} + \mathbf{y}_\mathbf{f}'\mathbf{b}_\mathbf{f} + \mathbf{d}_\mathbf{f}'\mathbf{x}_\mathbf{f}^\mathbf{o} + \boldsymbol{\lambda}_\mathbf{f}'(\mathbf{x}_\mathbf{f}^\mathbf{o} + \boldsymbol{\varepsilon}) + \boldsymbol{\alpha}_\mathbf{f}\mathbf{x}_\mathbf{f}^{\mathbf{o}\prime}\mathbf{V}\mathbf{x}_\mathbf{f}^\mathbf{o} - \boldsymbol{E}(\tilde{\mathbf{p}}_\mathbf{f})'\mathbf{x}_\mathbf{f}^\mathbf{o} \qquad (2.56)$$

subject to

$$\mathbf{d}_\mathbf{f} + \boldsymbol{\alpha}_\mathbf{f}\mathbf{V}\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{A}_\mathbf{f}'\mathbf{y}_\mathbf{f} + \boldsymbol{\lambda}_\mathbf{f} \geq \mathbf{E}(\tilde{\mathbf{p}}_\mathbf{f}) \qquad (2.57)$$

$$\mathbf{d}_\mathbf{f} + \boldsymbol{\lambda}_\mathbf{f} = \mathbf{Q}\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{u}_\mathbf{f} \qquad (2.58)$$

with non-negativity constraints for $\mathbf{y}_\mathbf{f}$, $\boldsymbol{\lambda}_\mathbf{f}$ and $\boldsymbol{\alpha}_\mathbf{f}$. Matrix $\mathbf{V}$ is an exogenous $(I \times I)$ variance-covariance matrix of output prices, that is common to all farms and that is derived based on monthly crop price time series. The vector $\boldsymbol{\varepsilon}$ represents a perturbance that has the sole purpose to prevent linear dependency between the resource and the calibration constraints, as in the early PMP models[6]. $\boldsymbol{\alpha}_\mathbf{f}$ is a coefficient representing the farmer's absolute risk aversion. The concept of risk aversion plays a crucial role for the incorporation of uncertainty in behavioural models. According to classical theory of risk aversion as formalized by Pratt (1964) and Arrow (1965) an agent is risk averse if for every uncertain wealth the expected utility of this wealth is lower than the utility of the expected wealth. In other words: A risk averse agent prefers a lower expected profit with low variance over a higher expected profit with high variance. It can easily be seen from (2.56) that for a risk-neutral decision maker $\alpha_f = 0$, as the (co-)variances of activity prices do not influence her decision. A risk-averse individual, however, strives to minimize the variance of revenues, implying that $\boldsymbol{\alpha}_\mathbf{f} > \mathbf{0}$.[7] Constraint (2.55) states that marginal costs must be larger or equal to the (expected) marginal revenue and is equivalent to constraint (2.23) of the model by Arfini and Donati (2011) for a risk neutral agent. Constraint (2.58) connects the linear and the non-linear model and is equivalent to constraint (2.20). The estimates $\widehat{\mathbf{Q}}$, $\widehat{\mathbf{u}}_\mathbf{f}$ and $\widehat{\boldsymbol{\alpha}}_\mathbf{f}$ are then employed in a non-linear model that calibrates activity levels to the base year and can be used for simulation:

$$max_{\mathbf{x_f}} EU(\tilde{\pi})_f = \mathbf{p}_\mathbf{f}'\mathbf{x}_\mathbf{f} - \widehat{\mathbf{u}}_\mathbf{f}'\mathbf{x}_\mathbf{f} - \tfrac{1}{2}\mathbf{x}_\mathbf{f}'\widehat{\mathbf{Q}}\mathbf{x}_\mathbf{f} - \tfrac{1}{2}\hat{\alpha}_f\mathbf{x}_\mathbf{f}'\mathbf{V}\mathbf{x}_\mathbf{f} \quad (2.59)$$

subject to $$\mathbf{A}_\mathbf{f}\mathbf{x}_\mathbf{f} \leq \mathbf{b}_\mathbf{f} \qquad [\mathbf{y}_\mathbf{f}] \quad (2.60)$$

with non-negative activity levels $\mathbf{x}_\mathbf{f}$. Note that for a risk-neutral decision maker where $\hat{\alpha}_f = 0$ problem (2.59) – (2.60) will be equivalent to model (2.26) – (2.27) from Arfini and Donati (2011).

Paris (2015) incorporates risk in a PMP framework by setting up a model similar to the one that was already outlined in equations (2.53) – (2.55) (Paris, 2017). Different from Arata *et al.* (2017), he explicitly includes a coefficient capturing relative risk aversion, besides the coefficient for absolute risk aversion, as the exponents of the gross revenue and of the variance of the revenue, respectively. This allows the model to reflect all possible combinations of increasing, constant or decreasing absolute and relative risk aversion, whereas Arata *et al.* (2017) assumes constant absolute risk aversion. Finally, the model by Paris (2015) includes a wealth parameter $\mathbf{w}_\mathbf{f}^\mathbf{0}$ measuring exogenous income that does not depend on the produced activity levels. This feature makes it possible to capture, for example, the effect of decoupled direct payments on production decisions, which might be present under non-constant risk aversion.

[6] The authors do not provide any further interpretation and refer to the paper by Paris and Howitt (1998), where the perturbance term $\boldsymbol{\varepsilon}$ takes on an arbitrary small, but positive values, like in Howitt (1995).
[7] By imposing non-negativity for $\boldsymbol{\alpha}_\mathbf{f}$ Arata *et al.* (2017) rule out the possibility of risk-affinity.

2.1.4. Preliminary Conclusions

The previous sections demonstrate that the “original” PMP approach formalized by Howitt (1995) ignited the evolution of a strand of research with significant relevance for the development of agricultural policy simulation models. Actually, a vast array of relevant regional, national or international models for policy evaluation is at least partly based on PMP, among them CAPRI for the EU, DRAM for Canada, and REAP for the USA (Heckelei *et al*., 2012). While PMP was at the beginning mainly a calibration procedure that often relied on ad-hoc solutions (e.g., for the problem of underdetermination), with the time it evolved increasingly towards a framework for parameter estimation based on econometric assumptions and procedures. This development was initiated by Paris and Howitt (1998) who made it possible to consider more than one observation and who introduced GME as an econometric criterion. The cornerstone for *consistent* parameter estimation was then laid by Heckelei and Wolff (2003) whose paper certainly coins the PMP literature until today. However, it is noteworthy that, even though they illustrated how cost function parameters can be estimated in one single step, many researchers still employ a multi-step procedure (e.g., Arfini and Donati (2011), Paris (2015, 2017) or Arata *et al*. (2017)). They first set up the first-order conditions of a linear model and use them as the constraints of an error term minimization function. Arfini and Donati (2011) and Arata *et al*. (2017) estimate the non-linear cost function parameters already in this step by equating the marginal costs of the linear model with the marginal costs of the non-linear model, just like in Howitt (1995).[8] Paris (2015, 2017) estimates only the error terms for the activity levels and the fixed-input shadow prices in the first step and uses them to estimate the parameters of a Leontief cost function in a second step. He therefore also relies on the questionable assumption that the derivative of the costs with respect to the fixed-input shadow price corresponds to the fixed-input demand function.

Apart from the differences in estimation procedures, various interpretations of the error terms that are to be minimized exist: Paris (2015, 2017), for example, considers the error terms of the activity levels and the fixed-input shadow prices as positive or negative deviations of optimal activity levels or fixed-input prices from observed activity levels or fixed-input prices, respectively. This interpretation is equivalent to the one suggested by Heckelei and Wolff (2003) and also provides a coherent explanation for why the squares of the error terms should be minimized: If error terms are idiosyncratic and the cost function specification is correctly determined, the cost function parameters that require the smallest error terms to fulfil equations (2.44) – (2.47) for the sample observations should be the best approximation of the population’s cost function. Other researchers, such as Arfini and Donati (2011) or Arata *et al*. (2017), minimize the error terms $\mathbf{u_f}$ that in fact can be interpreted as the linear cost

[8] However, in opposite to the early PMP models they do not face the problem of an underdetermined model, as they estimate the cost function parameters with multiple observations and therefore should have sufficient degrees of freedom.

function parameter in the final model (see equation (2.20) and (2.58)). Furthermore, Arata *et al*. (2017) employ the "perturbance" terms $\boldsymbol{\varepsilon}$ that were used in the very early PMP specifications to avoid linear dependency between constraints. As argued by Heckelei and Wolff (2003) and later elaborated by Paris (2017), there should not be any necessity for this arbitrary, user determined and barely interpretable perturbances if one captures all deviations between optimal and observed activity levels in a vector of errors, in this thesis denoted as $\mathbf{h_f}$. While the discussion in this thesis focusses on methodologies conceptualized and applied in scientific literature, Heckelei *et al.* (2012) review some of the existing models that were built for repeated use in policy analysis. Interestingly, they conclude that many models still use a multi-step procedure with a linear model in the first phase, leading to biased shadow prices and myopic calibration.

2.2. A One-Step Cost Function Estimation and Calibration Procedure

The model specification that is further developed and tested in this thesis was conceptualized by Henry de Frahan (2019). It bases upon the work that was conducted by researchers in the field of PMP within the last three decades and bridges several of the conceptions that were presented in section 2.1, especially the contributions by Heckelei and Wolff (2003) and by Paris (2017). This approach allows to estimate in only one step the cost function parameters of a non-linear model that is able to reproduce base-year values without the need of a calibration constraint. Like suggested by Paris (2015, 2017) the first-order conditions of a primal maximization problem and of the respective dual minimization problem are used as the constraints of a (squared) error term minimization problem. However, an important distinction between the framework proposed by Henry de Frahan (2019) and many of the approaches published in recent years (Arata *et al*. (2017), Arfini and Donati (2011), Paris (2015, 2017)) is that no vestiges of a linear "first phase" model are required. The definition and interpretation of the error terms follows the stance taken by Heckelei and Wolff (2003). The observed values of the decision variables deviate from the optimal values by an idiosyncratic error term that may represent both measurement errors made by the econometrician as well as imperfect production decisions made by the farmer (see constraint 2.63). As in Paris (2015, 2017), this reasoning is also applied on shadow prices of the resource constraint, assuming that some exogenous, "observed" fixed-input prices are available. The resulting primal total net revenue maximization problem, following Henry de Frahan (2019) is as follows:

$$max_{\mathbf{x_f},\mathbf{h_f}} AUX_f = \mathbf{p_f'x_f} - \mathbf{c'x_f} - \tfrac{1}{2}\mathbf{x_f'Qx_f} - \tfrac{1}{2}\mathbf{h_f'W_fh_f} \qquad (2.61)$$

subject to

$$\mathbf{A_fx_f} \leq \mathbf{b_f} \qquad [\mathbf{y_f}] \qquad (2.62)$$

$$\mathbf{x_f} = \mathbf{x_f^o} + \mathbf{h_f} \qquad [\boldsymbol{\lambda}_\mathbf{f}] \qquad (2.63)$$

where activity levels $\mathbf{x_f}$ are non-negative. For better readability (and to remain consistent with the notation in the previous models) the index $\mathrm{f} = 1, \dots, \mathrm{F}$ is the only index explicitly stated, representing all farms in the sample. However, one must be aware that, when using panel data, all variables also have a dimension $\mathrm{t} = 1, \dots, \mathrm{T}$ representing the points in time (usually years) for which data is available. Using the same notation as above, the matrix $\mathbf{Q}$ is symmetric, (semi-) positive definite, of dimension $(I \times I)$ and represents the quadratic cost function parameters, while $\mathbf{c}$ is a vector of linear cost function parameters. An important feature of this model specification is that no vector of accounting costs $\mathbf{d_f}$ is introduced, as all costs that increase linearly in activity levels, whether they are explicit or implicit, are captured by the vector $\mathbf{c}$. As in Paris (2015, 2017), an auxiliary maximization function is formed (2.61) that strives to minimize the (squared) error terms $\mathbf{h_f}$. The diagonal weight matrix $\mathbf{W_f}$ is of dimension $(I \times I)$ and ensures consistency of the units of measurements. For each time and farm its diagonal elements are $w_{f,ii} = p_{f,i}/x^o_{f,i}$. The Lagrangian and the respective first order conditions are as follows:

$$LMAX_f(\mathbf{x_f}, \mathbf{h_f}, \mathbf{y_f}, \boldsymbol{\lambda}_\mathbf{f}) = \mathbf{p_f'x_f} - \mathbf{c'x_f} - \frac{1}{2}\mathbf{x_f'Qx_f} - \frac{1}{2}\mathbf{h_f'W_fh_f} + \mathbf{y_f'}[\mathbf{b_f} - \mathbf{A_fx_f}] + \boldsymbol{\lambda}_\mathbf{f}'[\mathbf{x_f^o} + \mathbf{h_f} - \mathbf{x_f}]$$
(2.64)

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{x_f}} = \mathbf{p_f} - \mathbf{c} - \mathbf{Qx_f} - \mathbf{A_f'y_f} - \boldsymbol{\lambda}_\mathbf{f} \leq 0 \qquad (2.65)$$

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{h_f}} = -\mathbf{W_fh_f} + \boldsymbol{\lambda}_\mathbf{f} = 0 \qquad (2.66)$$

$$\frac{\partial \mathbf{LMAX_f}}{\partial \mathbf{y_f}} = \mathbf{b_f} - \mathbf{A_fx_f} \geq 0 \qquad (2.67)$$

$$\frac{\partial \mathbf{LMAX_f}}{\partial \boldsymbol{\lambda}_\mathbf{f}} = \mathbf{x_f^o} + \mathbf{h_f} - \mathbf{x_f} = 0 \qquad (2.68)$$

And, by substituting (2.66) in (2.65):

$$\mathbf{p_f} \leq \mathbf{c} + \mathbf{Qx_f} + \mathbf{A_f'y_f} + \mathbf{W_fh_f} \qquad (2.69)$$

Note that the only – but fundamental – difference between constraints (2.65) – (2.68) and constraints (2.34) – (2.37) by Paris (2017) is that the marginal costs, which are on the right-hand side of (2.69), are the marginal costs of the final, non-linear model. Henry de Frahan (2019) then sets up the respective dual problem with an auxiliary minimization function containing the weighted squares of the errors $\mathbf{u_f}$ related to the resource prices as in Paris (2015, 2017):

$$min_{\mathbf{y_f},\mathbf{u_f}} AUX_f = \mathbf{b_f'y_f} + \frac{1}{2}\mathbf{x_f'Qx_f} + (\mathbf{x_f^o} + \mathbf{h_f})'\boldsymbol{\lambda}_\mathbf{f} + \frac{1}{2}\mathbf{u_f'V_fu_f} \qquad (2.70)$$

subject to

$$\mathbf{A_f'y_f} + \mathbf{c} + \mathbf{Qx_f} + \boldsymbol{\lambda}_\mathbf{f} \geq \mathbf{p_f} \qquad [\mathbf{x_f}] \qquad (2.71)$$

$$\mathbf{y_f} = \mathbf{y_f^o} + \mathbf{u_f} \qquad [\boldsymbol{\psi}_\mathbf{f}] \qquad (2.72)$$

where the matrix $\mathbf{V_f}$ is a diagonal weight matrix of dimension $(J \times J)$ with the elements $v_{f,jj} = b_{f,j}/y^{o}_{f,j}$. It is the equivalent of the matrix $\mathbf{W_f}$ in the primal problem and ensures consistency of the units of measurement. Then, the Lagrangian function and its first-order conditions are stated:

$$LMIN_f(\mathbf{y_f}, \mathbf{x_f}, \mathbf{u_f}, \boldsymbol{\psi}_\mathbf{f}) = \mathbf{b_f}'\mathbf{y_f} + \frac{1}{2}\mathbf{x_f}'\mathbf{Q}\mathbf{x_f} + (\mathbf{x_f^o} + \mathbf{h_f})'\boldsymbol{\lambda}_\mathbf{f} + \frac{1}{2}\mathbf{u_f'}\mathbf{V_f}\mathbf{u_f} + \mathbf{x_f}'[\mathbf{p_f} - \mathbf{c} - \mathbf{Q}\mathbf{x_f} - \mathbf{A_f'}\mathbf{y_f} - \boldsymbol{\lambda}_\mathbf{f}] + \boldsymbol{\psi}_\mathbf{f}'[\mathbf{y_f^o} + \mathbf{u_f} - \mathbf{y_f}] \tag{2.73}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \mathbf{y_f}} = \mathbf{b_f} - \mathbf{A_f}\mathbf{x_f} - \boldsymbol{\psi}_\mathbf{f} \geq \mathbf{0} \tag{2.74}$$

$$\frac{\partial \mathbf{LMIN_t}}{\partial \mathbf{x_t}} = \mathbf{p_t} - \mathbf{c} - \mathbf{Q}\mathbf{x_t} - \mathbf{A_t'}\mathbf{y_t} - \boldsymbol{\lambda}_\mathbf{t} \ \leq \ \mathbf{0} \tag{2.75}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \mathbf{u_f}} = \mathbf{V_f}\mathbf{u_f} + \boldsymbol{\psi}_\mathbf{f} = \mathbf{0} \tag{2.76}$$

$$\frac{\partial \mathbf{LMIN_f}}{\partial \boldsymbol{\psi}_\mathbf{f}} = \mathbf{y_f^o} + \mathbf{u_f} - \mathbf{y_f} = \mathbf{0} \tag{2.77}$$

substituting (2.76) in (2.74) we can rewrite (2.74) as:

$$\mathbf{b_f} \geq \mathbf{A_f}\mathbf{x_f} - \mathbf{V_f}\mathbf{u_f} \tag{2.78}$$

Equivalently to equations (2.46) – (2.49) as formulated by Paris (2017), Henry de Frahan (2019) sets up a system of four equations by combining the first-order conditions from the primal and the dual problems:

$$\mathbf{A_f'}(\mathbf{y_f^o} + \mathbf{u_f}) \ + \mathbf{W_f}\mathbf{h_f} + \mathbf{c} + \mathbf{Q}(\mathbf{x_f^o} + \mathbf{h_f}) \geq \mathbf{p_f} \tag{2.79}$$

$$\mathbf{A_f}(\mathbf{x_f^o} + \mathbf{h_f}) - \mathbf{V_f}\mathbf{u_f} \leq \mathbf{b_f} \tag{2.80}$$

$$\mathbf{x_f^o} + \mathbf{h_f} \geq \mathbf{0} \tag{2.81}$$

$$\mathbf{y_t^o} + \mathbf{u_t} \ \geq \mathbf{0} \tag{2.82}$$

Finally, the cost function parameters $\mathbf{c}$ and $\mathbf{Q}$ can be estimated by minimizing the squared errors $\mathbf{h_f}$ and $\mathbf{u_f}$ with the following objective function, subject to constraints (2.79) – (2.82):

$$min_{\mathbf{h_f},\mathbf{u_f},\mathbf{c},\mathbf{Q}} LS = \sum_{f=1}^{F} \frac{1}{2}(\mathbf{h_f'}\mathbf{W_f}\mathbf{h_f} + \mathbf{u_f'}\mathbf{V_f}\mathbf{u_f}) \tag{2.83}$$

It is crucial to recall that the objective of this minimization is the estimation of nonlinear cost function parameters, i.e., those parameters that the modeler believes to be best representing the true data generating process. In opposite to Paris (2017) and building upon the reasoning developed by Heckelei and Wolff (2003), no second phase is necessary.

Henry de Frahan (2019) then suggests to impose that $\mathbf{h_f}$ and $\mathbf{u_f}$ be equal to a vector of zero for a year t serving as the base year for calibration. Under this assumption it is possible to set up the following simulation model for each farm f in the reference year t, similar to model (2.53) – (2.55) by Paris (2017), using the cost function parameters $\hat{\mathbf{c}}$ and $\widehat{\mathbf{Q}}$ that were estimated for the entire sample:

$$min_{\boldsymbol{\rho}_{\mathbf{pf}}, \boldsymbol{\rho}_{\mathbf{df}}, \mathbf{x_f}, \mathbf{y_f}} CSC_f = \boldsymbol{\rho}'_{\mathbf{df}}\mathbf{x_f} + \boldsymbol{\rho}'_{\mathbf{pf}}\mathbf{y_f} \tag{2.84}$$

subject to

$$\mathbf{A}'_{\mathbf{f}}\mathbf{y_f} + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}\mathbf{x_f} = \mathbf{p_f} + \boldsymbol{\rho}_{\mathbf{df}} \tag{2.85}$$

$$\mathbf{A_f}\mathbf{x_f} + \boldsymbol{\rho}_{\mathbf{pf}} = \mathbf{b_f} \tag{2.86}$$

with non-negativity imposed for $\boldsymbol{\rho}_{\mathbf{pf}}$, $\boldsymbol{\rho}_{\mathbf{df}}$, $\boldsymbol{y}_{\mathbf{f}}$ and $\mathbf{x_f}$. As in equations (2.53) – (2.55), the vectors $\boldsymbol{\rho}_{\mathbf{pf}}$ and $\boldsymbol{\rho}_{\mathbf{df}}$ represent slack variables and are related to the primal and the dual constraints that take the value of zero if constraints (2.80) and (2.79) are binding, respectively. According to Henry de Frahan (2019), model (2.84) – (2.86) may then be used for policy simulation, by evaluating, e.g., the impact of an output price change or availability of limiting inputs. However, as we will show in section 3.3.1, it is not necessarily possible to simply equate the vectors $\mathbf{h_f}$ and $\mathbf{u_f}$ of the error terms to a vector of zeros for an arbitrary base year. Consequently, model (2.84) – (2.86) must be amended by incorporating the error terms $\mathbf{h_f}$ and $\mathbf{u_f}$, as it will be done in section 3.3.1.

3. Methodological Framework

Henry de Frahan (2019) conceptualizes a methodology to estimate and calibrate the parameters of a non-linear cost function in a one-step procedure, as we have seen in the previous chapter. This chapter prepares the methodological ground for the practical implementation of the approach with farm data in chapters 4 and 5. First, some general, methodological foundations of non-linear programming models that were touched upon in the previous chapters are outlined in more detail, such as the Karush-Kuhn-Tucker conditions and the Cholesky Factorization. Then, we will provide suggestions for amendments and extensions of the approach, notably a within transformation to capture fixed effects, the introduction of a time trend and a more complex cost function specification, based on a methodological note provided by Henry de Frahan for this thesis.

3.1. Methodological Foundations

3.1.1. Convexity of the Cost Function

As discussed in section 2.1.2.1, non-linearity of a PMP or EMP model requires that either the marginal revenue is decreasing in outputs or that marginal costs are increasing in outputs. In the approach that is scrutinized in this thesis non-linearity is on the cost side, requiring that the cost function be convex in output activity levels. Only then there is a unique maximum of the TNR for which marginal costs are equal to marginal revenue (see figure 1). Convexity of the cost function can be ensured by forcing the non-linear parameters of the matrix $\mathbf{Q}$ to be symmetric positive (semi)definite. One method to impose symmetric positive semi definiteness on a matrix that has frequently been used in the PMP literature is the Cholesky factorization, an approach that was first published by Benoit (1924). The rationale behind it is that a symmetric, positive semidefinite matrix $\mathbf{Q}$ may be decomposed in the product of a unit lower triangular matrix $\mathbf{L}$, its transposed, a unit upper triangular matrix $\mathbf{L}'$, and a diagonal matrix with nonnegative elements, $\mathbf{H}$:

$$\mathbf{Q} = \mathbf{LHL}' \tag{3.1}$$

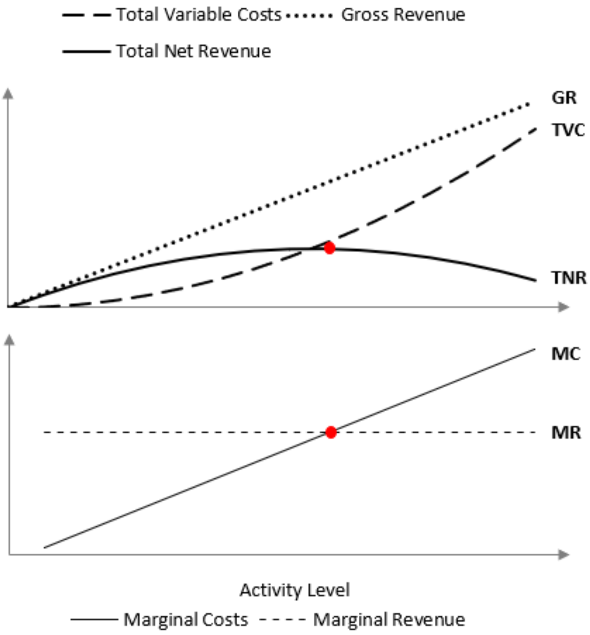

Figure 1: Relation between variable costs, gross revenue and total net revenue for a nonlinear, convex cost function. *The optimum corresponds to the activity level for which the TNR is maximized and MC=MR (red dot). Source: own visualization*

This means that, instead of estimating the $I(I+1)/2$ elements of matrix $\mathbf{Q}$ directly, the $I(I+1)/2 - I$ elements of $\mathbf{L}$ and the I elements of $\mathbf{H}$ are estimated, subject to non-negativity of the diagonal elements of $\mathbf{H}$. Therefore, we append model (2.79) – (2.83) by[9]:

$$\mathbf{Q} = \mathbf{LHL}' \tag{3.1}$$

$$l_{\mathrm{i,i}'} = 1 \qquad \forall\, i = i' \tag{3.2}$$

$$l_{i,i'} = 0 \qquad \forall\, i < i' \tag{3.3}$$

$$h_{i,i'} > 0 \qquad \forall\, i = i' \tag{3.4}$$

$$h_{i,i'} = 0 \qquad \forall\, i \neq i' \tag{3.5}$$

It should be noted that the Cholesky factorization can also be implemented without a diagonal matrix $\mathbf{H}$, but with the following constraints, instead:

$$\mathbf{Q} = \mathbf{LL}' \tag{3.6}$$

$$l_{i,i'} = 0 \qquad \forall\, i < i' \tag{3.7}$$

where the diagonal elements of $\mathbf{L}$ are free (note that equations (3.7) = (3.3)). This latter factorization was applied for example by Heckelei and Wolff (2003), whereas Paris and Howitt (1998) and Paris (2015, 2017) use the former factorization.

[9] We restate equation (3.1) for the sake of completeness.

3.1.2. The Karush-Kuhn-Tucker Conditions and Duality

Figure 1 visualizes the TNR for a linear revenue function and a convex cost function. The TNR maximizing activity levels can be found where marginal costs are equal to marginal revenues, in consistency with economic theory. In figure 1 we implicitly assume that no constraint exists that might impact the optimal activity level. Figure 2 depicts the same situation (1) but complements it by introducing a fixed-input constraint, e.g., a constraint for farmland, for situations (2) and (3). In situation (2) this constraint forces the farm to produce at an activity level for which $\mathbf{c} + \mathbf{Qx} < \mathbf{p_f}$. If the farm is endowed with one more unit of this limited fixed-resource, it could increase its TNR by $\mathbf{A_f^{-1}}(\mathbf{p_f} - \mathbf{c} - \mathbf{Q_x})$. The parameters of matrix $\mathbf{A_f}$ are coefficients that indicate how many units of this fixed-resource are required to produce an additional unit of the respective activity. We denote this implicitly derived value by which the TNR could be increased, the "shadow price" of the respective fixed-resource, with $\mathbf{y_f}$. In the cost function estimation framework by Henry de Frahan (2019) it is assumed that the elements of $\mathbf{y_f}$ must be "close" to the observed price of the fixed-resource $\mathbf{y_f^o}$, a feature that connects values endogenously derived in the model with exogenous information. Figure 2 also indicates why the shadow price is strictly positive only if the resource constraint is binding: In situation (3) a constraint exists and would allow production up to a point where MC ≥ MR, but it would be irrational for a farm to exploit this fixed-resource beyond the level for which MC = MR. An additional unit of this fixed-resource would be of no value to the farm, its shadow price $\mathbf{y_f}$ therefore must be zero.

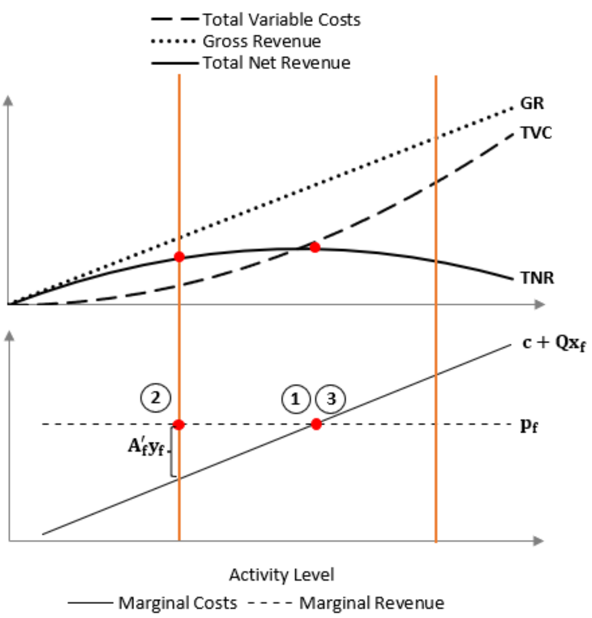

Figure 2: Impact of resource constraints on the optimal activity level and shadow price. *Situation (1) is equivalent to the simplified situation in figure 1 where constraints are ignored. Situation (2) shows the impact of a constraint at a level where MC<p (left orange line) and situation (3) the impact of a constraint at a level where MC>p. Source: own visualization*

Transferring the situation depicted in figure 2 to model (2.79) – (2.83), one can state that situation (2) indicates equality of constraint (2.80), whereas situation (3) indicates inequality.

However, in both cases, constraint (2.79) is an equality, as MC=MR in situations (2) and (3). This raises the question under which circumstances constraint (2.75) is an inequality. To answer it, it is useful to recall that increasing marginal costs urge farms represented in a nonlinear model to diversify their production portfolio – a fundamental property of PMP and the practical reason for convexity of the cost function. However, it is possible that, for some activities, marginal costs are larger than or equal to marginal revenues already for the first unit. As marginal costs increase whereas marginal revenues remain constant, a rational decision maker would not produce a single unit of the respective activity. This implies that, for a concave TNR function with a global maximum, either the derivative with respect to the activity level is zero or the activity level itself is zero. Therefore, the first-order condition can be transformed from an inequality into an equality as follows:

$$\frac{\partial \mathbf{TNR}}{\partial \mathbf{x}} = \mathbf{MR} - \mathbf{MC} \leq \mathbf{0} \rightarrow \mathbf{x}' \frac{\partial \mathbf{TNR}}{\partial \mathbf{x}} = \mathbf{x}'(\mathbf{MR} - \mathbf{MC}) = \mathbf{0} \quad (3.8)$$

where $\mathbf{MR}$ and $\mathbf{MC}$ represent vectors of the activities' marginal revenues and the marginal costs, respectively. This theorem was discovered originally by Karush (1939) and then published by Kuhn and Tucker (1951), which is why it is often referred to as (one of) the Karush-Kuhn-Tucker (KKT) conditions (see for example Paris (2010). Making use of these KKT conditions, constraint (2.79) can be rewritten as

$$(\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{h}_\mathbf{f})(\mathbf{p}_\mathbf{f} - \mathbf{A}'_\mathbf{f}(\mathbf{y}_\mathbf{f}^\mathbf{o} + \mathbf{u}_\mathbf{f}) - \mathbf{W}_\mathbf{f}\mathbf{h}_\mathbf{f} - \mathbf{c} - \mathbf{Q}(\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{h}_\mathbf{f})) = \mathbf{0} \qquad (2.79')$$

Equivalently, and by applying the reasoning depicted in figure 2, the resource constraint (2.80) may be restated as

$$(\mathbf{y}_\mathbf{f}^\mathbf{o} + \mathbf{u}_\mathbf{f})(\mathbf{b}_\mathbf{f} - \mathbf{A}_\mathbf{f}(\mathbf{x}_\mathbf{f}^\mathbf{o} + \mathbf{h}_\mathbf{f}) + \mathbf{V}_\mathbf{f}\mathbf{u}_\mathbf{f}) = \mathbf{0} \qquad (2.80')$$

Thus, the KKT conditions allow to write the constraints in a way that they are correct regardless of whether they are binding or not. If one is willing to assume that all resource constraints are binding and all activity levels are strictly positive, constraints (2.79) and (2.80) are equalities, without the need to employ the KKT conditions. Paris (2017) concludes that for his first phase model both constraints can always be written as equalities: The resource constraint is binding by construction, because the fixed-input-output coefficients of matrix $\mathbf{A}_\mathbf{f}$ are calculated as b_f/x_f^o. As we will see in chapter 3.5.1, this is also the case when implementing the model by Henry de Frahan (2019). However, a fundamental difference between the models conceptualized by Paris (2017) and Henry de Frahan (2019) concerns the constraint that MC must be larger than or equal to MR. To see it, it is important to note that both approaches hinge upon the following assumption: The error terms $h_{f,i}$ and $u_{f,i}$ must be zero if $x_{f,i}^o$ and $y_{f,i}^o$ are zero, respectively, and the sum of $x_{f,i}^o$ and $y_{f,i}^o$ as well as the sum of $y_{f,i}^o$ and $u_{f,i}$ must be strictly positive, if $x_{f,i}^o$ and $y_{f,i}^o$ are strictly positive, respectively. In other words, the “true” or “optimal” activity level or fixed-input price is strictly positive if the observed activity level or fixed-input price is strictly positive, and vice versa. The intuition behind this assumption is particularly clear if the discrepancy between observed and “true” activity levels or fixed-input prices can be largely explained by measurement errors made by the analyst. It is likely that the observed value is an under- or overestimation of the “true” value, but it seems unreasonable that a strictly positive level is measured if the true level is zero or vice versa. Therefore, constraints (2.81) and (2.82) may be stated more precisely as:

$$\mathbf{x^o_{f,i}} + \mathbf{h_{f,i}} > \mathbf{0} \quad \forall\, x^o_{f,i} > 0 \quad \text{and} \quad \mathbf{x^o_{f,i}} = \mathbf{h_{f,i}} = \mathbf{0} \quad \forall\, x^o_{f,i} = 0 \qquad (2.81')$$

$$\mathbf{y^o_{f,j}} + \mathbf{u_{f,j}} > \mathbf{0} \quad \forall\, y^o_{f,j} > 0 \quad \text{and} \quad \mathbf{y^o_{f,j}} = \mathbf{u_{f,j}} = \mathbf{0} \quad \forall\, y^o_{f,j} = 0 \qquad (2.82')$$

Where indices i and j indicate that this is valid for each activity i and each fixed-input j. These rules that define under which circumstances $\mathbf{h_f}$ and $\mathbf{u_f}$ have strictly positive elements are also valid for model (2.46) – (2.49) by Paris (2017). Following the reasoning given above, MC can only be larger than MR if the respective activity level is zero, in all other cases they must be equal. However, constraint (2.81’) helps understand why, for the model of Paris (2017), MC equate to MR under all circumstances. If the observed level $x^o_{f,i}$ of an activity is zero, there is no information on the respective element of the fixed-input-output coefficient $a_{f,j,i}$, the unit accounting costs $d_{f,i}$, or the price $p_{f,i}$[10]. Consequently, these coefficients will take on the null value. Since $h_{f,i}$ is zero, the left-hand side and the right-hand side will both be zero, meaning that MC always equal MR and constraint (2.45) is in both cases an equality:

$$\sum_{j=1}^{J}\left[a'_{f,j,i}(y^o_{f,j} + u_{f,j})\right] + W_{f,i}h_{f,i} + d_{f,i} = p_{f,i} \qquad \forall\, x^o_{f,i} > 0 \qquad (2.47a)$$

$$\sum_{j=1}^{J}\left[0(y^o_{f,j} + u_{f,j})\right] + 0 + 0 = 0 \qquad \forall\, x^o_{f,i} = 0 \qquad (2.47b)$$

It is important to understand why this is not the case for the model developed by Henry de Frahan (2019): The estimate for the linear cost function parameter $\mathbf{c}$, that is estimated over all observations in the sample, constitutes a summand of the marginal costs and is on the left-hand side of constraint (2.75):

$$\sum_{j=1}^{J}(a'_{f,j,i}(y^o_{f,j} + u_{f,j})) + W_{f,i}h_{f,i} + \mathbf{c} + \mathbf{Q}\left(x^o_{f,i} + h_{f,i}\right) = p_{f,i} \quad \forall\, x^o_{f,i} > 0 \qquad (2.79a)$$

$$\sum_{j=1}^{J}(0(y^o_{f,j} + u_{f,j})) + 0 + \mathbf{c} + \mathbf{Q}(0 + 0) \geq 0 \quad \forall\, x^o_{f,i} = 0 \qquad (2.79b)$$

Constraint (2.79) must be fulfilled for all farms and years in the sample and for each activity. If for only one observation in the sample an observed activity level was 0, this would force the linear cost function parameter $\mathbf{c}$ to be positive, as (2.79b) shows. Besides, if information on activity prices is not available at farm level but only at a higher geographic scale, it is hard

[10] Paris (2017, p. 32) sets $p_{f,i} = 0$ if there is no information about non-produced crops.

to justify why $p_{f,i}$ should be 0 if $x_{f,i}^{o} = 0$. If a reasonable value is available for $p_{f,i}$ even if the respective activity level is 0, this would imply that (2.79b) becomes:

$$c_i \geq p_{f,i} \qquad \forall\, x_{f,i}^{o} = 0 \qquad \text{(2.79b')}$$

Allowing this would significantly impact the result, as the estimate $\hat{\mathbf{c}}$ would have to be at least as large as the highest output price of all activities in the sample for which a level of zero is observed. Consequently, by applying the KKT conditions and transforming the inequalities into equalities as in constraint (2.79'), this problem can be evaded when estimating the cost function parameters of the model by Henry de Frahan (2019).

3.2. Introduction of a Fixed Effect and a Time-Varying Component

The purpose of the EMP model developed by Henry de Frahan (2019) is to receive consistent estimates for the parameters of the "true" cost function, i.e., the cost function that describes the costs that the population of farms in the respective region faces. By using (unbalanced) panel data, one can capitalize on information from multiple years. So far, we have assumed that the objective function (2.83) can be minimized subject to constraints (2.79) – (2.82) for all farms in all available years at once. Hence, F farms in T years would be interpreted as $F * T$ independent observations with error terms that are minimized in one process, using pooled OLS. However, this only produces consistent estimates for the cost function parameters if there is no unobserved effect that is constant over time, specific for each unit (e.g., farm) and correlated with an explanatory variable (Wooldridge, 2016, p. 413). This unobserved effect would manifest itself as part of the activity level and fixed-input price-related error terms, so that, for example:

$$\mathbf{h_{t,f}} = \boldsymbol{\varepsilon}_{\mathbf{t,f}} + \boldsymbol{\mu}_{\mathbf{f}} \qquad \text{(3.9)}$$

where $\boldsymbol{\varepsilon}_{\mathbf{t,f}}$ are vectors of idiosyncratic error terms and $\boldsymbol{\mu}_{\mathbf{f}}$ indicate vectors of unobserved components that remain constant over time but vary between farms. It is reasonable to assume that a farm's cost function is influenced by geographic features such as soil quality or altitude, by the skills and education of the manager, or by any other factors that are constant through time but not explicitly included in the model. Those factors are captured by the farm fixed effect $\boldsymbol{\mu}_{\mathbf{f}}$. If there is reason to believe that $\boldsymbol{\mu}_{\mathbf{f}}$ might be different from 0, pooling all observations and minimizing their squared error terms would therefore lead to inconsistent parameter estimates. A potential solution is the so-called random effects estimator. However, it only is efficient if $\boldsymbol{\mu}_{\mathbf{f}}$ is not correlated with any explanatory variables (ibid.). In the case of farm models, it seems more likely that activity levels are correlated, e.g., with soil conditions or market access[11]. For example, dairy production activities seem to be correlated with land

[11] A classical theory that connects farmers' land use decisions with distance from a central market place is von Thünen's location theory (1842).

characteristics, sunk costs linked to buildings and equipment, available forage area and spatial agglomerations (Ievoli *et al*. (2017), Mosnier and Wieck (2010)). Therefore, a more adequate approach may be a within transformation: By subtracting the individual mean over time of each variable and error term from the annual value of the variables and error terms, the term $\mathbf{\mu_f}$ is eliminated, being constant in time. To incorporate the within transformation in the model conceptualized by Henry de Frahan (2019), constraint (2.79) is rewritten as a time-demeaned constraint for each farm f:

$$\widetilde{\mathbf{Ay}^{\circ}}_{\mathbf{t}} + \widetilde{\mathbf{Au}}_{\mathbf{t}} + \widetilde{\mathbf{Wh}}_{\mathbf{t}} + \mathbf{Q}\left(\tilde{\mathbf{x}}_{\mathbf{t}}^{\mathbf{o}} + \tilde{\mathbf{h}}_{\mathbf{t}}\right) = \tilde{\mathbf{p}}_{\mathbf{t}} \qquad (2.79\_FE)$$

subject to

$$\overline{\mathbf{Wh}} = \frac{1}{\mathbf{T}}\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{W_t h_t}$$

$$\overline{\mathbf{Au}} = \frac{1}{\mathbf{T}}\sum_{\mathbf{t=1,j=1}}^{\mathbf{T,J}} (\mathbf{A}'_{\mathbf{t,j}}\mathbf{u_t})$$

$$\bar{\mathbf{h}} = \frac{1}{\mathbf{T}}\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{h_t}$$

where tilde (~) indicates the demeaned value, thus the value for the respective year subtracted by the farm mean over all years (see Annex II for more details). Note that the vector $\mathbf{c}$ of linear cost parameters has vanished from equation (2.79_FE), because its parameters are fixed over time, and unlike the coefficients of matrix $\mathbf{Q}$, they are not multiplied by any variable. However, after having received the estimates $\hat{\mathbf{h}}$, $\hat{\mathbf{u}}$ and $\widehat{\mathbf{Q}}$, the vector $\mathbf{c}$ can be recovered as follows:

$$\hat{\mathbf{c}} = \overline{\mathbf{p}} - \left[\overline{\mathbf{A}}'_{\mathbf{j}}\left(\overline{\mathbf{y}}_{\mathbf{j}}^{\mathbf{o}} + \bar{\hat{\mathbf{u}}}_{\mathbf{j}}\right) + \overline{\mathbf{Wh}} + \widehat{\mathbf{Q}}\left(\overline{\mathbf{x}^{\mathbf{o}}} + \bar{\hat{\mathbf{h}}}\right)\right] \qquad (3.10)$$

It is important to note that the parameters of the linear cost function are now farm dependent, even though they remain constant over time. Writing out all subscripts, $\hat{c}_i$ (the *i*th element of a $(I \times 1)$ vector) now becomes $\hat{c}_{f,i}$ (an element of a $(F \times I)$ matrix). The estimates of vectors $\hat{\mathbf{h}}$, $\hat{\mathbf{u}}$, $\hat{\mathbf{c}}$ and of matrix $\widehat{\mathbf{Q}}$ may now be used in model (2.84) – (2.86) for simulation purposes as before, with the only difference that the estimated vectors $\hat{\mathbf{c}}$ are now farm-individual.

The within transformation is useful if differences between farms are presumed to be non-variant in time, thus, if a farm fixed effect exists. However, regardless of whether there is a relevant unobserved, farm specific component or not, it seems conceivable that some parameters change over time, such as technology. Frick and Sauer (2021), for example, measure a technological change of on average 1% per year for German dairy farms from 1995-2013. This can be captured by model (2.79) – (2.83) by including a vector $\mathbf{t}$ of time-varying components that take the values $t = 1, \dots, T$ in equation (2.79), for each farm f:

$$\mathbf{A}'_{\mathbf{t}}(\mathbf{y}_{\mathbf{t}}^{\mathbf{o}} + \mathbf{u_t}) + \mathbf{W_t h_t} + \mathbf{c_0} + \mathbf{c}'_{\mathbf{1}}\mathbf{t} + \mathbf{Q}(\mathbf{x}_{\mathbf{t}}^{\mathbf{o}} + \mathbf{h_t}) \geq \mathbf{p_t} \qquad (2.79\_t)$$

where we decompose the vector $\mathbf{c}$ in a $(I \times 1)$ vector $\mathbf{c_0} + \mathbf{c}'_{\mathbf{1}}\mathbf{t}$ where $\mathbf{c}'_{\mathbf{1}}\mathbf{t}$ may capture a linear time trend. If it appears reasonable to combine a farm fixed effect and a time trend, equation (2.79_FE) can be modified as follows, for each farm f:

$$\mathbf{c_1}\tilde{\mathrm{t}} + \widetilde{\mathbf{Ay}^{\circ}}_{\mathbf{t}} + \widetilde{\mathbf{Au}}_{\mathbf{t}} + \widetilde{\mathbf{Wh}}_{\mathbf{t}} + \mathbf{Q}(\tilde{\mathbf{x}}_{\mathbf{t}}^{\mathbf{o}} + \tilde{\mathbf{h}}_{\mathbf{t}}) = \tilde{\mathbf{p}}_{\mathbf{t}} \qquad \text{(2.79\_FE\_t)}$$

where $\tilde{t}$ is the time-demeaned time index and $\boldsymbol{\vartheta}\tilde{t}$ can be interpreted like a component of the linear cost function parameters of the vector $\mathbf{c}$ that varies in time and is farm-independent, as in model (2.79_t).

3.3. Validation and Simulation

3.3.1. Minimization of the Complementary Slackness Conditions

After having solved model (2.79) – (2.83) the cost function parameter estimates $\hat{\mathbf{c}}$ and $\widehat{\mathbf{Q}}$ may then be used for market or policy analysis. Under the assumption that the model reflects the true data generating process adequately, $\hat{\mathbf{c}}$ and $\widehat{\mathbf{Q}}$ provide sufficient information to simulate farms' responses to a change in output prices, coupled subsidies, or constraints. Paris (2017) suggests to implement this via a model that minimizes the complementary slackness conditions (CSC) of the resource constraint and the constraint that MC should be equal to MR (model (2.53) – (2.55)). Henry de Frahan (2019) proposes a similar approach (model (2.84) – (2.86)). The intuition behind this procedure is, for the example of an output price change, that farms will adjust their production so that MR again equals MC for every activity while complying with all resource constraints. Consequently, when keeping all variables unchanged compared to a reference year t_{ref} that is used for parameter estimation, the estimated, CSC minimizing output levels and fixed-input shadow prices are expected to be $\hat{\mathbf{x}}_{\mathbf{t_{ref}}} = \mathbf{x}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{h}}_{\mathbf{t_{ref}}}$ and $\hat{\mathbf{y}}_{\mathbf{t_{ref}}} = \mathbf{y}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{u}}_{\mathbf{t_{ref}}}$. In other words, the model is expected to calibrate perfectly to the optimal reference year values. From now on we refer to this phase in which we validate whether the model replicates the variable levels of the reference year as the "validation" phase. The phase in which the same model is applied with different exogenous variables is referred to as the simulation phase. However, we argue that the model suggested by Henry de Frahan (2019) (model 2.84 – 2.86) must be modified slightly before being able to perform this phase. The reason for this is that Henry de Frahan (2019) imposes the vectors $\mathbf{h_t}$ and $\mathbf{u_t}$ to be zero for the reference year t_{ref}. Therefore, unless the estimates $\hat{\mathbf{h}}_{\mathbf{t}}$ and $\hat{\mathbf{u}}_{\mathbf{t}}$ truly are zero, model (2.84) – (2.86) does not yield activity levels and fixed-input shadow prices that are equal to $\mathbf{x}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{h}}_{\mathbf{t_{ref}}}$ or $\mathbf{y}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{u}}_{\mathbf{t_{ref}}}$, respectively. To show this, let us recall that with the estimated error terms and cost function parameters constraint (2.79) of the estimation phase is an equality for positive activity levels in the year taken as the reference for the validation, for each farm:

$$\mathbf{A}'_{\mathbf{t_{ref}}}(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{u}}_{\mathbf{t_{ref}}}) + \mathbf{W}_{\mathbf{t_{ref}}}\hat{\mathbf{h}}_{\mathbf{t_{ref}}} + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}(\mathbf{x}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{h}}_{\mathbf{t_{ref}}}) = \mathbf{p}_{\mathbf{t_{ref}}} \qquad (3.11)$$

Equation (2.85) of the validation phase states the following:

$$\mathbf{A}'_{\mathbf{t_{ref}}}\mathbf{y}_{\mathbf{t_{ref}}} + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}\mathbf{x}_{\mathbf{t_{ref}}} - \mathbf{z}_{\mathbf{dt_{ref}}} = \mathbf{p}_{\mathbf{t_{ref}}} \qquad (3.12)$$

Consequently, if the validation truly yields $\hat{\mathbf{x}}_{\mathbf{t_{ref}}} = \mathbf{x}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{h}}_{\mathbf{t_{ref}}}$ and $\hat{\mathbf{y}}_{\mathbf{t_{ref}}} = \mathbf{y}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{u}}_{\mathbf{t_{ref}}}$ we can equate (2.79) and (2.85) and write it as follows:

$$\begin{aligned}
&\mathbf{A}'_{\mathbf{t_{ref}}}\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{u}}_{\mathbf{t_{ref}}}\right) + \mathbf{W}_{\mathbf{t_{ref}}}\hat{\mathbf{h}}_{\mathbf{t_{ref}}} + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}\left(\mathbf{x}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{h}}_{\mathbf{t_{ref}}}\right) \\
&= \mathbf{A}'_{\mathbf{t_{ref}}}\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{u}}_{\mathbf{t_{ref}}}\right) + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}\left(\mathbf{x}^{\mathbf{o}}_{\mathbf{t_{ref}}} + \hat{\mathbf{h}}_{\mathbf{t_{ref}}}\right) - \boldsymbol{\rho}_{\mathbf{dt_{ref}}} \\
&\rightarrow \mathbf{W}_{\mathbf{t_{ref}}}\hat{\mathbf{h}}_{\mathbf{t_{ref}}} = -\boldsymbol{\rho}_{\mathbf{dt_{ref}}} \qquad (3.13)
\end{aligned}$$

First, since $\boldsymbol{\rho}_{\mathbf{dt_{ref}}}$ is a vector of slack variables that must not take any negative values, (3.13) is insolvable if elements of $\hat{\mathbf{h}}_{\mathbf{t_{ref}}} > \mathbf{0}$, since $\mathbf{W}_{\mathbf{t_{ref}}}$ is always positive. Second, as the objective function (2.84) attempts to minimize the slack variable $\boldsymbol{\rho}_{\mathbf{dt_{ref}}}$, it is reasonable that model (2.84) – (2.86) finds values for $\mathbf{x}_{\mathbf{t_{ref}}}$ and $\mathbf{y}_{\mathbf{t_{ref}}}$ that will lead to $\left|\boldsymbol{\rho}_{\mathbf{dt_{ref}}}\right| < \mathbf{W}_{\mathbf{t_{ref}}}\hat{\mathbf{h}}_{\mathbf{t_{ref}}}$. Therefore, let us reformulate the CSC minimization problem so that the optimal solution is the "true" activity level or fixed-input shadow price, i.e., the error term received in the estimation phase added to the observed value. This can be done by simply restating the two constraints (2.79) and (2.80) of the estimation phase as an equality with slack variables, with all estimates $\hat{\mathbf{h}}_{\mathbf{f}}$, $\hat{\mathbf{u}}_{\mathbf{f}}$, $\hat{\mathbf{c}}$ and $\widehat{\mathbf{Q}}$ incorporated. Being a modification of the original CSC minimization model (2.84) – (2.86), equations (3.14) and (3.15) display constraints that are applied for the reference year t_{ref} only:

$$\mathbf{A}'_{\mathbf{f}}(\mathbf{y}_{\mathbf{f}}) + \mathbf{W}_{\mathbf{f}}\hat{\mathbf{h}}_{\mathbf{f}} + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}(\mathbf{x}_{\mathbf{f}}) = \mathbf{p}_{\mathbf{f}} + \boldsymbol{\rho}_{\mathbf{df}} \qquad (3.14)$$

$$\mathbf{A}_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}) - \mathbf{V}_{\mathbf{f}}\hat{\mathbf{u}}_{\mathbf{f}} + \boldsymbol{\rho}_{\mathbf{pf}} = \mathbf{b}_{\mathbf{f}} \qquad (3.15)$$

Together with the objective function (2.84), constraints (3.14) and (3.15) constitute a model that can be used to estimate $\mathbf{x}_{\mathbf{f}}$ and $\mathbf{y}_{\mathbf{f}}$ both to validate the results and to simulate a policy or price change. Note that for validation we expect the elements of $\boldsymbol{\rho}_{\mathbf{df}}$ and $\boldsymbol{\rho}_{\mathbf{pf}}$ to be zero if (2.79) and (2.80) of the estimation phase are equalities. If a time index is added in the estimation phase, instead of $\hat{\mathbf{c}}$ constraint (3.10) must contain the decomposition $\hat{\mathbf{c}}_{\mathbf{0}} + \hat{\mathbf{c}}'_{\mathbf{1}}\mathbf{t}$, as in (2.79_t).

As the subscripts of the objective function (2.84) indicate, the CSC are minimized for each farm individually. For some policy scenarios, like the one implemented in chapter 5, it might be necessary to impose constraints that encompass a group of farms, for example, all farms in a region or subregion. Besides, so far, we assume that each of the J resources is fixed at farm level. While this might be a reasonable assumption especially in the short run, there are certainly scenarios that would become more realistic if farmland, for example, could be exchanged among farmers. However, this necessitates that the resource constraint (3.15) is imposed at the respective geographic level and that some link between the resource requirements of each of the F individual farms is established. Fortunately, the CSC minimization model can easily be modified to accommodate multiple farms at once, by modifying the objective function (2.84) accordingly:

$$min_{\boldsymbol{\rho}_{\mathbf{pf}},\boldsymbol{\rho}_{\mathbf{df}},\mathbf{x}_{\mathbf{f}},\mathbf{y}_{\mathbf{f}}} CSC = \sum_{f=1}^{F}\left[\boldsymbol{\rho}'_{\mathbf{df}}\mathbf{x}_{\mathbf{f}} + \boldsymbol{\rho}'_{\mathbf{pf}}\mathbf{y}_{\mathbf{f}}\right] \qquad (3.16)$$

subject to constraints (3.14) and (3.15) if the resource constraint persists at farm level. Section 5.2 provides an example of how a resource constraint can be imposed at a higher geographic scale. The objective function (3.16) with the constraints (3.14) and (3.15) yields the same estimates for activity levels and fixed-input prices as objective function (2.84) This is because, as long as resources are confined at farm level, each farm allocates its own fixed-inputs to produce outputs up to a level for which MC equate to MR.

3.3.2. Maximization of the Total Net Revenue

The interpretation of the objective value of a CSC minimization model and, consequently, the interpretation of the shadow prices of the constraints, is not as obvious as in the case of a TNR maximization model. The slack variables $\boldsymbol{\rho}_{\mathbf{df}}$ and $\boldsymbol{\rho}_{\mathbf{pf}}$ indicate for each farm, how much larger the marginal costs are than the marginal revenue and how much more of a resource is available than what is used by the farm, respectively (see equations (3.14) – (3.15). By multiplying $\boldsymbol{\rho}_{\mathbf{df}}$ with the vector of activity levels and $\boldsymbol{\rho}_{\mathbf{pf}}$ with the vector of shadow prices of the resources in the objective function (2.84), the CSC indicate the loss that a farm incurs by deviating from optimality conditions. This is the reason why we would expect the CSC to be zero for validation, because the estimates $\hat{\mathbf{c}}$, $\widehat{\mathbf{Q}}$, $\hat{\mathbf{h}}_{\mathbf{f}}$ and $\hat{\mathbf{u}}_{\mathbf{f}}$ are estimated subject to the constraint that all optimality conditions be fulfilled. If an output price changes or a new constraint is introduced, the CSC would not indicate how this impacts the total net revenue. It merely presents a mean to mathematically express the objective of finding activity levels and shadow prices for which the optimality conditions are fulfilled as good as possible. A more intuitive approach might be to use the estimated cost function parameters and error terms to set up the cost function of a farm TNR maximization problem. To ensure that CSC minimization and TNR maximization yield the same solution, the TNR maximization should represent the primal problem whose first order conditions correspond to constraints (3.14) and (3.15) of the CSC minimization problem. As the CSC minimization, it is applied for the reference year t_{ref}:

$$max_{\mathbf{x}_{\mathbf{f}}} TNR_f = \mathbf{p}_{\mathbf{f}}{}'\mathbf{x}_{\mathbf{f}} - \hat{\mathbf{c}}'\mathbf{x}_{\mathbf{f}} - \frac{1}{2}\mathbf{x}_{\mathbf{f}}{}'\widehat{\mathbf{Q}}\,\mathbf{x}_{\mathbf{f}} - \mathbf{x}_{\mathbf{f}}{}'\mathbf{W}_{\mathbf{f}}\hat{\mathbf{h}}_{\mathbf{f}} \qquad (3.17)$$

subject to $$\mathbf{A}'_{\mathbf{f}}\mathbf{x}_{\mathbf{f}} - \mathbf{V}_{\mathbf{f}}\hat{\mathbf{u}}_{\mathbf{f}} \leq \mathbf{b}_{\mathbf{f}} \qquad [\mathbf{y}_{\mathbf{f}}] \qquad (3.18)$$

and subject to $\mathbf{x}_{\mathbf{f}} \geq \mathbf{0}$. Model (3.17) – (3.18) represents the problem that the farmer faces, where the only decision variable are the activity levels $\mathbf{x}_{\mathbf{f}}$. Note that, instead of having two equations with slack variables to minimize, as in model (3.14) – (3.15), the derivatives of model (3.17) – (3.18) are inequalities to which the KKT conditions may be applied to transform them into equalities. The vector $\mathbf{y_f}$ implicitly includes all shadow prices of the resource constraint. As we will see in chapter 5, an appealing feature of this simulation model is that it can be complemented by other constraints whose shadow prices can be interpreted intuitively: It is the value that the farm manager would be willing to pay to loosen the constraint by one unit. Like the CSC minimization model, the TNR maximization model can be adapted to maximize the TNR of multiple farms at once, by simply summing the objective function over all respective farms (as in objective function 3.16).

3.4. Monte Carlo Evidence for the Cost Function Estimation

The incorporation of econometrics into the traditional PMP framework entails the notion of a "true" model whose parameters can be *estimated* rather than calibrated, based on a representative sample. In their pioneering paper, Heckelei and Wolff (2003) illustrate with Monte Carlo simulations that their GME-based estimation approach truly yields consistent cost function parameters estimates for the non-linear parameters of matrix $\mathbf{Q}$. They first define a matrix $\mathbf{Q}$ of "true" parameters. Then, they generate a dataset with F random observations for gross margins and land availability and calculate the corresponding optimal activity levels. They add normally distributed errors to the activity levels and then estimate the parameters of matrix $\mathbf{Q}$. They repeat this procedure for increasing sample sizes one thousand times per sample size and show that the estimation error decreases in sample size.

Before applying the model conceptualized by Henry de Frahan (2019) with some of the amendments introduced in the methodology chapter to real farm data, we test it with a Monte Carlo simulation, similar to the setup by Heckelei and Wolff (2003). The obvious advantage over real world observations is that the "true" cost function parameters of the vector $\mathbf{c}$ and the matrix $\mathbf{Q}$ as well as the vectors $\mathbf{h}$ and $\mathbf{u}$ of the "true" error terms are known since they (or their range) are determined by the analyst. This allows to assess the precision of the estimation approach under different circumstances, such as increasing farm sample sizes. For different sample sizes F ranging from $F = 10$ to $F = 150$ we generate $R = 100$ datasets containing each F farms with random fixed-resource availability and random activity prices with a predefined mean (see Annex II for more details). To keep the model simple, three different activities $(I = 3)$ can be produced by farms with one fixed-resource $(J = 1)$. The fixed-input-output coefficient matrix $\mathbf{A}$ is set to a vector of **1**, so that b can be interpreted as farmland per farm and $\mathbf{x}$ is a vector indicating the activity level, which in this case is the farmland attributed to each activity. The vector $\mathbf{x}$ of optimal, "true" activity levels and the shadow price y_f for each farm in the respective sample are calculated by minimizing the CSC (model (2.84), (3.14) – (3.15)). Finally, the error term vectors $\mathbf{h}$ and $\mathbf{u}$ are drawn from a normal distribution with zero mean and a standard deviation of 5% of the average land allocation of each activity or the average shadow price.[12] By adding the error terms to the “true” activity levels and fixed-input shadow prices, for each farm in the sample we receive vectors $\mathbf{x_f^o}$ and $\mathbf{y_f^o}$, which are finally used to estimate the set of cost function parameters for each farm sample according to the model proposed by Henry de Frahan (2019) (model (2.79) – (2.83)).

Following the example set by Heckelei and Wolff (2003), we assess estimation precision with the root mean squared error (RMSE) of the parameter estimates, which we calculate for each sample size. In other words, the square root was taken of the mean squared distance between parameter estimates and their true values over the one hundred datasets for each sample size, so that the RMSE for parameters $\mathbf{c}$ and $\mathbf{Q}$ is calculated for each sample size F as:

$$RMSE\_Q_{i,ii} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{q}_{r,i,ii} - q_{i,ii})^2} \tag{3.19}$$

$$RMSE\_c_i = \sqrt{\frac{1}{\mathrm{R}}\sum_{\mathrm{r=1}}^{\mathrm{R}}(\hat{\mathrm{c}}_{\mathrm{r,i}} - \mathrm{c_i})^2} \tag{3.20}$$

Finally, as in Heckelei and Wolff (2003), we sum the RMSE of all elements of the matrix Q and vector c for each farm sample size F to receive one single measure per sample size for the estimation precision of matrix $\mathbf{Q}$ and vector $\mathbf{c}$, $SRMSE\_Q$ and $SRMSE\_c$, respectively.

$$SRMSE\_Q = \sum_{\mathbf{i=1}}^{\mathbf{I}}\sum_{\mathbf{ii=1}}^{\mathbf{I}} \mathrm{RMSE\_Q_{i,ii}} \tag{3.21}$$

$$SRMSE\_c = \sum_{\mathbf{i=1}}^{\mathbf{I}} \mathrm{RMSE\_c_i} \tag{3.22}$$

Figure 3 shows the extent to which $SRMSE\_Q$ and $SRMSE\_c$ decrease with increasing farm sample size F. The absolute value of the SRMSE is not particularly meaningful as it depends, among others, on the absolute values of the parameter elements. However, its change in farm sample size F reveals that for small farm samples with $F < 30$ the estimation error is significantly higher than for larger farm samples. While the improvement is much smaller for farm sample sizes with $F > 30$, it is noteworthy that the SRMSE seems to be continuously decreasing even for farm sample sizes of $F > 100$.

[12] Heckelei and Wolff (2003) generate their error terms in a similar way, but with a standard deviation of 2% instead of 5%.

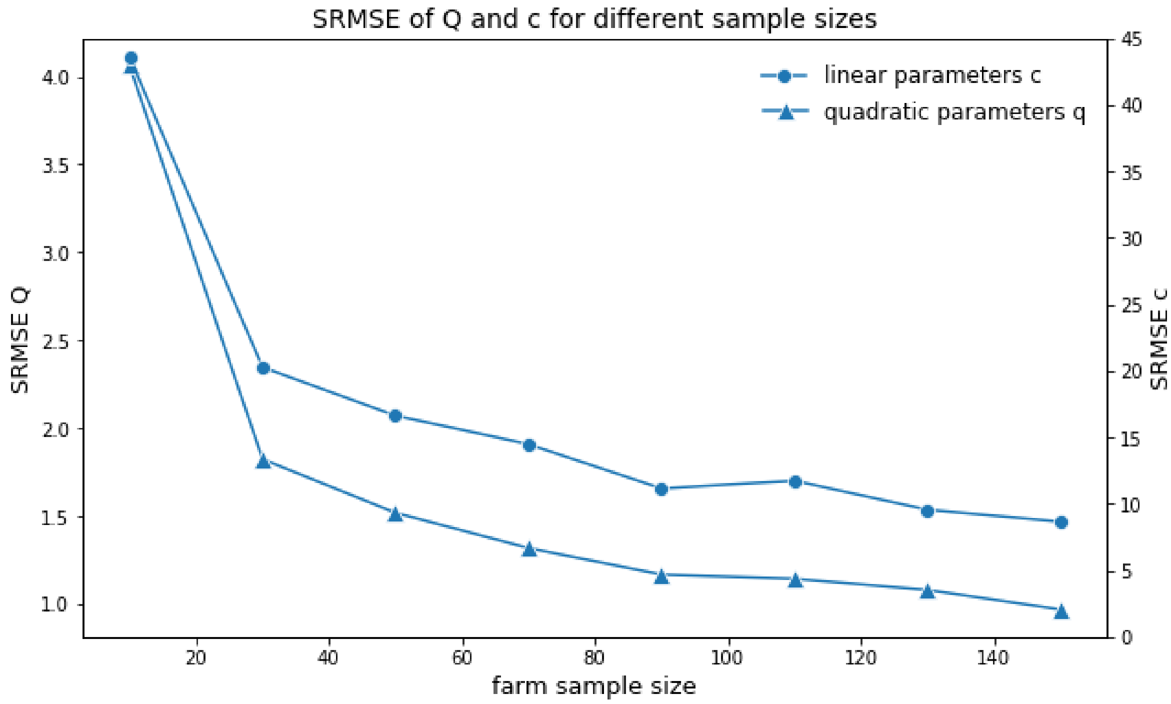

***Figure 3: SRMSE of matrix Q and vector c for increasing farm sample sizes based on estimations with model (2.79) – (2.83)**. The left vertical axis shows the SRMSE for the quadratic cost function parameter matrix **Q** and the right vertical axis indicates the SRMSE for the vector **c** of linear cost function parameters. The horizontal axis displays the farm sample size and each dot is the result of 100 random datasets of the respective farm sample size. The standard deviation of the error terms is set to 5% of the average activity level.*
Source: own calculations and visualizations.

Figure 4 depicts the individual estimates for each element $q_{i,ii}$ of the non-linear cost parameter matrix $\mathbf{Q}$. Each dot represents the estimate of element $\mathrm{q_{i,ii}}$ for one run r (out of $R = 100$ datasets) for the respective farm sample size F. To interpret figure 3 and 4, it is useful to recall that the mean standard error (MSE) of $\hat{q}_{,r,i,ii}$ is the sum of the variance of $\hat{q}_{r,i,ii}$ and its squared bias (Wooldridge, 2016, p. 680). Accordingly, figure 4 implies that the estimates $\hat{q}_{r,i,ii}$ seem to be largely unbiased (the blue horizontal line represents the true value $\mathbf{q_{i,ii}}$), while their variance seems to be particularly large for farm sample sizes of $F < 50$. This would suggest that the SRMSE in figure 3 is mainly determined by the standard deviation of $\hat{q}_{r,i,ii}$.

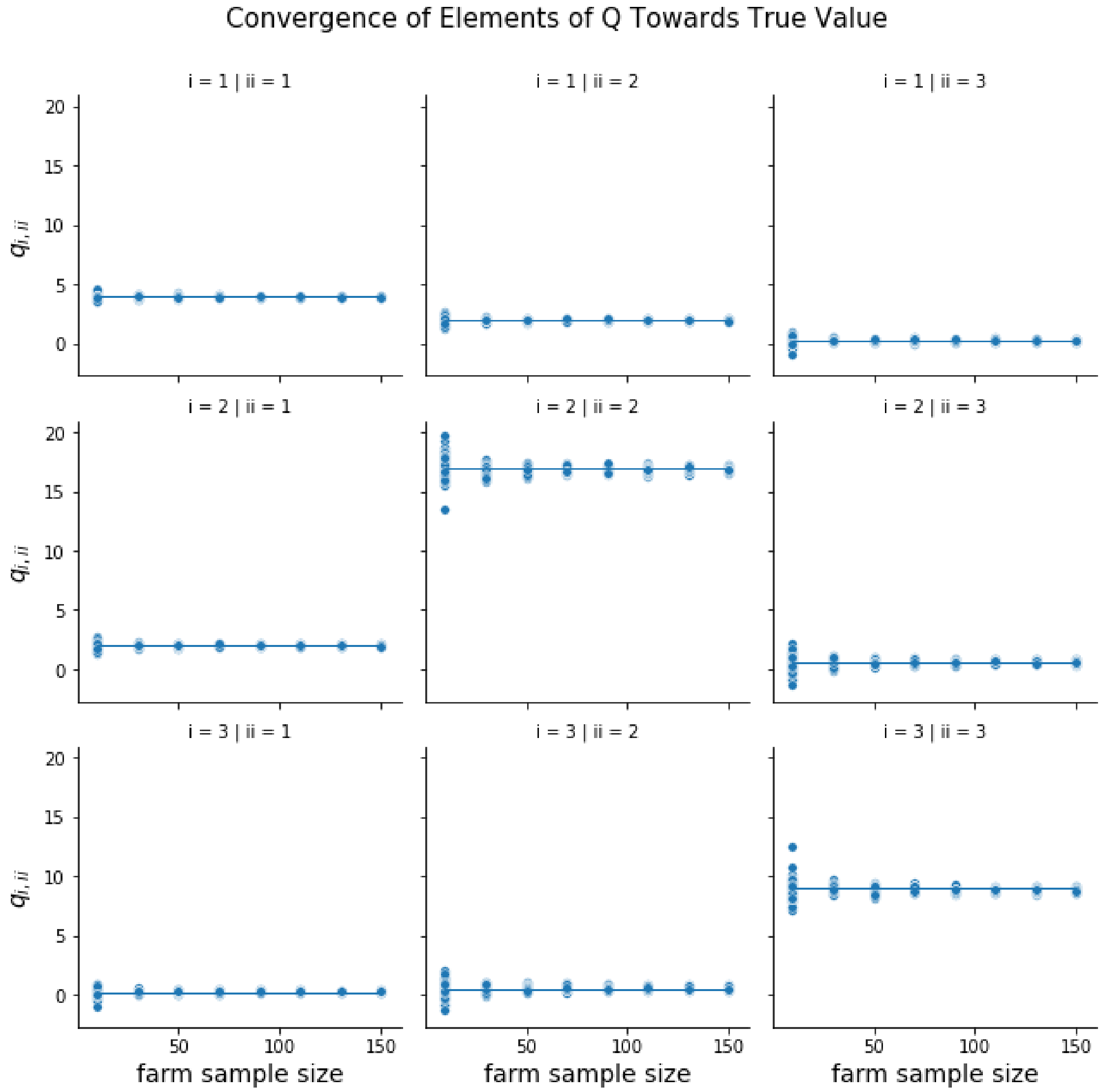

Figure 4: Convergence of estimates for the elements of matrix Q towards their true values with increasing sample sizes. *Note that matrix **Q** is a quadratic, positive (semi-) definite matrix and therefore its elements $q_{i,ii} = q_{ii,i}$. Each dot represents an estimate $\hat{q}_{r,i,ii}$. The blue horizontal line indicates the true value of the element $q_{i,ii}$. Source: own calculations and visualizations.*

3.5. Development of a Cost Function Specification Including Activity Levels and Fixed-Input Levels

3.5.1. Fixed Relationship Between Fixed-Inputs and Outputs

The Monte Carlo simulation underpins that the cost function parameter estimation approach suggested by Henry de Frahan (2019) with the amendments from the methodology section is a consistent procedure, in principle. We have also presented methodological remedies for the potential correlation between the errors of individual units (farm fixed effect) or trends in the underlying technology (time trend). However, a potentially significant shortcoming persists. The fixed-input-output coefficient matrix $\mathbf{A_{t,f}}$ gives the model some rigidity, as it presumes that input-output ratios cannot be changed. Besides, in many cases information is only available on the produced activity levels and the available quantities of fixed-resources, but not on the allocation of fixed-resources to individual activities. As we will see in the next chapter for real farm data, erroneous fixed-input-output coefficients may have devastating implications for the plausibility of the resulting estimates. To show this, let us again assume that we have the information that a farm has b_j units of J fixed-inputs and that it produces x_i^o units of I activities. A first challenge is that in many cases the quantity of fixed-input j allocated to the production of activity i, is not known to the analyst. Actually, the correct allocation of fixed costs to outputs may also be difficult for the farm manager, since the exact contribution of some fixed-inputs such as family labour or equipment to the production of individual activities is often hard to quantify (Mußhoff and Hirschauer, 2020). In section 4.2.1 we apply the following approach to calculate $a_{j,i}$:

$$a_{j,i} = \frac{b_j}{\sum_{i=1}^{I} \omega_{j,i} x_i^o} \tag{3.23}$$

Where the parameter $\omega_{j,i}$ is a dummy that takes the value 1 if the analyst has reason to believe that activity i uses fixed-input j and 0 otherwise. The coefficient $a_{j,i}$ then indicates the average fixed-resource requirements per activity level. In how far it is justifiable to assume that all activities require the same amount of fixed-input per produced level evidently depends also on how the activity levels are measured. In any case, often it is impossible for the analyst to validate her choice of fixed-resource attribution, which makes the calculation of matrix $\mathbf{A_{t,f}}$ prone to arbitrariness. Incorrect assumptions about the allocation of fixed-resources certainly may deteriorate the model. However, for some fixed-inputs, such as farmland, information about the allocation to activities is usually available. For example, in many cases farm data reveals the acreage of a certain crop. Therefore, it is possible to establish an unambiguous link between fixed-inputs (cropland) and activities (acreage of crops). If the analyst knows that $\beta_{j,i}$ units of fixed-resource j are dedicated to the production of x_i^o units of output activity i, the calculation of the respective fixe- input-output coefficient $a_{j,i}$ is straightforward:

$$a_{j,i} = \frac{\beta_{j,i}}{x_i^o} \tag{3.24}$$

If the activity level x_i^o is measured as acreage of activity i and if j is cropland, then $a_{j,i}$ would equal unity, as in the artificial example for the Monte Carlo simulations. If x_i^o is measured in quantities sold or harvested, $a_{j,i}$ would still be easy to calculate. However, one may question if $a_{j,i}$, that was calculated with observed activity levels of the farm in a given year, is a suitable fixed-input-output coefficient for policy simulation. If farmers can adapt the input intensity of production, $a_{j,i}$ would impose an unrealistically rigid relation between the fixed-input and the respective output. This rigidity becomes particularly problematic if a fixed-input is consumed by only one activity. Then the model does not allow an extension of this activity if the resource constraint is already binding for observed activity levels and estimated errors, implying an inelastic response to changing (especially increasing) prices. The significant shortcoming of the model in this situation is that the matrix $\mathbf{A_{t,f}}$ forces the marginal rate of technical substitution to be equal to zero. As we will see in section 4.2.1, an example for this feature is the fixed-input grassland for a farm that produces crops and dairy outputs. Evidently, the only activity that may possibly require grassland is the dairy output. Therefore, the fixed-input-output coefficient for grassland and dairy outputs is calculated as area of available grassland divided by the observed level of dairy outputs. Even if the matrix $\mathbf{A_{t,f}}$ acknowledges that other inputs (such as capital or labour) are required to produce dairy outputs, the resource constraint for grassland prohibits to extend dairy production compared to the level in the reference scenario without extending the available level of grassland. In practice, a farm could presumably substitute the grassland required for an extension of dairy outputs to some extent by other fixed-inputs, such as cropland (to substitute grass in a feed ratio by maize, for example) or a combination of labour and capital (e.g., to intensify grassland utilization).

3.5.2. Conceptualization of an Activity Level and Fixed-Input Level Dependent Cost Function Specification

As demonstrated in the previous section, the matrix of constant technical coefficients $\mathbf{A_{t,f}}$ is prone to misspecification and ignores potential substitutability between fixed-inputs. Accordingly, it is unlikely that a model that relies on the matrix $\mathbf{A_{t,f}}$ to connect activity levels with fixed-inputs correctly represents the true choices of producers. This is one of the reasons why the calibrating equilibrium model set up by Paris (2017) in the second phase contains an extended Leontief cost function specification, as this eliminates the matrix $\mathbf{A_{t,f}}$ which is still required in the first phase. However, an extension of the approach suggested by Henry de Frahan (2019) to remove some rigidity from the relationship between fixed-input and output levels would have to appear already in the primal problem (2.59) – (2.61), to comply with the fundamental tenet that the parameter estimation does not require two phases. Besides, a crucial requirement for a cost function is that the resulting model may adequately represent the true data generating process (Heckelei and Wolff, 2003). Consequently, it must avoid the rigidity entailed in a constant fixed-input-output coefficient matrix and at the same time be consistent with economic theory. Based on a methodological note by Henry de Frahan we start by including both activity levels and fixed-input levels in a cost function. To make a clear conceptual distinction between available fixed-input levels $\mathbf{b_f}$ and (utilized) fixed-input levels, we name the latter $\mathbf{z_{t,f}}$. According to Chambers (1988), a well behaved neo-classical cost function should be non-decreasing and convex in output quantities, non-increasing as well as convex in fixed-inputs, and regular. The two conditions for output quantities and the regularity condition are fulfilled for the activity level dependent cost function specification presented above and utilized by Henry de Frahan (2019). The Cholesky factorization ensures convexity in output levels and by multiplying the parameters $\mathbf{c}$ and $\mathbf{Q}$ with the activity levels the total variable costs will always be zero if all activity levels are null. Finally, the optimality condition which imposes MC to be larger than or equal to MR ensures that the total variable costs are not decreasing for positive output quantities. A cost function that also includes fixed-input levels must also fulfil the additional requirements: Convexity in fixed-inputs can be ensured by imposing positive semi-definiteness on a quadratic parameter that is multiplied with fixed-inputs, equivalent to the role that matrix $\mathbf{Q}$ plays for output levels. We denominate this quadratic $(J \times J)$ matrix $\mathbf{F}$, so that convexity in fixed-input levels of the new cost function is guaranteed by the term $0.5\mathbf{z}_{\mathbf{t,f}}'\mathbf{F}\mathbf{z}_{\mathbf{t,f}}$. The estimated elements of the matrix $\mathbf{F}$ provide information on the total *variable* costs for given levels of fixed-inputs. However, this requires that the term is null if all activity levels are null (regularity condition). This could be ensured by multiplying the term with a fixed-weight output quantity index $\boldsymbol{\phi}'\mathbf{x}_{\mathbf{t,f}}$, as in Henry de Frahan *et al.* (2011) where

$$\phi_{l_x} = T^{-1} \sum_{t=1}^{T} \left[\frac{\sum_{f=1}^{F} p_{t,f,l_x}}{\sum_{i=1}^{I} \sum_{f=1}^{F} p_{t,f,i}} \right] \tag{3.25}$$

where the index l_x denotes the respective activity. The vector of fixed-weight output quantity indices, which only has the purpose to maintain regularity, can be approximated to $\boldsymbol{\phi}'\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}$, so that it can already be employed in the estimation phase. The third requirement for the behaviour of a cost function is that costs must be non-increasing over the range of observed fixed-input levels. To understand the rationale behind this condition, it is important to recall that it refers to the derivative of total variable costs with respect to fixed-input levels, that is, a change in fixed-input levels while keeping activity levels constant. The criterion states that no rational farmer would employ an observed level z_j^o of fixed-input j leading to TVC^o if the same activity levels could be produced with input level $z_j^{o-1} < z_j^o$ at lower total variable costs $TVC^{o-1} < TVC^o$. Consequently, one can infer that $\partial TVC/\partial z_j \leq 0$, and if a minimum of TVC in fixed-input levels has been reached, then $\partial TVC/\partial z_j = 0$. However, in this thesis, we consider available quantities $\mathbf{b}_{\mathbf{f}}$ of fixed-inputs as truly invariable, an assumption that seems justifiable in the short- and mid-terms. Therefore, if the resource constraint is binding and the utilized fixed-input level $\mathbf{z}_{\mathbf{t,f}} = \mathbf{b}_{\mathbf{f}}$, it is possible that $\partial TVC/\partial z_j < 0$. While a farm could then produce the same outputs at lower TVC *if* it had a higher level $\mathbf{b}_{\mathbf{f}}$, it is impossible to achieve this long-run equilibrium in the short- or midterm.

Besides the quadratic parameter $\mathbf{F}$, let us define a J vector $\mathbf{g}$ of linear fixed-input-related cost function parameters. Just as the quadratic term, the product $\mathbf{g}'\mathbf{z}_{\mathbf{t,f}}$ must be multiplied by $\boldsymbol{\phi}'\mathbf{x}_{\mathbf{t,f}}$ to ensure that the regularity condition $TVC(\mathbf{x} = 0, \mathbf{z}) = 0$ holds.

A cost function capable of capturing the influence of fixed-input levels and activity levels on TVC may also be extended by a term that accounts for potential interactions between the two. We denominate the respective $(J \times I)$ matrix $\mathbf{H}$. Since it is multiplied by $\mathbf{x_{t,f}}$ and $\mathbf{z_{t,f}}$, the regularity condition is fulfilled. After having defined the additional cost function parameters and the conditions they need to comply with, the complete TNR maximization function containing the extended cost function specification can be stated for each farm in each year as follows:

$$max_{\mathbf{x_{t,f}},\mathbf{z_{t,f}},\mathbf{h_{t,f}},\mathbf{s_{t,f}}}\, AUX_{t,f} = \mathbf{p_{t,f}}'\mathbf{x_{t,f}} - \mathbf{c}'\mathbf{x_{t,f}} - \tfrac{1}{2}\mathbf{x_{t,f}}'\mathbf{Q}\,\mathbf{x_{t,f}} - (\boldsymbol{\phi}'\mathbf{x_{t,f}})\ \mathbf{g}'\,\mathbf{z_{t,f}}\ - \tfrac{1}{2}\mathbf{z_{t,f}}'\mathbf{F}\mathbf{z_{t,f}}\,(\boldsymbol{\phi}'\mathbf{x_{t,f}}) - \mathbf{z_{t,f}}'\mathbf{H}\,\mathbf{x_{t,f}} - \tfrac{1}{2}\mathbf{h}'_{\mathbf{t.f}}\mathbf{W_{t.f}}\mathbf{h_{t.f}} - \tfrac{1}{2}\mathbf{s}'_{\mathbf{t.f}}\mathbf{T_{t.f}}\mathbf{s_{t.f}} \qquad (3.26)$$

subject to

$$\mathbf{z_{t,f}} \leq \mathbf{b_{t,f}} \qquad [\mathbf{y_{t,f}}] \qquad (3.27)$$

$$\mathbf{x_{t,f}} = \mathbf{x^o_{t,f}} + \mathbf{h_{t,f}} \qquad [\boldsymbol{\lambda}_{\mathbf{t,f}}] \qquad (3.28)$$

$$\mathbf{z_{tmf}} = \mathbf{z^o_{t,f}} + \mathbf{s_{t,f}} \qquad [\boldsymbol{\kappa}_{\mathbf{t,f}}] \qquad (3.29)$$

with nonnegative elements of vectors $\mathbf{x_{t,f}}$ and $\mathbf{z_{t,f}}$ and the vectors of dual values indicated in square brackets. As it was already implemented for activity levels and fixed-input shadow prices, it is possible to account for potential disparities between observed and optimal utilized fixed-input-levels by introducing a vector of error terms $\mathbf{s_{t,f}}$ (constraint 3.29). The further steps follow the procedure for the generation of the cost function parameter estimation model (2.64) – (2.83) as presented by Henry de Frahan (2019): An auxiliary function (3.26) includes the squared error terms of vectors $\mathbf{h_{t,f}}$ and $\mathbf{s_{t,f}}$ multiplied with a diagonal weight matrix to harmonise the units of measurement. The matrix $\mathbf{W_{t,f}}$, again, has the elements $p_{t,f,i}/x^o_{t,f,i}$ on its main diagonal, and the diagonal elements of matrix $\mathbf{T_{t,f}}$ are $y^o_{t,f,j}/z^o_{t,f,j}$. Then, the Lagrangian function is set up with its corresponding first order conditions:

$$LMAX_{t,f}(\mathbf{x_{t,f}}, \mathbf{z_{t,f}}, \mathbf{h_{t,f}}, \mathbf{s_{t,f}}, \boldsymbol{\lambda}_{\mathbf{t,f}}, \boldsymbol{\kappa}_{\mathbf{t,f}}) = \mathbf{p}'_{\mathbf{t,f}}\mathbf{x_{t,f}} - \mathbf{c}'\mathbf{x_{t,f}} - \tfrac{1}{2}\mathbf{x}'_{\mathbf{t,f}}\mathbf{Q}\mathbf{x_{t,f}} - (\boldsymbol{\phi}'\mathbf{x_{t,f}})\ \mathbf{g}'\,\mathbf{z_{t,f}} - \tfrac{1}{2}\mathbf{z_{t,f}}'\mathbf{F}\mathbf{z_{t,f}}\,(\boldsymbol{\phi}'\mathbf{x_{t,f}})\ - \mathbf{z_{t,f}}'\mathbf{H}\,\mathbf{x_{t,f}} - \tfrac{1}{2}\mathbf{h}'_{\mathbf{t,f}}\mathbf{W_{t,f}}\mathbf{h_{t,f}} - \tfrac{1}{2}\mathbf{s}'_{\mathbf{t,f}}\mathbf{T_{t,f}}\mathbf{s_{t,f}} + \mathbf{y}'_{\mathbf{t,f}}[\mathbf{b_{t,f}} - \mathbf{z_{t,f}}] + \boldsymbol{\lambda}'_{\mathbf{t,f}}[\mathbf{x^o_{t,f}} + \mathbf{h_{t,f}} - \mathbf{x_{t,f}}] + \boldsymbol{\kappa}'_{\mathbf{t,f}}[\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}} - \mathbf{z_{t,f}}] \qquad (3.30)$$

$$\frac{\partial \mathbf{LMAX_{t,f}}}{\partial \mathbf{x_{t,f}}} = \mathbf{p_{t,f}} - (\mathbf{c} + \mathbf{Q}\mathbf{x_{t,f}} +\ \phi_{l_x}\,\mathbf{g}'\,\mathbf{z_{t,f}}\ +\ \tfrac{1}{2}\mathbf{z_{t,f}}'\mathbf{F}\mathbf{z_{t,f}}\,\phi_{l_x} +\ \mathbf{H}'\,\mathbf{z_{t,f}} + \boldsymbol{\lambda}_{\mathbf{t,f}}) \leq \mathbf{0} \qquad (3.31)$$

$$\frac{\partial \mathbf{LMAX_{t,f}}}{\partial \mathbf{z_{t,f}}} = -((\boldsymbol{\phi}'\mathbf{x_{t,f}})\,\mathbf{g} + \mathbf{F}\mathbf{z_{t,f}}(\boldsymbol{\phi}'\mathbf{x_{t,f}}) +\ \mathbf{H}\,\mathbf{x_{t,f}} + \mathbf{y_{t,f}} +\ \boldsymbol{\kappa}_{\mathbf{t,f}}) \leq \mathbf{0} \qquad (3.32)$$

$$\frac{\partial \mathbf{LMAX_{t,f}}}{\partial \mathbf{h_{t,f}}} = -\mathbf{W_{t,f}}\mathbf{h_{t,f}} + \boldsymbol{\lambda}_{\mathbf{t,f}} = \mathbf{0} \qquad (3.33)$$

$$\frac{\partial \mathbf{LMAX_{t,f}}}{\partial \mathbf{s_{t,f}}} = -\mathbf{T_{t,f}}\mathbf{s_{t,f}} + \boldsymbol{\kappa}_{\mathbf{t,f}} = \mathbf{0} \qquad (3.34)$$

$$\frac{\partial \mathbf{LMAX_{t,f}}}{\partial \boldsymbol{\lambda}_{\mathbf{t,f}}} = \mathbf{x^o_{t,f}} + \mathbf{h_{t,f}} - \mathbf{x_{t,f}} = \mathbf{0} \qquad (3.35)$$

$$\frac{\partial \mathbf{LMAX_{t,f}}}{\partial \boldsymbol{\kappa}_{\mathbf{t,f}}} = \mathbf{z^o_{t,f}} + \mathbf{s_{t,f}} - \mathbf{z_{t,f}} = \mathbf{0} \qquad (3.36)$$

With nonnegativity for the elements of vectors $\mathbf{x}_{t,f}$ and $\mathbf{z}_{t,f}$. Equations (3.31) and (3.32) may be transformed into equalities by applying the KKT conditions and multiplying them by vectors $\mathbf{x}_{t,f}$ and $\mathbf{z}_{t,f}$, respectively. Note that equations (3.33) and (3.34) allow to replace the vectors of dual values $\boldsymbol{\lambda}_{t,f}$ and $\boldsymbol{\kappa}_{t,f}$ in constraints (3.31) and (3.32) by the products $\mathbf{W}_{t,f}\mathbf{h}_{t,f}$ and $\mathbf{T}_{t,f}\mathbf{s}_{t,f}$, respectively. Paris (2010, p. 69) describes how the dual problem of any nonlinear maximization problem can be derived from the Lagrangian function of the primal problem. By following this guideline, we can write the dual problem as

$$min_{\mathbf{x}_{t,f},\mathbf{z}_{t,f},\mathbf{y}_{t,f},\mathbf{h}_{t,f},\mathbf{s}_{t,f}\mathbf{u}_{t,f}}\ AUX_{t,f} = \tfrac{1}{2}\mathbf{x}_{t,f}'\mathbf{Q}\mathbf{x}_{t,f} + \phi_{l_x}\mathbf{x}_{t,f}\,\mathbf{g}'\,\mathbf{z}_{t,f} + \tfrac{1}{2}\mathbf{x}_{t,f}\,\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\,\phi_{l_x} + \mathbf{z}_{t,f}'\mathbf{H}\mathbf{x}_{t,f} + \tfrac{1}{2}\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right) + \mathbf{y}_{t,f}'(\mathbf{b}_{t,f}) + \boldsymbol{\lambda}_{t,f}'\left[\mathbf{x}_{t,f}^{o} + \mathbf{h}_{t,f}\right] + \boldsymbol{\kappa}_{t,f}'\left[\mathbf{z}_{t,f}^{o} + \mathbf{s}_{t,f}\right] + \tfrac{1}{2}\mathbf{u}_{t,f}'\mathbf{V}_{t,f}\mathbf{u}_{t,f} \quad (3.37)$$

subject to

$$\mathbf{c} + \mathbf{Q}\mathbf{x}_{t,f} + \phi_{l_x}\,\mathbf{g}'\,\mathbf{z}_{t,f} + \tfrac{1}{2}\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\,\phi_{l_x} + \mathbf{H}'\,\mathbf{z}_{t,f} + \boldsymbol{\lambda}_{t,f} \geq \mathbf{p}_{t,f} \quad [\mathbf{x}_{t,f}] \quad (3.38)$$

$$\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right)\mathbf{g} + \mathbf{F}\mathbf{z}_{t,f}\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right) + \mathbf{H}\,\mathbf{x}_{t,f} + \boldsymbol{\kappa}_{t,f} \geq -\mathbf{y}_{t,f} \quad [\mathbf{z}_{t,f}] \quad (3.39)$$

$$\mathbf{y}_{t,f} = \mathbf{y}_{t,f}^{o} + \mathbf{u}_{t,f} \quad [\boldsymbol{\psi}_{t,f}] \quad (3.40)$$

With the elements of vectors $\mathbf{x}_{t,f}$, $\mathbf{y}_{t,f}$ and $\mathbf{z}_{t,f}$ all being non-negative, and the elements of the error term vectors $\mathbf{h}_{t,f}$, $\mathbf{s}_{t,f}$ and $\mathbf{u}_{t,f}$ free. Again, it is possible to form an auxiliary function which contains the sum of the squared deviations of vector $\mathbf{u}_{t,f}$ weighted by the diagonal matrix $\mathbf{V}_{t,f}$. The Lagrangian function with its first-order conditions is as follows:

$$LMIN_{t,f}\left(\mathbf{y}_{t,f}, \mathbf{x}_{t,f}, \mathbf{z}_{t,f}, \mathbf{u}_{t,f}, \boldsymbol{\psi}_{t,f}\right) = \tfrac{1}{2}\mathbf{x}_{t,f}'\mathbf{Q}\mathbf{x}_{t,f} + \phi_{l_x}\mathbf{x}_{t,f}\,\mathbf{g}'\,\mathbf{z}_{t,f} + \tfrac{1}{2}\mathbf{x}_{t,f}\,\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\,\phi_{l_x} + \mathbf{z}_{t,f}'\mathbf{H}\mathbf{x}_{t,f} + \tfrac{1}{2}\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right) + \mathbf{y}_{t,f}'(\mathbf{b}_{t,f}) + \boldsymbol{\lambda}_{t,f}'\left[\mathbf{x}_{t,f}^{o} + \mathbf{h}_{t,f}\right] + \boldsymbol{\kappa}_{t,f}'\left[\mathbf{z}_{t,f}^{o} + \mathbf{s}_{t,f}\right] + \tfrac{1}{2}\mathbf{u}_{t,f}'\mathbf{V}_{t,f}\mathbf{u}_{t,f} + \mathbf{x}_{t,f}'\Big[\mathbf{p}_{t,f} - \mathbf{c} - \mathbf{Q}\mathbf{x}_{t,f} - \phi_{l_x}\,\mathbf{g}'\,\mathbf{z}_{t,f} - \tfrac{1}{2}\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\,\phi_{l_x} - \mathbf{H}'\,\mathbf{z}_{t,f} - \boldsymbol{\lambda}_{t,f}\Big] + \mathbf{z}_{t,f}'\left[-\mathbf{y}_{t,f} - \left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right)\mathbf{g} - \mathbf{F}\mathbf{z}_{t,f}\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right) - \mathbf{H}\,\mathbf{x}_{t,f} - \boldsymbol{\kappa}_{t,f}\right] + \boldsymbol{\psi}_{t,f}'\left[\mathbf{y}_{t,f}^{o} + \mathbf{u}_{t,f} - \mathbf{y}_{t,f}\right] \quad (3.41)$$

$$\frac{\partial \mathbf{LMIN}_{t,f}}{\partial \mathbf{y}_{t,f}} = \mathbf{b}_{t,f} - \mathbf{z}_{t,f} - \boldsymbol{\psi}_{t,f} \geq \mathbf{0} \quad (3.42)$$

$$\frac{\partial \mathbf{LMIN}_{t,f}}{\partial \mathbf{x}_{t,f}} = \mathbf{p}_{t,f} - \mathbf{c} - \mathbf{Q}\mathbf{x}_{t,f} - \phi_{l_x}\,\mathbf{g}'\,\mathbf{z}_{t,f} - \tfrac{1}{2}\mathbf{z}_{t,f}'\mathbf{F}\mathbf{z}_{t,f}\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right) - \mathbf{H}'\,\mathbf{z}_{t,f} - \boldsymbol{\lambda}_{t,f} \leq \mathbf{0}$$

(3.43)≡(3.31)

$$\frac{\partial \mathbf{LMIN}_{t,f}}{\partial \mathbf{z}_{t,f}} = -\mathbf{y}_{t,f} - \left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right)\mathbf{g} - \mathbf{F}\mathbf{z}_{t,f}\left(\boldsymbol{\phi}'\mathbf{x}_{t,f}\right) - \mathbf{H}\,\mathbf{x}_{t,f} - \boldsymbol{\kappa}_{t,f} \leq \mathbf{0} \quad (3.44)\equiv(3.32)$$

$$\frac{\partial \mathbf{LMIN}_{t,f}}{\partial \mathbf{u}_{t,f}} = \mathbf{V}_{t,f}\mathbf{u}_{t,f} + \boldsymbol{\psi}_{t,f} = \mathbf{0} \quad (3.45)$$

$$\frac{\partial \mathbf{LMIN}_{t,f}}{\partial \boldsymbol{\psi}_{t,f}} = \mathbf{y}_{t,f}^{o} + \mathbf{u}_{t,f} - \mathbf{y}_{t,f} = \mathbf{0} \quad (3.46)$$

with nonnegativity imposed on the elements of vectors $\mathbf{x}_{t,f}$, $\mathbf{y}_{t,f}$ and $\mathbf{z}_{t,f}$. Again, it is possible to transform constraints (3.43) and (3.44) into equalities by making use of the KKT conditions. The vector $\boldsymbol{\psi}_{t,f}$ of dual variables in constraint (3.42) may be replaced by the product $-\mathbf{V}_{t,f}\mathbf{u}_{t,f}$ based on condition (3.45). According to the conditions (3.35), (3.36) and (3.46) it is possible to replace the "optimal" activity levels, fixed-input shadow prices and fixed-input levels by

$\left(\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{h}_{\mathbf{t,f}}\right)$, $\left(\mathbf{y}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{u}_{\mathbf{t,f}}\right)$ and $\left(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}}\right)$, respectively. Finally, by combining the first-order conditions from the two Lagrangian functions the following optimality conditions may be derived:

$$\mathbf{c}+\mathbf{Q}\left(\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{h}_{\mathbf{t,f}}\right)+\phi_{l_x}\,\mathbf{g}'\left(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}}\right)+\frac{1}{2}(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}})'\mathbf{F}(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}})\,\phi_{l_x}+\mathbf{H}'\left(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}}\right)+\mathbf{W}_{\mathbf{t,f}}\mathbf{h}_{\mathbf{t,f}}\geq\mathbf{p}_{\mathbf{t,f}} \quad (3.47)$$

$$\left(\boldsymbol{\phi}'(\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{h}_{\mathbf{t,f}})\right)\mathbf{g}+\mathbf{F}(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}})\left(\boldsymbol{\phi}'(\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{h}_{\mathbf{t,f}})\right)+\mathbf{H}\left(\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{h}_{\mathbf{t,f}}\right)+\mathbf{T}_{\mathbf{t,f}}\mathbf{s}_{\mathbf{t,f}}\geq-(\mathbf{y}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{u}_{\mathbf{t,f}}) \quad (3.48)$$

$$\left(\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}}\right)-\mathbf{V}_{\mathbf{t,f}}\mathbf{u}_{\mathbf{t,f}}\leq\mathbf{b}_{\mathbf{t,f}} \quad (3.49)$$

$$\mathbf{x}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{h}_{\mathbf{t,f}}\geq\mathbf{0} \quad (3.50)$$

$$\mathbf{z}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{s}_{\mathbf{t,f}}\geq\mathbf{0} \quad (3.51)$$

$$\mathbf{y}_{\mathbf{t,f}}^{\mathbf{o}}+\mathbf{u}_{\mathbf{t,f}}\geq\mathbf{0} \quad (3.52)$$

Besides, the Cholesky factorization is implemented for the quadratic matrices $\mathbf{Q}$ and $\mathbf{F}$ to ensure their positive (semi) definiteness, as described in section 3.1.1. These conditions serve as the constraints of the following weighted squared error minimization function, which allows to estimate the vectors and matrices $\mathbf{c}$, $\mathbf{Q}$, $\mathbf{g}$, $\mathbf{F}$ and $\mathbf{H}$ of cost function parameters as well as the vectors of error terms $\mathbf{h}_{\mathbf{t,f}}$, $\mathbf{s}_{\mathbf{t,f}}$ and $\mathbf{u}_{\mathbf{t,f}}$:

$$min_{\mathbf{h_{t,f}},\mathbf{s_{t,f}},\mathbf{u_{t,f}},\mathbf{c},\mathbf{Q},\mathbf{g},\mathbf{F},\mathbf{H}}LS=\frac{1}{2}\sum_{t=1}^{T}\sum_{f=1}^{F}\left(\mathbf{h}_{\mathbf{t,f}}'\mathbf{W}_{\mathbf{t,f}}\mathbf{h}_{\mathbf{t,f}}+\mathbf{s}_{\mathbf{t,f}}'\mathbf{T}_{\mathbf{t,f}}\mathbf{s}_{\mathbf{t,f}}+\mathbf{u}_{\mathbf{t.f}}'\mathbf{V}_{\mathbf{t,f}}\mathbf{u}_{\mathbf{t,f}}\right) \quad (3.53)$$

Constraint (3.42) states that MC should be larger or equal to MR, equivalent to constraint (2.79) of the model with only activity levels as arguments of the cost function, as introduced by Henry de Frahan (2019). Following the reasoning outlined in section 3.1.2 and applying the KKT conditions, MC equal MR for strictly positive activity levels. According to constraint (3.48) the marginal total variable costs with respect to the fixed-input levels of vector $\mathbf{z}_{\mathbf{t,f}}$ must be larger than or equal to the element of the negative fixed-input shadow price vector $\mathbf{y}_{\mathbf{t,f}}$. Again, by applying the KKT conditions this constraint becomes an equality for positive fixed-input levels. This implies that, if vector $\mathbf{z}_{\mathbf{t,f}}>\mathbf{0}$, total variable costs are non-increasing in fixed-input levels, as required by the monotonicity condition for well-behaved cost functions. Constraint (3.48) and the fixed-resource constraint (3.49) designate a fundamental difference between the cost function specified only in activity levels and this more complete cost function: In model (2.79) – (2.83) the vector of available fixed-resources, $\mathbf{b}_{\mathbf{t,f}}$, confines the extension of activity levels $\mathbf{x}_{\mathbf{t,f}}$ beyond a certain point. The fixed-resource constraint (3.49) constitutes a limit for the fixed-input levels of vector $\mathbf{z}_{\mathbf{t,f}}$, but the maximal admissible activity level of the vector $\mathbf{x}_{\mathbf{t,f}}$ is only determined by the optimality condition that MC equals MR. When performing simulations with the cost function as estimated in model (3.47) – (3.53), the optimal activity levels of vector $\mathbf{x}_{\mathbf{f}}$ depend on the estimated cost function parameters and not on the

exogenous fixed-input-output coefficient matrix $\mathbf{A_{t,f}}$. The liberation from the fixed technological coefficients is a pivotal advantage of the more complete cost function. However, it implicates that the simulated activity levels could become unrealistically large if the elements of matrix $\mathbf{Q}$ are all close to zero, as we will see in chapter 4.2.2. The different formulations of the fixed-resource constraint in the two model specifications also impact the fixed-input shadow prices of vector $\mathbf{y_{t,f}}$. As visualized by figure 2, an element of vector $\mathbf{y_f}$ is only strictly positive if the respective element of vector $(\mathbf{c} + \mathbf{Qx_{t,f}}) < \mathbf{p_{t,f}}$. The shadow price y_j therefore indicates by how much a farm could increase its TNR if an additional unit of the fixed-resource J permitted the production of higher activity levels. In contrast, the resource constraint (3.49) in model (3.47) – (3.53) never restricts the expansion of activity levels. However, by limiting $\mathbf{z_{t,f}}$ it prevents a further reduction of the total variable costs. The elements of $\mathbf{y_{t,f}}$ therefore are the values by which the TVC could be reduced if for given activity levels of vector $\mathbf{x_{t,f}}$ the farm could employ one more unit of the fixed-input J. In other words, since the elements of vector $\mathbf{b_{t,f}}$ do not confine an expansion of the activity levels but only a reduction in TVC, an additional unit of an element of vector $\mathbf{b_{t,f}}$ would not have an impact on the revenue but only on the cost side of the TNR function. Applying the KKT conditions, constraint (3.48) is an equality if the respective element of vector $\mathbf{z_{t,f}} > \mathbf{0}$. Then, the marginal costs w.r.t $\mathbf{z_{t,f}}$ must equal the shadow prices of vector $\mathbf{y_f}$.
If it is reasonable to assume that a time trend exists (e.g., thanks to technology changes over time), it is possible to extend the estimation model (3.47) – (3.53) to capture a time trend by decomposing the vector $\mathbf{c}$ into the vectors $\mathbf{c_0}$ and $\mathbf{c_1 t}$ and vector $\mathbf{g}$ into the vectors $\mathbf{g_0}$ and $\mathbf{g_1}$. Equations (3.47) and (3.48) are therefore modified as follows:

$$\mathbf{c_0} + \mathbf{c_1 t} + \mathbf{Q}(\mathbf{x^o_{t,f}} + \mathbf{h_{t,f}}) + \phi_{l_x}\mathbf{g}_0' \, (\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}}) + \phi_{l_x}\mathbf{g}_1' \, (\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}})\mathbf{t} + \frac{1}{2}(\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}})'\mathbf{F}(\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}})\, \phi_{l_x} + \mathbf{H}'(\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}}) + \mathbf{W_{t,f}h_{t,f}} \geq \mathbf{p_{t,f}} \qquad (3.47\_t)$$

$$(\boldsymbol{\phi}'(\mathbf{x^o_{t,f}} + \mathbf{h_{t,f}}))\, \mathbf{g_0} + (\boldsymbol{\phi}'(\mathbf{x^o_{t,f}} + \mathbf{h_{t,f}}))\, \mathbf{g_1 t} + \mathbf{F}(\mathbf{z^o_{t,f}} + \mathbf{s_{t,f}})(\boldsymbol{\phi}'(\mathbf{x^o_{t,f}} + \mathbf{h_{t,f}})) + \mathbf{H}(\mathbf{x^o_{t,f}} + \mathbf{h_{t,f}}) + \mathbf{T_{t,f}s_{t,f}} \geq -(\mathbf{y^o_{t,f}} + \mathbf{u_{t,f}}) \qquad (3.43\_t)$$

where now a distinction is made between $\mathbf{c_0}$ and $\mathbf{c_1}$ as well as $\mathbf{g_0}$ and $\mathbf{g_1}$ to differentiate between linear cost function parameters that remain constant over time and those that evolve over time. As in the case of the cost function specification with only activity levels, the estimated parameters may then be used for policy simulation either in a CSC minimization framework or in a TNR maximization function.
Equivalent to the objective function (2.84) with constraints (3.14) and (3.15), the CSC minimization model can be defined as follows for each farm in a reference year t:

$$min_{\boldsymbol{\rho}_{\mathbf{dx_{t,f}}}, \boldsymbol{\rho}_{\mathbf{dz_{t,f}}}, \boldsymbol{\rho}_{\mathbf{py_{t,f}}}, \mathbf{x_{t,f}}, \mathbf{z_{t,f}}, \mathbf{y_{t,f}}} CSC_{t,f} = \boldsymbol{\rho}_{\mathbf{dx_{t,f}}}{}'\mathbf{x_{t,f}} + \boldsymbol{\rho}_{\mathbf{dz_{t,f}}}{}'\mathbf{z_{t,f}} + \boldsymbol{\rho}_{\mathbf{py_{t,f}}}{}'\mathbf{y_{t,f}} \qquad (3.54)$$

subject to $\hat{\mathbf{c}} + \widehat{\mathbf{Q}}\mathbf{x_{t,f}} + \phi_{l_x} \hat{\mathbf{g}}'\mathbf{z_{t,f}} + \frac{1}{2}\mathbf{z_{t,f}}'\widehat{\mathbf{F}}\, \mathbf{z_{t,f}}\, \phi_{l_x} + \widehat{\mathbf{H}}'\mathbf{z_{t,f}} + \mathbf{W_{t,f}}\hat{\mathbf{h}}_{\mathbf{t,f}} = \mathbf{p_{t,f}} + \boldsymbol{\rho}_{\mathbf{dx_{t,f}}}$ (3.55)

$$\left(\boldsymbol{\phi}'(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,f}}+\mathbf{h}_{\mathbf{t,f}})\right)\ \hat{\mathbf{g}}+\hat{\mathbf{F}}\ \mathbf{z}_{\mathbf{t,f}}\left(\boldsymbol{\phi}'(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,f}}+\mathbf{h}_{\mathbf{t,f}})\right)+\ \hat{\mathbf{H}}\ \mathbf{x}_{\mathbf{t,f}}+\mathbf{T}_{\mathbf{t,f}}\hat{\mathbf{s}}_{\mathbf{t,f}}=-\mathbf{y}_{\mathbf{t,f}}+\boldsymbol{\rho}_{\mathbf{dz}_{\mathbf{t,f}}} \quad (3.56)$$

$$\mathbf{z}_{\mathbf{t,f}}-\mathbf{V}_{\mathbf{t,f}}\hat{\mathbf{u}}_{\mathbf{t,f}}+\ \boldsymbol{\rho}_{\mathbf{py}_{\mathbf{t,f}}}=\mathbf{b}_{\mathbf{t,f}} \quad (3.57)$$

with non-negativity imposed on the six decision variables. As for the CSC minimization in model (2.84) – (2.86), vector $\boldsymbol{\rho}_{\mathbf{py}_{\mathbf{t,f}}}$ is a slack-surplus variable vector that indicates the level of unused fixed-inputs and is therefore multiplied by vector $\mathbf{y}_{\mathbf{f}}$. Another vector of slack variables, $\boldsymbol{\rho}_{\mathbf{dx}_{\mathbf{t,f}}}$, is related to the primal constraint and denotes by how much MC with respect to activity levels exceed MR. The third vector of slack variables, $\boldsymbol{\rho}_{\mathbf{dz}_{\mathbf{t,f}}}$, is introduced for this cost function specification. It is related to the dual constraints and takes strictly positive values if the MC with respect to fixed-input levels are larger than the negative of the element of the shadow price vector $\mathbf{y}_{\mathbf{f}}$. If constraints (3.47) – (3.49) are fulfilled as equalities for the reference year, then the CSC minimization is equal or very close to zero and yields vectors $\mathbf{x}_{\mathbf{f}}=\mathbf{x}^{\mathbf{o}}_{\mathbf{f}}+\mathbf{h}_{\mathbf{f}}$, $\mathbf{y}_{\mathbf{f}}=\mathbf{y}^{\mathbf{o}}_{\mathbf{f}}+\mathbf{u}_{\mathbf{f}}$ and $\mathbf{z}_{\mathbf{f}}=\mathbf{z}^{\mathbf{o}}_{\mathbf{f}}+\mathbf{s}_{\mathbf{f}}$. As demonstrated for the cost function including only activity levels in the objective function (3.16), the CSC minimization with this more complete cost function specification may also be performed over multiple farms (for example, located within a region) at once, by summing the right-hand side of the objective function (3.54) over the respective farms f. If the parameter estimation model contains a time trend, the CSC minimization model can be slightly modified:

$$\hat{\mathbf{c}}_0+\hat{\mathbf{c}}_1\mathbf{t}+\hat{\mathbf{Q}}\mathbf{x}_{\mathbf{t,f}}+\ \phi_{l_x}\ \hat{\mathbf{g}}_{\mathbf{0}}'\mathbf{z}_{\mathbf{t,f}}+\phi_{l_x}\ \hat{\mathbf{g}}_{\mathbf{1}}'\mathbf{z}_{\mathbf{t,f}}\mathbf{t}\ +\tfrac{1}{2}\mathbf{z}_{\mathbf{t,f}}'\hat{\mathbf{F}}\ \mathbf{z}_{\mathbf{t,f}}\ \phi_{l_x}+\ \hat{\mathbf{H}}'\mathbf{z}_{\mathbf{t,f}}+\mathbf{W}_{\mathbf{t,f}}\hat{\mathbf{h}}_{\mathbf{t,f}}=\mathbf{p}_{\mathbf{t,f}}+\boldsymbol{\rho}_{\mathbf{dx}_{\mathbf{t,f}}} \quad \text{(3.55\_t)}$$

$$\left(\boldsymbol{\phi}'(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,f}}+\mathbf{h}_{\mathbf{t,f}})\right)\ \hat{\mathbf{g}}_0+\left(\boldsymbol{\phi}'(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,f}}+\mathbf{h}_{\mathbf{t,f}})\right)\ \hat{\mathbf{g}}_1\mathbf{t}+\hat{\mathbf{F}}\ \mathbf{z}_{\mathbf{t,f}}\left(\boldsymbol{\phi}'(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,f}}+\mathbf{h}_{\mathbf{t,f}})\right)+\ \hat{\mathbf{H}}\ \mathbf{x}_{\mathbf{t,f}}+\mathbf{T}_{\mathbf{t,f}}\hat{\mathbf{s}}_{\mathbf{t,f}}=-\mathbf{y}_{\mathbf{t,f}}+\boldsymbol{\rho}_{\mathbf{dz}_{\mathbf{t,f}}} \quad \text{(3.56\_t)}$$

If the modeler wishes to perform validation and simulation using a TNR maximization framework instead, for better interpretability of the dual values, the following model can be solved, like (3.17) – (3.18):

$$max_{\mathbf{x}_{\mathbf{t,f}},\mathbf{z}_{\mathbf{t,f}}}\ TNR_{t,f}=\mathbf{p}_{\mathbf{t,f}}'\mathbf{x}_{\mathbf{t,f}}-\hat{\mathbf{c}}'\mathbf{x}_{\mathbf{t,f}}-\tfrac{1}{2}\mathbf{x}_{\mathbf{t,f}}'\hat{\mathbf{Q}}\ \mathbf{x}_{\mathbf{t,f}}-\left(\boldsymbol{\phi}'\mathbf{x}_{\mathbf{t,f}}\right)\ \hat{\mathbf{g}}'\ \mathbf{z}_{\mathbf{t,f}}\ -\tfrac{1}{2}\mathbf{z}_{\mathbf{t,f}}'\hat{\mathbf{F}}\mathbf{z}_{\mathbf{t,f}}\left(\boldsymbol{\phi}'\mathbf{x}_{\mathbf{t,f}}\right)-\mathbf{z}_{\mathbf{t,f}}'\hat{\mathbf{H}}\ \mathbf{x}_{\mathbf{t,f}}-\mathbf{x}_{\mathbf{t,f}}'\mathbf{W}_{\mathbf{t,f}}\hat{\mathbf{h}}_{\mathbf{t,f}}-\mathbf{z}_{\mathbf{t,f}}'\mathbf{T}_{\mathbf{t,f}}\hat{\mathbf{s}}_{\mathbf{t,f}} \quad (3.58)$$

subject to $$\mathbf{z}_{\mathbf{t,f}}-\mathbf{V}_{\mathbf{t,f}}\hat{\mathbf{u}}_{\mathbf{t,f}}\leq\ \mathbf{b}_{\mathbf{t,f}} \qquad [\mathbf{y}_{\mathbf{t,f}}] \quad (3.59)$$

Again, if a time trend is included in the estimation model, the TNR maximization model can be adapted to distinguish between time-variant and time-invariant linear cost function parameters, by substituting vectors $\hat{\mathbf{c}}$ and $\hat{\mathbf{g}}$ in the objective function (3.58) by vectors $\hat{\mathbf{c}}_{\mathbf{0}}$ and $\hat{\mathbf{c}}_{\mathbf{1}}\mathbf{t}$ and vectors $\hat{\mathbf{g}}_{\mathbf{0}}$ and $\hat{\mathbf{g}}_{\mathbf{1}}\mathbf{t}$, respectively.

4. Model Implementation

In this chapter we apply the approach developed in the previous chapters to estimate an activity level dependent cost function as well as an activity level and fixed-input level dependent cost function using farm data from the EU's Farm Accountancy Data Network (FADN). After a concise description of the farm samples, we outline how the data was prepared and aggregated to be used in the modelling framework. Then, the results of the cost function estimation are presented, focussing on the estimated parameters themselves, the error terms, the results of the validation phase and the simulated elasticities of supply.

4.1. Farm Data

4.1.1. Descriptive Statistics

The farm dataset used for applying the EMP models in this thesis is an unbalanced panel of farms received from the FADN. The selected farm samples cover three of the most important agricultural regions of the EU, namely Central France for crop production, West France for dairy production and South-West France for beef production, in the years from 2005 to 2012. Table 1 provides an overview of several (sub)regional farm characteristics in 2010 and, for comparison, the French national average:

Region	Sub-Region	Farmland in region (UAA in 1000 ha)	Average farm size (UAA in ha)	Average standard output (in 1000 EUR)	share of cereal specialist farms (%)	share of dairy specialist farms (%)	share of beef specialist farms (%)	livestock units per ha UAA (LU/ha)
Central France	Champagne-Ardenne	1567	62.5	177.1	19	3	2	0.34
	Picardy	1347	95.8	166.7	28	8	4	0.39
	Upper Normandy	735	67.4	105.8	21	11	11	0.71
	Centre	2303	92.2	119.5	48	3	7	0.31
	Lower Normandy	1208	50.6	83.2	7	25	19	1.23
	Burgundy	1799	86.7	127.2	23	2	26	0.60
West France	Pays de la Loire	2131	61.2	150.8	10	18	16	1.62
	Brittany	1695	47.6	169.4	10	30	8	2.84
	Poitou-Charentes	1699	67.6	108.5	33	4	7	0.58
South-West France	Aquitaine	1392	34.2	90.5	19	4	9	0.77
	Midi-Pyrénées	2481	53.0	57.9	25	5	17	0.56
	Limousin	845	57.3	48.3	2	3	50	1.04
France		27814	53.9	98.3	17	10	12	0.81

***Table 1: Farm statistics at NUTS2 level for the main EU cropping, dairy and beef regions and the French average in 2010**. Source: Eurostat (2021d).*

As table 1 reveals, the three NUTS 1 regions together cover almost 70% of the entire French farmland. The share of cereal, dairy or beef specialist farms indicates the share of farms that generate more than two thirds of their standard output with cereals, dairy products, or beef carcasses. Especially the subregions Centre, Brittany, and Limousin exhibit a high degree of specialization for cereal, dairy, and beef production, respectively. Besides having highly specialized farms, all six subregions of the Central France region belong to the largest cereal, oilseed, and protein (COP) producers in the EU. Together, they account for ca. 10% of the total EU's cereal production and almost 22% of the total EU's pulses and oilseed production on average over the years 2005-2012 (Eurostat, 2021c). Likewise, on the 2005-2012 average, the West France region produces almost 6% of the EU's cow milk, with Brittany being the subregion with the highest cow milk production in the entire EU *(Eurostat, 2021g, 2021f)*. During the same period, more than 6% of all bovine animals in the EU are raised in West France, and almost 4% in the subregions Aquitaine, Midi-Pyrénées and Limousin in South-West France (Eurostat, 2021b). These figures together with table 1 highlight the weight of the Central, West, and South-West regions of France within the context of European agriculture, but they also illustrate the regional differences between farming activities. However, when estimating a cost function with a sample of farms, a crucial assumption is that all farms in this sample share the same cost function. If this is the case, one would expect a certain degree of similarity between farms with respect to the geographic, technological and climatic conditions and the agricultural specialization. This prompted us to a) split the French farm data at NUTS 1 level to receive three regional farm datasets and b) select for each region only those farms with the specialization that is dominant in the respective region. The rules according to which farms are defined as either crop farms, dairy farms or cattle farms are outlined in Annex III. Following this categorization, we receive three unbalanced panel datasets for the years 2005 - 2012:

- on average, 698 crop farms per year from Central France (Champagne-Ardenne, Picardy, Upper Normandy, Centre, Lower Normandy, Burgundy),[13]
- on average, 238 dairy farms per year from West France (Pays de la Loire, Brittany, Poitou-Charentes),
- on average, 238 cattle farms per year from South-West France (Aquitaine, Midi-Pyrénées, Limousin).

Table 2 depicts the fixed-resource endowment of the specialist crop, dairy and cattle farms in the selected farm samples on average over the years 2005-2012. Note that the size of farms measured as the sum of cropland and grassland in ha UAA is in each subregion significantly higher for the selected samples than for the corresponding farm population as depicted in table 1. This disparity results from the fact that only those farms which are large enough to be considered commercial are represented in the FADN (Neuenfeldt and Gocht, 2014). An obvious difference between all three farm types and regions is that crop farms in Central

[13] The crop farm sample is later split in two samples, one containing the crop farms in the northern subregions, the other the crop farms in the southern subregions of Central France (see chapter 4.1.2.4)

France have a fundamentally higher share of cropland compared to grassland, whereas dairy farms in the West have roughly equal shares and cattle farms in the South-West have considerably more grassland than cropland. Furthermore, the land-unpaid labour ratio reveals that dairy farms in the West have the highest unpaid labour intensity (relative to their total farmland), followed by cattle farms in the South-West, whereas the unpaid labour intensity of crop farms in the Centre region is rather low. Note that the standard deviation of grassland belonging to crop farms is larger than the mean, which indicates that some farms must have considerably more grassland than the average, while others do not have grassland, at all. Another important characteristic, even though not displayed by table 2, are the standard deviations within farms, that is, between different years for the same farm. On average over all farms in all three farm samples the standard deviation *within* a farm is 4.1 for cropland, 2.6 for grassland, and 0.08 for unpaid labour. This underpins that the largest share of the standard deviation presented in table 2 is caused by differences between farms, whereas the fixed-resource endowment of each farm does not change much between years.

Farm Type	Region	Sub-Region	Cropland		Grassland		Unpaid Labour		Land / Unpaid Labour
			mean in ha	std in ha	mean in ha	std in ha	mean in AWU	std in AWU	ha/AWU
Crops	Central France	Champagne-Ardenne	154.2	72.4	7.0	20.0	1.4	0.6	117.6
		Picardy	146.1	76.9	6.6	12.1	1.4	0.6	108.3
		Upper Normandy	148.1	81.9	12.8	15.5	1.4	0.6	118.0
		Centre	161.1	78.0	6.6	21.7	1.3	0.5	132.1
		Lower Normandy	129.8	66.6	12.5	20.5	1.4	0.6	98.9
		Burgundy	182.5	88.5	14.4	25.9	1.4	0.6	136.9
Dairy	West France	Pays de la Loire	39.4	28.7	46.4	24.4	1.9	0.8	45.8
		Brittany	39.2	25.2	33.8	18.4	1.8	0.7	41.4
		Poitou-Charentes	54.2	39.2	42.0	25.0	1.7	0.7	55.3
Cattle	South-West France	Aquitaine	12.5	12.9	57.7	33.1	1.4	0.5	51.1
		Midi-Pyrénées	12.6	11.4	77.7	51.0	1.3	0.5	68.4
		Limousin	17.2	17.4	99.2	45.1	1.5	0.7	78.4

Table 2: Means and standard deviations of fixed-resource endowments of all farms in the three farm samples, over the years 2005-2012. *The land labour ratio is the ratio between the sum of the subregional means of cropland and grassland (=total land) and the subregional mean of unpaid labour in AWU. Source: own calculation based on EU-FADN – DG AGRI.*

Farm Type	Region	Sub-Region	Share of farmland in % dedicated to the respective activities												
			common wheat	durum wheat	barley	oats	grain maize	other cereals	oilseeds	dry pulses	potatoes	sugar beet	other industrial crops	fresh vegetables	grassland
Crops	Central France	Champagne-Ardenne	29.4	0.0	23.9	0.1	1.9	0.1	16.7	2.3	1.2	8.1	0.9	0.2	4.3
		Picardy	45.9	0.0	9.9	0.3	2.7	0.0	11.2	5.3	3.5	9.9	1.0	1.6	4.3
		Upper Normandy	41.0	0.0	6.7	0.4	1.9	0.0	12.6	3.5	7.7	4.9	7.7	0.1	8.0
		Centre	35.7	6.2	12.1	0.3	6.3	1.2	22.0	2.2	0.8	1.4	0.2	0.7	4.0
		Lower Normandy	39.5	0.4	9.1	1.0	4.0	0.5	10.8	6.1	0.9	7.8	5.3	0.2	8.8
		Burgundy	33.8	0.2	22.0	0.5	2.1	0.2	25.2	1.6	0.4	0.6	0.1	0.4	7.3
Dairy	West France	Pays de la Loire	14.0	0.2	1.7	0.1	1.8	2.8	1.3	0.3	0.0	0.0	0.0	0.0	54.1
		Brittany	14.2	0.0	2.8	0.2	2.2	3.4	0.9	0.1	0.1	0.0	0.0	0.4	46.3
		Poitou-Charentes	14.0	0.5	2.5	0.1	5.0	2.1	3.6	0.4	0.0	0.0	0.0	0.0	43.7
Cattle	South-West-France	Aquitaine	1.1	0.0	0.8	0.2	3.8	3.5	0.3	0.0	0.0	0.0	0.0	0.0	82.2
		Midi-Pyrénées	2.2	0.0	2.7	0.1	1.4	2.0	0.2	0.1	0.0	0.0	0.0	0.0	86.1
		Limousin	2.7	0.0	2.0	0.3	0.4	4.3	0.0	0.0	0.0	0.0	0.0	0.0	85.2

Table 3: Share of farmland dedicated to the production of different crops in % of total farmland. *Average over all farms in the three farm samples from 2005-2012. The rows do not add up to 100% as a) only the most important crops are displayed and b) fodder crops are not displayed. Source: own calculation based on EU-FADN – DG AGRI.*

Table 3 indicates that the most important crops produced by the crop farms in the central France are common wheat, oilseeds, and barley, as well as sugar beets, for most subregions. The other industrial crops with a comparably high share in Upper and Lower Normandy are mainly flax for fibre production, while oilseeds are mainly rapeseed for all regions in Central France and Brittany and sunflowers for the subregions in South-West France, Pays de la Loire and Poitou-Charentes. As for crop farms, the most important cereal produced by dairy and cattle farms is common wheat, followed by barley and other cereals (such as rye), even though their shares are rather small.

The rows in table 3, especially for the dairy farms, do not add up to 100. The reason for this is mainly that only those cropland outputs that are sold are considered in this study. Therefore, fodder crops that are used on the farm, such as maize for silage, are ignored.

Table 4 shows the average number of livestock units (LU) per farm in each subregion. While it might appear surprising that in all subregions the specialized cattle farms have, on average, more dairy cows (in LU) than cattle, it is important to note that the categorization of the farm type depends on the (economic) standard output generated by the respective activities and not on the number of LU. Besides, one should keep in mind that calves, dairy cows, cattle for fattening and heifers are, by nature, to some extent complements in production. Finally, table 4 indicates that, in all subregions, the rearing of non-bovine animals plays only a marginal role for the farms in the samples.

Farm Type	Region	Sub-Region	Animals in LU per farm								
			calves	dairy cows	cattle	heifers	pigs	poultry	sheep	goats	horses
Crops	Central France	Champagne-Ardenne	0.0	5.0	4.6	1.6	0.0	0.4	0.5	0.0	0.0
		Picardy	0.0	5.9	5.9	1.4	0.9	1.6	0.3	0.0	0.1
		Upper Normandy	0.0	11.0	11.4	3.4	0.0	3.2	0.1	0.1	0.6
		Centre	0.0	3.0	2.2	0.5	1.8	3.2	0.5	0.2	0.1
		Lower Normandy	0.0	6.4	5.7	2.5	0.5	2.2	1.5	0.0	0.4
		Burgundy	0.1	5.9	6.1	1.7	1.3	0.7	1.5	0.0	0.3
Dairy	West France	Pays de la Loire	0.8	53.9	31.4	10.9	0.4	3.4	0.2	0.0	0.1
		Brittany	0.6	54.5	27.3	8.1	2.8	0.5	0.0	0.0	0.2
		Poitou-Charentes	0.3	62.6	26.0	8.9	0.1	0.0	0.0	0.0	0.1
Cattle	South-West France	Aquitaine	7.8	48.8	23.2	10.1	0.8	0.0	0.8	0.0	0.4
		Midi-Pyrénées	4.0	52.8	30.5	9.4	0.1	0.0	0.7	0.0	0.1
		Limousin	0.6	64.0	49.0	13.3	0.5	0.5	2.3	0.0	0.2

Table 4: Animals in livestock units (LU) per farm. *Average over all farms in the sample from 2005-2012. Source: own calculation based on EU-FADN – DG AGRI.*

4.1.2. Data Aggregation and Preparation

Before being able to apply the estimation procedures (2.79) – (2.83) or (3.47) – (3.53) with farm data, the modeller must specify several attributes, such as the level of aggregation for decision variables and output prices, the units of measurement for the decision variables and the fixed-inputs considered.

4.1.2.1. Output aggregation

The datasets received from the FADN distinguish between 22 different marketable crops and 27 different types of animal outputs. While some of them, such as common wheat or dairy cows, are of high importance for the farms in the sample, others, such as ornamental flowers or rabbits, are economically negligible. Let us also recall that for the cost function specification with only activity levels as decision variables, the number of parameters to be estimated is $I + I(I + 1)/2$, which is 1274, for 49 activities. For the fixed-input and output level dependent cost function, which also considers fixed-input levels and the interactions between fixed-input levels and activity levels, this number is even higher. In other words, to avoid potential under-determinacy of the estimation problem, the number of activities to choose from must be drastically reduced, even for comparably large sample sizes with several hundreds or even thousands of observations. Besides, the rationale for estimating a flexible cost function is also to capture potential interactions between different activities. It is unlikely that, for example, the wheat production decision of a specialized crop farm is strongly influenced by price changes for dairy products (provided that the wheat price remains unaffected), or that the price for potatoes has a strong effect on the level of beef produced by a grassland-based cattle farm. We therefore deemed it necessary to aggregate outputs in a different way for crop farms and the livestock farm types, receiving the following aggregated categories:

Crop Farms	Dairy Farms	Cattle Farms
- animal outputs - pulses and oilseeds - industrial crops - cereals - other crops	- dairy products - other bovine animal outputs - other non-bovine animal outputs - crops	- dairy products - other bovine animal outputs - other non-bovine animal outputs - crops

***Table 5: Aggregated output categories.** Source: own presentation based on Henry de Frahan et al. (2015).*

Annex III provides more details on the aggregation scheme. While for crop farms we distinguish between four different crop categories and aggregate all animal outputs, for dairy and cattle farms all crop outputs are aggregated. The output category “dairy products” aggregates all outputs from dairy cows, but dairy cows as an output themselves (e.g., if they are sold) are categorized as “other bovine animal outputs”. This distinction may be important for some policy evaluation scenarios, as we will see in chapter 5 in the case of greenhouse gas emission constraints. At the same time, it points towards the more general challenge of specifying decision variables in a coherent way for crops, animal products and living animals.

4.1.2.2. Output Level and Output Price Specification and Retrieval

For the Monte Carlo simulation in chapter 3.4 it was possible to interpret activity levels as land allocations. Britz and Arata (2019) use the share of land (instead of absolute values) dedicated to the production of a crop as the decision variable. Using the absolute or relative land allocation certainly is appealing, as it reflects the decision that a farmer makes on her farm, especially if variable inputs are ignored, as in this thesis. Britz and Arata (2019) focussed only on specialized crop farms and removed all observations of farms with livestock. However, the absolute value or relative share of land dedicated to a certain crop is unlikely to represent animal output levels adequately and therefore is unsuitable for this thesis, as two of the three selected farm types are specialized on livestock activities. Therefore, we measure activity levels, and thus the decision variable, as the *aggregated sales quantity in base-year values*. Their computation requires that a measure for output prices is available. Figure 5 provides a schematic overview on how observed activity levels and output prices are retrieved from FADN data or, if output prices are missing, imputed with prices from Eurostat or the OECD:

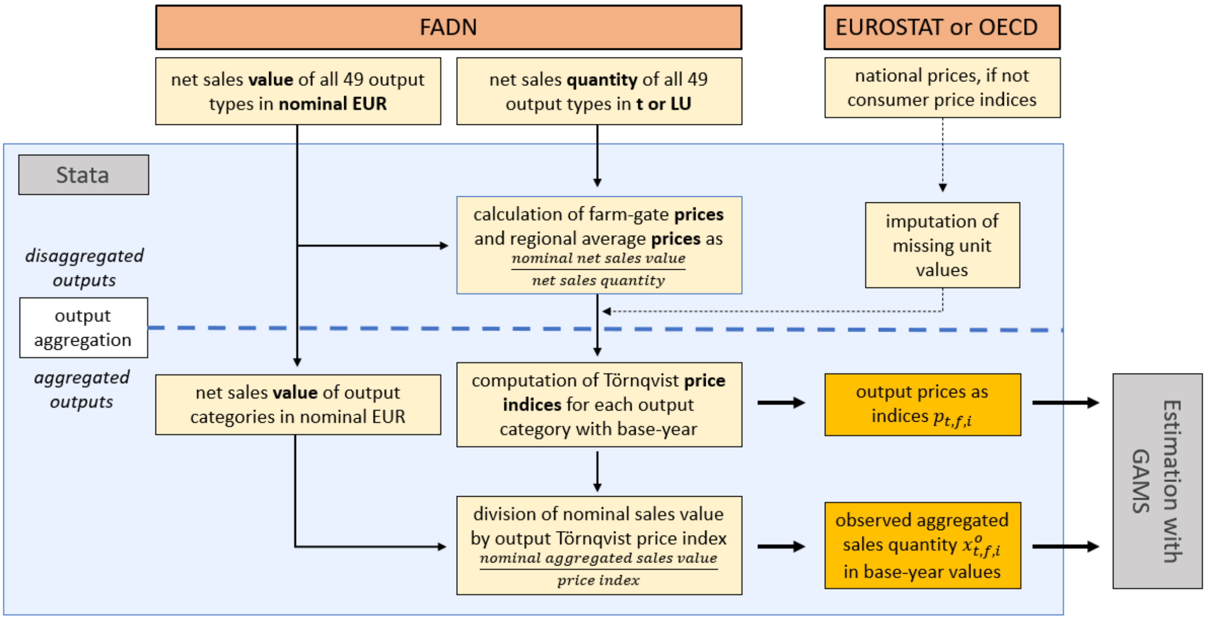

Figure 5: Schematic representation of the activity level and output price retrieval from raw data. *Calculation of the aggregated sales quantity in base-year values and the output price index. Source: own visualization based on Henry de Frahan et al. (2015).*

The calculation and the retrieval of observed activity levels and output prices is performed within an existing FADN data preparation framework that uses the software package Stata and that was developed and documented by Henry de Frahan *et al.* (2015). The observed decision variable $x^o_{t,f,i}$ is calculated as the aggregated observed output sales value divided by the Törnqvist output price index. The net sales value is the sales value corrected for variation in output stock values, to ensure that the sales value truly reflects output production of the respective year. The Törnqvist index is a changing-weight index (see, for example, Dean *et al.* (1996)) that is used in the context of this thesis to aggregate activity prices at regional level. The reason for this aggregation is that farm-level aggregate price indices prove too erratic, as explained by Henry de Frahan *et al.* (2015), who use the same farm data. In their Stata routine, the Törnqvist price index for outputs is constructed as follows:

$$\tau_{mrt} = \prod_{n=1}^{N_m} \left(\frac{p_{nrt}}{p_{nrt_0}}\right)^{\frac{g_{nrt}+g_{nrt_0}}{2}} \tag{4.1}$$

where $$g_{nrt} = \frac{V_{nrt}}{\sum_{k=1}^{N_m} V_{krt}} \quad \text{and} \quad g_{nrt_0} = \frac{V_{nrt_0}}{\sum_{k=1}^{N_m} V_{krt_0}} \tag{4.2}$$

The index τ_{mrt} is the Törnqvist price index for the aggregated output category m in region r at time t. The number N_m represents the number of output components n which constitute the output aggregate m. The variable p_{nrt} denotes the average farm-gate price of output component n in the respective region r at time t. The subscript t_o represents the base-year, which is the year 2005, in our case. The variable V_{nrt} is the total revenue generated by all farms in region r with output component n in year t. This variable is divided by the revenue generated in this region r and year t with all output components of output aggregate m. Con-

sequently, the variable g_{nrt} is the share of output aggregate m's revenue that is generated with output component n in a region r in year t, while the variable g_{nrt_0} is this share for the reference year. As outlined by figure 5, the annual average farm-gate price p_{nrt} in region r is obtained by dividing the nominal value of the total production of output n in this region r by the number of units sold. The Stata routine developed by Henry de Frahan *et al.* (2015) comprises a mechanism for the removal of extreme farm-gate prices and their replacement by regional or national prices, along with the imputation of missing values following a scheme that is described in Annex III. The computed price index τ_{mrt} is then used to calculate the aggregated activity level, as depicted in figure 5, but it also serves as the output price $p_{t,f,i}$ for the cost function parameter estimation. The rationale behind this calculation process is, besides aggregation, to receive an inflation-adjusted measure for observed activity levels that can be used for all types of farm outputs. If the nominal aggregated output price increases, this would also raise the nominal aggregated sales value, even if produced quantities remain constant. However, by dividing the sales value by the Törnqvist price index, the activity level $x^o_{t,f,i}$ would only increase if the relative raise in sales value is larger than the relative increase of prices, indicating that a farm produces higher quantities of an activity.
Before the calculated, "observed" activity levels $x^o_{t,f,i}$ are provided as inputs for other applications, the Stata program scales them, to ensure numerical stability for the solver:

$$x^{o\,(scaled)}_{t,f,i} = x^o_{t,f,i} * s_{t,f,i} \tag{4.3}$$

$$s_{t,f,i} = 10^{\left(-int\left[log_{10}\sqrt{E[x^o_{t,f,i}]^2 + Var[x^o_{t,f,i}]}\right]\right)} \tag{4.4}$$

where $\mathrm{int}[\,]$ is the nearest integer, so that by multiplying $x_{t,f,i}$ with $s_{t,f,i}$ the mean squared error of the scaled activity lies between 1 and 10. Consequently, $s_{t,f,i}$ may be different for each activity. Note that, when maximizing the TNR according to equations (3.58) – (3.59), it is necessary to maximize the *unscaled* TNR, to receive interpretable shadow prices. However, the maximization of the unscaled TNR would only yield the same optimal activity levels as the CSC minimization (3.54) – (3.57) if the scale factor is the same for all activities. Therefore, we scale all activities by the same factor before estimating the activity level and fixed-input dependent cost function.[14] Table A.4 in Annex III provides an overview of the aggregated activity levels $x^o_{t,f,i}$ in each region.

4.1.2.3. Retrieval of Fixed-Input Levels and Fixed-Input Prices

The fixed-inputs considered in this thesis are cropland, grassland, and unpaid labour. Again, the data preparation was implemented with the Stata code developed by Henry de Frahan *et*

[14] For the output cost function that depends on activity levels only, we just apply the CSC minimization, as we are not using it for the policy simulation. Therefore, for the estimation of this cost function it is tolerable if different activities are scaled by different scale factors.

al. (2015). The information on all fixed-input levels is retrieved from the FADN. Cropland is the total utilized agricultural area (UAA) of a farm subtracted by the area that is leased to others and by temporary grassland, permanent pasture, as well as rough grazing land. The sum of the latter three constituents is the fixed-input grassland. Accordingly, if a farm leases out some of its land, then the sum of the two land types considered in the simulation (cropland and grassland) is smaller than the total UAA owned by a farm. Excluding leased-out farmland from the resources available in the mid-term can be justified by the fact that French tenant protection laws stipulate long-term land contracts (Ciaian *et al.*, 2012). The third fixed-input, unpaid labour, is measured in annual work units (AWU) and refers to the family labour that is remunerated by a farm's total net revenue, like other fixed-inputs.
As mentioned above (e.g., table 3), the sum of the land allocations for each crop (or for all aggregates) is in some cases smaller than the total available cropland, since fodder crops are not incorporated in either output category. To ensure that the cropland constraint is binding for observed activity levels (to justify strictly positive observed resource prices $y^o_{t,f,cropland}$) the available resource level $b_{t,f,cropland}$ is defined as the sum of the cropland allocated to the aggregates. The available resource levels $b_{t,f,grassland}$ and $b_{t,f,unpaid\ labour}$ are equal to the available grassland and unpaid labour, as retrieved from FADN and as explained above. The cost function specification that involves activity levels and fixed-input levels requires values for the observed used fixed-input levels $z^o_{t,f,j}$. In practice, usually it is impossible to distinguish between observed used fixed-input levels (which appear as $z^o_{t,f,j}$ in the model) and available fixed-input levels ($b_{t,f,j}$). Therefore, we set $z^o_{t,f,j} = b_{t,f,j}$, lacking any other information on fixed-resource usage. Remember that it is nevertheless important to make a conceptual distinction between available fixed-resources and utilized fixed-resources. We must allow for the case that, in simulation, not all of the available fixed-input levels are used, leading to simulated shadow prices of zero. One may wonder why the differentiation between "observed" and "true" utilized fixed-input levels is not equally applied to available fixed-input levels $b_{t,f,j}$, given that utilized and available fixed-input levels are prone to the same kind of measurement errors. In fact, we conceptualized and tentatively implemented a model in which observed and true available fixed-input levels are linked via an error term $r_{t,f,j}$, so that $b_{t,f,j} = b^o_{t,f,j} + r_{t,f,j}$. However, this extension is not further considered in this thesis.[15]
In the activity level-dependent cost function specification (model 2.79 – 2.83) the product $\mathbf{A_{t,f,j}x_{t,f}}$ indicates the fixed-input usage and $\mathbf{A_{t,f,j}}$ is defined so that $\mathbf{A_{t,f,j}x^o_{t,f}} = b_{t,f,j}$. In section

[15] The results of the first implementation trials are not convincing enough to pursue this extension further for this thesis. Besides, a general reservation concerns the correlation between the error terms $\mathbf{r_{t,f,j}}$ (fixed-resource availability) and $\mathbf{s_{t,f,j}}$ (fixed-resource usage). It seems likely that, if the fixed-resource usage is under- or overestimated, so is the fixed-resource availability, especially if the data does not allow to distinguish between fixed-resource usage and availability.

3.5.1 we already described the calculation of the individual elements of $\mathbf{A_{t,f,j}}$ (equations 3.23 and 3.24). FADN provides information on allocation of cropland to (disaggregated) activities, as depicted in table 3. The land allocated to each of the crop-related aggregated activities can therefore easily be calculated. The fixed-input-output coefficient $a_{t,f,cropland,i}$ for each farm in a given year is then computed as the cropland allocated to aggregate i divided by the observed activity level $x^o_{t,f,i}$ of aggregate i, so that $a_{t,f,cropland,i}$ represents the cropland required to generate one quantity unit of output i measured in base-year values. As fodder crops are not incorporated in either output category (since their sales value is unknown if they are used on the farm, as it usually is the case) cropland is not allocated to either of the animal output categories. Instead, the entire grassland is allocated to the aggregated animal outputs. For crop farms this implies that all grassland is used for the generation of just one output category (animal outputs), whereas for cattle and dairy farms grassland must be distributed over the production of dairy products, other bovine animals and non-bovine animals (see equation 3.23). Consequently, the requirement of grassland *per activity level* is identical for all animal outputs, while their absolute usage is proportional to their activity levels, since $\mathbf{A_{t,f,j}x^o_{t,f}} = b_{t,f,j}$. It certainly is a bold assumption that animal outputs do not require any cropland. However, tables 2 and 3 indicate that, especially for cattle farms in the South-West region, the production of bovine animals is indeed mainly grassland-based. The fixed-resource unpaid labour is required by all output activities and, as for grassland, the resource requirements per activity level are identical.

When selecting suitable "observed" fixed-input prices $\mathbf{y^o_{t,f}}$ for cropland and grassland, information typically comes either in the form of (annual) rental prices or sale prices. In the EMP approach suggested by Henry de Frahan (2019) $\mathbf{y^o_{t,f}}$ is interpreted as the observed value of the shadow price of the resource constraint. The shadow price of a constraint in a TNR maximization problem that is solved annually (we assume that farmers make their production decision on an annual basis) indicates by how much the *annual* TNR could increase if the constraint was loosened by one unit. In other words, the shadow price of the land constraint is the *annual* value a farm would be willing to pay for one more unit of land. In a perfectly competitive market this value would be the rental price. Rental prices for grassland and cropland are obtained from Eurostat (2018) at national level, which is why $\mathbf{y^o_{t,f}}$ is identical for every farm. Presumably, this aggregation represents the most important driver of a potential deviation between the observed rental price and the "true" fixed-input price.[16]
The price for unpaid labour could be interpreted as the opportunity costs of a family member's labour, which is assumed to be at least the national minimum wage. We calculate it as

[16] Note that a potential pitfall might stem from the fact that *all* farms in a region might face systematically higher or lower "true" fixed-input prices compared to the observed national rental price, as one would expect land prices in a region to be correlated.

the minimum annual wage for France in the years 2005-2012 (OECD, 2021), divided by the national purchasing power parity, as retrieved from the Penn World Table.

4.1.2.4. Final Sample Preparation

The scaled observed activity levels $x^o_{t,f,i}$, the output price indices $p_{t,f,i}$, the fixed-input availability $b_{t,f,j}$, the observed fixed-input usage $z^o_{t,f,j}$, the observed fixed-input price $y^o_{t,f,j}$ and the matrix $\mathbf{A}_{\mathbf{t,f,}}$ of fixed-input-output coefficients can finally be used for the cost function estimation, following the procedure described in the previous chapters.

While some of the subregions of Central France, such as Upper and Lower Normandy, are located in the maritime north of France, other subregions, such as Centre and Burgundy, extend several hundreds of kilometres further south (IGN, 2021). To account for the geographic and potentially climatic heterogeneity, we decide to split the crop farm sample in farms from northern Central France (Champagne-Ardenne, Picardy, Upper Normandy, and Lower Normandy) and from the southern subregions of Central France (Centre and Burgundy). Besides, we detect observations with invalid values: In every sample some farms exhibit negative “observed” activity levels, which presumably results from sale of stock without production in the respective year. Every observation for which at least one activity level is strictly negative is discarded from the initial farm sample, which affects per year on average 2.1 farms of the crop farms from northern Central France, 0.6 crop farms from southern Central France, 11.5 dairy farms from West France and 54.6 cattle farms from South-West France. The following table indicates the size of the samples that is finally used for parameter estimation:

year	Central France (south)	Central France (north)	West France	South-West France
	crop farms	crop farms	dairy farms	cattle farms
2005	276	378	188	152
2006	303	428	229	186
2007	291	428	213	200
2008	291	442	229	173
2009	284	445	229	182
2010	265	409	244	196
2011	260	412	241	176
2012	244	408	235	203
sum	2214	3350	1808	1468
mean	276.75	418.75	226	183.5
(std)	(19.42)	(21.77)	(18.01)	(16.80)

Table 6: Sample size of data sets over the years used for parameter estimation. *Source: own calculation based on EU-FADN – DG AGRI.*

4.2. Results

This section presents the results of the cost function estimation, applied on the crop, dairy, and cattle farm samples from Central, West and South-West France, respectively, over the years 2005-2012. A core objective of this thesis is the implementation of the approach suggested by Henry de Frahan (2019) to estimate a quadratic cost function that depends on activity levels. Therefore, we first apply four different model specifications based on this estimation framework, namely:

- The “pure” estimation approach, as suggested by Henry de Frahan (2019) and as outlined in equations (2.79) – (2.83).
- A farm fixed effect extension with equation (2.79_FE) instead of equation (2.79).
- A time index extension with equation (2.79_t) instead of equation (2.79).
- A combination of farm fixed effect and time index with equation (2.79_FE_t) instead of equation (2.79).

Second, we use the cost function specification that is based on activity levels *and* fixed-input levels, as presented in model (3.47) – (3.53), for all three farm types with and without a time index.

First, the results of the activity level dependent cost function are presented, followed by the results of the activity level and fixed-input level dependent cost function. For each cost function specification, three different phases can be distinguished:

1. In the estimation phase we estimate the cost function parameters, as well as the error term vectors $\mathbf{h}_{\mathbf{t},\mathbf{f}}$, $\mathbf{u}_{\mathbf{t},\mathbf{f}}$ and $\mathbf{s}_{\mathbf{t},\mathbf{f}}$ for activity levels, fixed-input prices, and fixed-input levels, respectively, with the optimization software GAMS, using the CONOPT4 solver. This solver is suitable for large-scale nonlinear optimization problems and uses the generalized reduced gradient method (McCarl *et al.*, 2016). As described in the previous chapters, the estimation procedure consists of the minimization of the sum of the squared error terms. To facilitate the finding of a feasible solution, initial values for the cost function parameters were given (usually 1). The optimization process takes several minutes for each of the different specifications of the output cost function, and multiple hours for the estimation of each of the activity level and fixed-input level dependent cost functions.
   For the artificial Monte Carlo simulation, the assessment of the estimation quality is straightforward, as the cost function parameters of the population are known. This is not the case for this simulation with actual farm data. However, the coefficient of determination (R^2) may serve as an indicator to evaluate the extent to which an estimated cost function specification fits the data. We calculate R^2 for each activity level, fixed-input price and fixed-input level as follows, for the example of activity levels:

$$R\_activity_i^2 = 1 - \frac{\sum_{f=1}^{F}\sum_{t=2005}^{2012}(\hat{h}_{t,f,i})^2}{\sum_{f=1}^{F}\sum_{t=2005}^{2012}(y_{t,f,i}^{o} - \bar{y}_{t,f,i}^{o})^2} \qquad (4.5)$$

2. In the validation phase we test if the estimated models satisfactorily replicate the observed values of a base-year, adjusted by the error terms. Calibration has been a deci-

sive property of the PMP related models since the first formalization by Howitt (1995). To validate that the estimated model calibrates, Paris (2017) introduces a CSC minimization as a “calibrating equilibrium model”. This approach is also adopted by Henry de Frahan (2019) (see equations 2.84, 3.14 and 3.15 or equations 3.54 – 3.57). For this thesis, it is implemented with the CSC minimization and with the TNR maximization model (3.17 – 3.18 or 3.58 – 3.59), which both yield identical results.
The measure of validation used in this section is the Theil’s inequality coefficient, which typically is used for forecast evaluation and compares true values with simulated values (Cook, 2019). We calculated it as proposed by Pindyck and Rubinfeld (1998). The indicator $Theil\_x_{t,i}$ is the Theil inequality coefficient for activity levels in the reference year 2012:

$$Theil\_x_{t,i} = \frac{\sqrt{\frac{1}{F}\sum_{f=1}^{F}[\left(x_{t,f,i}^{o}+\hat{h}_{t,f,i}\right)-\hat{x}_{t,f,i}]^2}}{\sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(x_{t,f,i}^{o}+\hat{h}_{t,f,i}\right)^2}+\sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(\hat{x}_{t,f,i}\right)^2}} \tag{4.6}$$

Note that the simulated values $\hat{x}_{t,f,i}$ are compared with the observed values plus the estimated error term $\hat{h}_{t,f,i}$ as, according to the interpretation of the error term, $x_{t,f,i}^{o} + \hat{h}_{t,f,i}$ represents the “true” activity level.

3. In the simulation phase we examine the response of the estimated model to a price change by calculating supply elasticities to output prices. Replication of base-year values is a necessary but not a sufficient condition for a PMP or EMP model to be satisfactory. It is equally important that its responses to changes (e.g., output price changes) largely reflect the behaviour of the farms represented by the model. In fact, a main problem of the early ad-hoc solutions that were developed to cope with the under-determinacy of PMP (see section 2.1.2) was the often arbitrary simulation behaviour, even if calibration is guaranteed (Heckelei and Britz (2000); Heckelei and Wolff (2003)). We calculate supply elasticities to prices for the two cost function specifications by simulating a 10% output price increase for one activity, while keeping the prices of all other activities constant. The percentage change of activity levels is then calculated relative to the *simulated* activity levels of the reference scenario (i.e., the simulated activity levels of the validation phase), so that the supply elasticity of output i with respect to a price change of output k is:

$$\varepsilon_{ik} = \frac{100 * \frac{\sum_{f=1}^{F}(\hat{x}_{t,f,i}-\hat{x}_{t,f,i}^{*})}{\sum_{f=1}^{F}\hat{x}_{t,f,i}^{*}}}{\%\Delta p_k} \tag{4.7}$$

where $\hat{x}_{t,f,i}^{*}$ indicates the activity level that is optimal in the reference scenario, in our case the year 2012, without price changes. The denominator $\%\Delta p_k$ represents the percentage change of the price of activity k, and is determined by the analyst.

4.2.1. Activity Level-Dependent Cost Function

Estimation Phase

Table 7 displays the coefficients of determination (R^2) for different specifications of the cost function that depends on activity levels, for all four farm samples. The R^2 values are one or close to one for all activity levels in all farm samples, implying that the non-linear model with its vectors $\mathbf{c}$ and $\mathbf{Q}$ can explain a significant share of the deviation in the observed activity levels. However, it is important to remark that the model contains a second error term, $\hat{u}_{t,f,j}$, which is why it cannot be guaranteed that the entire explanatory power suggested by the high R^2 values for activity levels is solely due to the estimated cost function parameters. The negative R^2 value for grassland in the crop farm samples and for cropland in the dairy and cattle farm samples can partly be explained by the low variance in fixed-input prices, which only vary between years, but not between farms. However, a closer look at those farms for which $(\hat{u}_{t,f,j})^2$ is particularly high reveals that $\hat{u}_{t,f,j}$ equals the negative observed shadow price, meaning that the sum of $y^{o}_{t,f,j}$ and $\hat{u}_{t,f,j}$ is zero. For crop farms, this is more the case for farms which are employed with grassland but do not have any animal outputs in the respective year, for whatever reason. For dairy and cattle farms, this situation emerges if a farm is employed with cropland but does not produce any crop outputs in the respective year. The right column in table 7 indicates the least squares (LS) values of the estimation problem, which is the sum of the weighted squared errors. The absolute LS values are not comparable between different farm samples, as a sample with more observations also has more error terms. However, for different model specifications *within* a farm sample these LS values are meaningful: They reveal that in all cases the squared error terms could be reduced by employing a farm fixed effect, a time index, or a combination of the two. The lowest LS value is given for all farm samples by the cost function specification that combines farm fixed effects and time indices.

Coefficient of Determination (R^2)												
Region	Farm Type	Specification of the Cost Function		activity levels					fixed-input prices			Least Squares (LS)
		Fixed Effect	Time Index	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour	
Central France (south)	crops	no	no	1.00	1.00	1.00	0.99	1.00	0.50	-331.54	0.83	237,395
		yes	no	1.00	0.98	0.99	0.99	1.00	0.71	-331.54	0.90	221,075
		no	yes	1.00	1.00	1.00	0.99	1.00	0.77	-331.54	0.92	215,647
		yes	yes	1.00	0.99	1.00	0.99	1.00	0.83	-331.54	0.94	210,561
Central France (north)	crops	no	no	1.00	1.00	1.00	0.99	1.00	0.63	-312.22	0.93	249,068
		yes	no	1.00	0.98	1.00	0.99	1.00	0.76	-312.26	0.95	238,690
		no	yes	1.00	1.00	1.00	1.00	1.00	0.77	-312.22	0.96	234,220
		yes	yes	1.00	0.99	1.00	0.99	1.00	0.84	-312.26	0.97	230,670
Region	Farm Type	Specification of the Cost Function		activity levels					fixed-input prices			Least Squares (LS)
		Fixed Effect	Time Index	dairy products	other bovine animal outputs	non-bovine animal outputs	crops		cropland	grassland	unpaid labour	
West France	dairy	no	no	1.00	1.00	1.00	1.00		-2.56	1.00	1.00	3,469
		yes	no	1.00	1.00	1.00	1.00		-2.56	1.00	1.00	3,394
		no	yes	1.00	1.00	1.00	1.00		-2.55	1.00	1.00	3,348
		yes	yes	1.00	1.00	1.00	1.00		-2.55	1.00	1.00	3,321
South-West France	cattle	no	no	1.00	1.00	1.00	1.00		-16.17	1.00	1.00	11,463
		yes	no	1.00	1.00	1.00	1.00		-16.04	1.00	1.00	10,990
		no	yes	1.00	1.00	1.00	1.00		-16.09	1.00	1.00	11,225
		yes	yes	1.00	1.00	1.00	1.00		-15.99	1.00	1.00	10,890

***Table 7: Precision of the estimation of the activity-level dependent cost function.** The coefficient of determination is calculated for each activity and for each fixed-input price. The right column indicates the least squares (LS) value of the estimation function. Source: own estimation.*

Figure 6 visualizes the vector of error terms $\hat{h}_{t,f,i}$ estimated with the activity-level cost function for crop farms in southern Central France. Each dot represents an estimated error term for one farm in one year and for one activity. On the horizontal axis, each point represents one farm, so that all dots on a vertical trajectory are the error terms of the same farm, over all years for which the farm produced any activities. Farms are sorted from left to right according to the mean of all error terms for this farm. For the model specification without farm-fixed effects and a time index, it appears that for some farms most of the error terms are negative (on the left side of the horizontal axis), whereas for others most activity related error terms are positive (on the right side of the horizontal axis). This indicates that there exists a systematic effect which biases *all* activity related error terms of a farm. Indeed, the use of a farm-fixed effect specification removes some farm-specific bias of the error terms.

Figure 7 shows all activity level error terms that are estimated over all years, all farms, and all activities, but sorted by year. As indicated by the sum of weighted squared errors in the right column of table 7, the incorporation of a time index seems to have a stronger effect on the LS value for most farm samples than the use of a farm fixed effect. The use of the time index seems to slightly reduce the time-specific variations in error terms, as figure 7 suggests, and as one would expect.

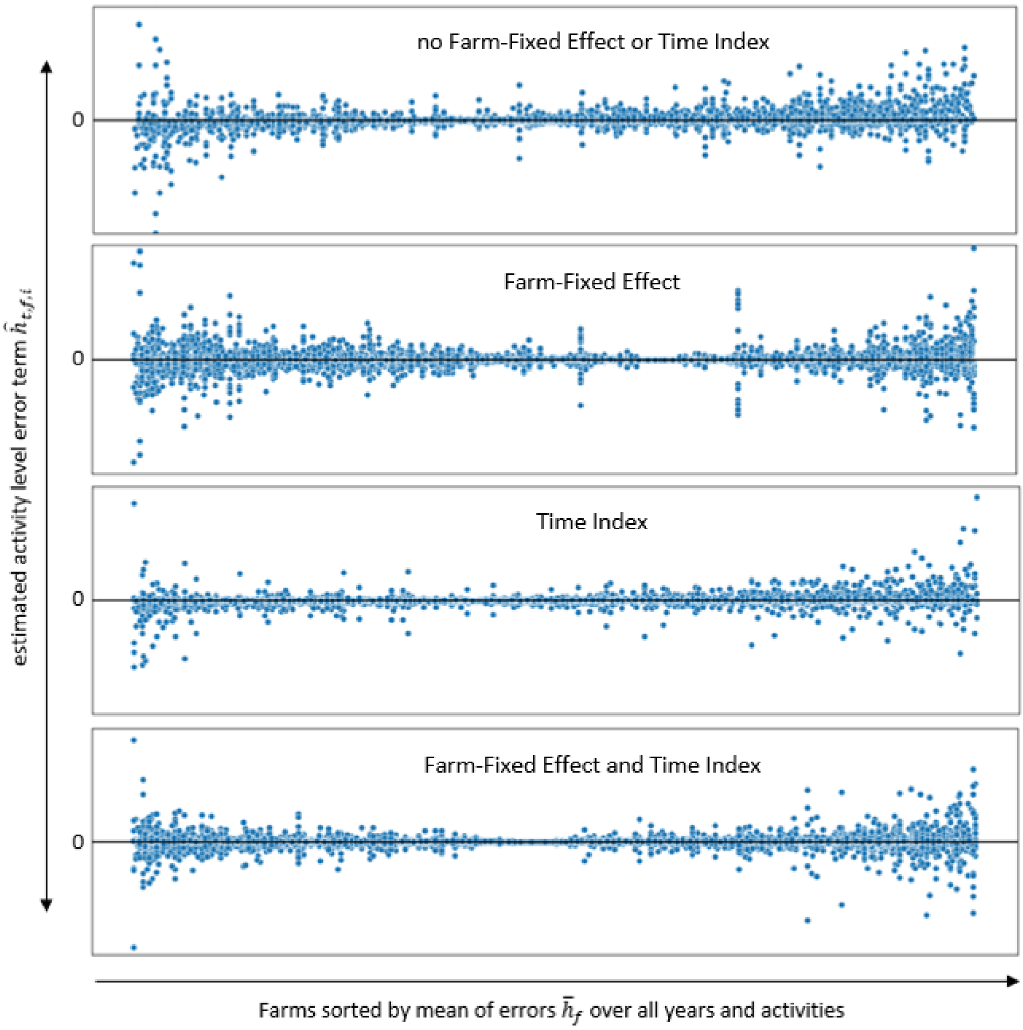

Figure 6: Estimated activity-level error terms $\hat{h}_{t,f,i}$ per farm. *Estimation for crop farms in southern Central France for different specifications of the activity-level cost function. Source: own visualization.*

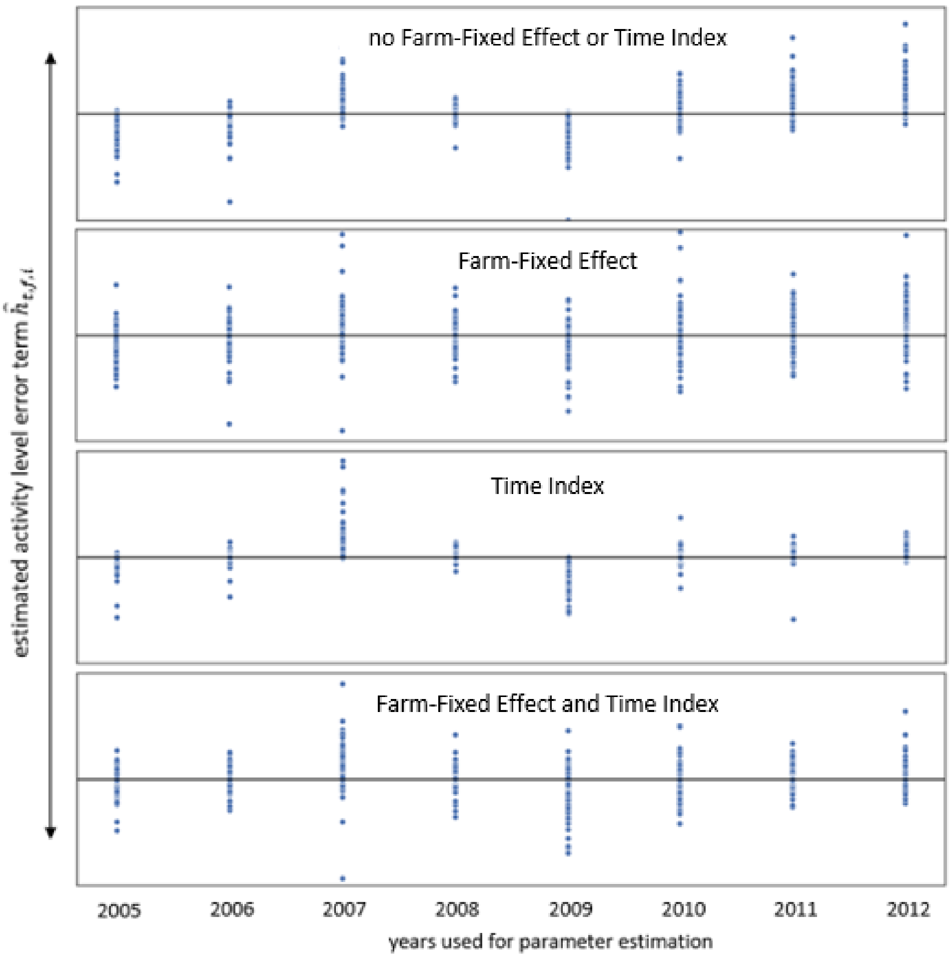

Figure 7: Estimated activity-level error terms $\hat{\mathbf{h}}_{t,f,i}$ per year*. Estimation for crop farms in southern Central France for different specifications of the activity-level cost function. Source: own visualization.*

The purpose of the entire estimation process is to receive estimates for vector $\mathbf{c}$ and matrix $\mathbf{Q}$ of the cost function parameter. The parameters determine the relationship between the different activities, which is scrutinized in the "simulation phase" section. The estimated parameters themselves are presented in Annex III.

Validation Phase

Table 8 depicts the Theil inequality coefficient for the activity levels and fixed-input prices, simulated using the activity-level dependent cost function. The closer the value is to zero, the smaller is the difference between the simulated values and the "true" values (the observed value plus the estimated error term) for the exact same exogenous parameters and variables, indicating calibration. The reference year is 2012, which means that the output prices, the fixed-input-output coefficient matrices, the fixed-input limits, the error terms and the weight matrices $\mathbf{p}_{2012,f}$, $\mathbf{A}_{2012,f}$, and $\mathbf{b}_{2012,f}$, $\hat{\mathbf{h}}_{2012,f}$, $\hat{\mathbf{u}}_{2012,f}$, $\mathbf{W}_{2012,f}$ and $\mathbf{V}_{2012,f}$ are used in the CSC minimization model (2.84), (3.14) and (3.15).

The low Theil inequality coefficient for almost all activities indicates that the model seems largely capable of replicating the "true" activity levels of a reference year. Interestingly, it appears that including solely a time index decreases the model's ability to replicate reference activity levels in some cases, while it improves the replication of fixed-input prices (e.g., for the crop farm samples).

Theil Inequality Coefficient											
				activity levels					fixed-input prices		
Region	Farm Type	Specification of the Cost Function: Fixed Effect	Time Index	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour
Central France (south)	crops	no	no	0.00	0.00	0.00	0.00	0.00	0.04	0.28	0.03
		yes	no	0.00	0.00	0.00	0.00	0.00	0.04	0.28	0.03
		no	yes	0.00	0.00	0.11	0.00	0.17	0.01	0.28	0.02
		yes	yes	0.01	0.03	0.08	0.01	0.13	0.09	0.28	0.09
Central France (north)	crops	no	no	0.00	0.01	0.04	0.00	0.20	0.04	0.31	0.03
		yes	no	0.00	0.02	0.02	0.00	0.12	0.05	0.31	0.03
		no	yes	0.00	0.00	0.02	0.01	0.17	0.01	0.31	0.01
		yes	yes	0.01	0.12	0.03	0.02	0.13	0.09	0.51	0.06
				activity levels					fixed-input prices		
Region	Farm Type	Specification of the Cost Function: Fixed Effect	Time Index	dairy products	other bovine animal outputs	non-bovine animal outputs	crops		cropland	grassland	unpaid labour
West France	dairy	no	no	0.00	0.00	0.00	0.00		0.07	0.39	0.08
		yes	no	0.00	0.00	0.00	0.00		0.05	0.39	0.09
		no	yes	0.00	0.00	0.00	0.00		0.05	0.33	0.07
		yes	yes	0.00	0.02	0.05	0.00		0.06	0.35	0.08
South-West France	cattle	no	no	0.00	0.00	0.00	0.00		0.47	0.33	0.18
		yes	no	0.00	0.00	0.00	0.00		0.47	0.34	0.20
		no	yes	0.00	0.00	0.00	0.00		0.40	0.30	0.18
		yes	yes	0.00	0.00	0.01	0.00		0.30	0.28	0.17

Table 8: Theil inequality coefficients comparing simulated and "true" values for activity levels and fixed-input prices, using the model based on the activity-level dependent cost function for the year 2012. *A coefficient of 0 indicates that all simulated values are equal to the "true" values. Source: own simulation.*

Simulation Phase

For each farm sample and for each specification of the activity-level dependent cost function, a 10% output price increase of each activity was simulated consecutively, while keeping the other output prices constant, as observed for the year 2012. Table 9 displays the resulting supply elasticities to prices for crop farms in southern Central France and dairy farms in West France.

The supply elasticities to prices for crop farms from northern Central France and for cattle farms can be found in Annex III (tables A.7 and A.8). However, the limited usefulness of the estimated models for simulation purposes becomes apparent when inspecting table 9. According to the simulated model, the supply of crops is entirely irresponsive to its own price for dairy farms. The reason for this is the fixed matrix of fixed-input-output coefficients, $\mathbf{A_{t,f}}$, which prevents any extension of crop production, as crops are the only outputs that require cropland. For the sample of crop farms the same limitation exists, even if it is expressed less obviously. As remarked above, animal outputs are the only activity that may utilize grassland. Thus, a farm that has grassland cannot extend its animal activity level. The reason why the supply elasticity to its own price is still positive is that farms without grassland (e.g., pork or poultry farms) may increase the level of animal activities. This significant shortcoming of the activity-level dependent cost function is the reason why from now on we apply the activity level and fixed-input level dependent cost function.

Cost Function Specification	Supply Elasticities																								
	fixed effect: no time index: no					fixed effect: no time index: yes					fixed effect: yes time index: no					fixed effect: yes time index: yes					mean over all cost function specifications				
	Crop Farms from southern Central France																								
price change	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
animal outputs	0.33	0.00	-0.16	0.00	0.00	0.30	0.01	-0.14	0.00	0.00	0.58	0.03	-0.54	0.00	1.57	0.48	0.03	-0.54	-0.01	1.04	0.42	0.02	-0.35	0.00	0.65
pulses and oilseeds	-0.17	3.86	0.40	-1.49	-0.69	-0.32	7.88	-0.04	-2.94	-0.42	-1.11	6.92	1.73	-3.09	2.13	-0.33	14.09	-0.07	-5.54	-1.15	-0.48	8.19	0.51	-3.26	-0.04
industrial crops	-0.11	0.09	2.27	-0.10	-4.99	-0.11	-0.06	1.58	-0.06	-2.99	-0.44	0.09	1.79	-0.29	-1.41	-0.81	-0.09	3.42	-0.16	-6.43	-0.37	0.01	2.26	-0.15	-3.96
cereals	-0.16	-3.50	-0.39	1.33	-0.80	-0.29	-6.76	-0.34	2.42	-0.66	0.09	-5.66	-4.24	2.32	1.80	-0.40	-9.05	-1.07	3.42	-0.63	-0.19	-6.24	-1.51	2.37	-0.07
other crops	-0.31	-0.18	-2.54	-0.18	10.95	-0.46	-0.11	-0.89	-0.04	4.53	-0.73	0.11	-3.77	-0.25	26.22	-0.46	-0.17	-3.43	-0.10	15.31	-0.49	-0.09	-2.66	-0.14	14.25

Cost Function Specification	fixed effect: no time index: no				fixed effect: no time index: yes				fixed effect: yes time index: no				fixed effect: yes time index: yes				mean over all cost function specifications			
	Dairy Farms from West France																			
price change	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops
dairy products	5.11	-0.47	-0.01	0.00	2.72	-0.25	-0.01	0.00	6.23	-0.58	1.15	0.00	2.77	-0.26	0.00	0.00	4.20	-0.39	0.28	0.00
other bovine animal outputs	-4.82	0.53	-9.29	0.00	-2.63	0.34	-10.00	0.00	-5.98	0.59	-4.44	0.00	-2.68	0.32	-8.28	0.00	-4.03	0.45	-8.00	0.00
non-bovine animal outputs	-0.16	-0.44	49.36	0.00	-0.98	-1.18	136.70	0.00	0.32	-0.13	10.80	0.00	-0.11	-0.34	38.56	0.00	-0.23	-0.52	58.85	0.00
crops	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 9: Elasticity of supply for crop farms in southern Central France and dairy farms in West France, calculated with the cost function that depends on activity levels only ("simple" cost function). *Elasticities were calculated for a 10% increase of the price of the activities indicated on the vertical axis. Source: own calculation*

4.2.2. Activity Level and Fixed-Input Level Dependent Cost Function

Estimation Phase

By minimizing the weighted sum of squared errors according to model (3.47) – (3.53), the parameters of the activity level and fixed-input level dependent cost function are estimated. As outlined in chapter 2, the matrix $\mathbf{Q}$ of the quadratic activity level dependent cost function parameters has always been of particular interest for modellers using PMP or EMP, as its elements are the second derivatives of the cost function with respect to the activity levels. In other words, the diagonal elements q_{ii} of this matrix $\mathbf{Q}$ indicate by how many Euros the marginal costs of activity i change if a farm changes its output of that activity level i by one unit.

Annex III provides an overview of the parameter estimates for the four farm samples, with and without a time index. Strikingly, many of the diagonal elements of matrix $\mathbf{Q}$ are close to zero (see tables A.11-A.14). This implies that marginal costs raise only negligibly if a farm increases the production of an activity, e.g., dairy outputs for the dairy farm sample. This has serious implications. A small change of the marginal revenue (i.e., output prices) would require a much more drastic change in outputs to re-establish the relationship that marginal costs equal marginal revenue[17]. This can be seen when solving the relation MC=MR (equation 3.47) for activity level $x_{t,f,i}$ and then taking the derivative of $x_{t,f,i}$ with respect to its own output price $p_{t,f,i}$:

$$\frac{\partial x_{t,f,i}}{\partial p_{t,f,i}} = \frac{1}{q_{ii}} \tag{4.8}$$

The elasticity of supply for a change of activity i's own price can then be calculated as:

$$\varepsilon_{t,i,i} = \frac{1}{q_{ii}} \frac{\bar{p}_{t_{ref},i}}{\bar{x}_{t_{ref},i}} \tag{4.9}$$

Where the variable $\varepsilon_{t,i,i}$ is the supply elasticity of activity i to its own price, the value $\bar{p}_{t_{ref},i}$ the annual average output price of activity i in the reference scenario, and the value $\bar{x}_{t_{ref},i}$ the annual average level of activity i in the reference scenario. Note that the elasticity calculated in equation (4.9) only approximates the true, simulated elasticity, as it disregards the off-diagonal element of matrix $\mathbf{Q}$. Using the cost function estimates for dairy farms with a time index and the price-activity level ratio for 2012, the elasticity for dairy outputs as calculated with equation (4.9) would take a value of almost 900 000. Therefore, to ensure that elasticities are in a more reasonable range, we impose a lower boundary q_min_{ii} for the diagonal elements of matrix $\widehat{\mathbf{Q}}$ as an *ad-hoc* solution:

[17] In principle, it is also possible that marginal costs with respect to activity levels are adapted through a change of utilized fixed-input levels $\mathbf{z}$. However, as the resource constraint usually is binding for the reference scenario, it is not possible to extend $\mathbf{z}$. Reducing $\mathbf{z}$ would imply that the resource constraint is no longer binding, which would require that marginal costs with respect to fixed-input levels must become zero. According to our test, the model responds to a change in output prices mainly by adjusting output levels.

$$q_{ii} \geq q\_min_{ii} \tag{4.10}$$

The rationale behind this intervention is that, if $q_{ii} \geq q\_min_{ii}$, the own-price elasticity of supply, $\varepsilon_{t,i,i}$, could not become unrealistically large, according to equation (4.9) and for a given ratio $\bar{p}_{t_{ref},i}/\bar{x}_{t_{ref},i}$. This raises the question what value the modeller should attribute to q_min_{ii} to yield realistic own-price elasticities. If one has a notion of a realistic value for the own-price elasticity of supply of activity i, for example, because exogenous supply elasticities $\varepsilon\_exog_{i,i}$ are available, it is possible to calculate q_min_{ii} by solving equation (4.9) for q_{ii} and employing condition (4.10):

$$q\_min_{ii} \leq \frac{1}{\varepsilon\_exog_{i,i}} \frac{\bar{p}_{t_{ref},i}}{\bar{x}_{t_{ref},i}} \tag{4.11}$$

To maintain comparability between the farm samples, we decide to use one single value q_min_{ii} that is imposed as a minimum value for the diagonal of matrix $\widehat{\mathbf{Q}}$ of all four farm samples for all I activities. For each farm sample, we gather exogenous information on the own-price elasticities of supply for the most important activity in each sample, being cereals for crop farms, dairy products for dairy farms, and other bovine outputs for cattle farms. We retrieve the following values for $\varepsilon\_exog_{i,i}$ from literature:

- 1.50 for cereals (Jansson and Heckelei (2011, p. 147); average of the supply elasticity estimate for soft wheat, maize, and barley, computed with CAPRI)
- 0.41 for dairy products (Jongeneel and Tonini (2009, p. 275); average of dairy supply elasticities used by AGMEMOD, CAPSIM and EDIM for France)
- 1.06 for other bovine outputs (Balkhausen *et al.* (2005, p. 578), supply elasticity for beef calculated with ESIM-GAMS)

Using these values as $\varepsilon\_exog_{i,i}$ in equation (4.11) with information on $\bar{p}_{t_{ref},i}/\bar{x}_{t_{ref},i}$ for the respective farm sample and the respective main activity i, we calculate q_min_{ii} for each of the four farm samples and then calculate the average over these four values, which is 0.2. We therefore add to the cost function estimation model (3.47) – (3.53) the following constraint:

$$q_{ii} \geq q\_min_{ii} = 0.2 \tag{4.12}$$

Table 10 shows the estimates for the activity level and fixed-input level dependent cost function with a time index for the dairy farms in West France, having imposed constraint (4.10) in the estimation process. As the scale factor for dairy farms is 10 000 for all activities, the diagonal elements of matrix $\mathbf{Q}$ indicate by how many base-year Euros the marginal costs of activity i increase if a farm increases its output of i by a quantity that is worth 10 000 base-year Euros, while the levels of all other outputs remain equal. However, a change in the level of activity i also impacts the marginal costs with respect to activity levels of all other activities via the off-diagonal elements of $\mathbf{Q}$, and the marginal costs with respect to fixed-input levels via the elements of matrix $\mathbf{H}$. The net effect of a price change of activity i on activity levels and fixed-input prices is therefore hard to evaluate by looking at table 10. This is why only the cost function parameters for dairy farms are exemplarily depicted, while the remaining estimated parameters are presented in Annex III. Instead, we discuss the corresponding elasticities in more detail in the next section, where they are calculated by simulating a price change.

Estimated Cost Function Parameters Dairy Farms							
Activity Level and Fixed-Input Level Dependent Cost Function with Time Index and $q\_min_{ii} = 0.2$	activity level				fixed-input level		
	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour
		$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$	
dairy products	0.200	-0.064	0.015	-0.032	-0.009	0.015	0.509
other bovine animal outputs	-0.064	0.200	0.023	-0.020	0.006	0.012	0.538
non-bovine animal outputs	0.015	0.023	0.200	-0.048	0.001	0.027	0.558
crops	-0.032	-0.020	-0.048	0.200	-0.011	0.029	1.024
		$\hat{\mathbf{c}}_0$				$\hat{\mathbf{g}}_0$	
	-0.711	0.104	-0.861	0.716	0.016	-0.076	-3.406
		$\hat{\mathbf{c}}_1$				$\hat{\mathbf{g}}_1$	
	-0.093	0.039	-0.192	0.188	0.000	0.000	-0.045
						$\hat{\mathbf{F}}$	
cropland					0.000	0.000	0.000
grassland					0.000	0.000	0.002
unpaid labour					0.000	0.002	0.221

***Table 10**: **Estimated parameters of the activity level and fixed-input level dependent cost function for dairy farms for the cost function specification with a time index**. Lower boundary for the diagonal $q\_min_{ii} = 0.2$. Source: own estimation.*

Nevertheless, a look at the signs of the parameters in table 10 already reveals something about the relationships between the four different activities and the three fixed-inputs. It is important to recall that positive off-diagonal elements of matrix $\hat{\mathbf{Q}}$ indicate a substitutive relationship between two activities, whereas a negative off-diagonal element of matrix $\hat{\mathbf{Q}}$ implies a complementary relationship. The second element of the $\hat{\mathbf{Q}}$ matrix indicates that the marginal costs of other bovine animal outputs decrease by -0.064 base-year Euros, if the production of dairy products is increased by a value of 10 000 base-year Euros ceteris paribus, and vice versa. This suggests that an increase of dairy outputs simultaneously leads to an increase of other-bovine animal outputs, to maintain the relationship that MC=MR. A potential explanation is the complementarity, e.g., between milk production and calves that are born and sold[18]. However, whether the cross-price elasticity of supply is positive between dairy products and other bovine outputs, as we would expect for complementary goods, also depends on the interrelation with other outputs.

Matrix $\widehat{\mathbf{H}}$ indicates how the marginal costs of the outputs change if an additional unit of a fixed-input is used. However, the observed utilized fixed-input levels $z^{o}_{t,f,j}$ are equal to available fixed-input levels $b_{t,f,j}$ and $b_{t,f,j}$ varies much more between farms than between years. Consequently, the elements of matrices $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{F}}$ do not necessarily indicate how the marginal costs of a farm change if the farm utilized more of a fixed-input, but rather how marginal costs differ between farms with either a small or a large fixed-input endowment over the years in the sample. By imposing positive (semi-) definiteness on matrix $\widehat{\mathbf{F}}$ marginal costs with respect to fixed-inputs inevitably increase in fixed-inputs. The reason why the value for unpaid labour on the diagonal of matrix $\widehat{\mathbf{F}}$ is so much larger than the other values may simply be that the average levels of cropland and grassland are considerably higher than the levels of unpaid labour (see table 2). The interplay between matrices $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{H}}$ complicates the interpretation of the elements of the two matrices, but by taking the derivative of the two MC functions with respect to fixed-input levels it is possible to see how a change of fixed-input levels impacts marginal costs:

$$\frac{\partial^2 TVC}{\partial x_i \partial z_j} = \phi_{l_x}\hat{g}_{0_j} + \mathbf{t}\,\phi_{l_x}\hat{g}_{1_j} + \phi_{l_x}\sum_{k=1}^{J}\left(f_{jk}z_k\right) + h_{ij} \quad (4.13) \qquad \frac{\partial^2 TVC}{\partial z_j \partial z_j} = f_{jj} \quad (4.14)$$

The average dairy farm in Brittany, for example, has 39.2 ha of cropland, 33.8 ha of grassland and 1.8 AWU of unpaid labour (see table 2). The value the time index takes for the year 2012 is 8, since 2012 is the 8th year for which we have observations in the farm sample. Taking dairy outputs as an example, with $\overline{\phi_{l_{dairy}}} = 0.22$, the change in marginal costs with respect to dairy outputs if one more AWU of unpaid labour was available is $0.22 * (-3.406) + 8 * 0.22 * (-0.045) + 0.22 * (0.002 * 33.8 + 0.221 * 1.8) + 0.509 = -0.217$, according to equation (4.13). With the output price index for dairy outputs being 1.109 in 2012, this implies that marginal costs for dairy outputs would decrease by almost 20% for the average dairy farm in Brittany, if one more family AWU was available.

Marginal costs with respect to the fixed-input unpaid labour would change by 0.221, according to equation (4.14). The price for the fixed-input labour would *decrease* by this value.

[18] It is also conceivable that dairy farms increase dairy outputs by culling dairy cows later or by increasing the herd size with heifers raised on the farm, instead of selling heifer calves. This could lead to a short-term *decrease* in other bovine animal outputs. However, De Vries et al. (2008) argue that an increasing number of dairy farms specializes on dairy cow milking and buys pregnant dairy heifers instead of rearing them. For those farms, calves are mere by-products that are sold. If such a specialized farm increases dairy outputs by extending its herd size with purchased pregnant heifers, it would indeed be reasonable to expect that other bovine outputs increase with an increase of dairy outputs.

The vector $\hat{\mathbf{c}}_1$ of linear, time-dependent cost function parameters indicates that MC with respect to the output levels of dairy products and non-bovine animal outputs decrease over the years 2005-2012, while the MC with respect to the levels of other bovine animal outputs and crops increase. A likely explanation for the decreasing MC of dairy outputs over time is the steady rise of dairy activity levels while output prices remained largely stable (see figures A.3 and A.4 in Annex III). Presumably, the increase of dairy output levels over time is at least partly due to the phasing out of the dairy quota, as well as some technological progress.

Table 11 depicts the coefficients of determination (R^2) and the sum of the weighted squares of errors in the right column. All R^2 values are one or close to one, except for crop outputs in the cattle sample, and the right column indicates that the sum of the weighted squared errors is much lower than in the case in which the activity level dependent cost function is used (see table 7). For dairy farms, the inclusion of the time index allows to reduce the least squares value quite significantly. For cattle farms it notably increases the R^2 value for crops. The coefficient of determination R^2 reveals that, overall, all error terms are comparably small.

Coefficient of Determination (R^2)

Region	Farm Type	Specification of the Cost Function / Time Index	activity levels: animal outputs	activity levels: pulses and oilseeds	activity levels: industrial crops	activity levels: cereals	activity levels: other crops	fixed-input prices: cropland	fixed-input prices: grassland	fixed-input prices: unpaid labour	fixed-input levels: cropland	fixed-input levels: grassland	fixed-input levels: unpaid labour	Least Squares (LS)
Central France (south)	crops	no	1.00	1.00	1.00	0.95	0.86	1.00	1.00	1.00	1.00	1.00	1.00	424.29
		yes	1.00	1.00	1.00	0.95	0.86	1.00	1.00	1.00	1.00	1.00	1.00	378.23
Central France (north)	crops	no	0.99	1.00	0.97	0.98	0.95	1.00	1.00	1.00	1.00	1.00	1.00	2,899.50
		yes	0.99	1.00	0.97	0.98	0.95	1.00	1.00	1.00	1.00	1.00	1.00	2,828.54

Region	Farm Type	Specification of the Cost Function / Time Index	activity levels: dairy products	activity levels: other bovine animal outputs	activity levels: non-bovine animal outputs	activity levels: crops		fixed-input prices: cropland	fixed-input prices: grassland	fixed-input prices: unpaid labour	fixed-input levels: cropland	fixed-input levels: grassland	fixed-input levels: unpaid labour	Least Squares (LS)
West France	dairy	no	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	1.00	75.54
		yes	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00	1.00	37.85
South-West France	cattle	no	1.00	0.99	1.00	0.85		1.00	1.00	1.00	1.00	1.00	1.00	154.87
		yes	1.00	0.99	1.00	0.92		1.00	1.00	1.00	1.00	1.00	1.00	149.14

Table 11: Precision of the estimation of the activity level and fixed-input level dependent cost function with a minimum value of 0.2 for the diagonal elements of matrix Q . *The coefficient of determination is calculated for each activity, each fixed-input price and for each fixed-input level. The far-right column indicates the least squares value of the minimized objective value of the estimation function. Source. own simulation.*

However, table 11 does not inform about potential relationships between errors stemming from systematic deviations for individual farms or for different years. Figures A.5 - A.16 in Annex III visualize the distribution of all three types of error terms for the four farm samples. While the activity level related error terms $\hat{h}_{t,f,i}$ do not seem to be strongly correlated within farms, deviations from observed fixed-input levels ($\hat{s}_{t,f,j}$) and fixed-input prices ($\hat{u}_{t,f,i}$) appear to be correlated at the farm level, for the four farm samples. It is quite striking that, e.g., for crop farms in southern Central France, both $\hat{s}_{t,f,j}$ and $\hat{u}_{t,f,j}$ are systematically strictly positive for cropland or strictly negative for grassland. A potential remedy would be the inclusion of a farm-fixed effect, as it was done for the activity level dependent cost function. However, this would presumably make the matrix $\mathbf{F}$ redundant: As the variation of fixed-input levels *within* farms is minimal, $\mathbf{F}$ largely captures differences in costs *between* farms.[19]

Validation Phase

Table 12 displays the Theil inequality index for each sample, each activity and for each fixed-input, for a lower boundary of 0.2 for the diagonal elements of matrix $\mathbf{Q}$. Again, a Theil inequality index of zero indicates that the simulated values equal the observed value plus the error term (the “true” value) or are at least very close. Therefore, the size of the error terms $\hat{h}_{t,f,i}$, $\hat{u}_{t,f,i}$ and $\hat{s}_{t,f,i}$ that was discussed in the previous section should not impact the quality of the calibration. The highest value the index can take is one, which means, for our application, that there is no calibration, at all. The validation is performed with the parameters and error terms that are estimated in the estimation phase, with the output prices that are observed for 2012. We implement this step by using the TNR maximization function (3.58) – (3.59), as the TNR maximization is also used for the policy simulation in section 5.

[19] Actually, the first tentative trials with the inclusion of a farm fixed effect into the activity level and fixed-input level dependent cost function confirm that matrix $\mathbf{F}$ becomes redundant if a within transformation is performed.

Theil Inequality Coefficient													
			activity levels					fixed-input prices			fixed-input levels		
Region	Farm Type	Specification of the Cost Function / Time Index	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour	cropland	grassland	unpaid labour
Central France	(south) crops	no	0.01	0.08	0.03	0.08	0.04	0.33	0.29	0.12	0.07	0.00	0.04
		yes	0.00	0.00	0.00	0.00	0.00	0.10	0.04	0.01	0.00	0.00	0.00
	(north) crops	no	0.03	0.06	0.06	0.06	0.08	0.11	0.75	0.27	0.06	0.00	0.04
		yes	0.07	0.05	0.03	0.01	0.00	0.29	0.31	0.39	0.00	0.04	0.02
			activity levels					fixed-input prices			fixed-input levels		
Region	Farm Type	Specification of the Cost Function / Time Index	dairy products	other bovine animal outputs	non-bovine animal outputs	crops		cropland	grassland	unpaid labour	cropland	grassland	unpaid labour
West France	dairy	no	0.03	0.08	0.00	0.07		0.04	0.05	0.03	0.01	0.05	0.03
		yes	0.00	0.00	0.00	0.00		0.01	0.05	0.02	0.00	0.00	0.00
South-West France	cattle	no	0.00	0.00	0.03	0.00		0.33	0.04	0.14	0.00	0.00	0.00
		yes	0.00	0.00	0.02	0.01		0.19	0.05	0.01	0.00	0.00	0.00

Table 12: Theil inequality coefficients comparing simulated and "true" values for activity levels, fixed-input prices, and fixed-input levels for the reference year 2012. A minimum boundary of 0.2 for the diagonal elements of matrix Q is imposed. *Values are simulated with the TNR maximization function. A coefficient of 0 indicates that all simulated values are equal to the "true" values. Source: own simulation.*

The Theil inequality index is useful because it expresses the quality of the calibration for each activity level, fixed-input level, and fixed-input price in a single value. However, a high Theil inequality index does not inform about how simulated values deviate from "true" values. Besides, for the crop farms in northern Central France and the cattle farms, the Theil inequality coefficients do not univocally indicate if the application of the time index leads to a better calibration. The share of simulated values that deviate very little (≤1%) or very much (≥100%) is therefore presented in table 13. In most cases the application of a time index reduces the share of extreme deviations. For activity levels and fixed-input levels, almost all deviations seem to be either very small or very large. In many cases, a deviation of at least 100% indicates that the simulated activity or fixed-input level is erroneously zero.

Share of Simulated Values that Deviate by ≤ 1% (≥ 100%) from True Values													
			activity levels					fixed-input prices			fixed-input levels		
Region	Farm Type	Specification of the Cost Function: Time Index	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour	cropland	grassland	unpaid labour
Central France (south)	crops	no	0.98 (0.00)	0.96 (0.01)	0.96 (0.00)	0.96 (0.02)	0.95 (0.05)	0.00 (0.05)	0.68 (0.01)	0.04 (0.03)	0.96 (0.04)	1.00 (0.00)	0.98 (0.02)
		yes	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.24 (0.00)	0.70 (0.00)	0.89 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Central France (north)	crops	no	0.98 (0.02)	0.97 (0.01)	0.97 (0.02)	0.97 (0.01)	0.90 (0.06)	0.10 (0.03)	0.52 (0.18)	0.01 (0.04)	0.97 (0.02)	1.00 (0.00)	0.97 (0.03)
		yes	0.96 (0.03)	0.97 (0.01)	0.97 (0.01)	0.99 (0.00)	1.00 (0.00)	0.81 (0.01)	0.51 (0.14)	0.07 (0.03)	1.00 (0.00)	0.95 (0.00)	1.00 (0.00)
			activity levels					**fixed-input prices**			**fixed-input levels**		
Region	**Farm Type**	**Specification of the Cost Function: Time Index**	**dairy products**	**other bovine animal outputs**	**non-bovine animal outputs**	**crops**		**cropland**	**grassland**	**unpaid labour**	**cropland**	**grassland**	**unpaid labour**
West France	dairy	no	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)		0.37 (0.00)	0.14 (0.00)	0.97 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
		yes	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)		0.83 (0.00)	0.23 (0.00)	0.67 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
South-West France	cattle	no	0.93 (0.00)	0.99 (0.00)	0.93 (0.07)	0.99 (0.01)		0.21 (0.02)	0.61 (0.00)	0.10 (0.00)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)
		yes	1.00 (0.00)	0.99 (0.00)	0.93 (0.06)	0.98 (0.02)		0.23 (0.01)	0.18 (0.00)	0.91 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

Table 13: Share of simulated values that deviate by less than 1% and by at least 100% (in brackets) for the reference year 2012. *A minimum boundary of 0.2 for the diagonal elements of matrix Q is imposed. Values are simulated with the TNR maximization function. Source: own calculation.*

Simulation Phase

The employment of a time index seems to improve the validation of the model for all farm samples. It allows to capture a potentially existing time trend, which is particularly helpful for the dairy sample, as table 11 and figures A.5 - A.16 in the Annex reveal. However, also for the three other farm samples it appears reasonable to believe that there is a trend in costs per outputs over the observed period. Therefore, all simulations are performed using the cost function estimates with the time index. Table 14 depicts the simulated supply elasticities to prices for the four farm samples, using 2012 as the reference year. The own-price elasticity is in a reasonable range for the major activities of each farm type (cereals for crop farms, dairy products for dairy farms, and other bovine animal outputs for cattle farms).

Supply Elasticities

price change	Crop farms from southern Central France: animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	Crop farms from northern Central France: animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
animal outputs	1.94	0.02	-0.01	-0.02	0.00	3.06	0.55	-0.47	0.00	-0.88
pulses and oilseeds	0.17	3.93	-0.02	0.04	-0.99	1.76	8.41	-0.62	1.19	-2.63
industrial crops	-0.03	0.00	1.18	0.05	-0.19	-5.25	-1.11	2.28	-0.23	-0.12
cereals	-0.48	0.10	0.52	1.50	-0.27	1.29	4.57	-0.54	2.30	-2.08
other crops	0.00	-0.08	-0.06	-0.01	1.36	-0.37	-0.20	-0.05	-0.11	1.32

price change	Dairy farms from West France: dairy products	other bovine animal outputs	non-bovine animal outputs	crops	Cattle farms from South-West France: dairy products	other bovine animal outputs	non-bovine animal outputs	crops
dairy products	0.53	0.61	-0.83	0.83	0.56	0.00	-0.50	0.04
other bovine animal outputs	0.18	1.80	-1.26	0.66	0.00	0.79	-10.00	0.80
non-bovine animal outputs	0.00	0.00	14.43	0.02	-0.28	-1.47	166.64	-1.16
crops	0.19	0.50	4.59	7.44	0.02	0.11	-7.30	23.26

Table 14: Supply elasticities to prices, simulated using the activity level and fixed-input level dependent cost function with a time index and a lower boundary of 0.2 for the diagonal elements of matrix Q for 2012. *Elasticities are simulated for a 10% increase in output prices of the activity indicated on the vertical axis. Source: own simulation.*

The own-price elasticities of non-bovine animal outputs appear unrealistically high for dairy farms, but especially for cattle farms. The reason for this is that MC of other bovine animal outputs increase quite significantly if the activity level of non-bovine animal outputs is increased, as the element of the matrix $\widehat{\mathbf{Q}}$ that connects the two is quite large (see table 10 in Annex III). Therefore, other bovine animal outputs and non-bovine animal outputs seem to be strong substitutes in production: If the price of other bovine animal outputs increases, the total level of non-bovine animal outputs drops to zero for farms that produce both in the reference scenario, and vice versa. However, the quantity of non-bovine animal products produced by dairy and cattle farms is negligible (see table A.4 in Annex III) and the number of observations is low, which may explain the drastic supply response.

Using the TNR maximization model (3.54) – (3.55), fixed-input prices can be retrieved as the dual values of the resource constraints. Figure 8 depicts the simulated fixed-input prices for an increase of the price of the most important output for each farm type by 5%, 10%, and 15%.

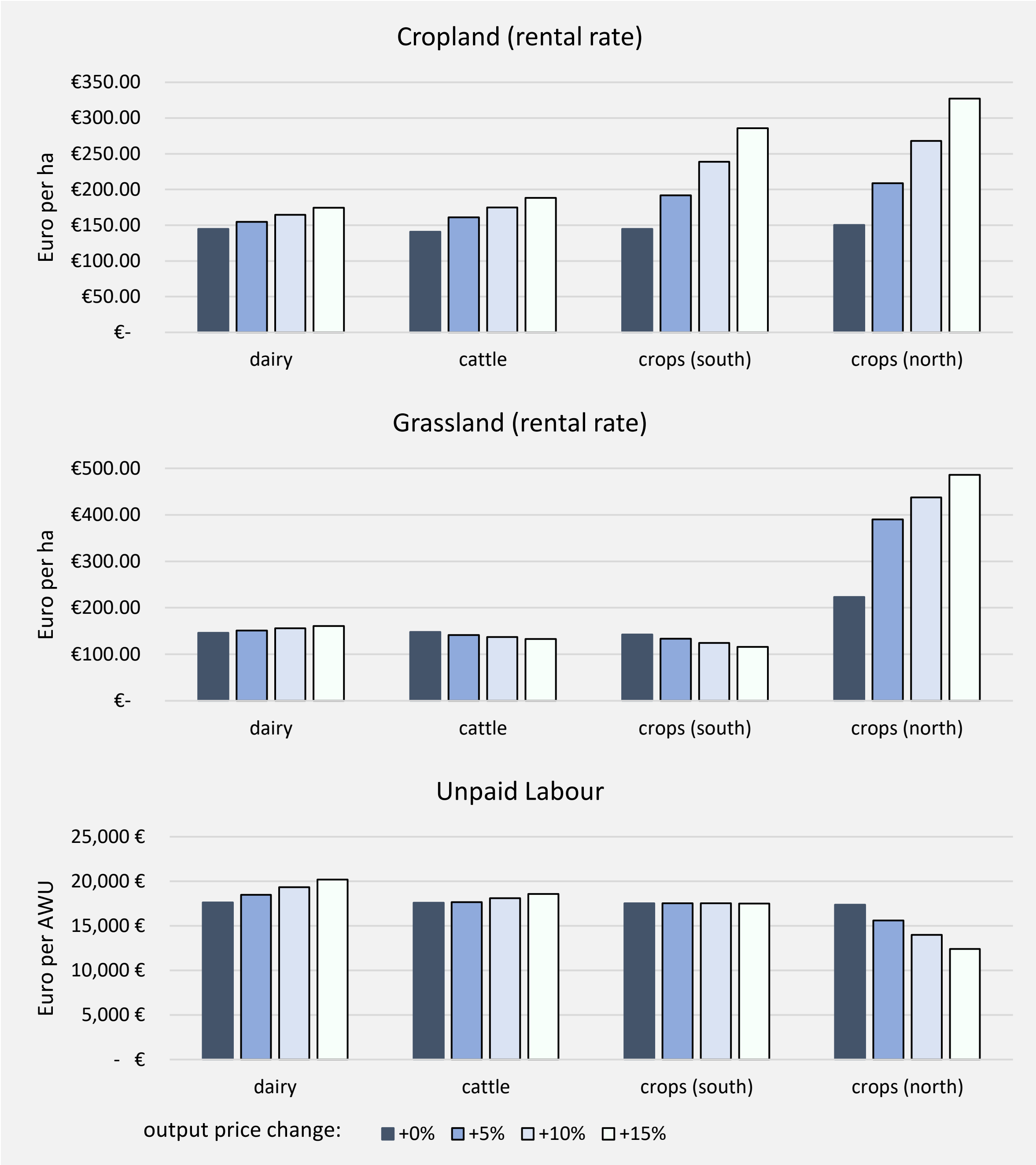

***Figure 8: Simulated fixed-input prices for an increase of the output price of the most important activity per sample**. Price increase of 5%, 10%, and 15% for dairy products for dairy farms, other bovine animal products for cattle farms, and cereals for crop farms. Fixed-input prices are simulated as the dual values of the resource constraint (3.59). The simulation is performed using the estimated activity level and fixed-input level dependent cost function with a time index. Source: own simulation.*

Surprisingly, the rental rate of grassland increases quite significantly for crop farms in northern Central France if the price of cereals increases, whereas the rental rate of unpaid labour decreases. A potential explanation for the rise of the grassland rental rate is that the cross-price elasticity between cereals and animal outputs for a price increase of cereals is positively (see table 14), indicating a complementarity between cereals and animal outputs. As the increase of the TNR significantly enhances the remuneration for cropland and grassland, it induces a decrease of the remuneration for unpaid labour. This is possible as the model accounts for the impact of activity levels on MC with respect to fixed-input levels – whether this specific reaction is realistic or not is debatable.

5. Application for a Greenhouse Gas Emission Cap Simulation

In this chapter the estimated parameters of the activity level and fixed-input level dependent cost function with a time-index are applied to simulate the introduction of a greenhouse gas (GHG) emission constraint at regional level for dairy farms in West France, cattle farms in South-West France, and crop farms in southern Central France[20]. First, a theoretical background on agricultural GHG emission mitigation and abatement costs is provided. Then, we modify our model for the incorporation of a GHG emission constraint. We describe the compilation and the preparation of GHG emission data to generate a link between the simulated production decisions and farms' GHG emissions. Finally, we simulate regional GHG emission caps of different magnitude, up to a reduction of 20% compared to the GHG emissions in the reference scenario. The simulated marginal abatement costs, activity level changes and fixed-input price changes are presented and discussed.

5.1. Agricultural Greenhouse Gas Emission Mitigation and Abatement Cost Estimation in Literature

5.1.1. Agricultural Greenhouse Gas Emissions

According to the United Nation's Intergovernmental Panel on Climate Change (IPCC) human influence has unequivocally warmed the global climate and continues to do so, posing unprecedented risk mainly to vulnerable populations and ecosystems (see for example IPCC (2014b, 2021)). Against this backdrop, a first legally binding international agreement attempting to limit global warming to well below 2 degrees Celsius was signed by 196 parties in Paris in 2016 (United Nations, 2021). The EU's approaches to comply with the Paris Agreement and to tackle climate change were bundled and communicated by the European Commission in 2019 in the so-called "Green Deal" strategy (European Commission, 2019). In 2021 the European Climate Law was adopted by the European Parliament and the Council, a regulation that establishes the legislative framework for achieving climate neutrality by 2050 and cutting greenhouse gas emissions by at least 55% for 2030 compared to 1990 levels (Regulation (EC) 2021/1119, n.d.). Within the Green Deal, the "Farm-to-Fork Strategy" constitutes the overarching concept to ensure that the agri-food sector contributes "appropriately" to the objectives set out by the European Climate Law (European Commission, 2020, p. 3)[21].

---

[20] We do not present the simulation results of crop farms from northern Central France, as they proofed too erratic. The reason for this seems to be that the optimal solution found in the validation phase for many farms is only a local maximum, and therefore a TNR that is larger than in the reference scenario was found for many farms if the emission constraint is imposed. Below, we describe the same problem and its consequences in more detail for cattle farms but disregard the results for crop farms from northern Central France for reasons of brevity.

[21] To our knowledge, no binding emission reduction targets have been set so far for the agricultural sector.

Agriculture is affected by climate change, but at the same time it contributes to global warming through the emission of greenhouse gases. In the EU-28, agriculture accounted for 10% to 12% of all GHG emissions over the years 2010-2020. Methane (CH_4) emissions from enteric fermentation of livestock and nitrous oxide (N_2O) emissions from agricultural soils contributed 38% and 32% of the total agricultural emissions in 2016, respectively. The use of mineral and organic nitrogen fertilizers as well as livestock manure management are additional important contributors of non-CO_2 GHG emissions (EEA, 2019, 2021). In total, when accounting also for the production, the transport, and the processing of feed, the livestock sector is responsible for more than 80% of the agricultural GHG emissions. Almost 60% of the emissions from livestock are caused by cattle production (dairy and beef with roughly equal shares), and about 25% by pork production (Leip *et al.* (2010), Peyraud and MacLeod (2020)).

The absolute value of the agricultural sector's emissions in the EU has significantly decreased in the early 1990s, followed by a moderate decline until 2012. Presumably at least partly because of the phasing out of the dairy quota, agricultural emissions then increased again slightly until 2017, before they declined in 2019 back to the level of 2009 (EEA, 2021; Klootwijk *et al.*, 2016).

In France, agriculture emitted about 84,000 kilotons of CO_2 equivalents in 2019, which corresponds to a share of almost 20% of the country's total emissions (12% on average in the EU-28).[22] France's emissions from agriculture were approximately 10% lower in 2019 than in 1990, whereas the average reduction in the EU-28 over the same time period was 20% (EEA, 2021). These figures highlight the particular importance of the agricultural sector for the design of effective emission reduction strategies for France and the other EU member states.

5.1.2. Marginal Emission Abatement Costs

While the contribution of agriculture to anthropogenic GHG emissions is indisputable, its central function for food production and the provision of ecosystem services complicates the role it may play for emission mitigation (Eory *et al.* (2018)). Besides, ill-designed GHG mitigation strategies for the agricultural sector in Europe may deteriorate food security in low- and middle-income countries quite significantly, if they induce food price increases. Further on, especially if mitigation policies are implemented solely in the EU, this would impact producers' global competitiveness and limit their income. Emission savings would be partly offset through carbon leakage, i.e., the relocation of GHG emissions to countries outside the EU (Beckman *et al.* (2020), Hasegawa *et al.* (2018), Himics *et al.* (2018)). To minimize the ad-

[22] The global warming potential of a ton of CH_4 or N_2O is considerably higher than that of a ton of CO_2. To calculate the total emissions of a country or a sector it is common practice to account for non-CO_2 emissions by expressing their global warming potential in CO_2 equivalents. See, for example, IPCC (2014a, p. 714).

verse side-effects of emission mitigation, Fellmann *et al.* (2018) suggest a more flexible implementation of mitigation efforts that would allow to reduce agricultural emissions where the reduction is least costly, instead of a strict adherence to national mitigation targets.

A commonly used framework to assess and compare the cost effectiveness of different GHG mitigation options is the marginal abatement cost (MAC) curve approach. Marginal abatement costs are the costs of eliminating an additional unit of GHG emissions. They can be plotted against a corresponding emission reduction amount to obtain the MAC curve. The area under the MAC curve then indicates the total abatement costs, i.e., the total costs of reducing a certain quantity of emissions with the respective mitigation option (Eory *et al*., 2018; Morris *et al*., 2012).[23] However, the variety of production types, farm sizes, and general farming conditions in the EU causes significant heterogeneity of the mitigation potential and the MAC within the agricultural sector (Domínguez and Fellmann (2015), Fellmann *et al*. (2018), Macleod *et al*. (2015)). Vermont and De Cara (2010) remark that MAC are rarely observable in an agricultural context, often because mitigation measures have not yet been implemented, which is the reason why one has to resort to model simulations. In their meta-analysis, they distinguish three categories of models that have been used for MAC assessment.

First, in a strand of literature abatement potential and costs are analysed with engineering approaches. Based on information about available technology, a set of potential emission mitigation options is compiled along with the mitigation potential and the implementation costs of each option. Options with their total mitigation potential are then ranked according to their costs per unit of emission reduction to obtain the MAC curve. A recent example of the engineering approach is the analysis of the abatement potential of GHG emissions in the Irish agriculture (Lanigan *et al*., 2018). Various measures, such as the improvement of beef genomics, extended grazing, or the selection of different fertilizer types, are ranked according to their abatement costs per ton of CO_2-equivalents, so that the total costs for a given emission reduction can be derived.

Second, partial or general equilibrium models with endogenous equilibrium prices are used to assess the effects of a mitigation policy. If an emission tax is introduced, this particularly affects the supply-side directly and indirectly, through the change in equilibrium prices. The economic assessment of GHG mitigation options for EU agriculture by Van Doorslaer *et al*. (2015) represents such an approach, where the Common Agricultural Policy Regionalised Impact (CAPRI) model is extended by options to account for GHG emissions and implement technological mitigation options.

Third, micro-economic supply-side models are used to describe the behaviour of a sample of representative farms who maximize their TNR subject to resource constraints. MAC can then

[23] The term “welfare loss” might be more precise than “total costs”. If “producers” of mitigated emissions can exploit a producer’s rent, the costs paid e.g., by a government agency, will be higher than the area under the curve.

be computed by introducing an emission tax, as it is done, for example, by De Cara *et al.* (2005) who also use the FADN as their primary source of data. However, a second option to compute MAC with farm models is by imposing an emission constraint whose dual value represents the MAC.

The calculation of MAC as the dual value of an emission constraint is pursued for example by Morris *et al*. (2012), although their approach does not focus on the agricultural sector and uses a computable general equilibrium model, instead of a farm supply model. Their interpretation of the dual value is that it reflects the market price that would be obtained for the right to emit one unit of GHG if emission allowances are traded under a cap-and-trade system. The European Commission operates such an emission trading system (EU ETS) since 2005, where a cap is set on the total amount of GHG emitted by all market participants of certain sectors, who may trade allowances with one another. The cap is set low enough to be a binding constraint and is reduced each year (European Commission, 2021). Although the EU ETS does not comprise the agricultural sector, De Cara and Jayet (2011) argue that a cap-and-trade system for agricultural emissions could significantly lower the total abatement costs for the EU, compared to inflexible national mitigation targets. Given their advocacy for the method and the theoretical appeal of a market for emission allowances, we apply the model developed and tested in the previous chapters for the simulation of a GHG emission constraint.

5.2. Incorporation of a Greenhouse Gas Emission Constraint into the Initial Model

In chapter 3.3.2 we developed a total net revenue maximization model as an alternative to the minimization of the complementary slackness conditions (CSC) for the validation of the estimated cost function and for the simulation of output price changes. While TNR maximization and CSC minimization yield identical optimal fixed-input and activity levels, the optimal fixed-input prices are obtained in a different way. In the CSC minimization framework, they (minus the value of a slack variable) are equated to the negative of the marginal costs with respect to the fixed-input (equation 3.52). In the TNR maximization framework, they are retrieved as the shadow prices of the respective resource constraint and do not explicitly appear in the model as the optimization software calculates them (equation 3.59). This property renders the TNR maximization particularly useful for policy simulation, as it allows to interpret the dual values of constraints that were not included in the estimation phase. In the case of an emission constraint, their dual values indicate the change in TNR for the elimination of an additional unit of GHG emissions.

The crux of a cap-and-trade system in a perfectly competitive market for GHG emission allowances is that the market price indicates the lowest costs per unit of GHG emissions at which a desired amount of GHG emission savings can be achieved. For this market price to be positive, the total quantity of GHG emissions granted to all farms in the market via allowances must be lower than the total GHG emissions of farms in a refence scenario without a GHG emission cap. We define the total annual GHG emissions of the reference scenario TE_t^* as

$$TE_t^* = \sum_{f=1}^{F_r} \omega_{t,f} (\mathbf{e_{t,f}}'\mathbf{x^*_{t,f}}) \tag{5.1}$$

where the coefficient $\omega_{t,f}$ is a farm weight that is provided by the FADN dataset and that indicates how many farms in the population are represented by each farm in the sample (Neuenfeldt and Gocht, 2014). The elements of vector $\mathbf{e_{t,f}}$ serve as a GHG emission factor, representing the emissions of CO_2 equivalents for each of the I activities per unit of output. It may vary between farms and years. The vector $\mathbf{x^*_{t,f}}$ represents the activity levels of each farm that are optimal in the reference scenario without emission cap and F_r stands for all farms participating in the market. Note that it is possible that F_r does not comprise all F farms considered for the cost function estimation, if the modeller imposes the emission constraint, for example, at subregional level. As a perfectly competitive GHG emission market allows to maximize the total welfare subject to the GHG emission constraint, the optimization problem (3.58 – 3.59) must be rewritten to maximize the *sum* of the individual $TNR_{t,f}$. We can therefore express the optimization problem as

$$MAX_{\mathbf{x_{t,f}},\mathbf{z_{t,f}}}\, TNR_t = \sum_{\mathbf{f=1}}^{\mathbf{F_r}} \omega_{t,f}[\mathbf{p_{t,f}}'\mathbf{x_{t,f}} - \mathbf{\hat{c}}'\mathbf{x_{t,f}} - ½\mathbf{x_{t,f}}'\mathbf{\widehat{Q}}\,\mathbf{x_{t,f}} - (\boldsymbol{\phi}'\mathbf{x_{t,f}})\ \mathbf{g}'\ \mathbf{z_{t,f}}\ - ½\mathbf{z_{t,f}}'\mathbf{\hat{F}}\mathbf{z_{t,f}}\ (\boldsymbol{\phi}'\mathbf{x_{t,f}}) - \mathbf{z_{t,f}}'\mathbf{\widehat{H}}\,\mathbf{x_{t,f}} - \mathbf{x_{t,f}}'\mathbf{W_{t,f}}\mathbf{\hat{h}_{t,f}} - \mathbf{z_{t,f}}'\mathbf{T_{t,f}}\mathbf{\hat{s}_{t,f}}] \tag{5.2}$$

subject to
$$\mathbf{z_{t,f}} - \mathbf{V_{t,f}}\mathbf{u_{t,f}} \leq\ \mathbf{b_{t,f}}\ \ [\mathbf{y_{t,f}}] \tag{5.3}$$

$$\sum_{\mathbf{f=1}}^{\mathbf{F_r}} \omega_{t,f}(\mathbf{e'_{t,f}}\mathbf{x_{t,f}}) \leq (1-\alpha) \sum_{\mathbf{f=1}}^{\mathbf{F_r}} \omega_{t,f}(\mathbf{e'_{t,f}}\mathbf{x^*_{t,f}}) = (1-\alpha)\, TE_t^*\ \ [\mu_t] \tag{5.4}$$

and subject to non-negativity for $\mathbf{x_{t,f}}$ and $\mathbf{z_{t,f}}$. Note that the individual farm's $TNR_{t,f}$ is weighted by the farm weight coefficient $\omega_{t,f}$. The parameter α [0,1] is a GHG emission reduction factor that is set by the policy analyst to impose an upper boundary on the total GHG emissions of the F_r farms. The dual value μ_t indicates the marginal abatement costs (MAC) per unit of GHG emissions for the reference year *t*, in the same currency units as used in the objective function. When taking the derivative of the national (or regional) TNR maximization function (5.2) with respect to the activity levels, the dual value μ of the GHG emission constraint appears in the new optimality condition for each farm and each activity as follows:

$$\mathbf{MC_{t,f}} + \mu_t \mathbf{e_{t,f}} \geq \mathbf{p_{t,f}} \tag{5.5}$$

where $\mathbf{MC_{t,f}}$ summarizes all terms that constitute the MC with respect to activity levels before the introduction of the emission constraint. Equation (5.5) can be solved for μ_t, and for positive activity levels it can be written as the following equality:

$$\mu_t = \left(\mathbf{e_{t,f}}\right)^{-1}\left(\mathbf{p_{t,f}} - \mathbf{MC_{t,f}}\right) \qquad (5.6)$$

In the cap-and-trade system for the agricultural sector as modelled by Van Doorslaer *et al*. (2015) all tradeable GHG emission permits are distributed at the beginning of the simulation among market participants (either homogenously or according to distribution factors). The EU ETS also grants a certain number of initial free allowances to some market participants, which they may then trade (European Commission, 2021). Transferring this concept to our model would imply that each farm initially receives allowances for $(1-\alpha)\, TE^*_{t,f}$ units of GHG emissions. The value of an additional emission allowance may now vary between farms: For some farms, the gain in TNR they could generate if they had one more allowance may be larger than the decline in TNR that other farms would incur if they had one allowance less, indicating that the MAC of the former farms are higher than that of the latter farms. The former farms would become net buyers whereas the latter farms would become net sellers of emission allowances. However, our model does not calculate the MAC for each farm, but only the MAC of the last allowance that is traded on the market before market clearing is completed, which is seen as the market price. The MAC that each individual farm faces for the market equilibrium may lie below the market price, if the farm is a net suppler of emission allowances, or above the market price, if the farm is a net receiver of emission allowances. The sum of the supplier rents and the receiver rents is the welfare gain generated by the emission allowance trade, compared to the situation where each farm must reduce its GHG emissions by the factor α.

5.3. Policy Scenario Simulation

5.3.1. Greenhouse Gas Emission Data Preparation and Aggregation

The vectors and matrices $\hat{\mathbf{c}}$, $\widehat{\mathbf{Q}}$, $\hat{\mathbf{g}}$, $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{H}}$ of the cost function parameters as well as the vectors of error terms $\hat{\mathbf{h}}_{\mathbf{t,f}}$ and $\hat{\mathbf{s}}_{\mathbf{t,f}}$ were estimated in the estimation phase. The vectors and matrices $\boldsymbol{\phi}$, $\mathbf{p_{t,f}}$, $\mathbf{W_{t,f}}$ and $\mathbf{T_{t,f}}$ are exogenous to the model. The vectors of optimal activity levels $\mathbf{x^*_{t,f}}$ of the reference scenario can be easily estimated by running model (5.2) – (5.4) with the GHG emission factor α set to zero. For policy simulation, the GHG emission reduction factor α is set by the modeller. The only parameters of the simulation model with the emission constraint that still need to be defined before being able to run model (5.2) – (5.4) is $\mathbf{e}$, vector of emission factors for each activity.

An obvious shortcoming of the calculation of the total GHG emissions as the product of a vector $\mathbf{e}$ of emission factors and the vector of activity levels is that it pretends that GHG emissions depend exclusively on activity levels. The very important implication of this is that farmers can reduce emissions only (and linearly) by lowering activity levels. This certainly does not reflect reality since, for example, manure management, N-fertilizer application or soil management also have a significant impact on GHG emissions from farm practices, as outlined in section 5.1.1. However, as our model does not take variable-input levels or potentially available information on farm-specific technology into account, we resort to the use of emission factors for activities, while being aware of the limitations of this approach.

To construct the vector $\mathbf{e}$ of GHG emission factors, that indicates GHG emissions per level of the *aggregated* activity, we start by collating data on the GHG emissions of each of the 49 *disaggregated* activities. The IPCC elaborates guidelines for the estimation of agricultural GHG emissions and distinguishes three tiers of detail for calculation: Tier 1 is the simplest method and appears suitable under the conditions of this thesis, at least for the estimation of GHG emissions stemming from livestock activities. The IPCC (2006) provides GHG emission factors for different categories of livestock in different production systems (mainly categorized by geographic region) that indicate average values of methane emissions from enteric fermentation and manure management per head of animal. We transform these values into CO_2-equivalents, based on data on the global warming potential of methane from IPCC (2014a, p. 714), and use them as the starting point for the calculation of GHG emissions from livestock.

GHG emission factors for crops and dairy products are retrieved from the freely accessible and simplified database of Agribalyse (2020a, 2020b) as tons of CO_2-equivalents emitted per ton of the product, measured at the farm-gate for the average farm in France.[24]

In the following step, the total GHG emissions per farm in the sample, for each year and for each *disaggregated* activities are calculated by multiplying the emission factors of each of the 49 disaggregated activities with the quantities produced in each year by each farm. As indicated in figure 5 in section 4.1.2, produced quantities in the FADN dataset are expressed in tons or LU. Consequently, GHG emission factors from IPCC have to be converted from GHG emissions per head of animal to emissions per LU, using data on heads of animals per LU from Eurostat (2020).

Then, the total GHG emissions per *aggregated* activity in each year for each farm are calculated by summing all the emissions of the disaggregated activities that belong to the same aggregate. To avoid double counting of emissions from dairy outputs and the output "dairy cows" (usually referring to animals sold alive), we assumed that all emissions from a dairy cow are captured by the dairy outputs. Finally, the total GHG emissions per aggregated activity are divided by $x_{t,f,i}^{o}$, the observed level of the aggregated activity measured in base-year monetary values, to receive $e_{t,f,i}$. Note that this GHG emission factor varies between years and farms, as the composition of an aggregated activity also varies along these two dimensions. In some cases, the produced quantities of a disaggregated activity in weight

[24] As the freely accessible datasets from Agribalyse do not provide information on GHG emissions from oat production, we use the estimates from Rajaniemi et al. (2011).

units are not provided by the FADN dataset. This concerns the aggregated activity "other crops" in the crop farm samples, and one (four) observation in the dairy (cattle) sample. The reason seems to be that for vegetables, fruits, flowers or mushrooms only the sales value or the cultivation area are provided. While this allows calculation of aggregated activity levels (using exogenous output prices), it is not possible to calculate emission factors for aggregated activities according to the procedure applied for the other farms and activities. Therefore, missing values of $e_{t,f,i}$ are imputed with the subregional average $\bar{e}_{t,subregion,i}$, if available, or with the annual farm average for all other crops, $\bar{e}_{t,f}$. Table 15 displays the mean, the median and the standard deviation of the emission factors $e_{t,f,i}$ for each sample in 2012.

GHG Emissions per Produced Activity Level (kg of CO_2-equivalents/base-year €)						
	Crop farms from southern Central France					
	mean	median	std			
animal outputs	4.56	4.76	2.40			
pulses and oilseeds	2.32	2.30	0.28			
industrial crops	0.78	0.67	0.61			
cereals	3.09	3.06	0.35			
other crops	2.00	1.88	0.45			
farm's weighted average	2.73	2.75	0.63			
	Dairy farms from West France			Cattle farms from South-West France		
	mean	median	std	mean	median	std
dairy products	4.13	4.18	0.36	4.25	4.27	0.17
other bovine animal outputs	4.97	4.34	4.48	5.47	5.29	2.59
non-bovine animal outputs	1.72	1.82	1.62	5.80	4.85	5.83
crops	11.53	3.46	62.90	20.30	8.50	40.61
farm's weighted average	4.12	4.11	0.43	5.39	5.17	2.29

Table 15: Mean, median and standard deviation of the emission factor $e_{t,f,i}$ for three samples of farms in France in the year 2012*. The farm's weighted average is the average emission factor per farm weighted by the farms activity levels. Source: own calculation based on Agribalyse (2020a, 2020b), EU-FADN – DG AGRI, Eurostat (2020), IPCC (2006) and Rajaniemi et al. (2011).*

Strikingly, the means of the GHG emission factors for crops in the dairy and the cattle farm samples are large compared with the GHG emission factors of other activities of dairy and cattle farms, but also compared with the crop activities of the crop farms. At the same time, the standard deviation of the GHG emission factor for crops is very large for dairy and cattle farms. It is important to recall that the GHG emission factor $e_{t,f,i}$ indicates the GHG emissions per output *value* and not per output weight. The differences between the GHG emission factors for crops, both within and between the dairy and cattle farm samples, result almost entirely from a different average value per ton of produced crops: Dairy farms generate on average 92.29 base-year Euros per ton of crop output and cattle farms only 52.64 base-year Euros. The emissions per weight unit of crop output do not vary significantly between the two samples.

However, the different emission factors of non-bovine animal outputs can be explained by the types of non-bovine animals kept by farms: In the cattle farm sample sheep and pigs dominate, whereas the few non-bovine animal outputs produced by dairy farms in 2012 are mainly poultry.

Table 15 also indicates the mean of each farm's weighted average emission factor, which indicates how many kg of CO_2-equivalents are emitted per average activity level in base-year Euros. Notably, crop farms have lower average emissions per value unit than dairy or cattle farms.

5.3.2. Policy Simulation Results

This section presents the results of the simulation of the introduction of a GHG emission cap at regional level. To receive a MAC *curve*, eight reduction scenarios are simulated, ranging from a 2.5% reduction compared to the reference level in 2012, to a 20% reduction. Besides the MACs, which are presented first, other important simulation results are the impact of the GHG emission cap on activity levels and fixed-input prices. All farms are weighted by the farm weight that the FADN dataset provides and that indicates how many farms in the population are represented by each farm in the sample (Neuenfeldt and Gocht, 2014).

5.3.2.1. Estimated Marginal Abatement Costs

Figure 9 depicts the simulated MAC for dairy farms, cattle farms, and crop farms for eight different emission caps at the regional level. The MAC displayed on the vertical axis are the equilibrium prices of an emission trading system that would establish if the admissible GHG emissions of the region are reduced by the value on the horizontal axis. The area under each curve from the origin to a point on the horizontal axis represents the total GHG emission abatement costs incurred by the region for the respective emission reduction. For a 10% emission reduction the simulated MAC are 10.14 €/tCO_2eq for cattle farms, 17.51 €/tCO_2eq for dairy farms, and 29.99 €/tCO_2eq for crop farms in southern Central France. In section 5.4 we demonstrate that these values appear reasonable, when comparing our MAC with MAC estimates from other researchers. The MAC curves for dairy farms and crop farms increase linearly, so that the MAC for a 20% emission reduction are twice as high as the MAC for a 10% reduction. For cattle farms, the MAC curve has a slightly increasing upward slope, with an interruption for an emission reduction between 15% and 17.5%.

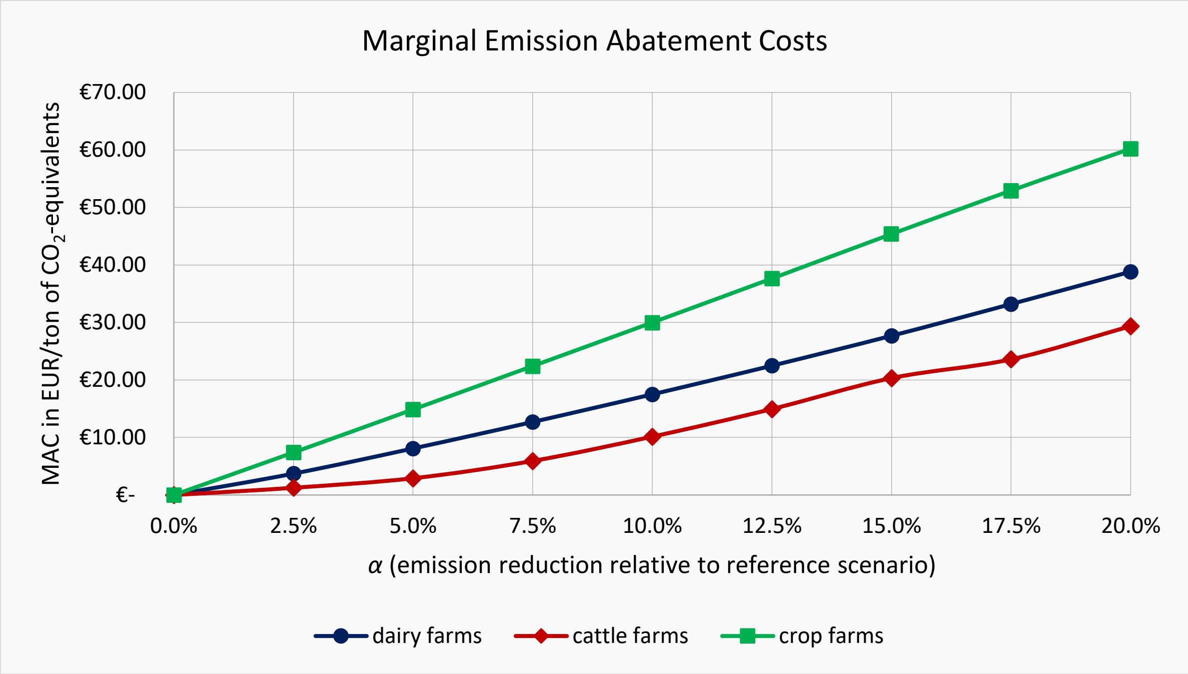

Figure 9: Marginal Emission Abatement Costs (MACs) for GHG emissions from dairy, cattle, and crop farms. *The horizontal axis displays the emission reduction scenario.*

A likely explanation for the differences of the simulated MAC between the three samples are the average GHG emissions generated by farms in each sample per activity level. As indicated by table 15, crop farms emit less GHG per average activity level than dairy farms, and dairy farms emit less GHG per average activity level than cattle farms. This implies that the abatement of a ton of GHG on average requires a stronger activity level reduction for crop farms than for dairy or cattle farms. It therefore appears reasonable that abatement costs per ton of GHG are highest for crop farms, followed by dairy farms, while cattle farms have lowest MAC.

The regional MAC do not reveal how farms change their supply behaviour or how the emission reduction is distributed between individual farms. In section 5.2 we briefly elaborated that some farms would become net buyers, whereas others would become net sellers of GHG emission certificates. Before assessing the simulated behaviour of individual farms, let us recapitulate the process represented by the model. Farms can only reduce GHG emissions by reducing outputs, and the reduction of the level of activity i *ceteris paribus* leads to a decrease in the MC with respect to the level of activity i, as the underlying cost function is upward sloping. In other words, the existence of a binding GHG emission constraint makes it impossible (at least for some farms) to fulfill the economic optimality condition $MC_{t,f,i} = MR_{t,f,i}$ and, hence, $MC_{t,f,i} < MR_{t,f,i}$ for the scenario.[25] According to equation (5.6) the MAC is equal to the difference between $MR_{t,f,i}$ and $MC_{t,f,i}$, divided by the emission factor $e_{t,f,i}$. Importantly, the dual value μ_t in equation (5.6) is identical for each farm and activity, as it

[25] Note that what we refer to as MC does not include the emission costs per activity unit, see equation (5.5).

stems from the national (or regional) GHG emission constraint. Besides, for our application $MR_{t,f,i}$ does not vary between farms, because the reference output prices are identical for all farms within the same region. The coefficient $e_{t,f,i}$ is exogenous to the model, so that farms can fulfil constraint (5.6) only by adjusting their $MC_{t,f,i}$. Figure 10 shows exemplarily for dairy outputs of dairy farms in West France that farms' marginal costs are indeed (anti-) proportional to their GHG emission factor $e_{f,i}$.

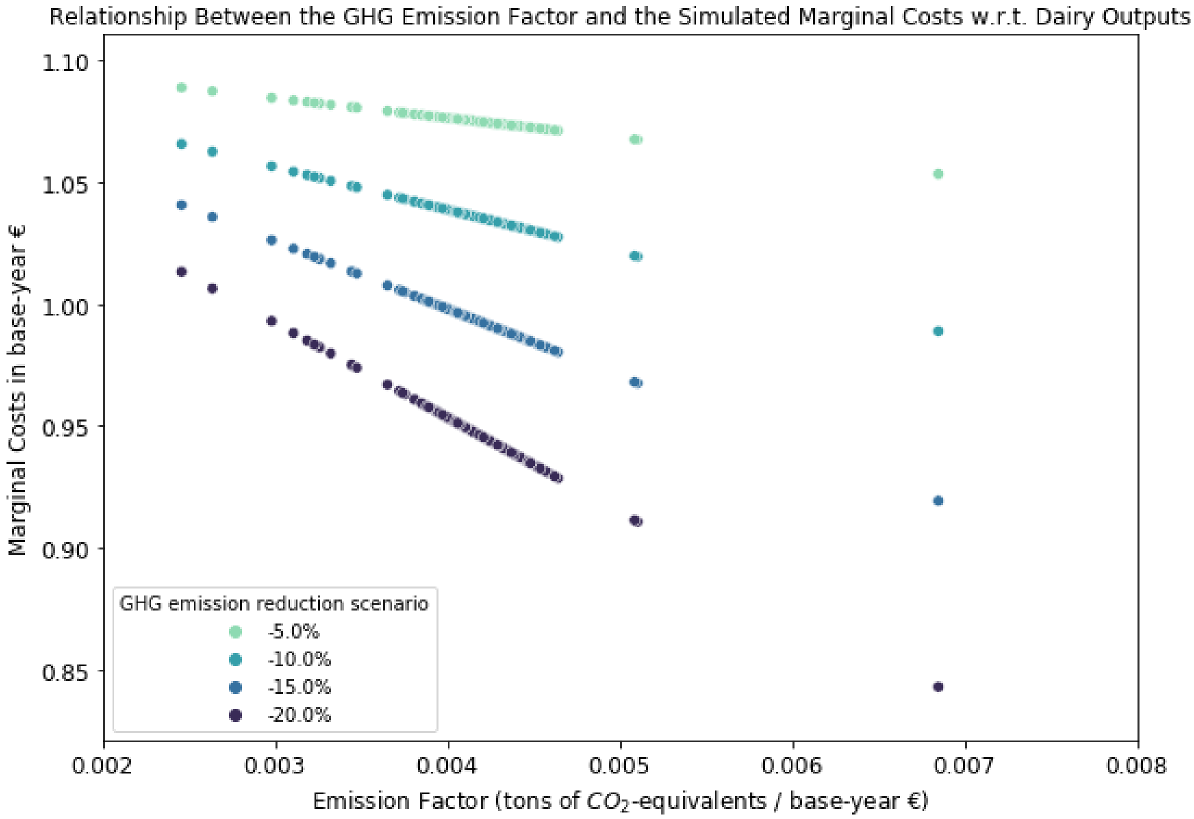

Figure 10: Simulated marginal costs with respect to dairy outputs for each of the dairy farms from West France, represented by the respective GHG emission factor $e_{t,f,dairy}$. *Four different emission reduction scenarios are displayed. Source: own simulation and visualization.*

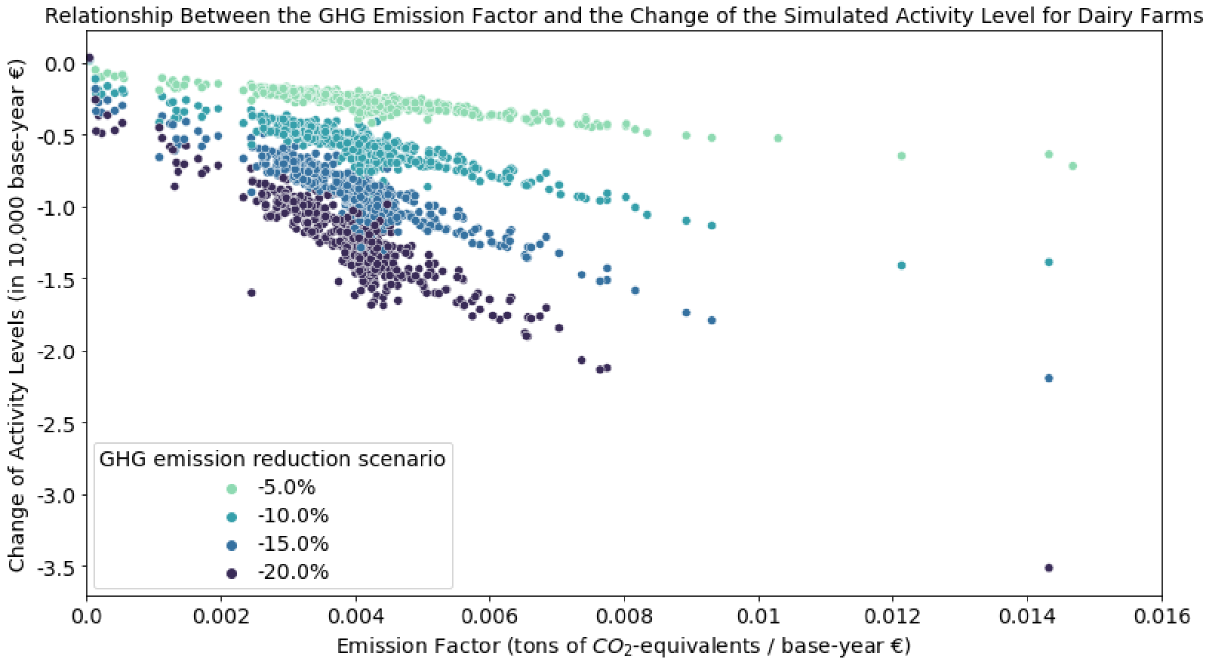

Figure 11: Relationship between the change of the simulated activity levels and the respective emission factors $e_{t,f,i}$ for dairy farms. *Four different emission reduction scenarios are displayed. Source: own simulation and visualization.*

The MC with respect to the other three outputs of dairy farms are displayed in Annex IV and exhibit the same relation towards the emission factor. Figure 11 shows that, by reducing the levels of most activities, farms adjust their MC according to the total required GHG emission reduction and according to the GHG emission factor of the respective activity. However, figure A.21 in Annex IV shows that the relation between the change of the activity level and the emission factor is not equally distinct for all activities. The reason for this is that the MC of dairy outputs, for example, is also affected by the levels of all other three activities, through the matrix $\mathbf{Q}$ and other interaction parameters. This explains why it is possible that the activity level of other non-bovine animal outputs even increases for one farm if an emission cap is imposed (top left corner of figure 11, where one observation exhibits a positive activity level change). Nevertheless, the absolute change in activity levels clearly depends on the GHG emission factor, overall. This dependency has important implications. A farm that has high activity levels in the reference scenario can adjust the MC with a *relatively* smaller change of activity levels, compared to a farm that has the same GHG emission factor but that has low activity levels in the reference scenario. This is because the change of marginal costs with respect to activity levels in activity levels is determined by matrix $\mathbf{Q}$, implying that a certain change in MC requires the same absolute activity level change for each farm. The reader be reminded that the GHG emissions caused by an activity are largely proportional to its simulated activity level (see figure A.17 in Annex IV). For an emission cap at 10% below the regional reference GHG emissions we hypothesize that the GHG emissions of an individual activity are reduced by more than 10% if the reference activity level is rather small and if the emission factor $e_{t,f,i}$ is comparably large. Figure 12 confirms this hypothesis:

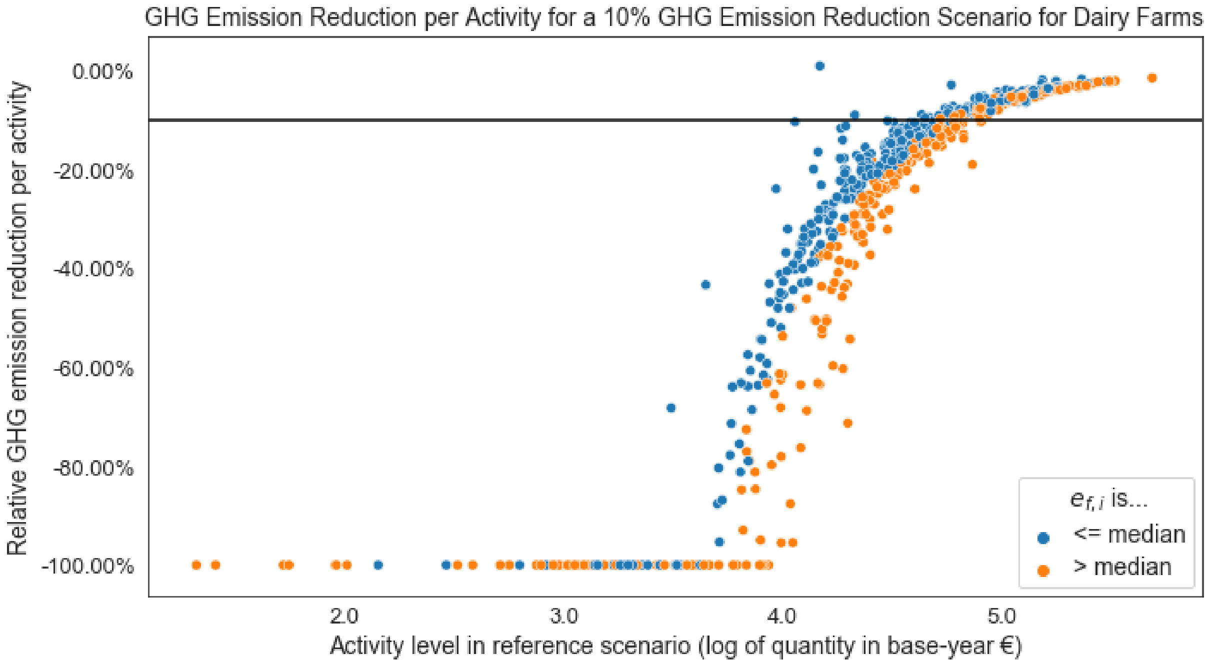

Figure 12: Emission reduction of dairy farms per activity for a 10% emission cap, for all four activities. *Each dot represents one of the four outputs produced by dairy farms. Dots that lie above the horizontal line are activities for which the emission reduction is less than 10%. The color of the dots indicates if the emission factor for this activity of the respective farm is above or below the median over all farms and over all activities. Source: own simulation and visualization.*

Figure 12 represents the relative GHG emission reduction for each of the I activities for all F farms, which is calculated as

$$relative\_GHG\_emission\_reduction_{t,f,i} = 100\frac{\left(e_{t,f,i}x_{t,f,i}-e_{t,f,i}x^*_{t,f,i}\right)}{e_{t,f,i}x^*_{t,f,i}} \qquad (5.7)$$

where $x^*_{t,f,i}$ is the optimal activity level in the reference scenario without a GHG emission cap. Figure 12 shows that the relative GHG emission reduction is stronger the lower the reference activity level $x^*_{t,f,i}$. The GHG emissions of all activities for which the level in the reference scenario was below a certain threshold (ca. $10^{3.5}$ base-year €) are reduced by 100%, indicating that farms cease the production of these activities. Figure A.22 in annex IV shows that this concerns almost exclusively the cropping activities of dairy farms. It also indicates that most of the activities in the top right corner, i.e., activities with a high reference activity level $x^*_{t,f,i}$ and a comparably low relative emission reduction $e_{t,f,i}$, are dairy outputs.

To answer the question which farms would become net buyers of emission certificates and which farms would become net sellers, the relative farm GHG emission change over all activities matters, not the relative farm GHG emission change per activity. Nevertheless, as the total relative GHG emission change of a farm is composed of the emission changes of each activity supplied by the farm, the pattern observed for individual activities is also largely valid for entire farms. Table 16 indicates that the average net GHG emission buyer (i.e., a farm that reduces its GHG emissions relatively little) has considerably higher activity levels in the reference scenario and emits less GHG per activity level than a net GHG emission seller.

Differences Between Net Sellers and Net Buyers of Emission Rights						
determinants of emission right trading behaviour	Dairy farms from West France net emission right...		Cattle farms from South-West France net emission right...		Crop farms from southern Central France net emission right...	
	buyers	sellers	buyers	sellers	buyers	sellers
mean of the sum of actvitiy levels per farm in the reference scenario (quantity in base-year €)	231,497	119,631	109,153	45,146	153,871	75,394
mean of farms' weighted average emission factor $\bar{e}_f$ (kg of CO_2-eq/base-year €)	4.00	4.24	4.65	7.09	2.62	3.00

Table 16: Characteristics of net GHG emission right buyers and sellers for a 10% GHG emission reduction scenario. *Net buyers are farms that reduce their emissions by less than 10%. The first row indicates the mean of each farm's total activity level in the reference scenario (the sum of the levels of each activity produced in 2012). The mean of farms' weighted average emission factor in the second row indicates how many kg of GHG are emitted per output level on average. Source: own simulation.*

In conclusion, farms with higher activity levels reduce their GHG emissions proportionally less than farms with lower activity levels. Already in the reference scenario farms contribute unevenly to GHG emissions (see Lorenz curve in figure 13). This uneven contribution becomes even more pronounced the more restrictive the regional GHG emission cap, which indicates that farms with comparably high GHG emissions in the reference scenario buy GHG emission rights from others, usually farms with lower reference activity level. Besides, the levels of activities with high GHG emissions per reference output value are stronger reduced than the levels of activities with a comparably smaller GHG emission factor.

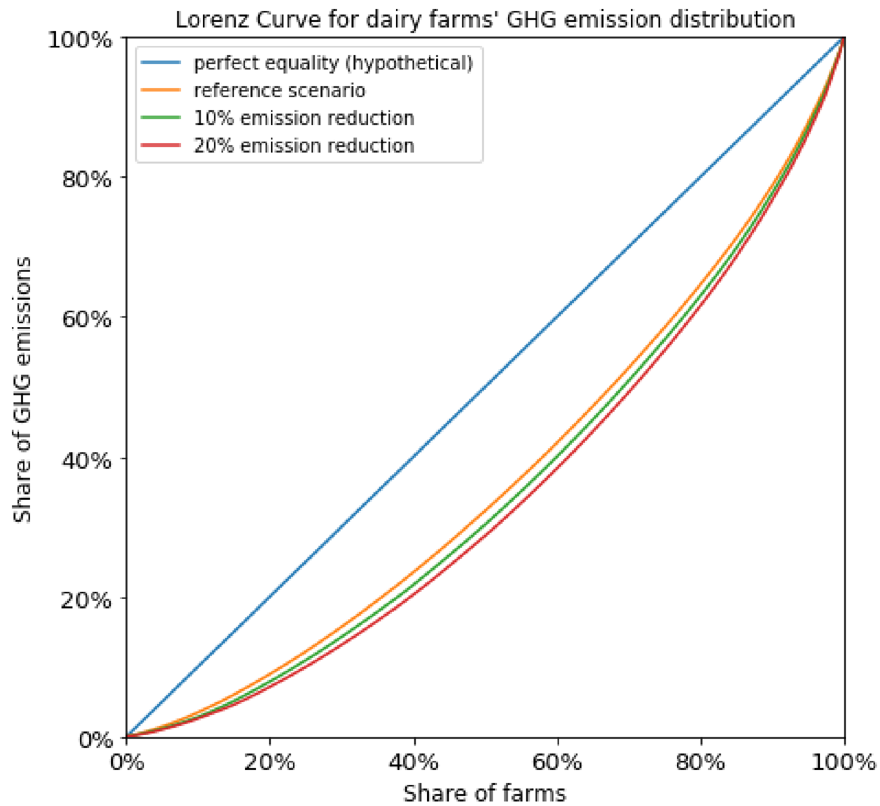

Figure 13: Lorenz curve indicating that emissions become more unevenly distributed if an emission cap is imposed. *Source: own simulation and visualization.*

5.3.2.2. Effect of Greenhouse Gas Emission Reduction on Activity Levels, Shadow Prices and Farm Total Net Revenues

Figure 14 depicts the total levels of activities produced by the three farm samples in the reference scenario, in a scenario with a GHG emission cap at 90% of the reference emissions and 80% of the reference GHG emissions. The levels of almost all activities decrease, as farms can only reduce their emissions by reducing their activity levels. Importantly, the strongest relative decrease of activity levels in every sample is observed for activities that already have a comparably low level in the reference scenario: crop activities in the dairy and in the cattle farm sample, and animal outputs in the crop farm sample. The contribution of the main activities to GHG emission reduction is less than proportional for dairy farms and crop farms (see figure A.25 in Annex IV).

As elaborated in the previous section, the explanation for this seems to be that the adjustment of marginal costs requires an *absolute* change of activity levels that is relatively less important for the most dominant activities. Surprisingly, the levels of other non-bovine animal outputs produced by cattle farms increase by 529% if a regional GHG emission constraint at 10% below the reference level is imposed, and even more for a 20% emission reduction. A look at the impact of the simulation on individual farms reveals that not all farms that produced other non-bovine animal outputs in the reference scenario increase their activity levels. An important determinant for the supply response of non-bovine animals seems to be the relation between the GHG emission factor $e_{t,f,i}$ for other non-bovine animal outputs and the farm's GHG emission factor for other bovine outputs (see figure A.24 in Annex IV). If a farm has higher GHG emissions per base-year € of other bovine outputs than per base-year € of non-bovine animal outputs, it reduces only the former outputs. As the element of matrix $\mathbf{Q}$ that connects the two activities is positive and quite large, a reduction of other bovine outputs induces an increase of other non-bovine animal outputs for the respective farm. A closer look at the disaggregated activities produced by the individual farms reveals that those cattle farms that increase their outputs of other non-bovine animals mostly fatten pigs while those that decrease the level of other non-bovine animals predominantly produce sheep. According to our farm data, sheep cause more emissions per value unit than pigs, which is a likely explanation for why the change of the levels of other non-bovine animal outputs is so different between farms.

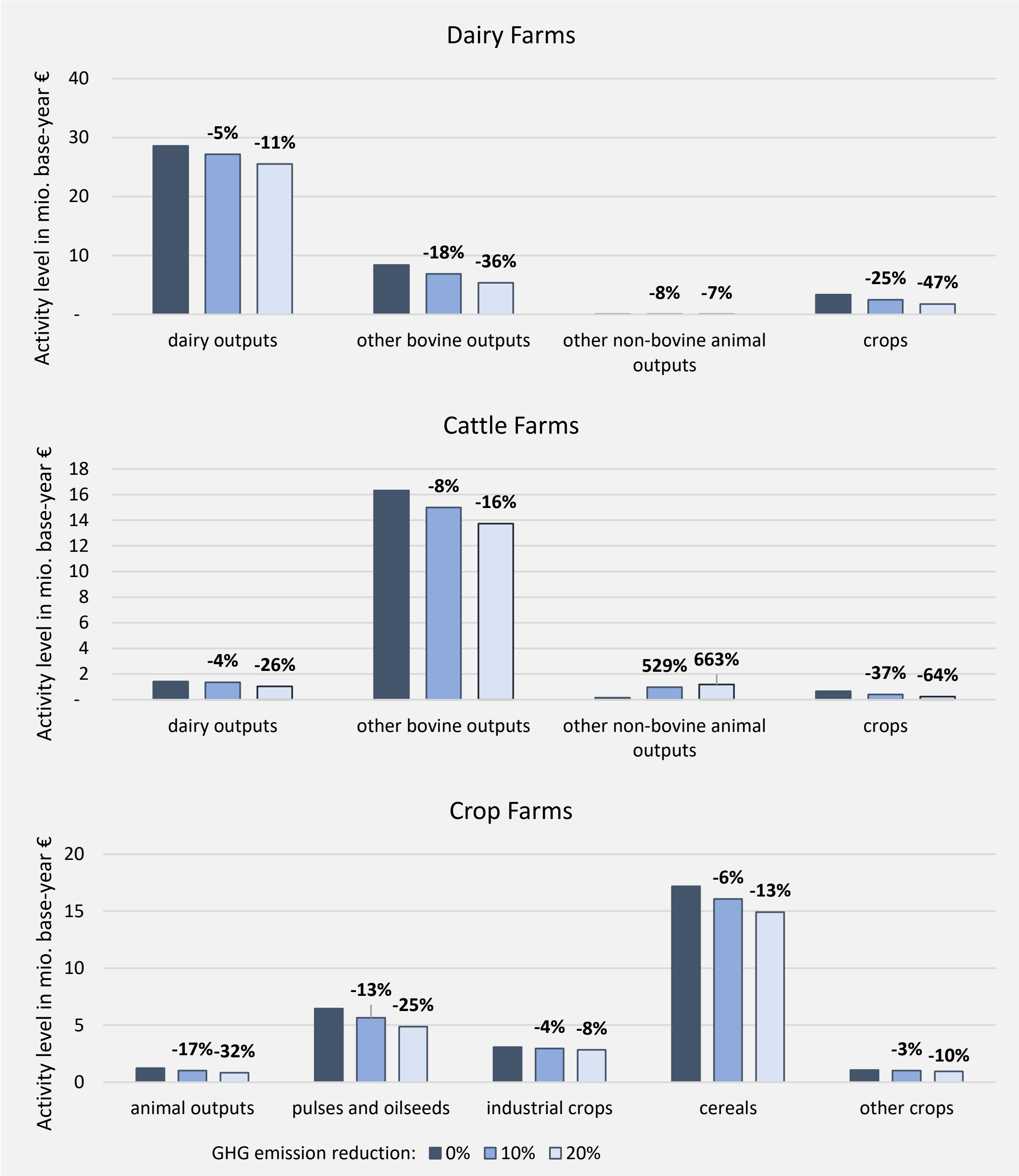

Figure 14: Activity level change for dairy, cattle and crop farms induced by an emission cap. *The height of each columns indicates the total quantity of the output produced by all farms in the sample. The reference scenario and two emission reduction scenarios (10% and 20% reduction) are displayed. The percentage change refers to the output level in the reference scenario. Source: own simulation*

Table 17 shows that the vectors $\mathbf{z_f}$ of the fixed-input levels employed by dairy and crop farms are not impacted by the emission cap. However, this is different for the cropland and the unpaid labour of cattle farms. For a GHG emission cap at 10% below the reference GHG emission nine farms abandon their cropland entirely and give up the activity crop outputs. For a GHG emission cap at 20% below the reference GHG emission this number grows to 14 farms. If the regional GHG emission is reduced by 17.5% or more, three farms entirely cease their employment of unpaid labour.

Change of Fixed-Input Levels and Prices for Different GHG Emission Reduction Scenarios							
farm type	GHG emission reduction	fixed-input level			fixed-input price		
		cropland	grassland	unpaid labour	cropland	grassland	unpaid labour
crop farms from southern Central France	10%	0%	0%	0%	-36%	12%	0%
	20%	0%	0%	0%	-71%	23%	-1%
dairy farms from West France	10%	0%	0%	0%	5%	-7%	-12%
	20%	0%	0%	0%	13%	-16%	-25%
cattle farms from South-West France	10%	-7%	0%	0%	-12%	12%	8%
	20%	-11%	0%	-3%	-28%	23%	8%

Table 17: Change of fixed-input prices and fixed-input levels induced by an emission cap at 10% and 20% below the reference emissions. *The change of the fixed-input price is the weighted (by farm weight and fixed-input level) average of individual farm's fixed-input price change. Source: own simulation*

Strikingly, by reducing the level of employed cropland or unpaid labour to zero, these farms *increase* their TNR. This explains the raise of the TNR generated by all farms together in the cattle sample for some scenarios in figure 15. However, the fact that the objective value of the TNR maximization problem is larger with a GHG emission constraint than without it indicates that the optimum found in the reference scenario is only a local optimum for the cattle sample. We will discuss the reasons and implications of this in chapter 6. It is important to remark that the increase in TNR and the reduction of the utilized fixed-input levels to zero is only observed for 15 out of 203 cattle farms. Consequently, for the large majority of farms the TNR decreases, and the employed fixed-input levels remain constant at their reference level. The impact of the GHG emission constraint on fixed-input prices as represented in table 17 is different for each sample. The displayed fixed-input price changes are averaged over all farms in the respective sample and weighted by the farm's population weight and the fixed-input levels. As activity levels influence the MC with respect to fixed-input levels, the shadow price of the fixed-input constraint is also affected by activity levels via matrix $\mathbf{H}$. Again, a closer look at the fixed-input prices of individual farms is informative. For example, the shadow price of unpaid labour of cattle farms increases only for 36 out of 203 farms. Noteworthily, almost all farms for which the level of non-bovine animal outputs increases also exhibit increasing prices of unpaid labour. On the other hand, for those farms that reduce their produced levels of non-bovine animal outputs the shadow price of unpaid labour decreases. This relation is explained by the strongly negative element of matrix $\mathbf{H}$ that connects non-bovine animal outputs and unpaid labour for cattle farms (table A.10 in Annex IV). If non-bovine animal output levels increase, the MC with respect to the level of unpaid labour decrease (i.e., they become more negative), and the shadow price of unpaid labour consequently increases. Similar interrelations determine the changes of the prices of the other fixed-inputs.

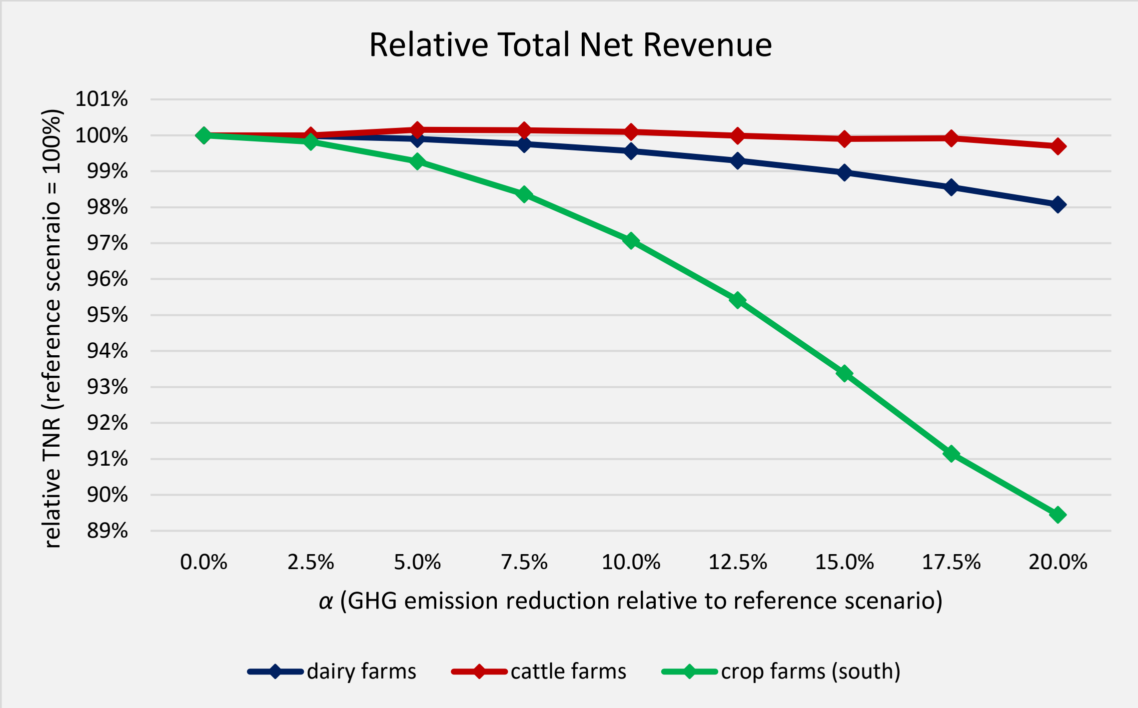

Figure 15: Relative Total Net Revenue generated together by all farms of each farm sample (reference scenario=100%). *Source: own simulation.*

While we have outlined that activities contribute differently to the total GHG emission reduction, it is important to point out that GHG emission mitigation also varies between subregions within one region. Figure A.26 in Annex IV shows that GHG emissions from farms from Poitou-Charentes (West France), Midi-Pyrénées and Aquitaine (South-West France) as well as from Burgundy (southern Central France) decrease more than the average regional GHG emissions. Even though the subregional differences are not substantial, they underpin that the most efficient GHG emission reduction is not necessarily the most even reduction.

5.4. Discussion of Policy Scenario Simulation Results

To assess the reasonability of the simulated MAC presented in the previous section (figure 9) the values estimated by other studies may represent a suitable benchmark. However, we acknowledge that GHG emission abatement from agriculture constitutes a comprehensive and complex field of research. As our simulation is based on very simplified assumptions and mainly serves as a practical test for the developed methodology, comparability with the results found in literature is limited. Nevertheless, the literature provides a range for the order of magnitude that can be expected for MAC:

One of the most comprehensive meta-analyses up to this date was published by Vermont and DeCara (2010) and reviews the assessments of marginal abatement costs of non-CO_2 GHG emission from agriculture based on 21 studies. The reviewed studies include bottom-

up, engineering approaches, supply-side mathematical models, and general or partial equilibrium models. The authors find that, on average over all 21 studies, a 10% emission abatement rate corresponds to an emission price of 20 €/tCO_2eq. This suggests that our estimated MAC for a 10% emission reduction of 10.14 €/tCO_2eq for cattle farms, 17.51 €/tCO_2eq for dairy farms, and 29.99 €/tCO_2eq for crop farms are in a justifiable range. Vermont and DeCara (2010) also conclude that the estimates for MAC vary significantly between studies: An emission price of 20 €/tCO_2eq may lead to an emission abatement between 1% and 57%. Abatement rates for given emission prices are lower for supply-side models than for engineering approaches, as the latter often detect measures that allow to reduce a share of agriculture's GHG emissions at low or even negative costs (Vermont and De Cara (2010)). This is also the case for a more recent engineering study by Pellerin *et al.* (2017), who focussed on the GHG mitigation potential of agriculture in France: They find that about 10% of the total agricultural emissions in France in 2010 could be abated at negative costs, while another 10% would cause costs of less than 25 €/tCO_2eq.
Mosnier *et al.* (2017) analyze the impact of several GHG reduction scenarios on beef and dairy farms in France with an ex-ante simulation for the year 2035. An emission tax of 40 €/tCO_2eq would lead to a GHG emission reduction for grassland-based dairy farms of about 20%. As figure 9 indicates, also according to our model a 20% GHG emission reduction corresponds to an emission price of about 40 €/tCO_2eq for dairy farms. For dairy and cattle farms with temporary grassland and annual crops Mosnier *et al*. (2017) simulate a reduction of more than 70% for dairy farms and more than 50% for cattle farms. The grassland-based cattle farms would cease beef production entirely, while all other farms either produce organically or implement other changes of the production system.

Overall, literature confirms that the MAC estimated in this thesis are plausible and in some cases very close to the values estimated by other researchers. The significant variance between the values calculated in different studies also underpins that the estimation approach, the study region, and the analyzed activity types crucially impact the results. As farms in our model can only reduce their emissions by lowering their output quantity, it seems likely that we overestimate MAC. For example, if farms have the option to convert to organic production as in Mosnier *et al*. (2017) or to adjust their nitrogen fertilization as in the engineering approach implemented by Pellerin *et al*. (2017), emission mitigation might be cheaper. On the other hand, an equilibrium model that accounts for market feedback on output prices would presumably also yield lower MAC. If the GHG emission cap induced a significant decrease of market supply this would most likely have positive effects on output prices, a feed-back that our model disregards.

We have addressed the shortcoming that our model only allows to reduce emissions by lowering activity levels. A further reason for scepticism might concern the calculation of the vector $\mathbf{e}$ of emission factors, which admittedly is oversimplified. For example, the selection of average emission factors for the disaggregated output quantities may in some cases only poorly reflect the true emissions of an activity produced under the respective conditions in West, South-West, or Central France. Besides, the aggregation of emission factors based on observed values pretends that the composition of an activity aggregate cannot be changed. For example, a dairy farm in our model would rather cease crop production entirely than to switch to the production of less GHG emission intensive crops.

Finally, a problematic feature of the model is that the TNR for cattle farms increases if an emission cap is imposed. The reason for this seems to be that some cattle farms modify the levels of utilized fixed-inputs to change their marginal costs. While it is reasonable that farms reduce their levels of employed cropland to zero if they stop producing crop outputs, it has the illogical implication of an increasing TNR. Therefore, the maximum TNR found by the solver in the reference scenario cannot be a global maximum. As this seems to be a methodological issue, we will elaborate on it in the discussion in the next chapter.

However, regardless of the accuracy of the estimated MAC or the modelled supply response, the emission cap simulation performed in this thesis reveals some important properties of the mechanism behind agricultural emission mitigation. Whenever the supply of outputs is associated with costs per ton of emissions, this reduces a farm's marginal costs with respect to activity levels i by $\mu_t e_{t,f,i}$, where μ_t is the cost per emission unit and $e_{t,f,i}$ the GHG emissions per activity level, as in equation (5.5). If the underlying cost function is upward sloping in activity levels, it is likely that farms reduce their MC by reducing their activity level. Notably, this is even the case if farms could reduce GHG emissions by changing technology without reducing activity levels, as long as $\mu_t e_{t,f,i}$ is larger than the increase in marginal revenue.

The reduction of activity levels that is necessary to achieve a certain reduction of marginal costs depends only on the elements of matrix $\mathbf{Q}$, provided that the fixed-input levels remain constant. We demonstrated that, consequently, the *relative* activity level reduction is anti-proportional to the absolute level of an activity supplied by a farm in the reference scenario. We have also seen that GHG emissions are unevenly distributed between farms and that the farms that already accounted for a disproportionally high share of GHG emissions in the reference scenario reduce their emissions relatively less. As emissions are largely proportional to activity levels, it seems that smaller farms reduce their activity levels relatively more than larger farms. This finding should be independent of the model complexity.

Additionally, our model and the simulation show that the change in marginal costs crucially depends on the GHG emissions per activity unit, with implications for the activity levels. For example, according to our simulation, the change of other non-bovine animal outputs produced by some cattle farms hinges on the relationship between the GHG emissions per activity level of other bovine outputs and non-bovine animal outputs (see figure A.24 in Annex

IV). However, the strong *increase* of other non-bovine animal outputs supplied by some cattle farms underpins that one must not ignore how the marginal costs of one activity are impacted by the levels of other activities, through the off-diagonal elements of matrix **Q**. The acknowledgement of these interactions between different activities represents an important strength of the modelling approach analysed throughout this thesis.

6. Conclusion

This chapter concludes by compiling the strengths and weaknesses of the design, the development and the application of the model. First, we discuss and attempt to answer the research questions defined in the introduction of this thesis. Then, the remaining challenges are carved out to identify potential pathways for future research.

6.1. General Discussion

The first objective of this thesis was to test the validity and the applicability of the cost function estimation and calibration approach developed by Henry de Frahan (2019). As in Paris (2017), the differentiation between observed and true values is not only made for activity levels, but also for fixed-input prices. This proved particularly useful as the observed fixed-input prices used for our simulation are national averages, with no distinction made between grassland and cropland, or proxies such as the national minimum wage in the case of unpaid labour. It is very likely that those observed values are not the true fixed-input prices faced by the individual farm. A crucial characteristic of the procedure suggested by Henry de Frahan (2019) is that it performs parameter estimation and calibration in a one-step procedure, in opposite to, e.g., Paris (2015), Petsakos and Rozakis (2015), and Paris (2017). The cost function parameter estimates are those values that allow to fulfil the first-order conditions of the final, *non-linear* TNR maximization problem, by minimizing the error terms of variables that are prone to error. Therefore, the approach accounts for the principle findings of Heckelei and Wolff (2003), who proofed that the traditional PMP two-step procedure does not allow consistent parameter estimation. Instead, Henry de Frahan (2019) follows their econometric perspective that the estimated parameters should belong to the model that best represents the true data-generating process – in this case assumed to be a non-linear cost function depending on activity levels. With the Monte-Carlo simulation performed in section 3.4 we show that the parameter estimates indeed converge to their true values, provided that the error terms have a mean of zero. We extend the model by Henry de Frahan (2019) with a time index and the possibility to capture farm fixed effects. This could improve the independence of the error terms (see section 4.2.1). However, a fundamental weakness becomes apparent when simulating a price change with the FADN data set: The fixed-input-output matrix $\mathbf{A_{t,f}}$ renders the relationship between fixed-inputs and activity levels so rigid that the supply elasticities of some activities are estimated to be zero – ignoring potential substitutability between fixed-inputs. This prompts us to modify the underlying model and to employ the same estimation approach for a cost function that does not impose fixed-input-output coefficients.

The development and the exploration of a cost function that depends on activity levels and fixed-input levels is therefore a second objective of this thesis, based on a methodological note provided by Henry de Frahan for this thesis. Importantly, the resulting model not just allows to simulate activity levels and fixed-input prices, but also fixed-input levels. This is why a third vector $\mathbf{s}_{\mathbf{t,f}}$ of fixed-input level-related error terms is introduced that allows the true utilized fixed-input levels to deviate from the observed levels. A crucial property of this more complete cost function is that it no longer imposes a strict relationship between fixed-input and activity levels. Therefore, the resource constraint no longer confines the extension of output levels. Instead, a farm only stops to produce more of an activity if the marginal costs with respect to it becomes larger than its marginal revenue. As we have seen, this leads to unrealistically large supply elasticities to prices if marginal costs increase too slowly in activity levels, i.e., if the elements of the matrix $\mathbf{Q}$ are very small. To circumvent this problem, we impose minimum values for the diagonal elements of the matrix $\mathbf{Q}$.
By confining the extension of fixed-input levels the resource constraint ensures that the fixed-input price can be positive: In the reference scenario farms find themselves on the part of the convex TVC function that is downward sloping in fixed-input levels. Consequently, the dual value of a fixed-input is positive, as an additional unit of the fixed-input allows the farm to decrease its TVC. Without the fixed-input constraint farms would not enhance their fixed-input utilization infinitely, but until the MC with respect to fixed-input levels are zero.
We demonstrate that with the estimated cost function parameters and error terms it is possible to set up a model that replicates reference values and that can be used for simulation. Instead of simulating activity levels, fixed-input levels and fixed-input prices by minimizing the complementary slackness conditions as proposed by Paris (2017) and Henry de Frahan (2019), we perform a TNR maximization using the estimated cost function parameters and error terms. This allows a more intuitive interpretation of the objective value and dual values. The simulated response to an output price increase is in a reasonable range.

Finally, we test the applicability of the model for agricultural or environmental policy analysis, based on the activity level and fixed-input level dependent cost function. As variable inputs are assumed constant, the range of policies that can be modelled is confined to interventions that affect or are directly linked to fixed-input levels, fixed-input prices, activity levels, or output prices. We deem GHG emissions of agricultural outputs, especially dairy and beef, sufficiently tied to activity levels to apply our model for the simulation of an emission cap at regional level. A comparison of the simulated MAC with MAC estimated by other researchers reveals that our MAC are in a very reasonable range. However, the strong simplification of agricultural emission processes in our model might impair trustworthiness of the accuracy of the results. Researchers or policy actors interested in precise estimates for emission mitigation costs may therefore find models more useful that account for the complex interrelations between emissions and technology, farming systems or variable inputs. However, in many cases the objective is *not* to accurately predict MAC curves, but rather to provide a “coherent forum for the (…) complex discussions surrounding agricultural GHG mitigation” (Eory *et al*., 2018, p. 714). We believe that, from this perspective, the model developed and discussed in this thesis can serve as a truly helpful tool:

First, being a farm supply model, it bases on the assumption that economic optimality conditions are fulfilled. Thereby the model draws attention to the impact of an emission price on marginal costs and, potentially, marginal revenue. If marginal costs are not constant in activity levels, farms adjust their output levels accordingly. In other words, farms do not only reduce their activity levels to lower their GHG emissions, but because the pricing of GHG emissions impacts the MC and therefore requires a change of activity levels to maintain the condition that MC equal MR. As we have seen, the direction and the magnitude of the output level change is not homogenous across activities, because of the interactions between activities. Our model allows to simulate the net effect of GHG emission pricing by acknowledging these interactions between activity levels.

Second, the model illustrates that an efficient reduction of GHG emissions requires to look at an activity’s emissions relative to its value, and not so much at the emissions relative to its quantity. For example, dairy and cattle farms reduce their crop activity relatively more than their livestock activities. Admittedly, it is possible that a more elaborated calculation of emission factors may lead to different conclusions in this specific case. Nevertheless, our results highlight that the “low hanging fruits” in terms of GHG emission reduction are not necessarily those activities with the highest GHG emissions per quantity.

Third, the same is true with regard to which farms reduce GHG emissions. We show that the most efficient allocation of GHG emission rights may not always lead to even GHG emission reduction across farms and subregions. This supports at a regional level the results that De Cara and Jayet (2011) as well as Fellmann *et al*. (2018) found for the EU as a whole: Market-based emission reduction may lead to considerably lower mitigation costs compared to inflexible national reduction targets. Presumably, the MAC calculated in this thesis would be lower and the GHG emission reduction even more efficient if the emission cap was imposed at national or European level. Besides, the fact that MAC vary between crop farms, dairy farms and cattle farms implies that a trade of GHG emission certificates between farms with different specialization would very likely reduce total GHG emission abatement costs. However, some researchers have pointed out that efficient, but uneven GHG emission reduction may have undesired effects on distributional justice and may be perceived as unfair (Caney and Hepburn (2011)). For example, our simulation results suggests that a GHG emission cap at regional level would lead to a relatively stronger reduction of GHG emissions and activity levels by smaller farms. The model could therefore help quantify the benefits of market-based environmental protection and identify potentially problematic distributional inequalities, so that policies can be designed that sufficiently balance efficiency and social acceptability.

6.2. Challenges and Future Improvements

Several limitations and challenges remain that must be addressed in the future for the model to fully unfold the potential described in the previous section. A first reservation concerns the compliance with a fundamental precept of PMP-related models, namely the inclusion of all available information. As suggested by Paris (2017) and Henry de Frahan (2019), we incorporate observed fixed-input prices into the model as shadow prices of the resource constraint. However, for all farm samples, except for the dairy sample, the fixed-input price-related error terms $\mathbf{u_{t,f}}$ seem to be systematically strictly positive for cropland and systematically strictly negative for grassland (see figures A.6, A.9 and A.15 in the Annex IV). Recall that the observed price for both grassland and cropland in France is the national average rental price of land. It is likely that the "true" rental rates systematically deviate from this observed value at the spatial dimension (e.g., between regions) and between the types of usage (cropland or grassland). From Eurostat (2021a) we can retrieve the information that land sales prices in all years from 2011 to 2016 are above the national mean in Central France (Bassin Parisien) and South-West France and that, across the EU, cropland is more expensive than permanent grassland. It is conceivable that these relations also exist between rental rates of land in the years 2005 - 2012. This raises the question how the model could be adapted to acknowledge justified assumptions, for example that rental rates for cropland are higher than for grassland, even if no certain and precise information on rental rates for cropland and grassland for the respective years and regions is available. The ignorance of these likely (but uncertain) relations seems to violate the independency of the error terms.
A second weakness of the model based on the fixed-input and activity level dependent cost function became apparent when implementing the GHG emission cap simulation for cattle farms: the simulated regional TNR for some GHG emission reduction scenarios is larger than in the reference scenario, which confirms that the optimal solution found in the reference scenario is a local optimum. The fixed-input and activity levels simulated for the reference scenario are very close to the "true" values given as the sum of the observed value and its error term. This shows that there are other admissible values that allow to generate a higher TNR than the TNR obtained with the "true" values, which by definition fulfil the optimality conditions. To understand this apparent paradox, it is important to see that the individual farms whose TNR is larger if a GHG emission constraint is imposed than their TNR in the reference scenario all reduce at least one of their fixed-input levels to zero. In the estimation

phase we set the vector of fixed-input levels $\mathbf{z^o_{t,f}}$ equal to the vector of fixed-input constraints $\mathbf{b_{t,f}}$ and therefore state that the elements of the vectors $\mathbf{y^o_{t,f}}$ and $\mathbf{y_{t,f}}$ are strictly positive for every observed fixed-input. We thereby assume that the fixed-input constraint is binding in the optimal solution, and only under *this* assumption the generated TNR is optimal. As long as output prices or constraints are identical or very close to their values in the estimation phase, the GAMS solver finds a local optimum where simulated activity and fixed-input levels are equal to their "true" values. This is the conventional idea of calibration and, in our case, it can be improved by using the "true" values as initial values for the solver. However, if output price changes or modifications of the constraints require the solver to search for optimal values that are away from the reference values, it detects that setting at least one of the fixed-input levels to zero may lead to an optimal solution. Note that this means that the MC with respect to fixed-input levels no longer need to be equal to the shadow price, which becomes zero. While this leads to the illogical result of a TNR value that is higher with a GHG emission constraint than without, it also means that the total level of fixed-inputs used at the regional scale decreases, i.e., that crop- or grassland falls idle. At the same time, the shadow price of cropland and grassland is positive for many other farms, indicating that there is still demand for crop- and grassland. A potential pathway for further improvements of the model could therefore be to allow trade of fixed-inputs between farms within one subregion, so that the fixed-resource constraint at the farm level is replaced by a fixed-resource constraint at the subregional level. This means that the shadow price of each resource is positive for *all* farms as long as there is at least one farm which could decrease its TVC if more fixed-resources are available at the subregional level than in the reference scenario. Without a fixed-input constraint at the farm level that prevents the extension of a farm's crop- or grassland it is possible that all except one farm exit the market, while the remaining farm uses the entire crop- and grassland of a subregion. However, in the reference scenario the model would calibrate to optimal fixed-input levels that are equal to the observed fixed-input level $z^o_{t,f,j}$ plus a minimized error term $s_{t,f,j}$, which should prevent the agglomeration of fixed-inputs. Besides, thanks to the upward sloping MC with respect to activity levels it seems unlikely that MC would still equal MR if one single farm drastically increased its activity levels. If one restricts the level of unpaid labour at the farm level, which appears reasonable, the growth of individual farms would be limited even stronger. The change of the level at which the fixed-resource constraint is imposed would have to be implemented already in the estimation phase and the number of fixed-input price-related error terms would decrease to the number of subregions multiplied by the number of years. This would help address another problem: Currently, constraint (3.49) imposes a rigid relationship between the vector $\mathbf{s_{t,f}}$ of fixed-input level related error terms and the vector $\mathbf{u_{t,f}}$ of fixed-input prices related error terms. Consequently, the fixed-input level related error terms are impacted by the same independency problems that we already discussed for the fixed-input price related error terms. If the fixed-input price related error terms indicate deviations between observed and true fixed-input prices at the subregional level, the rigid linkage to the fixed-input level related error terms, which are defined at the farm level, would cease to exist. It remains to be seen how this modification would impact the problem of a TNR that is larger with a GHG emission constraint than without it.

A third obvious shortcoming of the simulation performed in this thesis is the fact that we impose minimum boundaries for the diagonal of the quadratic cost function matrix $\mathbf{Q}$ to receive estimates for the elasticities that are in a reasonable range. This is certainly an important deficiency, as an explicit objective of bridging PMP and econometrics is to derive the necessary information on the cost function parameters from the data. Besides, as the supply response to an emission constraint is significantly determined by matrix $\mathbf{Q}$, all interventions that affect matrix $\mathbf{Q}$ are likely to also impact the simulated MAC.

Presumably the reason for why the diagonal elements of $\mathbf{Q}$ are so close to zero when there is no lower boundary is that the output prices for an activity received by a farm are identical for all farms in the region. Consequently, there are only eight different observations (one per year) for the marginal revenue and therefore the marginal costs of each activity. As the elements of $\mathbf{Q}$ indicate the change of marginal costs for a change of an activity level, the estimation of $\mathbf{Q}$ requires many sufficiently different observations for marginal costs. We therefore expect that by using output prices with stronger variance the imposition of lower boundaries for the diagonal elements of matrix $\mathbf{Q}$ would become redundant. However, one must keep in mind that the law of a single market prevents large variations between the activity prices faced by different farms within the same region. Therefore, the inclusion of observations from more years in the estimation phase might be a promising approach to obtain stronger variance of marginal revenues across the full panel data set.

We have shown that with the EMP approach developed and tested throughout this thesis it is possible to estimate cost function parameters that can replicate base-year values and that can be used to simulate policy scenarios. This EMP approach allows to bridge the PMP procedure with the econometric conception of a “true” underlying model that can be estimated if sufficient data is available. Even though the methodological foundations for EMP have already been laid in the early 2000s, there are still only few applications of econometric mathematical programming for agricultural or environmental policy analysis. Contrarily, the set-up of a first-stage linear model as well as other vestiges of the original PMP method persist in younger literature. The approach presented and applied in this thesis therefore represents an easily implementable alternative belonging to the class of EMP models. Some of the flaws that we identified in the previous section could presumably be remedied by using exogenous variables that exhibit stronger variance between observations. Other shortcomings might require some model adjustments, such as the introduction of the possibility to trade some fixed-inputs between farms. Further developments and applications of the model are necessary to shed more light on the identified flaws and to explore the model's potential for the analysis of different agricultural or environmental policies.

References

Agribalyse. (2020a). *Tableur pour les produits agricoles bruts (à la sortie de la ferme) pour les productions bio*. https://doc.agribalyse.fr/documentation/acces-donnees

Agribalyse. (2020b). *Tableur pour les produits agricoles bruts (à la sortie de la ferme) pour les productions conventionnels*. https://doc.agribalyse.fr/documentation/acces-donnees

Arata, L., Donati, M., Sckokai, P., & Arfini, F. (2017). Incorporating risk in a positive mathematical programming framework: A dual approach. *Australian Journal of Agricultural and Resource Economics*, *61*(2), 265–284. https://doi.org/10.1111/1467-8489.12199

Arfini, F., & Donati, M. (2011). The impact of the Health Check on structural change and farm efficiency: A comparative assessment of three European agricultural regions. In C. Moreddu (Ed.), *Disaggregated Impacts of CAP Reforms* (pp. 75–90). OECD. https://doi.org/10.1787/9789264097070-7-en

Arfini, F., Donati, M., Grossi, L., & Paris, Q. (2008). Revenue and Cost Functions in PMP: A Methodological Integration for a Territorial Analysis of CAP. *Paper Presented at the 107th EAAESeminar 'Modelling of Agricultural and Rural Development Policies'. Sevilla, Spain*. https://doi.org/10.22004/AG.ECON.6636

Arrow, K. (1965). *Aspects of the Theory of Risk Bearing*. Yrjö Jahnsson Foundation.

Balkhausen, O., Banse, M., Grethe, H., & Nolte, S. (2005). *Modelling the Effects of Partial Decoupling on Crop and Fodder Area And Ruminant Supply in the EU: Current State and Outlook*. *proceedings of the 89th EAAE Seminar*, 565–587. https://doi.org/10.22004/AG.ECON.231832

Beckman, J., Ivanic, M., Jelliffe, J., Baquedano, F. G., & Scott, S. G. (2020). *Economic and Food Security Impacts of Agricultural Input Reduction Under the European Union Green Deal's Farm to Fork and Biodiversity Strategies* (EB-30). U.S. Department of Agriculture, Economic Research Service.

Benoit, C. (1924). Notes sur une Méthode de Résolution des Équations Normales Provenant de l' Application de la Méthode des Moindres Carrés a un Systéme d'Équations Linéaires en Nombre Inférieur a Celui des Inconnues.-Application de la Méthode a la Résolution d'un Systéme Defini d'Équations Linéaires, (Procédé du Commandant Cholesky). *Bulle Géodésique*, *2*, 67–77.

Britz, W., & Arata, L. (2019). Econometric mathematical programming: An application to the estimation of costs and risk preferences at farm level. *Agricultural Economics*, *00*, 1–16. https://doi.org/10.1111/agec.12476

Butault, J.-P., Zardet, G., Mathias, L., Delame, N., Desbois, D., Rousselle, J.-M., Kleinhanss, W., & Offermann, F. (2011). *User guide: The FACEPA model software (costs of production: FADN)* (FACEPA Deliverable No. D4.2).

Buysse, J., Van Huylenbroeck, G., & Lauwers, L. (2007). Normative, positive and econometric mathematical programming as tools for incorporation of multifunctionality in agricultural policy modelling. *Agriculture, Ecosystems & Environment*, *120*(1), 70–81. https://doi.org/10.1016/j.agee.2006.03.035

Caney, S., & Hepburn, C. (2011). Carbon Trading: Unethical, Unjust and Ineffective? *Royal Institute of Philosophy Supplement*, *69*, 201–234. https://doi.org/10.1017/S1358246111000282

Carpentier, A., Gohin, A., Sckokai, P., & Thomas, A. (2015). Economic modelling of agricultural production: Past advances and new challenges. *Revue d'Etudes En Agriculture et Environnement - Review of Agricultural and Environmental Studies*, *96*(1), 131–165.

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. Cambridge University Press.

Ciaian, P., Kancs, D., Swinnen, J. F. M., Van Herck, K., Vranken, L., Ciaian, P., Kancs, D., Swinnen, J. F. M., Van Herck, K., & Vranken, L. (2012). *Institutional Factors Affecting Agricultural Land Markets*. https://doi.org/10.22004/AG.ECON.120251

Cook, S. (2019). *Forecast Evaluation using Theil's Inequality Coefficients*. The Economics Network. https://doi.org/10.53593/n3168a

Cypris, C. (2000). *Positive Mathematische Programmierung (PMP) im Agrarsektormodell RAUMIS*. Forschungsgesellschaft für Agrarpolitik und Agrarsoziologie.

De Cara, S., Houzé, M., & Jayet, P.-A. (2005). Methane and Nitrous Oxide Emissions from Agriculture in the EU: A Spatial Assessment of Sources and Abatement Costs. *Environmental and Resource Economics*, *32*(4), 551–583. https://doi.org/10.1007/s10640-005-0071-8

De Cara, S., & Jayet, P.-A. (2011). Marginal abatement costs of greenhouse gas emissions from European agriculture, cost effectiveness, and the EU non-ETS burden sharing agreement. *Ecological Economics*, *70*(9), 1680–1690. https://doi.org/10.1016/j.ecolecon.2011.05.007

De Vries, A., Overton, M., Fetrow, J., Leslie, K., Eicker, S., & Rogers, G. (2008). Exploring the Impact of Sexed Semen on the Structure of the Dairy Industry. *Journal of Dairy Science*, *91*(2), 847–856. https://doi.org/10.3168/jds.2007-0536

Dean, E. R., Harper, M. J., & Sherwood, M. S. (1996). Productivity measurement with changing-weight indexes of outputs and inputs. *OECD, Industry Productivity: International Comparison Measurement Issues*.

Domínguez, I. P., & Fellmann, T. (2015). The Need for Comprehensive Climate Change Mitigation Policies in European Agriculture. *EuroChoices*, *14*(1), 11–16. https://doi.org/10.1111/1746-692X.12076

EEA. (2019). *Climate change adaptation in the agriculture sector in Europe* (EEA Report No 4/2019, European Environment Agency). https://www.eea.europa.eu/publications/cc-adaptation-agriculture

EEA. (2021). *EEA greenhouse gases—Data viewer*. https://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer

Eory, V., Pellerin, S., Carmona Garcia, G., Lehtonen, H., Licite, I., Mattila, H., Lund-Sørensen, T., Muldowney, J., Popluga, D., Strandmark, L., & Schulte, R. (2018). Marginal abatement cost curves for agricultural climate policy: State-of-the art, lessons learnt and future potential. *Journal of Cleaner Production*, *182*, 705–716. https://doi.org/10.1016/j.jclepro.2018.01.252

European Commission. (2019). *COM(2019) 640 final—Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions—The European Green Deal*.

European Commission. (2020). *COM(2020) 381 final—Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—A Farm to Fork Strategy for a fair, healthy and environmentally-friendly food system*.

European Commission. (2021). *EU Emissions Trading System (EU ETS)*. https://ec.europa.eu/clima/eu-action/eu-emissions-trading-system-eu-ets_en

Eurostat. (2018). land prices and rents—Annual data (1985-2016). *[Apri_ap_aland]*.

Eurostat. (2020). *Glossary: Livestock unit (LSU)*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Livestock_unit_(LSU)

Eurostat. (2021a). Agricultural land prices by region (2011-2020). *[Apri_lprc]*.

Eurostat. (2021b). Animal populations by NUTS 2 regions (2005-2012). *[Agr_r_animal]*.

Eurostat. (2021c). Crop production in EU standard humidity by NUTS 2 regions (2005-2012). *[Apro_cpshr]*.

Eurostat. (2021d). Farm indicators by agricultural area, type of farm, standard output, legal form and NUTS2 regions (2010). *[Ef_m_farmleg]*.

Eurostat. (2021e). *NUTS - GISCO: Geographische Informationen und Karten*. https://ec.europa.eu/eurostat/de/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

Eurostat. (2021f). Production and utilization of milk on the farm—Annual data (2005-2012). *[Apro_mk_farm]*.

Eurostat. (2021g). Production of cow's milk on farms by NUTS 2 regions (2005-2012). *[Agr_r_milkpr]*.

Fellmann, T., Witzke, P., Weiss, F., Van Doorslaer, B., Drabik, D., Huck, I., Salputra, G., Jansson, T., & Leip, A. (2018). Major challenges of integrating agriculture into climate change mitigation policy frameworks. *Mitigation and Adaptation Strategies for Global Change*, *23*(3), 451–468. https://doi.org/10.1007/s11027-017-9743-2

Frick, F., & Sauer, J. (2021). Technological Change in Dairy Farming with Increased Price Volatility. *Journal of Agricultural Economics*, *72*(2), 564–588. https://doi.org/10.1111/1477-9552.12417

Gohin, A., & Chantreuil, F. (1999). *La programmation mathématique positive dans les modèles d'exploitation agricole: Principes et importance du calibrage*. https://doi.org/10.22004/AG.ECON.211096

Golan, A., Judge, G. G., & Miller, D. (1996). *Maximum entropy econometrics: Robust estimation with limited data*. Wiley.

Graveline, N., & Merel, P. (2014). Intensive and extensive margin adjustments to water scarcity in France's Cereal Belt. *European Review of Agricultural Economics*, *41*(5), 707–743. https://doi.org/10.1093/erae/jbt039

Hasegawa, T., Fujimori, S., Havlík, P., Valin, H., Bodirsky, B. L., Doelman, J. C., Fellmann, T., Kyle, P., Koopman, J. F. L., Lotze-Campen, H., Mason-D'Croz, D., Ochi, Y., Pérez Domínguez, I., Stehfest, E., Sulser, T. B., Tabeau, A., Takahashi, K., Takakura, J., van Meijl, H., … Witzke, P. (2018). Risk of increased food insecurity under stringent global climate change mitigation policy. *Nature Climate Change*, *8*(8), 699–703. https://doi.org/10.1038/s41558-018-0230-x

Heckelei, T. (2002). *Calibration and Estimation of Programming Models for Agricultural Supply Analysis*. University of Bonn, Germany.

Heckelei, T., & Britz, W. (2000). Positive Mathematical Programming with Multiple Data Points: A Cross-Sectional Estimation Procedure. *Cahiers d'Economie et de Sociologie Rurales, INRA Editions*, *57*, 27–50. https://doi.org/10.22004/AG.ECON.206148

Heckelei, T., & Britz, W. (2005). Models Based on Positive Mathematical Programming: State of the Art and Further Extensions. *Conference Paper for European Association of Ag-*

ricultural Economists, 89th Seminar, Parma, Italy, 48–73. https://doi.org/10.22004/AG.ECON.234607

Heckelei, T., Britz, W., & Zhang, Y. (2012). Positive Mathematical Programming Approaches – Recent Developments in Literature and Applied Modelling. *Bio-Based and Applied Economics Journal*, *01*(1), 109–124. https://doi.org/10.22004/AG.ECON.125722

Heckelei, T., Mittelhammer, R. C., & Jansson, T. (2008). *A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models*. https://doi.org/10.22004/AG.ECON.56973

Heckelei, T., & Wolff, H. (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy. *European Review of Agriculture Economics*, *30*(1), 27–50. https://doi.org/10.1093/erae/30.1.27

Helming, J. F. M., Peeters, L., & Veenendaal, P. J. J. (2001). *Assessing the consequences of environmental policy scenarios in Flemish agriculture* (T. Heckelei, W. Britz, & W. Henrichsmeyer, Eds.). Wiss.-Verl. Vauk.

Henry de Frahan, B. (2019). Towards Econometric Mathematical Programming for Policy Analysis. In S. Msangi & D. MacEwan (Eds.), *Applied Methods for Agriculture and Natural Resource Management* (Vol. 50, pp. 11–36). Springer International Publishing. https://doi.org/10.1007/978-3-030-13487-7_2

Henry de Frahan, B., Baudry, A., De Blander, R., Polome, P., & Howitt, R. (2011). Dairy farms without quotas in Belgium: Estimation and simulation with a flexible cost function. *European Review of Agricultural Economics*, *38*(4), 469–495. https://doi.org/10.1093/erae/jbr013

Henry de Frahan, B., Buysse, J., Polomé, P., Fernagut, B., Harmignie, O., Lauwers, L., Huylenbroeck, G., & Meensel, J. (2007). Positive Mathematical Programming for Agricultural and Environmental Policy Analysis: Review and Practice. In A. Weintraub, C. Romero, T. Bjørndal, R. Epstein, & J. Miranda (Eds.), *Handbook Of Operations Research In Natural Resources* (Vol. 99, pp. 129–154). Springer US. https://doi.org/10.1007/978-0-387-71815-6_8

Henry de Frahan, B., Dong, J., & De Blander, R. (2015). *Multi-input multi-output cost function for IFM-CAP model* (Deliverable 8: Final Report; Earth and Life Institute, Université Catholique de Louvain).

Himics, M., Fellmann, T., Barreiro-Hurlé, J., Witzke, H.-P., Pérez Domínguez, I., Jansson, T., & Weiss, F. (2018). Does the current trade liberalization agenda contribute to greenhouse gas emission mitigation in agriculture? *Food Policy*, *76*, 120–129. https://doi.org/10.1016/j.foodpol.2018.01.011

Howitt, R. E. (1995). Positive Mathematical Programming. *American Journal of Agricultural Economics*, *77*(2), 329–342. https://doi.org/10.2307/1243543; an accompanying paper that is not available to the author of this thesis is: Howitt, R. E. (1995b). A Calibration Method for Agricultural Economic Production Models. *Journal of Agricultural Economics*, *46*(2), 147–159. https://doi.org/10.1111/j.1477-9552.1995.tb00762.x

Ievoli, C., Basile, R. G., & Belliggiano, A. (2017). The Spatial Patterns of Dairy Farming In Molise. *European Countryside*, *9*(4), 729–745. https://doi.org/10.1515/euco-2017-0041

IGN. (2021). géoportail—Cartes. *Géoportail, Le Portail National de La Connaissance Du Territoire*. https://www.geoportail.gouv.fr/carte

IPCC. (2006). Emissions from livestock and manure management. In *IPCC Guidelines for National Greenhouse Gas Inventories: Vol. 4: Agriculture, Forestry and Other Land Use*.

IPCC. (2014a). *Climate change 2013: The physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

IPCC. (2014b). *Summary for Policymakers* (Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change; pp. 1–32). IPCC.

IPCC. (2021). *Summary for Policymakers* (Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change). IPCC.

Jansson, T., & Heckelei, T. (2011). Estimating a Primal Model of Regional Crop Supply in the European Union: Regional Crop Supply in the EU. *Journal of Agricultural Economics*, *62*(1), 137–152. https://doi.org/10.1111/j.1477-9552.2010.00270.x

Jongeneel, R. A., & Tonini, A. (2009). *The impact of quota rent and supply elasticity estimates for EU dairy policy evaluation: A comparative analysis*. *58*(05–06), 269–278. https://doi.org/10.22004/AG.ECON.134880

Karush, W. (1939). *Minima of functions of several variables with inequalities as side conditions (M.Sc. Thesis)*. Department of Mathematics, University of Chicago.

Klootwijk, C. W., Van Middelaar, C. E., Berentsen, P. B. M., & de Boer, I. J. M. (2016). Dutch dairy farms after milk quota abolition: Economic and environmental consequences of a new manure policy. *Journal of Dairy Science*, *99*(10), 8384–8396. https://doi.org/10.3168/jds.2015-10781

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear Programming. *Proceeding of 2nd Berkeley Symposium*, 481–492.

Lanigan, G., Donnellan, T., Hanrahan, K., Carsten, P., Shalloo, L., Krol, D., Forrestal, P. J., Farrelly, N., O'Brien, D., Ryan, M., Murphy, P., Caslin, B., Spink, J., Finnan, J., Boland, A., Upton, J., & Richards, K. G. (2018). *An Analysis of Abatement Potential of Greenhouse Gas Emissions in Irish Agriculture 2021-2030* [Technical Report]. Teagasc. https://t-stor.teagasc.ie/handle/11019/2092

Leip, A., Weiss, F., Wassenaar, T., Perez, I., Fellmann, T., Loudjani, P., Tubiello, F., Monni, S., Grandgirard, D., & Biala, K. (2010). *Evaluation of the livestock sector's contribution to the EU greenhouse gas emissions (CGELS)*. Joint Research Center of the European Commission.

Macleod, M., Eory, V., Gruere, G., & Lankoski, J. (2015). *Cost-Effectiveness of Greenhouse Gas Mitigation Measures for Agriculture: A Literature Review* (No. 89; OECD Food, Agriculture and Fisheries Papers). https://doi.org/10.1787/5jrvvkq900vj-en

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91.

McCarl, B., Meeraus, A., van der Eijk, P., Bussiek, M., Dirkse, S., & Nelissen, F. (2016). McCarl Expanded GAMS User Guide Version 24.6. *McCarl GAMS User Guide*.

Mérel, P., & Bucaram, S. (2010). Exact calibration of programming models of agricultural supply against exogenous supply elasticities. *European Review of Agricultural Economics*, *37*(3), 395–418. https://doi.org/10.1093/erae/jbq024

Mérel, P., & Howitt, R. (2014). Theory and Application of Positive Mathematical Programming in Agriculture and the Environment. *Annual Review of Resource Economics*, *6*(1), 451–470. https://doi.org/10.1146/annurev-resource-100913-012447

Morris, J., Paltsev, S., & Reilly, J. (2012). Marginal Abatement Costs and Marginal Welfare Costs for Greenhouse Gas Emissions Reductions: Results from the EPPA Model. *Environmental Modeling & Assessment*, *17*(4), 325–336. https://doi.org/10.1007/s10666-011-9298-7

Mosnier, C., Duclos, A., Agabriel, J., & Gac, A. (2017). What prospective scenarios for 2035 will be compatible with reduced impact of French beef and dairy farm on climate change? *Agricultural Systems*, *157*, 193–201. https://doi.org/10.1016/j.agsy.2017.07.006

Mosnier, C., & Wieck, C. (2010). Determinants of spatial dynamics of dairy production: A review. *Agricultural and Resource Economics, Discussion Paper*, *2*. https://doi.org/10.22004/AG.ECON.162896

Mußhoff, O., & Hirschauer, N. (2020). *Modernes Agrarmanagement: Betriebswirtschaftliche Analyse- und Planungsverfahren* (5., überarbeitete und erweiterte Auflage). Verlag Franz Vahlen.

Neuenfeldt, S., & Gocht, A. (2014). *A handbook on the use of FADN database in programming models*. Johann Heinrich von Thünen-Institut. https://doi.org/10.3220/WP_35_2014

OECD. (2021). *Minimum wages at current prices in NCU*. https://stats.oecd.org/Index.aspx?DataSetCode=MW_CURP#

Paris, Q. (2001). Symmetric Positive Equilibrium Problem: A Framework for Rationalizing Economic Behavior with Limited Information. *American Journal of Agricultural Economics*, *83*(4), 1049–1061. https://doi.org/10.1111/0002-9092.00229

Paris, Q. (2010). *Economic Foundations of Symmetric Programming*. Cambridge University Press.

Paris, Q. (2015). Positive Mathematical Programming with Generalized Risk: A Revision. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2594573

Paris, Q. (2017). Cost function and positive mathematical programming. *Bio-Based and Applied Economics*, *6*(1), 19–35. https://doi.org/10.13128/BAE-18140

Paris, Q., & Howitt, R. E. (1998). An Analysis of Ill-Posed Production Problems Using Maximum Entropy. *American Journal of Agricultural Economics*, *80*(1), 124–138. https://doi.org/10.2307/3180275

Pe'er, G., Zinngrebe, Y., Moreira, F., Sirami, C., Schindler, S., Müller, R., Bontzorlos, V., Clough, D., Bezák, P., Bonn, A., Hansjürgens, B., Lomba, A., Möckel, S., Passoni, G., Schleyer, C., Schmidt, J., & Lakner, S. (2019). A greener path for the EU Common Agricultural Policy. *Science*, *365*(6452), 449–451. https://doi.org/10.1126/science.aax3146

Pellerin, S., Bamière, L., Angers, D., Béline, F., Benoit, M., Butault, J.-P., Chenu, C., Colnenne-David, C., De Cara, S., Delame, N., Doreau, M., Dupraz, P., Faverdin, P., Garcia-Launay, F., Hassouna, M., Hénault, C., Jeuffroy, M.-H., Klumpp, K., Metay, A., … Chemineau, P. (2017). Identifying cost-competitive greenhouse gas mitigation potential of French agriculture. *Environmental Science & Policy*, *77*, 130–139. https://doi.org/10.1016/j.envsci.2017.08.003

Petsakos, A., & Rozakis, S. (2011). Integrating risk and uncertainty in PMP models. *European Association of Agricultural Economists (EAAE); 2011 International Congress, August 30-September 2, 2011, Zurich, Switzerland*. https://doi.org/10.22004/AG.ECON.114762

Petsakos, A., & Rozakis, S. (2015). Calibration of agricultural risk programming models. *European Journal of Operational Research*, *242*(2), 536–545. https://doi.org/10.1016/j.ejor.2014.10.018

Peyraud, J.-L., & MacLeod, M. (2020). *Study on future of EU livestock how to contribute to a sustainable agricultural sector?: Final report*. https://doi.org/10.2762/3440

Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric models and economic forecasts* (4. ed). Irwin McGraw-Hill.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, *32*(1/2), 122–136. https://doi.org/10.1016/B978-0-12-214850-7.50010-3

Preckel, P. V. (2001). Least Squares and Entropy: A Penalty Function Perspective. *American Journal of Agricultural Economics*, *83*(2), 366–377. https://doi.org/10.1111/0002-9092.00162

Rajaniemi, M., Mikkola, H., & Ahokas, J. (2011). Greenhouse gas emissions from oats, barley, wheat and rye production. *Agronomy Research*, *Biosystem Engineering Special Issue 1*, 189–195.

Regulation (EC) 2021/1119, no. Regulation (EC) 2021/1119.

Röhm, O., & Dabbert, S. (2003). Integrating Agri-Environmental Programs into Regional Production Models: An Extension of Positive Mathematical Programming. *American Journal of Agricultural Economics*, *85*(1), 254–265. https://doi.org/10.1111/1467-8276.00117

Severini, S., & Cortignani, R. (2012). Modeling farmer participation to a revenue insurance scheme by means of Positive Mathematical Programming. *Agricultural Economics*, *Czech 58*(7), 324–331. https://doi.org/10.22004/AG.ECON.116001

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Thünen, J. H., von. (1842). *Der isolierte Staat in Beziehung auf Landwirtschaft und National-ökonomie*.

United Nations. (2021). *The Paris Agreement | UNFCCC*. UNFCC Sites and Platforms. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

Van Doorslaer, B., Witzke, P., Huck, I., Weiss, F., Fellmann, T., Salputra, G., Jansson, T., Drabik, D., & Leip, A. (2015). *An economic assessment of GHG mitigation policy options for EU agriculture*. https://doi.org/10.2791/180800

Vermont, B., & De Cara, S. (2010). How costly is mitigation of non-CO2 greenhouse gas emissions from agriculture?: A meta-analysis. *Ecological Economics*, *69*(7), 1373–1386. https://doi.org/10.1016/j.ecolecon.2010.02.020

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (Sixth edition, student edition). Cengage Learning.

Annex I – Theoretical Framework

Derivation of equation 10 based on Heckelei and Britz (2005)

First, state the Lagrangian formulation of the non-linear PMP model (equations (5) and (6)):

$$L(x) = \mathbf{p}'\mathbf{x} - \mathbf{c}'\mathbf{x} - 0.5\mathbf{x}'\mathbf{Q}\mathbf{x} + \mathbf{y}(\mathbf{b} - \mathbf{A}\mathbf{x}) \tag{I.1}$$

If all optimal activity levels are positive, the first order conditions are:

$$\frac{\partial \mathbf{L}}{\partial \mathbf{x}} = \mathbf{p} - \mathbf{c} - \mathbf{Q}\mathbf{x} - \mathbf{A}'\mathbf{y} = \mathbf{0} \tag{I.2}$$

$$\frac{\partial \mathbf{L}}{\partial \mathbf{y}} = \mathbf{b} - \mathbf{A}\mathbf{x} = \mathbf{0} \tag{I.3}$$

Solving (I.2) for x:

$$\mathbf{x} = \mathbf{Q}^{-1}(\mathbf{p} - \mathbf{c} - \mathbf{A}'\mathbf{y}) \tag{I.4}$$

Substituting (I.4) in (I.3):

$$\mathbf{b} - \mathbf{A}\left(\mathbf{Q}^{-1}(\mathbf{p} - \mathbf{c} - \mathbf{A}'\mathbf{y})\right) = \mathbf{0} \tag{I.5}$$

Solving (I.5) for y:

$$\mathbf{y} = \left(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\right)^{-1}(\mathbf{A}\mathbf{Q}^{-1}(\mathbf{p} - \mathbf{c}) - \mathbf{b}) \tag{I.6}$$

Substituting (I.6) in (I.2) and solving for x:

$$\mathbf{x} = \mathbf{Q}^{-1}(\mathbf{p} - \mathbf{c}) - \mathbf{Q}^{-1}\mathbf{A}'(\left(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\right)^{-1}(\mathbf{A}\mathbf{Q}^{-1}(\mathbf{p} - \mathbf{c}) - \mathbf{b}) \tag{I.7}$$

Finally, equation (10) can be derived as the derivative of (I.7) w.r.t. p:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{p}} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}')^{-1}\mathbf{A}\mathbf{Q}^{-1}$$

Cholesky factorization example

For a (3×3) symmetric, positive semidefinite matrix the Cholesky factorization would look as follows (Paris, 2010):

$$\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} h_{11} & 0 & 0 \\ 0 & h_{22} & 0 \\ 0 & 0 & h_{33} \end{bmatrix} \begin{bmatrix} 1 & l_{21} & l_{31} \\ 0 & 1 & l_{32} \\ 0 & 0 & 1 \end{bmatrix} \tag{I.8}$$

Annex II – Methodological Framework

Within Transformation for simple cost function

The within transformation of constraint (2.75) looks as follows (assuming positive activity levels):

$$\left[\mathbf{A}'_{\mathbf{t,k,m}}\mathbf{y}^{\mathbf{o}}_{\mathbf{t,k}} + \mathbf{A}'_{\mathbf{t,k,m}}\mathbf{u}_{\mathbf{t,k}} + \mathbf{W}_{\mathbf{t,m}}\mathbf{h}_{\mathbf{t,m}} + \mathbf{c}_m + \boldsymbol{Q}_{m,mm}\left(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,m}} + \mathbf{h}_{\mathbf{t,m}}\right)\right] - \left[\overline{\mathbf{A}\mathbf{y}^{\circ}}_{\mathbf{m}} + \overline{\mathbf{A}\mathbf{u}}_{\mathbf{m}} + \overline{\mathbf{W}_{,\mathbf{m}}\mathbf{h}}_{,\mathbf{m}} + \mathbf{c}_m + \boldsymbol{Q}_{m,mm}\left(\bar{\mathbf{x}}^{\mathbf{o}}_{,\mathbf{m}} + \bar{\mathbf{h}}_{\mathbf{m}}\right)\right] = \mathbf{p}_{\mathbf{t,m}} - \bar{\mathbf{p}}_{\mathbf{m}} \tag{I.9}$$

By indicating the time-demeaned variables with tilde (~) we get:

$$\widetilde{\mathbf{A}\mathbf{y}^{\circ}}_{\mathbf{t,m}} + \widetilde{\mathbf{A}\mathbf{u}}_{\mathbf{t,m}} + \widetilde{\mathbf{W}\mathbf{h}}_{\mathbf{t,m}} + \mathbf{Q}\left(\tilde{\mathbf{x}}^{\mathbf{o}}_{\mathbf{t,m}} + \tilde{\mathbf{h}}_{\mathbf{t,m}}\right) = \tilde{\mathbf{p}}_{\mathbf{t,m}} \qquad \text{(I.10))}$$

subject to

$$\overline{\mathbf{Wh}}_{\mathbf{m}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{W}_{\mathbf{t,m}}\mathbf{h}_{\mathbf{t,m}}$$
$$\overline{\mathbf{Au}}_{\mathbf{m}} = \frac{1}{T}\sum_{t=1,k=1}^{T,K} (\mathbf{A}_{\mathbf{t,k,m}}\mathbf{u}_{\mathbf{t,k}})$$
$$\bar{\mathbf{h}}_{\mathbf{m}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{h}_{\mathbf{t,m}}$$

Where:

$$\widetilde{\mathbf{Ay}}^{\circ}{}_{\mathbf{t,m}} = \mathbf{A}'_{\mathbf{t,k,m}}\mathbf{y}^{\mathbf{o}}_{\mathbf{t,k}} - \overline{\mathbf{Ay}^{\circ}}_{\mathbf{m}}$$

$$\overline{\mathbf{Ay}^{\circ}}_{\mathbf{m}} = \frac{1}{T}\sum_{t=1,k=1}^{T,K} (A_{t,k,m}\mathbf{y}^{\circ}{}_{\mathbf{t,k}})$$

$$\widetilde{\mathbf{Au}}_{\mathbf{t,m}} = \mathbf{A}'_{\mathbf{t,k}}\mathbf{u}_{\mathbf{t,k}} - \overline{\mathbf{Au}}_{\mathbf{m}}$$

$$\widetilde{\mathbf{Wh}}_{\mathbf{t,m}} = \mathbf{W}_{\mathbf{t,m}}\mathbf{h}_{\mathbf{t,m}} - \overline{\mathbf{Wh}}_{\mathbf{m}}$$

$$\tilde{\mathbf{x}}^{\mathbf{o}}_{\mathbf{t,m}} = \mathbf{x}^{\mathbf{o}}_{\mathbf{t,m}} - \bar{\mathbf{x}}^{\mathbf{o}}_{\mathbf{m}}$$

$$\bar{\mathbf{x}}^{\mathbf{o}}_{\mathbf{m}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}^{\circ}{}_{\mathbf{t,m}}$$

$$\tilde{\mathbf{h}}_{\mathbf{t,m}} = \mathbf{h}_{\mathbf{t,m}} - \bar{\mathbf{h}}_{\mathbf{m}}$$

$\widetilde{\mathbf{Ay}}^{\circ}{}_{\mathbf{t,m}}$, $\overline{\mathbf{Ay}^{\circ}}_{\mathbf{m}}$, $\tilde{\mathbf{x}}^{\mathbf{o}}_{\mathbf{t,m}}$ and $\bar{\mathbf{x}}^{\mathbf{o}}_{\mathbf{m}}$ must be defined and calculated before the estimation, because they consist only of exogenous parameters. Note that $\mathbf{c}_{m}$ vanishes from (I.9). However, after having estimated $\mathbf{Q}$, $\mathbf{h}_{\mathbf{t,m}}$ and $\mathbf{u}_{\mathbf{t,k}}$, we can recover $\mathbf{c}$ from (I.9) as follows:

$$\bar{\mathbf{p}}_{\mathbf{m}} - \left[\bar{\mathbf{A}}'{}_{\mathbf{k,m}}(\bar{\mathbf{y}}^{\mathbf{o}}_{\mathbf{,k}} + \bar{\hat{\mathbf{u}}}_{\mathbf{,k}}) + \overline{\mathbf{W}_{\mathbf{,m}}\hat{\mathbf{h}}_{\mathbf{,m}}} + \widehat{\mathbf{Q}}_{m,mm}\left(\bar{\mathbf{x}}^{\mathbf{o}}_{\mathbf{,m}} + \bar{\hat{\mathbf{h}}}_{\mathbf{m}}\right)\right] = \mathbf{p}_{\mathbf{t,m}} - \left[\mathbf{A}'_{\mathbf{t,k,m}}(\mathbf{y}^{\mathbf{o}}_{\mathbf{t,k}} + \hat{\mathbf{u}}_{\mathbf{t,k}}) + \mathbf{W}_{\mathbf{t,m}}\hat{\mathbf{h}}_{\mathbf{t,m}} + \widehat{\mathbf{Q}}_{m,mm}(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,m}} + \hat{\mathbf{h}}_{\mathbf{t,m}})\right]$$

From (2.75) we see that:

$$\mathbf{p}_{\mathbf{t,m}} - \left[\mathbf{A}'_{\mathbf{t,k,m}}(\mathbf{y}^{\mathbf{o}}_{\mathbf{t,k}} + \hat{\mathbf{u}}_{\mathbf{t,k}}) + \mathbf{W}_{\mathbf{t,m}}\hat{\mathbf{h}}_{\mathbf{t,m}} + \widehat{\mathbf{Q}}_{\mathbf{m,mm}}(\mathbf{x}^{\mathbf{o}}_{\mathbf{t,m}} + \hat{\mathbf{h}}_{\mathbf{t,m}})\right] = \hat{\mathbf{c}}_{\mathbf{m}}$$

So

$$\bar{\mathbf{p}}_{\mathbf{m}} - \left[\bar{\mathbf{A}}'{}_{\mathbf{k,m}}(\bar{\mathbf{y}}^{\mathbf{o}}_{\mathbf{,k}} + \bar{\hat{\mathbf{u}}}_{\mathbf{,k}}) + \overline{\mathbf{W}_{\mathbf{,m}}\hat{\mathbf{h}}_{\mathbf{,m}}} + \widehat{\mathbf{Q}}_{\mathbf{m,mm}}\left(\bar{\mathbf{x}}^{\mathbf{o}}_{\mathbf{,m}} + \bar{\hat{\mathbf{h}}}_{\mathbf{m}}\right)\right] = \hat{\mathbf{c}}_{\mathbf{f,m}}$$

Monte Carlo simulation

In a first step, the linear cost parameter c and Q were set by the modeller (while ensuring that Q is positive semi-definite), assuming that farmers can produce three activities i_1, i_2 and i_3:

$$\mathbf{c} = \begin{bmatrix} 1.5 \\ 5 \\ 2 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 4 & 2 & 0.2 \\ 2 & 17 & 0.5 \\ 0.2 & 0.5 & 9.02 \end{bmatrix}$$

Then, the range of sample sizes F_s was set as $[F_1 = 10, F_2 = 30, \dots, F_S = 150]$. For each sample size F_s farms were "created" by drawing F_s "land endowments" $\mathbf{b_f}$ from a random normal distribution with a mean $\mu = 100$ and a standard deviation $\sigma = 25$. The fixed-input availability matrix $\mathbf{b}$ therefore is an $(F_s \times 1)$ vector, since the number of fixed-inputs $J = 1$. Besides, since each farm can produce $I = 3$ activities, an $I \times F_S$ matrix of output prices $\mathbf{p}$ was generated. For each of the I activities a mean was defined around which its prices would be distributed, with $\mu_{p_1} = 340,\ \mu_{p_2} = 420$ and $\mu_{p_3} = 410$. To ensure that there is some heterogeneity in the prices farmers face, $\mathbf{p}$ was drawn from a discrete uniform distribution with a range of $\pm 10\%$ from the respective mean μ_{p_i}. Finally, for each farm f the optimal, "true" activity levels $\mathbf{x_f}$ and fixed-input prices $\mathbf{y_f}$ were calculated by minimizing the CSC according to model (2.80), (3.10) and (3.11) with the generated output prices $\mathbf{p_f}$ the fixed-input availabilities $\mathbf{b_f}$ and with the "true" cost function parameters $\mathbf{c}$ and $\mathbf{Q}$. In other words: Whereas the CSC minimization is used in the validation phase to estimate the values of $\mathbf{x_f}$ and $\mathbf{y_f}$ with the previously *estimated* cost function parameters, we now use the *true* cost function parameters to generate the true values of $\mathbf{x_f}$ and $\mathbf{y_f}$. This procedure mimics the decision-making process of farmers who know the true parameters of the cost function they face. Finally, error terms were added to the "true" values to create "observed" activity levels and fixed-input prices. These error terms were drawn from a standard normal distribution with μ being the mean of the "true" activity level or the "true" fixed-input price over all farms. The standard deviation σ was 5% of the mean of the respective activity level or fixed-input price. Heckelei and Wolff (2003) use a similar procedure (but a standard deviation of only 5%). For each sample size, this whole procedure was repeated $R = 100$ times. Then, with the "observed" values the cost function parameters were estimated for each of the 100 samples per sample size, using model (2.75) – (2.79) as proposed by Henry de Frahan (2019). To evaluate the estimates, the root mean squared errors (RMSE) and the summed root mean squared errors (SRMSE) were calculated, as described in chapter 3.4.

Annex III – Model Implementation

Regions

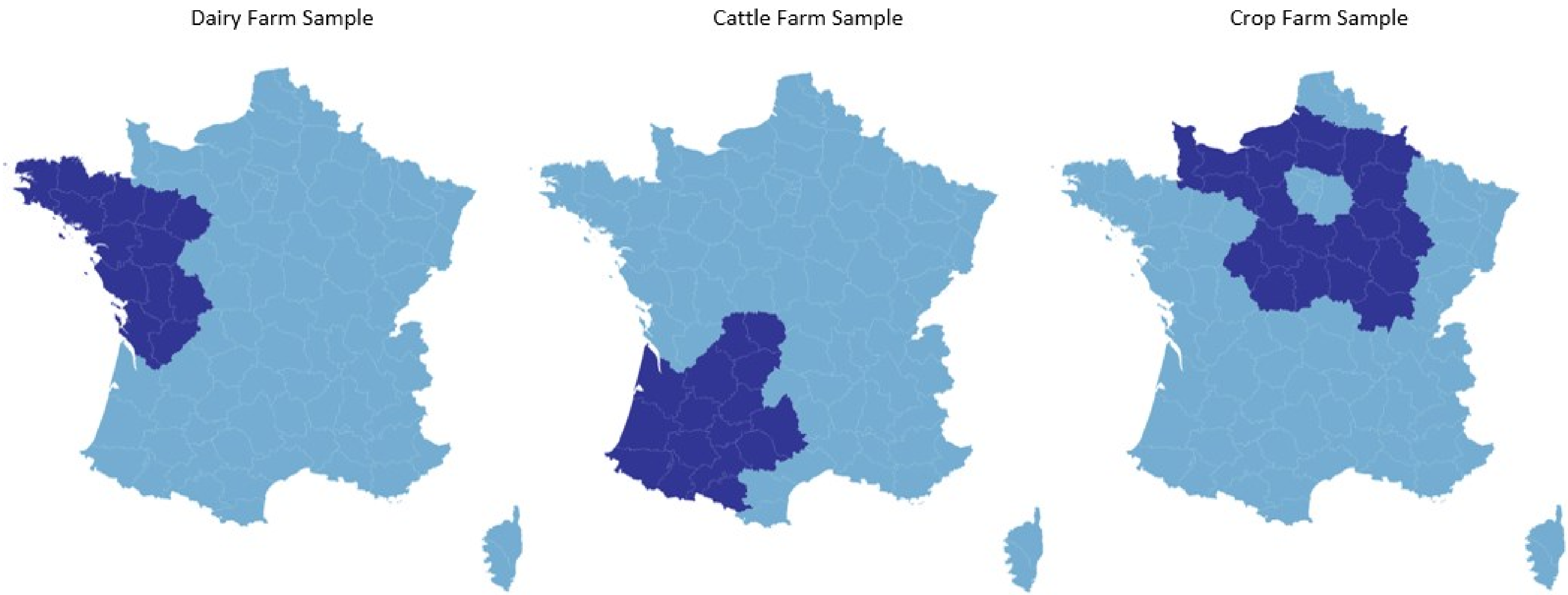

Figure A.1: Location of the study areas in France. *The dark blue area indicates the NUTS1 region that the farms in each of the farm samples are located in, being West France (dairy farms), South-West France (cattle farms) and Central France (crop farms). Source: Eurostat (2021e).*

Grouping farms according to their specialization

The cost function parameter estimation procedure presupposes that all farms in the sample for estimation share a common cost function. For this to be a realistic assumption we selected for each region only those farms that share common production characteristics, such as farm specialization. Farms were grouped as either "dairy farms", "cattle farms" or "crop farms" according to the following rules:

Farm type for cost function parameter estimation	Type of farming at recording in A30 FADN code	Description
Crop farms	1310; 1320; 1330; 1410; 1420; 1443;	Specialist cereals, oilseeds and proteins (COP); specialist rice; COP and rice combined; specialist root crops; cereal and root crops combined; various field crops combined;
Cattle farms	4210; 4220; 4310; 4320	Cattle rearing; cattle fattening; dairying with rearing & fattening; rearing & fattening with dairying
Dairy farms	4110; 4120	Milk; Milk & cattle rearing

Table A.1: Definition of farm types "crop farms", "dairy farms" and "cattle farms". *Source for description of FADN code: (Butault et al., 2011)*

Aggregation of farm outputs for each farm type

FADN Variable	FADN description	Aggregation category		
		Crop Farms	Dairy Farms	Cattle Farms
K162sa	Milk sales	animal outp.	dairy outp.	dairy outp.
K163sa	Milk products sold	animal outp.	dairy outp.	dairy outp.
N23sa	Calves/fattening	animal outp.	oth. bovine	oth. bovine
N24sa	Other cattle (<12m)	animal outp.	oth. bovine	oth. bovine
N25sa	Male cattle (12-24m)	animal outp.	oth. bovine	oth. bovine
N26sa	Female cattle (12-24m)	animal outp.	oth. bovine	oth. bovine
N27sa	Male cattle (>24m)	animal outp.	oth. bovine	oth. bovine
N28sa	Breeding heifers	animal outp.	oth. bovine	oth. bovine
N29sa	Heifers/fattening	animal outp.	oth. bovine	oth. bovine
N30sa	Dairy cows	animal outp.	oth. bovine	oth. bovine
N31sa	Cull dairy cows	animal outp.	oth. bovine	oth. bovine
N32sa	Other cows	animal outp.	oth. bovine	oth. bovine
N38sa	Goat (breeding female)	animal outp.	non-bovine	non-bovine

N39sa	Other goats	animal outp.	non-bovine	non-bovine
N40sa	Ewes	animal outp.	non-bovine	non-bovine
N41sa	Other sheep	animal outp.	non-bovine	non-bovine
N43sa	Piglets	animal outp.	non-bovine	non-bovine
N44sa	Breeding sows	animal outp.	non-bovine	non-bovine
N45sa	Pigs/fattening	animal outp.	non-bovine	non-bovine
N46sa	Other pigs	animal outp.	non-bovine	non-bovine
N47sa	Table chickens	animal outp.	non-bovine	non-bovine
N48sa	Laying hens	animal outp.	non-bovine	non-bovine
N49sa	Other poultry	animal outp.	non-bovine	non-bovine
N22sa	Horses	animal outp.	non-bovine	non-bovine
N33sa	Bees	animal outp.	non-bovine	non-bovine
N34sa	Rabbits (breeding female)	animal outp.	non-bovine	non-bovine
N50sa	Other animals	animal outp.	non-bovine	non-bovine
K129sa	Dry pulses	pulses & oils.	crops	crops
K132sa	Oil seeds	pulses & oils.	crops	crops
K120sa	Common wheat	cereals	crops	crops
K121sa	Durum wheat	cereals	crops	crops
K126sa	Grain maize	cereals	crops	crops
K127sa	Rice	cereals	crops	crops
K122sa	Rye	cereals	crops	crops
K123sa	Barley	cereals	crops	crops
K124sa	Oats	cereals	crops	crops
K125sa	Summer cereal mix.	cereals	crops	crops
K128sa	Other cereals	cereals	crops	crops
K130sa	Potatoes	indust. crops	crops	crops
K131sa	Sugar beet	indust. crops	crops	crops
K135sa	Other industrial crops	indust. crops	crops	crops
K133sa	Hops	other crops	crops	crops
K134sa	Tobacco	other crops	crops	crops
K136sa	Vegetables	other crops	crops	crops
K137sa	Vegetables	other crops	crops	crops
K138sa	Vegetables	other crops	crops	crops
K139sa	Mushrooms	other crops	crops	crops
K140sa	Flowers	other crops	crops	crops
K141sa	Flowers	other crops	crops	crops

Table A.2: Aggregation of farm outputs for each farm type. *Source: Henry de Frahan et al. (2015)*

Imputation of mission or extreme farm gate prices

The following description is based on Henry de Frahan et al. (2015), who provide an even more detailed explanation of the data preparation procedure. Average regional output prices are calculated by removing extreme farm-gate prices. A farm-gate price is considered extreme if it has a probability of occurrence that is less than one over twice the sample size:

$$\left|\left(p_{t,f,m} - p_{t,r,m}\right)\right| > \Phi^{-1}\left(1 - \tfrac{1}{2}F_r\right) * SD_{p_{t,f,m}}$$

where $p_{t,f,m}$ is the price of an output disaggregate m for each farm and $p_{t,r,m}$ the regional average, ϕ is a fixed weight input price index, as described in the text, F_r is the number of farms in the respective region, and $SD_{p_{t,f,m}}$ is the standard deviation of farm-gate prices for disaggregate m for each farm in the year and region. Extreme and missing values are replaced by their regional average if available, or otherwise, by the national average.

The following table demonstrates the process of the imputation of missing prices exemplary for dry pulses:

Price	Source	Category/Description	Variable name or formula
p_{ft}	FADN	Dry pulses / farm-gate price	$\frac{K129tp}{K129qq}$
$p_{region;t}$	FADN	Dry pulses / regional (NUTS1) average	$\sum_{region}(K129tp)/\sum_{region}(K129qq)$
$p_{country;t}$	FADN	Dry pulses / country average	$\sum_{country}(K129tp)/\sum_{country}(K129qq)$
p_{pls}	EUROSTAT	Pulses / country average	p41980
p_{oth}	EUROSTAT	Other fresh vegetables / country average	p41900
p_{crp}	EUROSTAT	Crop output / country average	p100000
cpi	OECD	Consumer price index	MEI-Prices8

Table A.3: Illustration of the price imputation scheme for the example of dry pulses. *Source: Henry de Frahan et al. (2015, p. 45).*

		mean (1 000 base year Euros)					standard deviation (1 000 base year Euros)				
		animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
Crop Farms	Champagne-Ardenne	12.69	19.61	40.75	60.77	0.80	46.62	14.47	54.31	32.20	5.23
	Picardy	13.45	16.80	61.48	67.59	8.03	36.57	14.47	84.02	38.00	29.96
	Upper Normandy	28.20	18.79	122.06	63.76	0.78	41.70	16.66	243.53	39.50	5.06
	Centre	6.60	25.57	14.45	74.30	5.39	23.36	21.09	35.48	47.01	26.98
	Lower Normandy	13.28	15.53	45.55	57.12	1.46	29.14	12.34	54.81	30.21	5.19
	Burgundy	11.86	31.28	6.03	68.90	2.85	29.02	21.83	22.26	38.43	13.13
		dairy products	other bovine animal outputs	non-bovine animal outputs	crops		dairy products	other bovine animal outputs	non-bovine animal outputs	crops	
Dairy Farms	Pays de la Loire	104.32	41.07	2.51	9.40		54.15	37.25	16.22	13.58	
	Brittany	107.61	37.79	0.58	10.37		52.43	31.10	6.92	10.83	
	Poitou-Charentes	125.68	28.99	0.14	14.30		69.34	27.28	0.91	17.88	
Cattle Farms	Aquitaine	1.89	79.64	1.18	2.62		11.99	79.77	6.01	5.43	
	Midi-Pyrénées	5.05	64.71	1.31	1.93		19.09	40.10	14.29	4.18	
	Limousin	6.84	77.47	1.90	1.68		28.37	45.59	7.20	4.00	

Table A.4: Aggregated activity levels in base year Euros, average values from 2005-2012. *Source: EU-FADN – DG AGRI.*

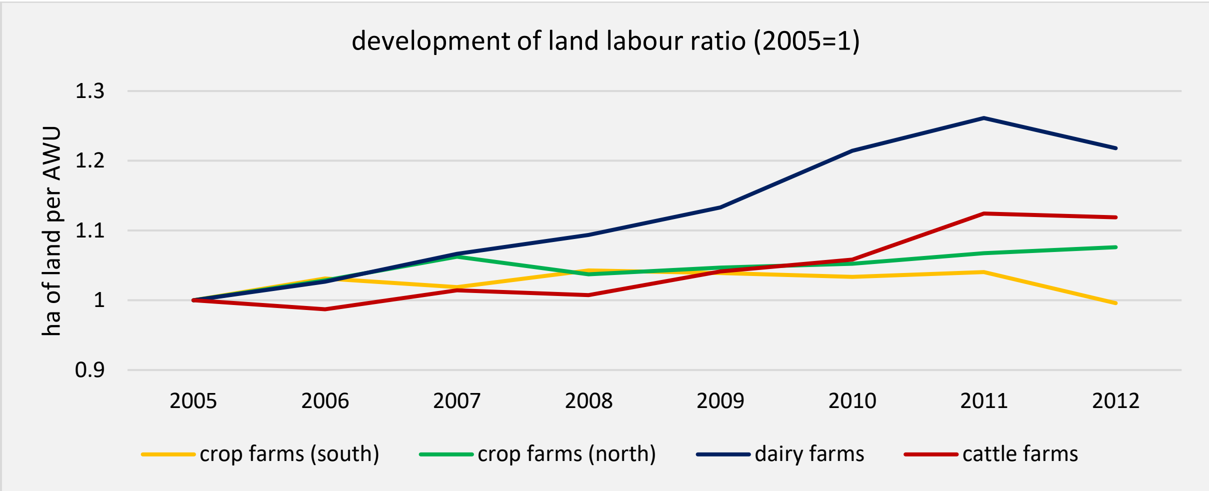

Figure A.2: relative land labour ration for all four samples in ha of land per AWU (2005=1). *Source: own visualization based on EU-FADN – DG AGRI.*

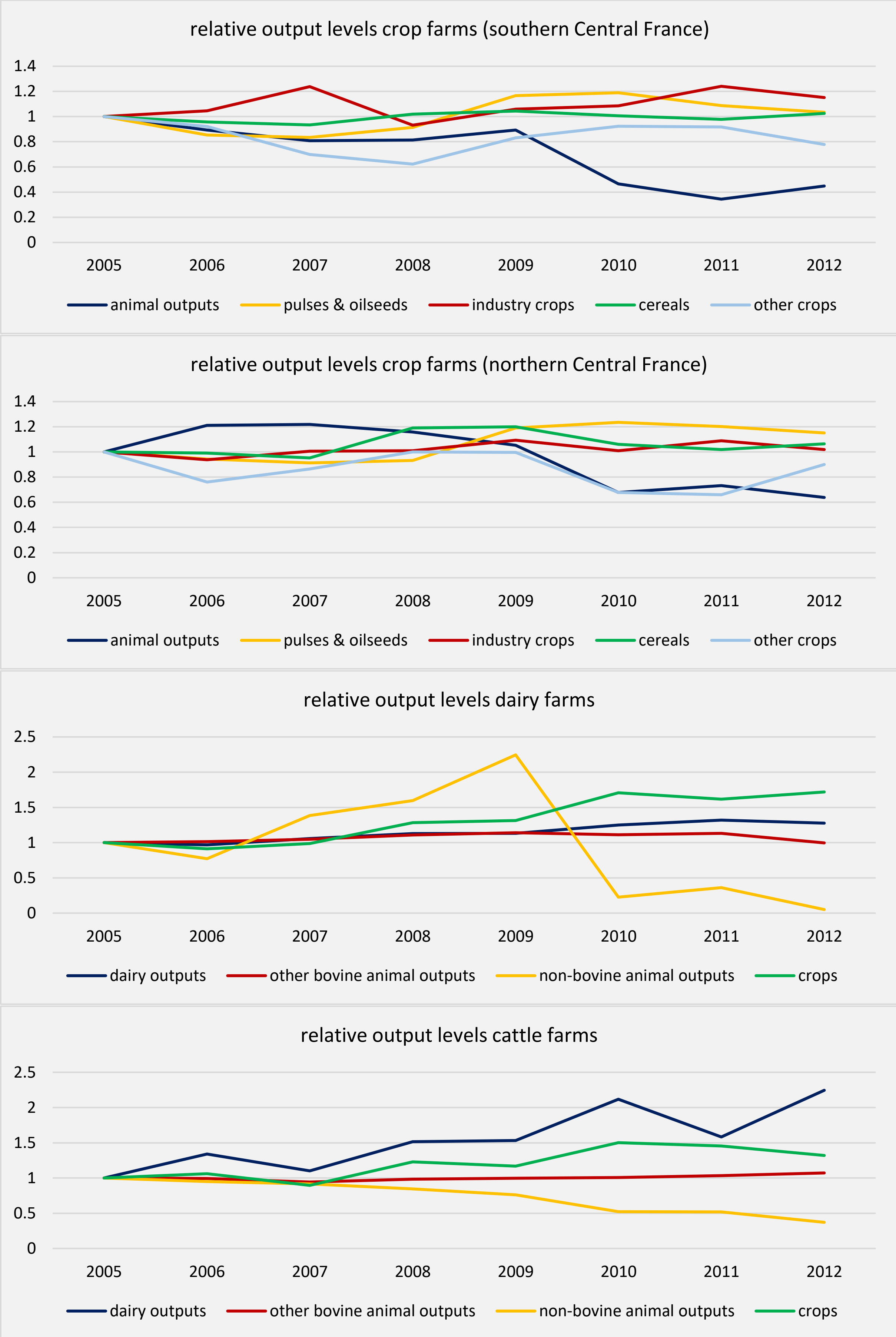

Figure A.3: relative average aggregated output levels per farm (2005=1). *Source: own visualization based on EU-FADN – DG AGRI.*

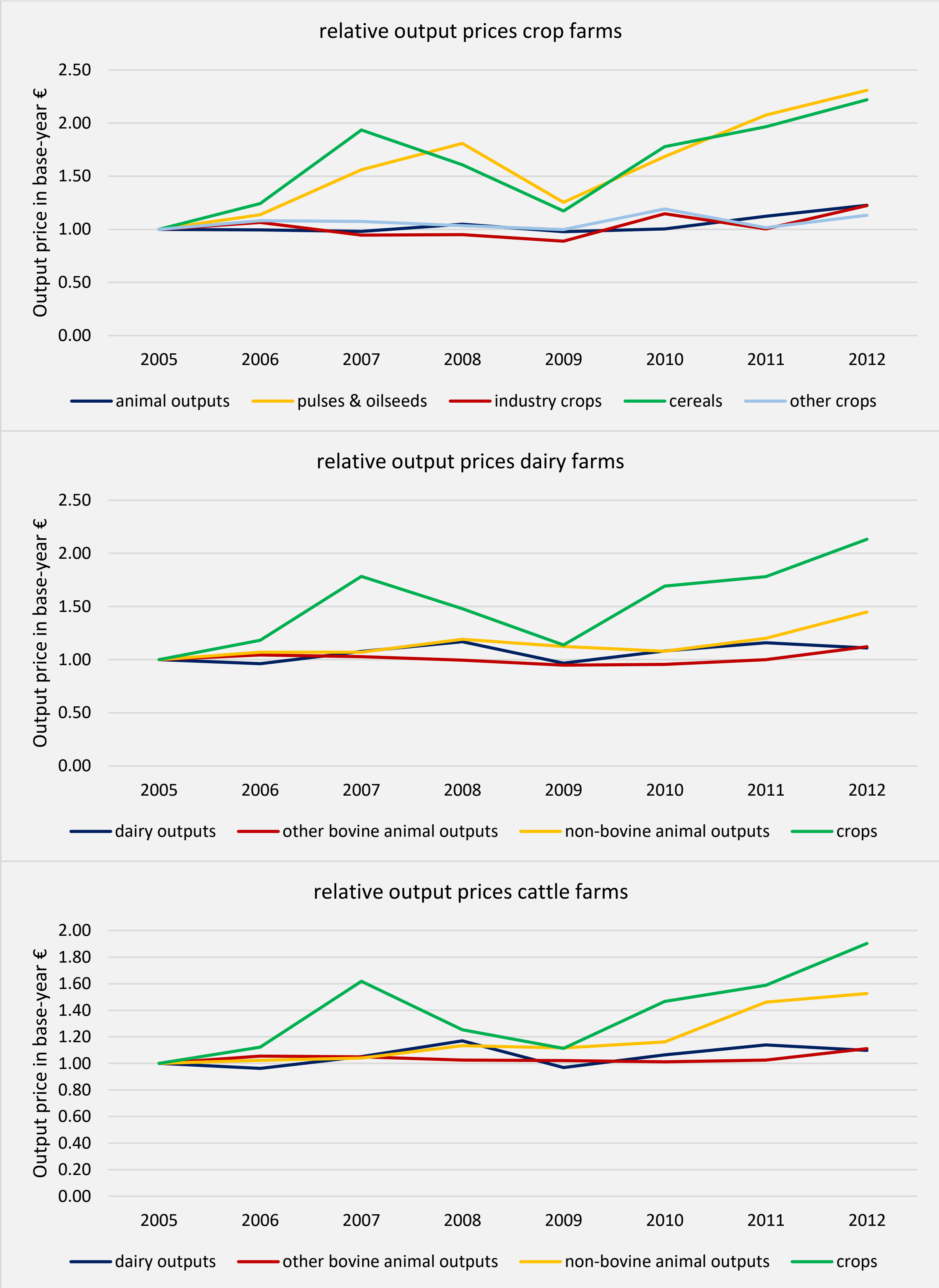

Figure A.4: relative output prices for dairy, cattle, and crop farms (2005=1,00€). *Source: own visualization based on EU-FADN – DG AGRI.*

Activity Level Dependent Cost Function

Cost Function Parameters

Crop Farms from southern Central France

	fixed effect: no time index: no					fixed effect: no time index: yes					fixed effect: yes time index: no					fixed effect: yes time index: yes				
	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
			Q					Q					Q					Q		
animal outputs	0.01	0.01	0.00	-0.01	0.00	0.01	-0.01	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
pulses and oilseeds	0.01	0.10	-0.03	-0.02	0.01	-0.01	0.05	-0.08	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
industrial crops	0.00	-0.03	0.22	-0.03	0.01	0.02	-0.08	0.18	0.00	0.03	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00
cereals	-0.01	-0.02	-0.03	0.02	0.00	0.00	-0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.04	0.00
other crops	0.00	0.01	0.01	0.00	0.00	0.01	-0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
			c					$\hat{c}_0$										$\hat{c}_1$		
	0.52	1.50	1.83	1.02	0.83	0.57	0.29	0.90	0.55	0.84						-0.04	0.12	0.22	0.13	-0.01
								$\hat{c}_1$												
						0.00	0.17	0.04	0.13	0.00										

Crop Farms from northern Central France

	fixed effect: no time index: no					fixed effect: no time index: yes					fixed effect: yes time index: no					fixed effect: yes time index: yes				
	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
			Q					Q					Q					Q		
animal outputs	0.01	-0.01	0.00	0.00	0.00	0.01	-0.01	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
pulses and oilseeds	-0.01	0.12	0.01	-0.05	0.00	-0.01	0.14	0.01	-0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
industrial crops	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
cereals	0.00	-0.05	0.00	0.03	0.00	0.00	-0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.01	0.00
other crops	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			c					$\hat{c}_0$										$\hat{c}_1$		
	0.61	2.46	1.00	1.05	0.93	0.84	1.04	0.88	0.54	0.92						-0.05	0.28	0.03	0.12	0.02
								$\hat{c}_1$												
						-0.02	0.19	0.03	0.13	0.01										

Table A.5: estimated parameters of the activity level dependent cost function for crop farms in Central France. *Source: own simulation.*

	Cost Function Parameters															
	fixed effect: no time index: no				fixed effect: no time index: yes				fixed effect: yes time index: no				fixed effect: yes time index: yes			
	Dairy Farms from West France															
	Q				*Q*				*Q*				*Q*			
	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops
dairy products	0.10	0.01	0.03	-0.03	0.06	0.00	0.02	-0.01	0.27	0.01	0.00	0.00	0.10	0.01	0.00	0.00
other bovine animal outputs	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00
non-bovine animal outputs	0.03	0.00	0.01	-0.01	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
crops	-0.03	-0.01	-0.01	0.08	-0.01	0.00	0.00	0.04	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.05
	c				$\hat{c}_0$								$\hat{c}_1$			
	0.73	0.81	0.74	1.03	0.68	0.78	0.76	0.40					0.01	0.00	0.02	0.14
					$\hat{c}_1$											
					0.02	0.00	0.01	0.13								
	Cattle Farms from South-West France															
	Q				*Q*				*Q*				*Q*			
	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops
dairy products	0.01	0.00	0.01	0.00	0.01	0.00	-0.03	0.01	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
other bovine animal outputs	0.00	0.01	0.01	0.00	0.00	0.01	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00
non-bovine animal outputs	0.01	0.01	0.07	-0.05	-0.03	0.02	0.32	-0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00
crops	0.00	0.00	-0.05	0.03	0.01	0.00	-0.09	0.03	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.03
	c				$\hat{c}_0$								$\hat{c}_1$			
	0.66	0.47	0.23	0.33	0.63	0.49	-1.60	-0.06					0.00	-0.01	0.03	0.09
					$\hat{c}_1$											
					0.01	0.00	0.09	0.09								

***Table A.6: estimated parameters of the activity level dependent cost function for dairy and cattle farms in West and South-West France**. Source: own simulation.*

Cost Function Specification		Supply Elasticities																							
		fixed effect: no time index: no					fixed effect: no time index: yes					fixed effect: yes time index: no					fixed effect: yes time index: yes					mean over all cost function specifications				
		Crop Farms from southern Central France																								
		animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
price change	animal outputs	0.33	0.00	-0.16	0.00	0.00	0.30	0.01	-0.14	0.00	0.00	0.58	0.03	-0.54	0.00	1.57	0.48	0.03	-0.54	-0.01	1.04	0.42	0.02	-0.35	0.00	0.65
	pulses and oilseeds	-0.17	3.86	0.40	-1.49	-0.69	-0.32	7.88	-0.04	-2.94	-0.42	-1.11	6.92	1.73	-3.09	2.13	-0.33	14.09	-0.07	-5.54	-1.15	-0.48	8.19	0.51	-3.26	-0.04
	industrial crops	-0.11	0.09	2.27	-0.10	-4.99	-0.11	-0.06	1.58	-0.06	-2.99	-0.44	0.09	1.79	-0.29	-1.41	-0.81	-0.09	3.42	-0.16	-6.43	-0.37	0.01	2.26	-0.15	-3.96
	cereals	-0.16	-3.50	-0.39	1.33	-0.80	-0.29	-6.76	-0.34	2.42	-0.66	0.09	-5.66	-4.24	2.32	1.80	-0.40	-9.05	-1.07	3.42	-0.63	-0.19	-6.24	-1.51	2.37	-0.07
	other crops	-0.31	-0.18	-2.54	-0.18	10.95	-0.46	-0.11	-0.89	-0.04	4.53	-0.73	0.11	-3.77	-0.25	26.22	-0.46	-0.17	-3.43	-0.10	15.31	-0.49	-0.09	-2.66	-0.14	14.25
		Crop Farms from northern Central France																								
		animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops
price change	animal outputs	1.01	0.10	-0.26	0.04	0.00	0.41	0.14	-0.14	0.00	0.59	1.22	0.12	-0.25	0.05	-0.75	0.48	0.08	-0.11	0.00	0.00	0.78	0.11	-0.19	0.02	-0.04
	pulses and oilseeds	0.55	4.28	-0.68	-1.04	1.80	0.02	21.51	-3.86	-4.84	9.44	0.88	4.22	-1.87	-0.88	10.57	0.02	32.36	-3.38	-9.07	3.01	0.37	15.59	-2.45	-3.96	6.21
	industrial crops	-6.36	-2.91	5.19	-1.92	-9.00	-3.35	-3.47	2.34	-0.31	-7.94	-6.69	-3.55	5.69	-2.76	-8.39	-3.79	-1.61	3.79	-1.88	-8.48	-5.05	-2.88	4.25	-1.72	-8.45
	cereals	0.51	-3.20	-1.02	1.54	-4.66	-0.06	-8.88	-0.93	2.94	9.13	0.49	-3.01	-1.24	1.53	-1.46	0.05	-9.38	-3.10	4.20	0.22	0.25	-6.12	-1.57	2.55	0.81
	other crops	-0.37	-0.07	-1.46	-0.16	15.93	-0.14	0.17	-1.94	-0.13	41.12	-0.36	-0.02	-1.91	-0.12	33.07	-0.16	0.18	-1.86	-0.16	30.99	-0.26	0.07	-1.79	-0.14	30.28

Table A.7: Elasticities of supply for crop farms in the northern and in the southern part of Central France, calculated with the activity level dependent cost function. *Elasticities are calculated for a 10% increase of the price of the activities indicated on the vertical axis. Source: own simulation*

Supply Elasticities

Dairy Farms from West France

Cost Function Specification	fixed effect: no, time index: no				fixed effect: no, time index: yes				fixed effect: yes, time index: no				fixed effect: yes, time index: yes				mean over all cost function specifications			
price change	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops
animal outputs	5.11	-0.47	-0.01	0.00	2.72	-0.25	-0.01	0.00	6.23	-0.58	1.15	0.00	2.77	-0.26	0.00	0.00	4.20	-0.39	0.28	0.00
pulses and oilseeds	-4.82	0.53	-9.29	0.00	-2.63	0.34	-10.00	0.00	-5.98	0.59	-4.44	0.00	-2.68	0.32	-8.28	0.00	-4.03	0.45	-8.00	0.00
industrial crops	-0.16	-0.44	49.36	0.00	-0.98	-1.18	136.70	0.00	0.32	-0.13	10.80	0.00	-0.11	-0.34	38.56	0.00	-0.23	-0.52	58.85	0.00
cereals	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Cattle Farms from South-West France

Cost Function Specification	fixed effect: no, time index: no				fixed effect: no, time index: yes				fixed effect: yes, time index: no				fixed effect: yes, time index: yes				mean over all cost function specifications			
price change	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	dairy products	other bovine animal outputs	non-bovine animal outputs	crops
dairy products	5.11	-0.47	-0.01	0.00	2.72	-0.25	-0.01	0.00	6.23	-0.58	1.15	0.00	2.77	-0.26	0.00	0.00	4.20	-0.39	0.28	0.00
other bovine animal outputs	-4.82	0.53	-9.29	0.00	-2.63	0.34	-10.00	0.00	-5.98	0.59	-4.44	0.00	-2.68	0.32	-8.28	0.00	-4.03	0.45	-8.00	0.00
non-bovine animal outputs	-0.16	-0.44	49.36	0.00	-0.98	-1.18	136.70	0.00	0.32	-0.13	10.80	0.00	-0.11	-0.34	38.56	0.00	-0.23	-0.52	58.85	0.00
crops	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A.8: Elasticity of supply for dairy farms in West France and cattle farms in South-West France, calculated with the activity level dependent cost function. *Elasticities are calculated for a 10% increase of the price of the activities indicated on the vertical axis. Source: own simulation*

Activity level and fixed-input level dependent cost funtion

Estimated cost function parameters

Cost Function Parameters																
Activity Level and Fixed-Input Level Dependent Cost Function with Time Index and $q\_min_{ii} = 0.2$	Crop Farms from southern Central France								Crop Farms from northern Central France							
	activity level					fixed-input level			activity level					fixed-input level		
	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour
			$\hat{Q}$				$\hat{H}$				$\hat{Q}$				$\hat{H}$	
animal outputs	0.200	-0.010	0.019	0.029	0.005	-0.002	0.000	-0.336	0.200	-0.037	0.128	-0.012	0.103	-0.005	0.037	-1.675
pulses and oilseeds	-0.010	0.200	0.007	-0.005	0.078	-0.006	-0.005	0.116	-0.037	0.200	0.012	-0.110	0.116	0.000	0.064	-1.105
industrial crops	0.019	0.007	0.200	-0.048	0.032	0.001	-0.009	-0.447	0.128	0.012	0.200	-0.002	0.076	-0.010	0.070	-2.454
cereals	0.029	-0.005	-0.048	0.200	0.013	-0.011	-0.003	0.010	-0.012	-0.110	-0.002	0.200	-0.008	-0.008	0.066	-1.272
other crops	0.005	0.078	0.032	0.013	0.200	-0.008	0.003	0.054	0.103	0.116	0.076	-0.008	0.200	-0.011	0.054	-0.953
			$\hat{c}_0$				$\hat{g}_0$				$\hat{c}_0$				$\hat{g}_0$	
	-1.461	0.746	-0.791	0.815	-1.439	0.004	0.032	-0.066	-0.026	0.798	-6.463	0.919	-0.755	-0.007	-0.261	5.235
			$\hat{c}_1$				$\hat{g}_1$				$\hat{c}_1$				$\hat{g}_1$	
	0.133	0.102	-0.062	0.119	0.094	0.000	-0.002	0.023	-0.075	0.154	0.383	0.145	0.071	0.000	-0.001	-0.027
							$\hat{F}$								$\hat{F}$	
cropland						0.000	0.000	-0.001						0.000	0.000	0.001
grassland						0.000	0.000	0.000						0.000	0.000	0.001
unpaid labour						-0.001	0.000	0.054						0.001	0.001	0.024

Table A.9: Parameter estimates used for all simulations – activity level and fixed-input level dependent cost function with time index and q_min_{ii} = 0.2 for crop farms.
Source: own simulation

Cost Function Parameters														
Activity Level and Fixed-Input Level Dependent Cost Function with Time Index and $q\_min_{ii} = 0.2$	Dairy Farms from West France							Cattle Farms from South-West France						
	activity level				fixed-input level			activity level				fixed-input level		
	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour
	$\hat{Q}$				$\hat{H}$			$\hat{Q}$				$\hat{H}$		
dairy products	0.200	-0.064	0.015	-0.032	-0.009	0.015	0.509	0.200	0.004	0.038	-0.004	-0.019	0.008	-0.807
other bovine animal outputs	-0.064	0.200	0.023	-0.020	0.006	0.012	0.538	0.004	0.200	0.195	-0.008	-0.015	0.007	-0.216
non-bovine animal outputs	0.015	0.023	0.200	-0.048	0.001	0.027	0.558	0.038	0.195	0.200	0.021	-0.008	0.005	-0.919
crops	-0.032	-0.020	-0.048	0.200	-0.011	0.029	1.024	-0.004	-0.008	0.021	0.200	-0.020	0.009	-0.040
	$\hat{c}_0$				$\hat{g}_0$			$\hat{c}_0$				$\hat{g}_0$		
	-0.711	0.104	-0.861	0.716	0.016	-0.076	-3.406	0.938	-1.144	-0.035	0.897	0.067	-0.029	-0.320
	$\hat{c}_1$				$\hat{g}_1$			$\hat{c}_1$				$\hat{g}_1$		
	-0.093	0.039	-0.192	0.188	0.000	0.000	-0.045	-0.127	-0.142	0.109	0.083	0.000	0.001	0.011
					$\hat{F}$							$\hat{F}$		
cropland					0.000	0.000	0.000					0.000	0.000	-0.014
grassland					0.000	0.000	0.002					0.000	0.000	0.000
unpaid labour					0.000	0.002	0.221					-0.014	0.000	0.432

Table A.10: Parameter estimates used for all simulations – activity level and fixed-input level dependent cost function with time index and $q\_min_{ii} = 0.2$ for dairy and cattle farms. *Source: own simulation*

Activity Level and Fixed-Input Level Dependent Cost Function with Time Index and $q\_min_{ii} = 0$	Cost Function Parameters															
	Crop Farms from southern Central France								Crop Farms from northern Central France							
	activity level					fixed-input level			activity level					fixed-input level		
	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour
			$\hat{Q}$				$\hat{H}$				$\hat{Q}$				$\hat{H}$	
animal outputs	0.001	-0.001	0.002	0.002	0.000	-0.001	-0.002	-0.046	0.000	0.000	0.000	0.000	0.000	-0.001	-0.008	-0.051
pulses and oilseeds	-0.001	0.001	0.001	-0.002	0.001	-0.001	-0.003	-0.047	0.000	0.029	0.002	-0.009	0.003	-0.002	-0.011	-0.076
industrial crops	0.002	0.001	0.013	-0.001	0.003	-0.001	-0.003	-0.132	0.000	0.002	0.000	-0.001	0.000	-0.001	-0.008	-0.062
cereals	0.002	-0.002	-0.001	0.009	-0.002	-0.002	-0.003	-0.043	0.000	-0.009	-0.001	0.003	-0.001	-0.002	-0.012	-0.046
other crops	0.000	0.001	0.003	-0.002	0.001	-0.001	-0.002	-0.031	0.000	0.003	0.000	-0.001	0.000	-0.001	-0.007	-0.041
			$\hat{c}_0$				$\hat{g}_0$				$\hat{c}$				$\hat{g}_0$	
	1.025	0.908	1.023	1.055	0.994	0.004	0.011	0.127	0.985	0.949	1.002	1.040	1.024	0.006	0.051	0.140
			$\hat{c}_1$				$\hat{g}_1$				$\hat{c}_1$				$\hat{g}_1$	
	0.014	0.155	0.018	0.128	0.014	0.000	0.000	0.013	0.028	0.159	0.033	0.133	0.015	0.000	0.000	-0.007
							$\hat{F}$								$\hat{F}$	
cropland						0.000	0.000	0.000						0.000	0.000	0.000
grassland						0.000	0.000	0.000						0.000	0.000	0.000
unpaid labour						0.000	0.000	0.004						0.000	0.000	0.013

Table A.11: Parameters of the activity level and fixed-input level dependent cost function with time index and $q\_min_{ii} = 0$ for crop farms. *Source: own simulation.*

Cost Function Parameters														
	Dairy Farms from West France							Cattle Farms from South-West France						
	activity level				fixed-input level			activity level				fixed-input level		
Activity Level and Fixed-Input Level Dependent Cost Function with Time Index and $q\_min_{ii} = 0$	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour
		$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$			$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$	
dairy products	0.000	0.000	0.000	0.000	-0.001	-0.003	-0.061	0.005	0.001	0.008	0.005	-0.010	0.003	-0.113
other bovine animal outputs	0.000	0.001	0.001	-0.002	-0.001	-0.002	0.021	0.001	0.000	0.002	-0.003	-0.009	0.003	-0.115
non-bovine animal outputs	0.000	0.001	0.001	-0.004	0.001	-0.003	0.030	0.008	0.002	0.013	0.002	-0.011	0.003	-0.240
crops	0.000	-0.002	-0.004	0.011	-0.001	-0.003	0.039	0.005	-0.003	0.002	0.110	-0.014	0.005	-0.110
		$\hat{\mathbf{c}}_0$				$\hat{\mathbf{g}}_0$			$\hat{\mathbf{c}}_0$				$\hat{\mathbf{g}}_0$	
	1.328	1.093	1.092	1.022	-0.001	0.007	-0.310	0.998	1.083	1.139	0.792	0.037	-0.015	0.440
		$\hat{\mathbf{c}}_1$				$\hat{\mathbf{g}}_1$			$\hat{\mathbf{c}}_1$				$\hat{\mathbf{g}}_1$	
	0.016	0.000	0.021	0.131	0.000	0.000	-0.001	0.025	0.013	0.076	0.109	0.000	0.000	-0.025
						$\hat{\mathbf{F}}$							$\hat{\mathbf{F}}$	
cropland					0.000	0.000	0.000					0.000	0.000	0.000
grassland					0.000	0.000	0.000					0.000	0.000	0.000
unpaid labour					0.000	0.000	0.053					0.000	0.000	0.000

Table A.12: Parameters of the activity level and fixed-input level dependent cost function with time index and q_min_{ii} = 0 for dairy and cattle farms. *Source: own simulation.*

Cost Function Parameters								
	Crop Farms from southern Central France							
	activity level					fixed-input level		
Activity Level and Fixed-Input Level Dependent Cost Function without Time Index and $q\_min_{ii} = 0$	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour
			$\hat{Q}$				$\hat{H}$	
animal outputs	0.000	0.000	0.000	0.000	0.000	-0.008	0.001	-0.219
pulses and oilseeds	0.000	0.012	-0.002	0.014	0.000	-0.012	0.002	-0.240
industrial crops	0.000	-0.002	0.012	-0.001	0.002	-0.008	0.000	-0.256
cereals	0.000	0.014	-0.001	0.017	0.000	-0.013	0.002	-0.246
other crops	0.000	0.000	0.002	0.000	0.000	-0.008	0.002	-0.116
			$\hat{c}$				$\hat{g}$	
	1.128	1.590	1.033	1.638	0.994	0.045	-0.008	1.059
							$\hat{F}$	
cropland						0.000	0.000	-0.001
grassland						0.000	0.000	-0.001
unpaid labour						-0.001	-0.001	0.115

Cost Function Parameters								
	Crop Farms from northern Central France							
	activity level					fixed-input level		
Activity Level and Fixed-Input Level Dependent Cost Function without Time Index and $q\_min_{ii} = 0$	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour
			$\hat{Q}$				$\hat{H}$	
animal outputs	0.001	0.000	0.000	-0.004	0.000	-0.004	-0.013	-0.550
pulses and oilseeds	0.000	0.014	0.001	0.008	0.001	-0.008	-0.018	-0.826
industrial crops	0.000	0.001	0.000	0.001	0.000	-0.005	-0.013	-0.525
cereals	-0.004	0.008	0.001	0.015	0.002	-0.008	-0.017	-0.780
other crops	0.000	0.001	0.000	0.002	0.000	-0.005	-0.010	-0.549
			$\hat{c}$				$\hat{g}$	
	1.074	1.745	1.105	1.656	1.156	0.029	0.073	2.907
							$\hat{F}$	
cropland						0.000	0.000	0.000
grassland						0.000	0.000	0.000
unpaid labour						0.000	0.000	0.036

Table A.13: Parameters of the activity level and fixed-input level dependent cost function without time index and q_min_{ii} = 0 for crop farms. *Source: own simulation.*

Cost Function Parameters														
	Dairy Farms from West France							Cattle Farms from South-West France						
	activity level				fixed-input level			activity level				fixed-input level		
Activity Level and Fixed-Input Level Dependent Cost Function without Time Index and $q\_min_{ii} = 0$	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour
	$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$			$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$		
dairy products	0.000	0.000	0.000	0.000	0.000	-0.002	-0.514	0.010	0.002	0.010	-0.001	-0.024	0.002	-0.426
other bovine animal outputs	0.000	0.001	0.000	-0.005	0.001	-0.002	-0.440	0.002	0.001	0.001	-0.012	-0.022	0.002	-0.407
non-bovine animal outputs	0.000	0.000	0.000	-0.001	0.002	-0.002	-0.485	0.010	0.001	0.013	0.034	-0.027	0.001	-0.538
crops	0.000	-0.005	-0.001	0.046	0.000	0.000	-0.697	-0.001	-0.012	0.034	0.343	-0.036	0.004	-0.494
	$\hat{c}$				$\hat{g}$			$\hat{c}$				$\hat{g}$		
	1.383	1.096	1.179	1.664	-0.004	0.005	1.794	1.097	1.151	1.399	1.221	0.095	-0.010	1.574
					$\hat{\mathbf{F}}$							$\hat{\mathbf{F}}$		
cropland					0.000	0.000	0.000					0.000	0.000	0.001
grassland					0.000	0.000	0.000					0.000	0.000	0.000
unpaid labour					0.000	0.000	0.040					0.001	0.000	0.012

Table A.14: Parameters of the activity level and fixed-input level dependent cost function without time index and q_min$_{ii}$ = 0 for dairy and cattle farms. *Source: own simulation.*

Cost Function Parameters																
	Crop Farms from southern Central France								Crop Farms from northern Central France							
	activity level					fixed-input level			activity level					fixed-input level		
Activity Level and Fixed-Input Level Dependent Cost Function without Time Index and $q\_min_{ii} = 0.2$	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour	animal outputs	pulses and oilseeds	industrial crops	cereals	other crops	cropland	grassland	unpaid labour
			$\hat{Q}$				$\hat{\mathrm{H}}$				$\hat{Q}$				$\hat{\mathrm{H}}$	
animal outputs	0.200	-0.004	0.018	0.020	0.022	-0.008	0.013	-0.683	0.200	-0.043	0.125	-0.019	0.111	-0.006	0.024	-1.877
pulses and oilseeds	-0.004	0.200	-0.002	0.022	0.106	-0.015	0.012	-0.300	-0.043	0.200	0.016	-0.069	0.117	-0.004	0.047	-1.490
industrial crops	0.018	-0.002	0.200	-0.044	0.033	-0.005	0.003	-0.738	0.125	0.016	0.200	0.001	0.077	-0.011	0.058	-2.653
cereals	0.020	0.022	-0.044	0.200	0.018	-0.020	0.016	-0.396	-0.019	-0.069	0.001	0.200	-0.001	-0.011	0.048	-1.633
other crops	0.022	0.106	0.033	0.018	0.200	-0.015	0.014	-0.103	0.111	0.117	0.077	-0.001	0.200	-0.014	0.042	-1.196
			$\hat{c}$				$\hat{g}$				$\hat{c}$				$\hat{g}$	
	-0.943	1.156	-1.070	1.331	-1.171	0.040	-0.049	1.743	-0.290	1.513	-4.803	1.585	-0.453	0.005	-0.192	6.400
							$\hat{\mathrm{F}}$								$\hat{\mathrm{F}}$	
cropland						0.000	0.000	-0.002						0.000	0.000	0.001
grassland						0.000	0.000	-0.001						0.000	0.000	0.001
unpaid labour						-0.002	-0.001	0.110						0.001	0.001	0.034

Table A.15: Parameters of the activity level and fixed-input level dependent cost function without time index and q_min_{ii} = 0.2 for crop farms. *Source: own simulation.*

Cost Function Parameters														
Activity Level and Fixed-Input Level Dependent Cost Function without Time Index and $q\_min_{ii} = 0.2$	Dairy Farms from West France							Cattle Farms from South-West France						
	activity level				fixed-input level			activity level				fixed-input level		
	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour	dairy products	other bovine animal outputs	non-bovine animal outputs	crops	cropland	grassland	unpaid labour
		$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$			$\hat{\mathbf{Q}}$				$\hat{\mathbf{H}}$	
dairy products	0.200	0.200	0.031	-0.042	-0.008	0.018	-0.066	0.200	0.003	0.044	-0.009	-0.028	0.000	-1.777
other bovine animal outputs	0.200	0.200	0.030	-0.042	-0.017	-0.001	-1.014	0.003	0.200	0.194	-0.016	-0.023	-0.001	-1.125
non-bovine animal outputs	0.031	0.030	0.200	-0.053	-0.011	0.029	-0.459	0.044	0.194	0.200	0.018	-0.015	-0.004	-1.959
crops	-0.042	-0.042	-0.053	0.200	-0.014	0.031	-0.074	-0.009	-0.016	0.018	0.346	-0.032	-0.001	-1.213
		$\hat{c}$				$\hat{g}$			$\hat{c}$				$\hat{g}$	
	-3.386	-0.470	-1.383	1.640	0.041	-0.077	-0.091	0.494	-1.794	0.397	1.237	0.102	0.009	3.645
						$\hat{\mathbf{F}}$							$\hat{\mathbf{F}}$	
cropland					0.000	0.000	-0.001					0.000	0.000	-0.016
grassland					0.000	0.000	0.000					0.000	0.000	0.000
unpaid labour					-0.001	0.000	0.254					-0.016	0.000	0.539

Table A.16**: **Parameters of the activity level and fixed-input level dependent cost function without time index and q_min$_{ii}$ = 0.2 for dairy and cattle farms. *Source: own simulation.*

.

Estimated error terms of the activity level and fixed-input level dependent cost function with q_min$_{iii}$=0.2 – comparison between estimation with and without time index

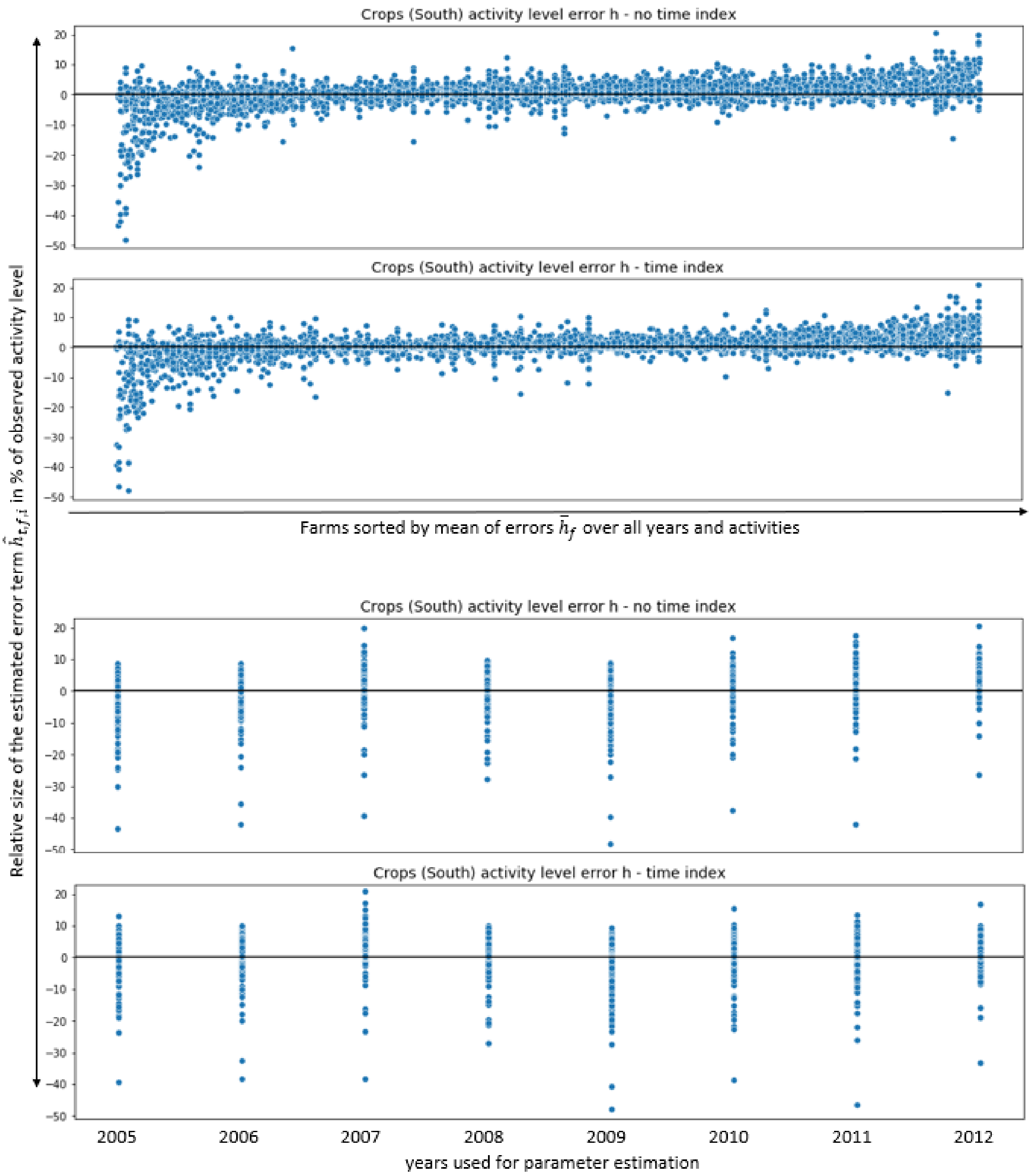

Figure A.5: Distribution of estimated activity level related error term $\hat{h}$ for crop farms in southern Central France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

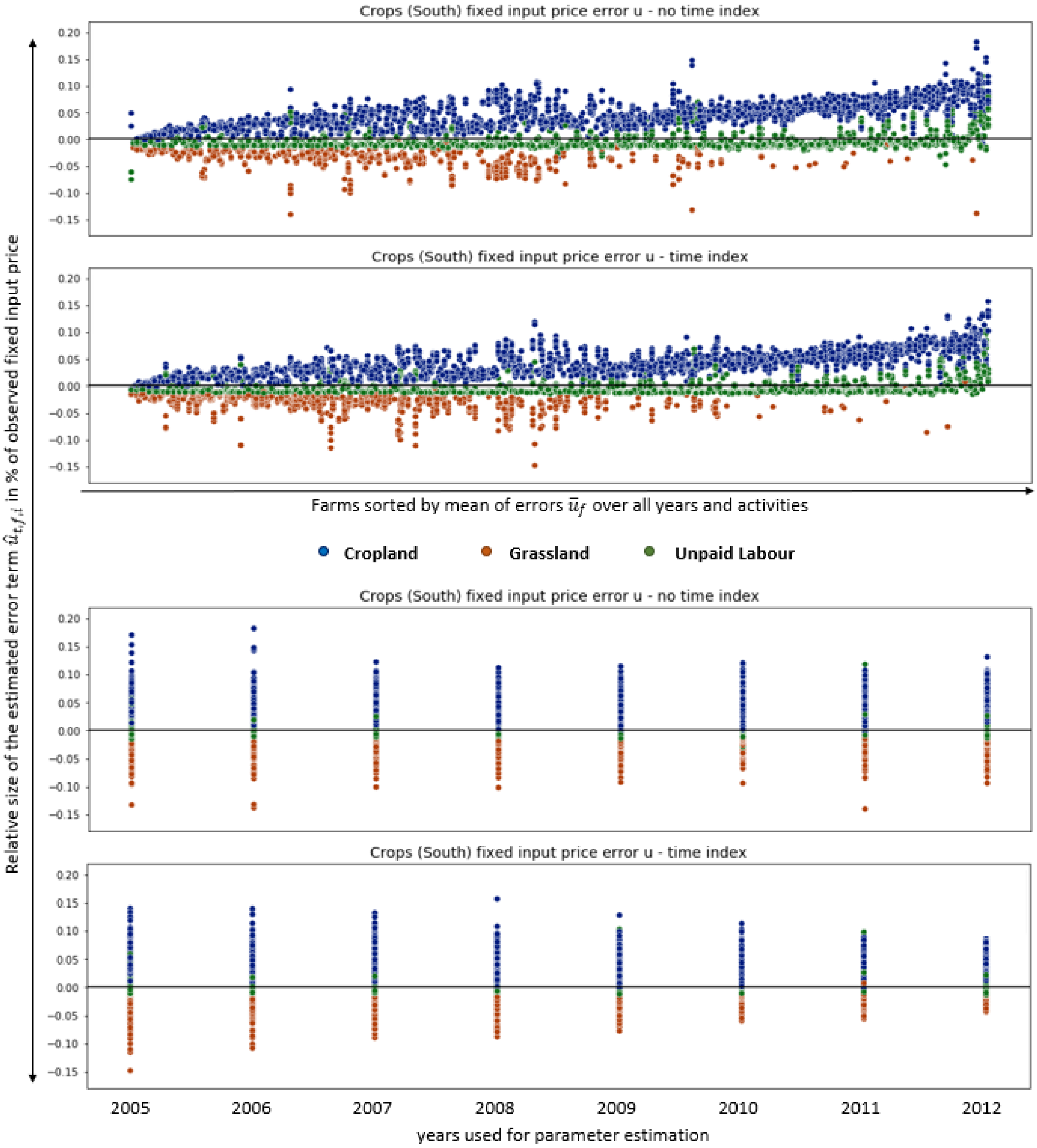

Figure A.6: Distribution of estimated fixed-input price related error term û for crop farms in southern Central France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

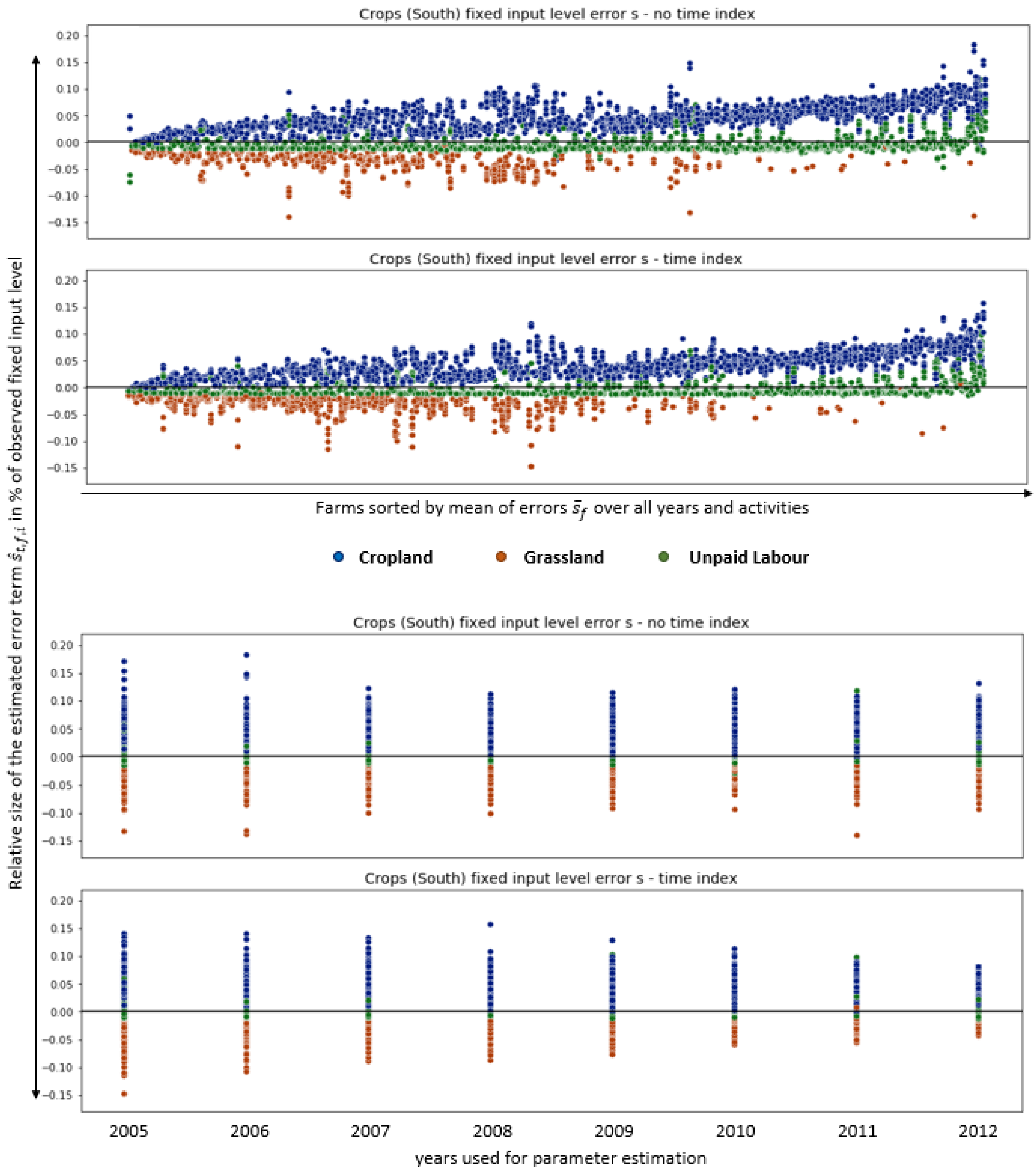

Figure A.7: Distribution of estimated fixed-input level related error term ŝ for crop farms in southern Central France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min_{ii} = 0.2. Source: own simulation and visualization.*

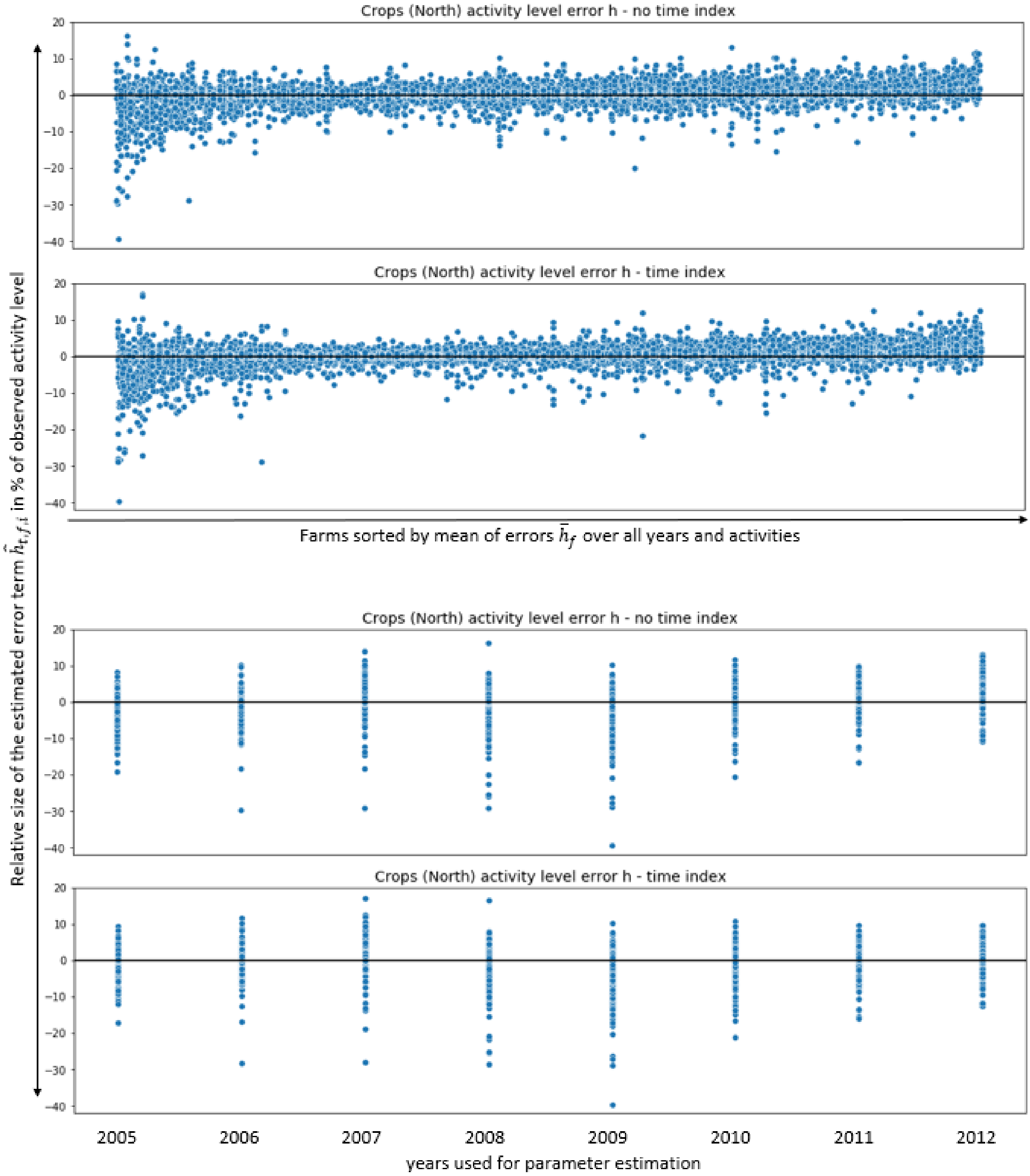

Figure A.8: Distribution of estimated activity level related error term $\hat{h}$ for crop farms in northern Central France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

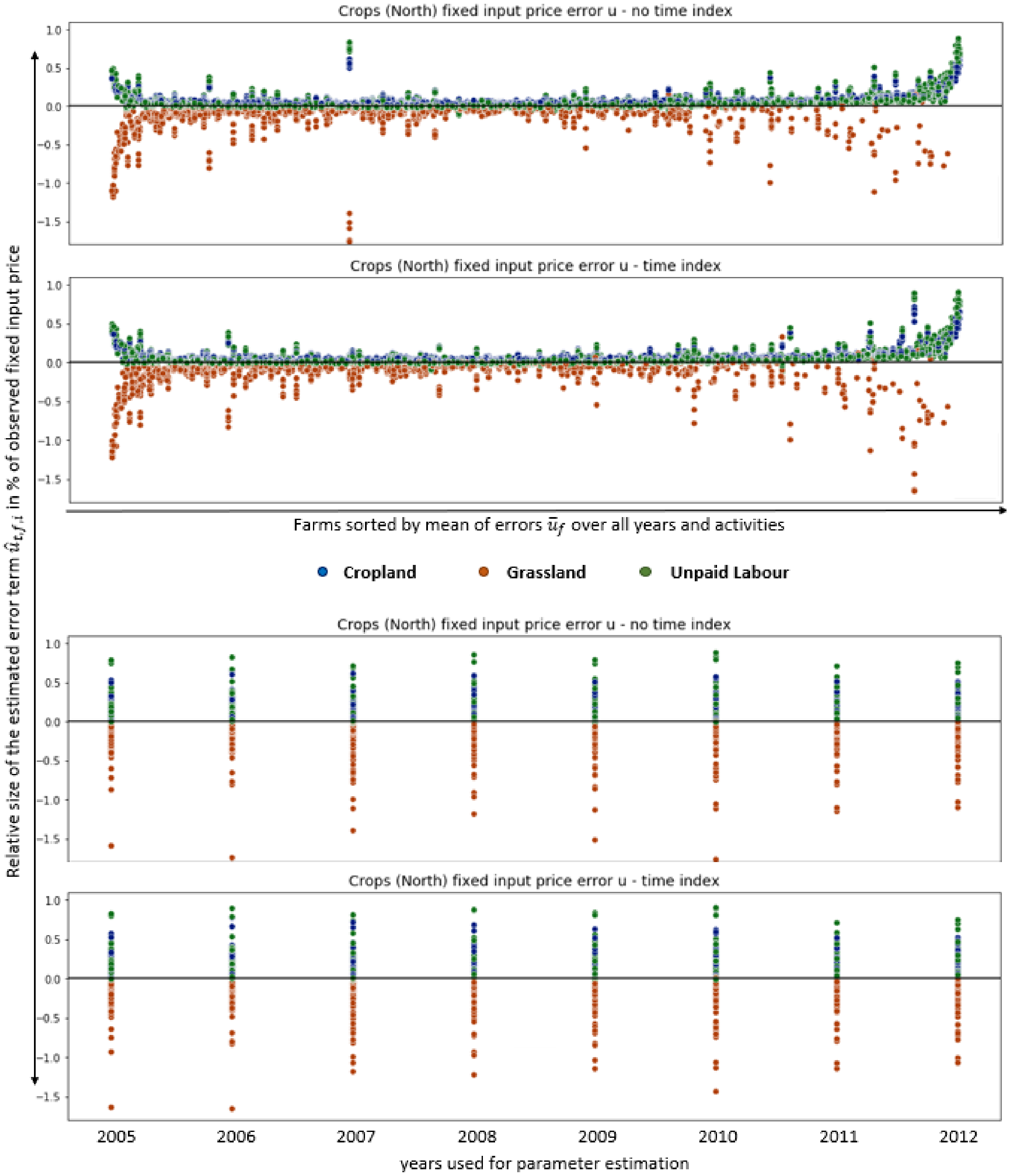

Figure A.9: Distribution of estimated fixed-input price related error term $\hat{u}$ for crop farms in northern Central France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min_{ii} = 0.2. Source: own simulation and visualization.*

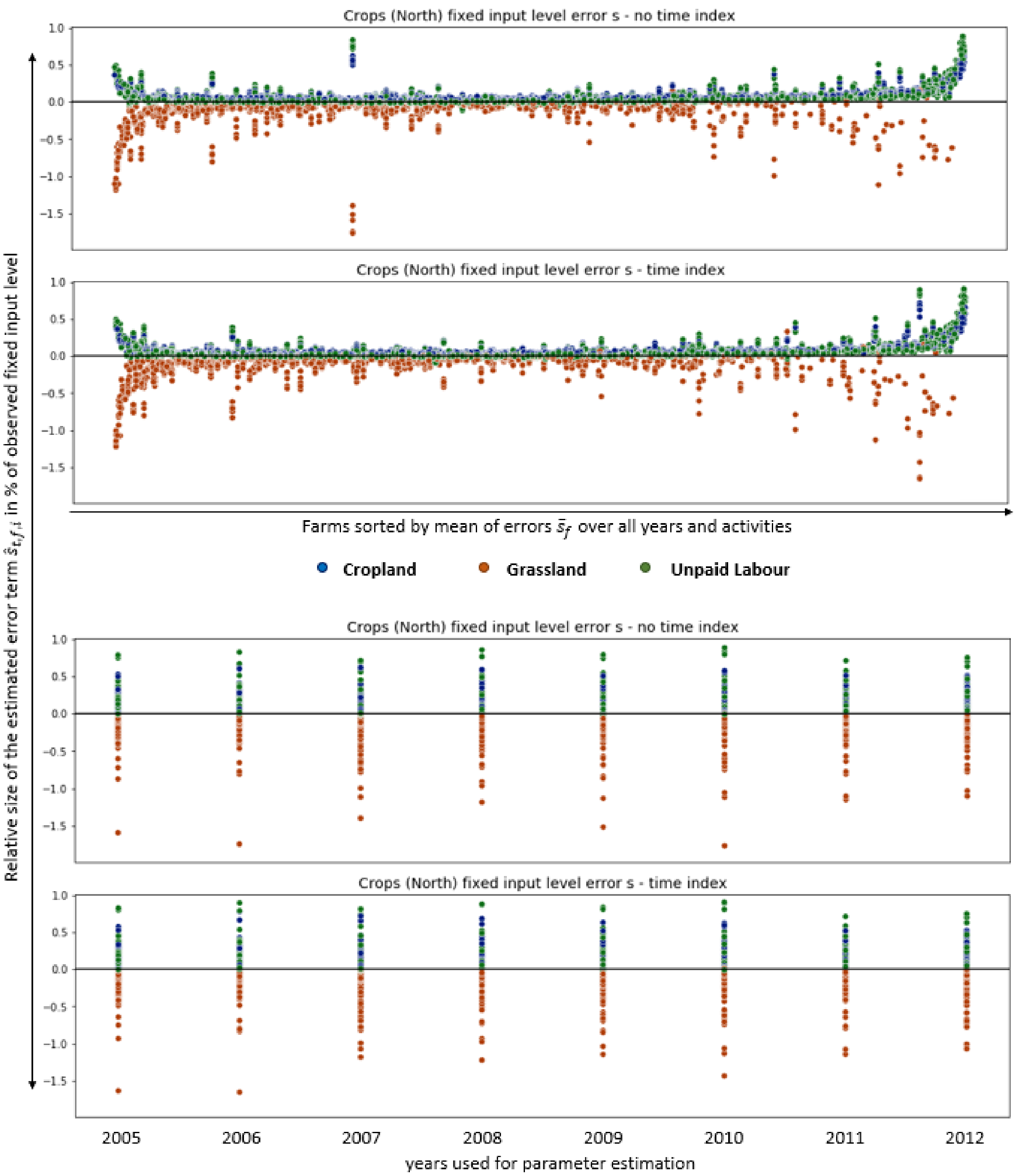

Figure A.10: Distribution of estimated fixed-input level related error term $\hat{s}$ for crop farms in northern Central France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min_{ii} = 0.2. Source: own simulation and visualization.*

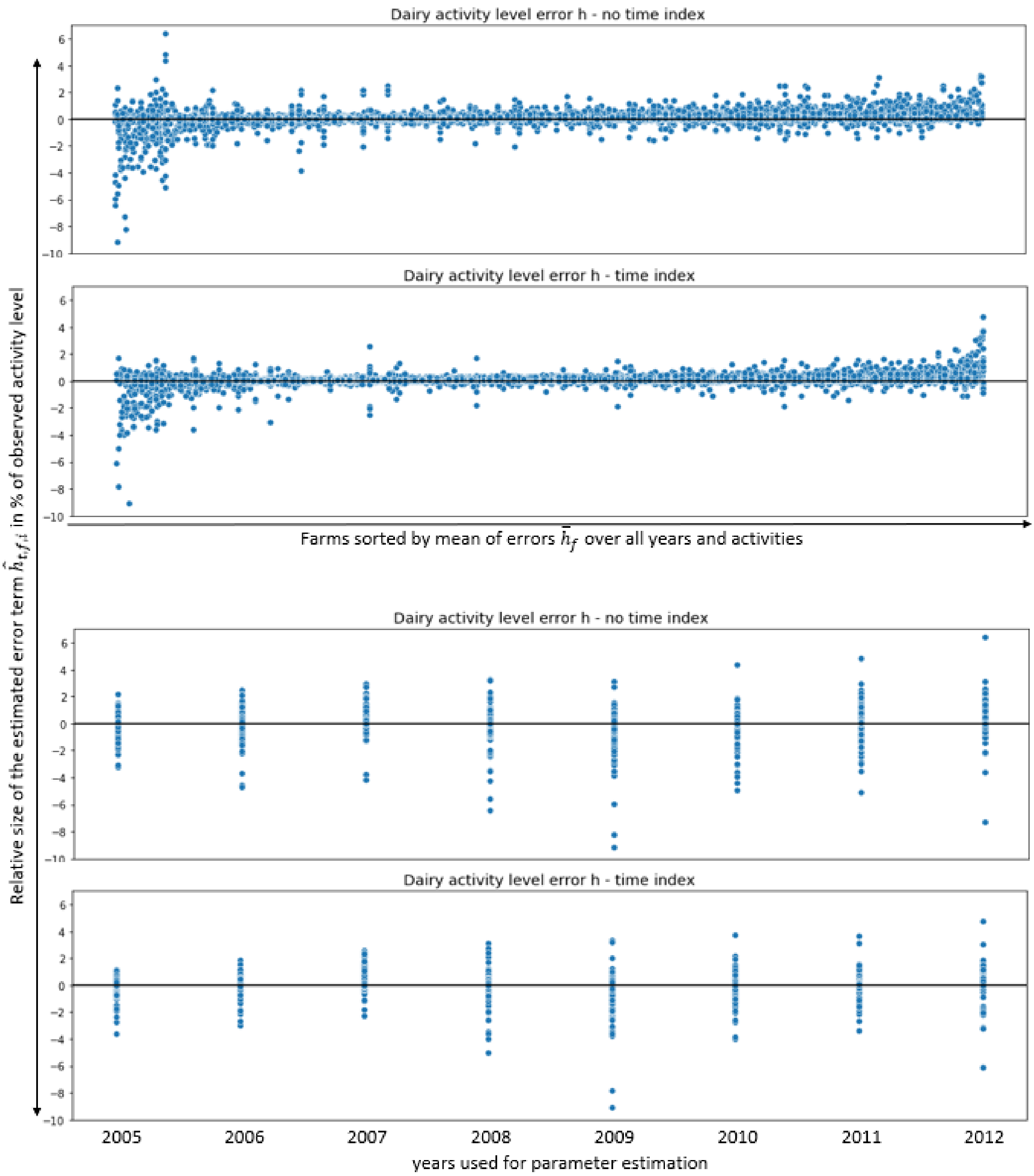

Figure A.11: Distribution of estimated activity level related error term $\hat{h}$ for dairy farms in West France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min_{ii} = 0.2. Source: own simulation and visualization.*

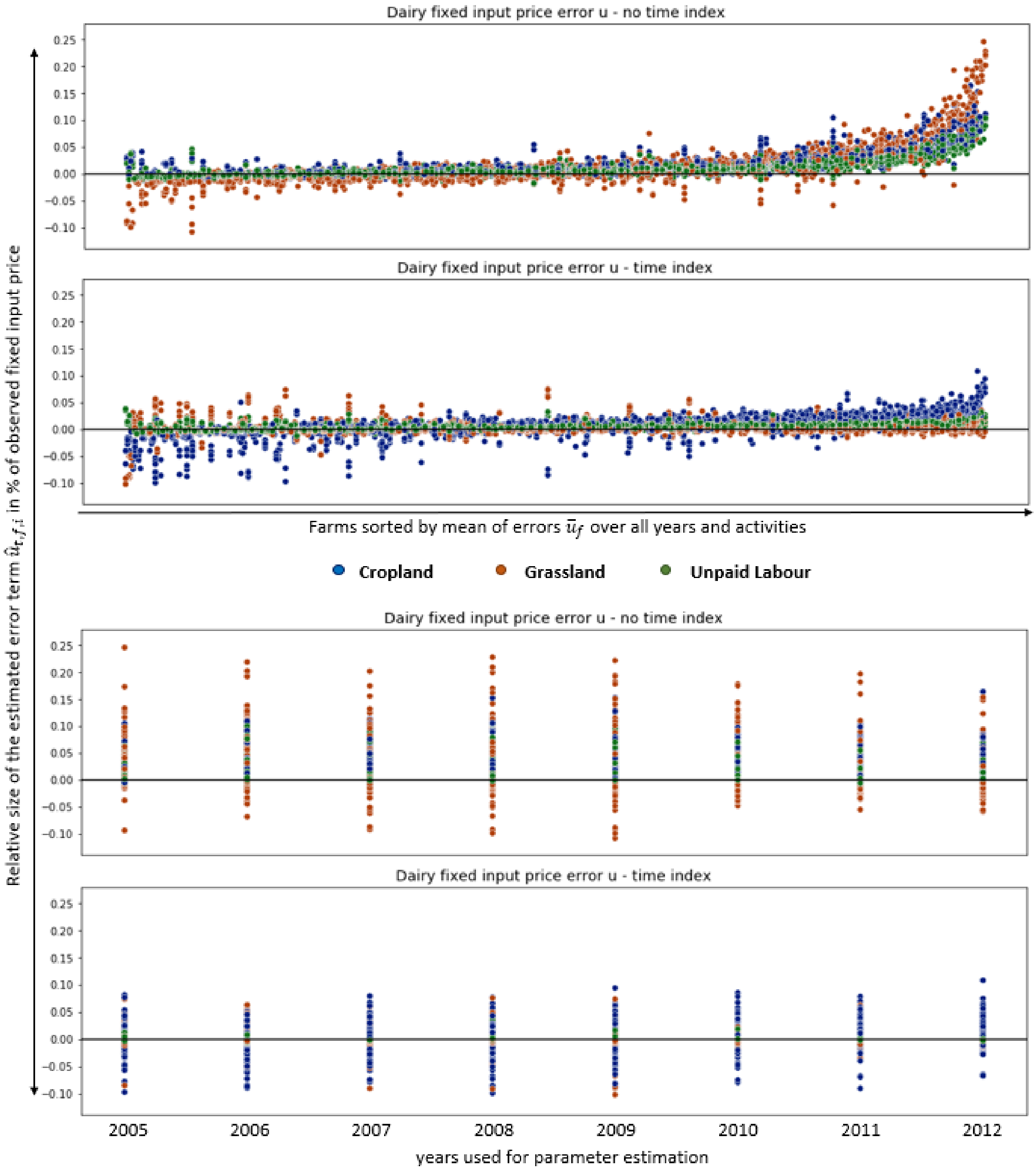

Figure A.12: Distribution of estimated fixed-input price related error term $\hat{u}$ for dairy farms in West France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

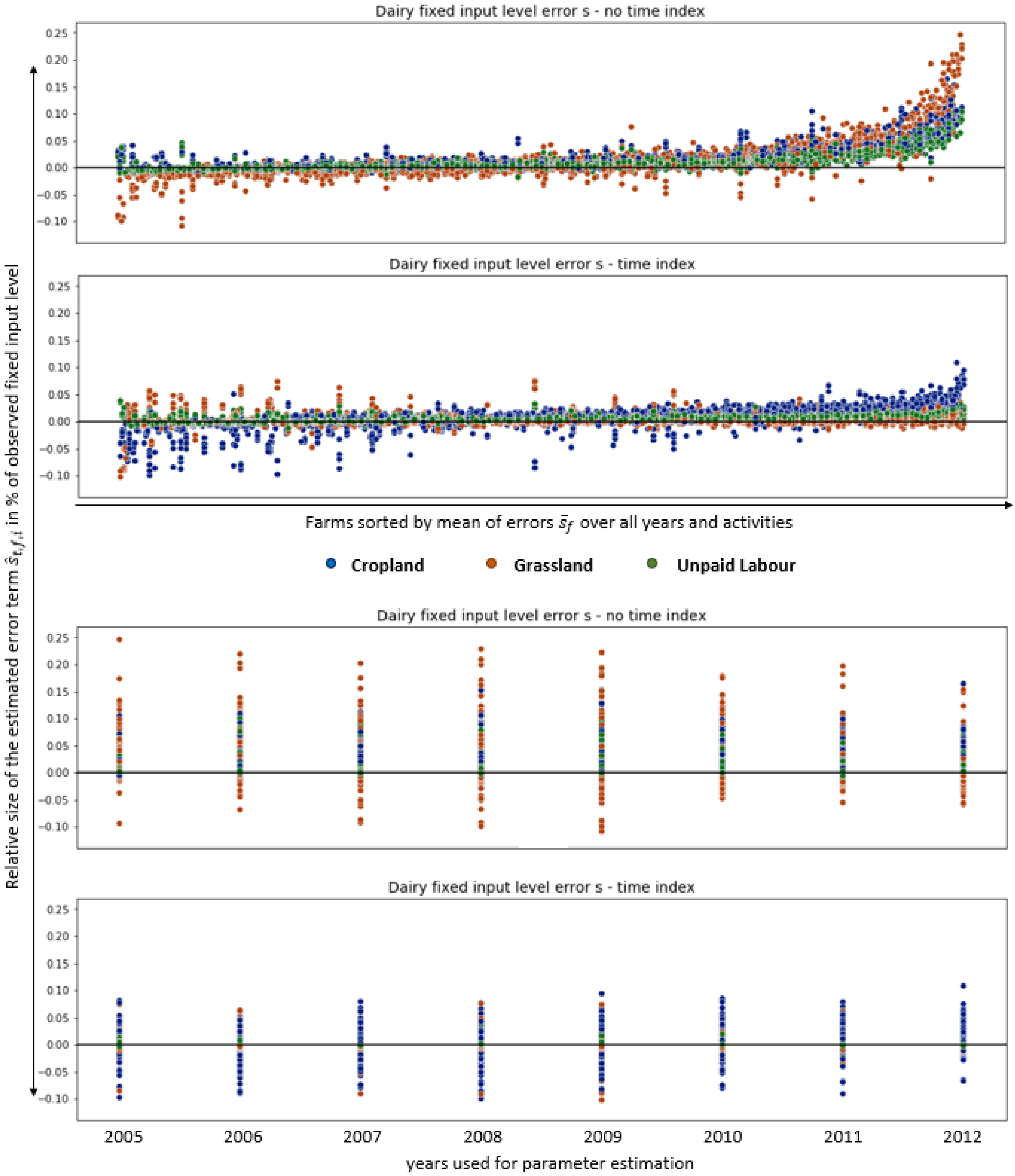

Figure A.13: Distribution of estimated fixed-input level related error term $\hat{s}$ for dairy farms in West France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

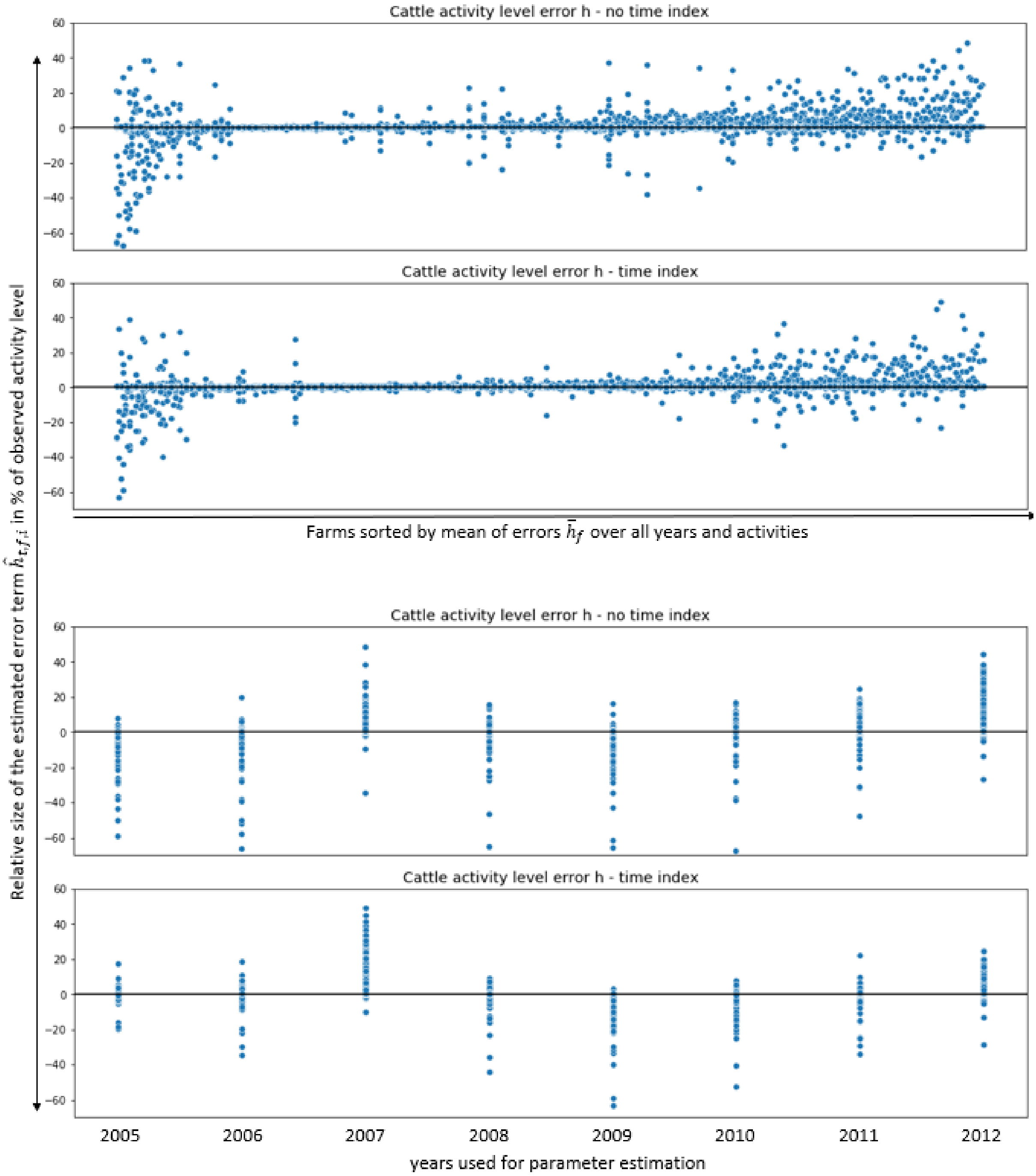

Figure A.14: Distribution of estimated activity level related error term $\hat{h}$ for cattle farms in South-West France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

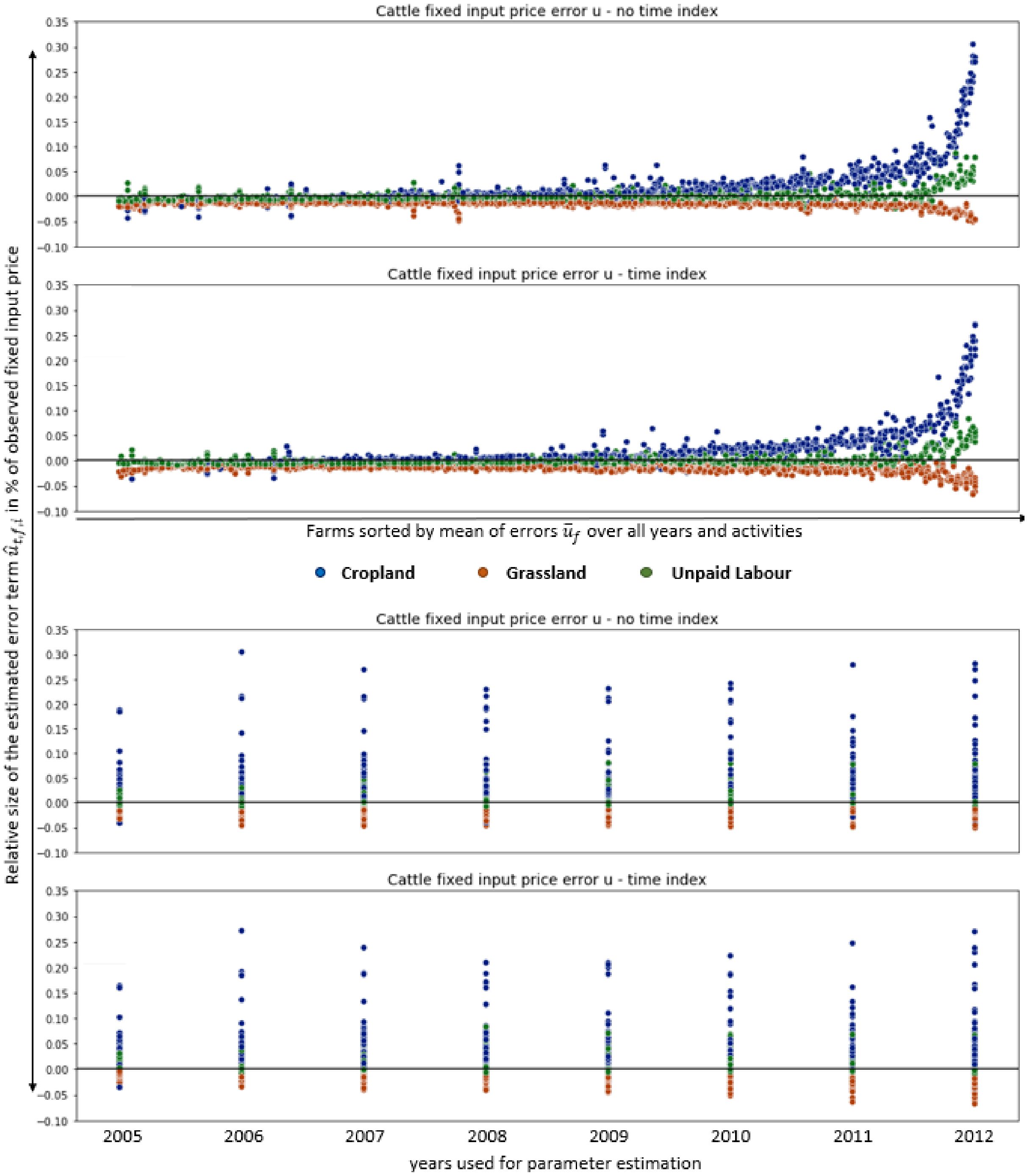

Figure A.15: Distribution of estimated fixed-input price related error term $\hat{u}$ for cattle farms in South-West France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

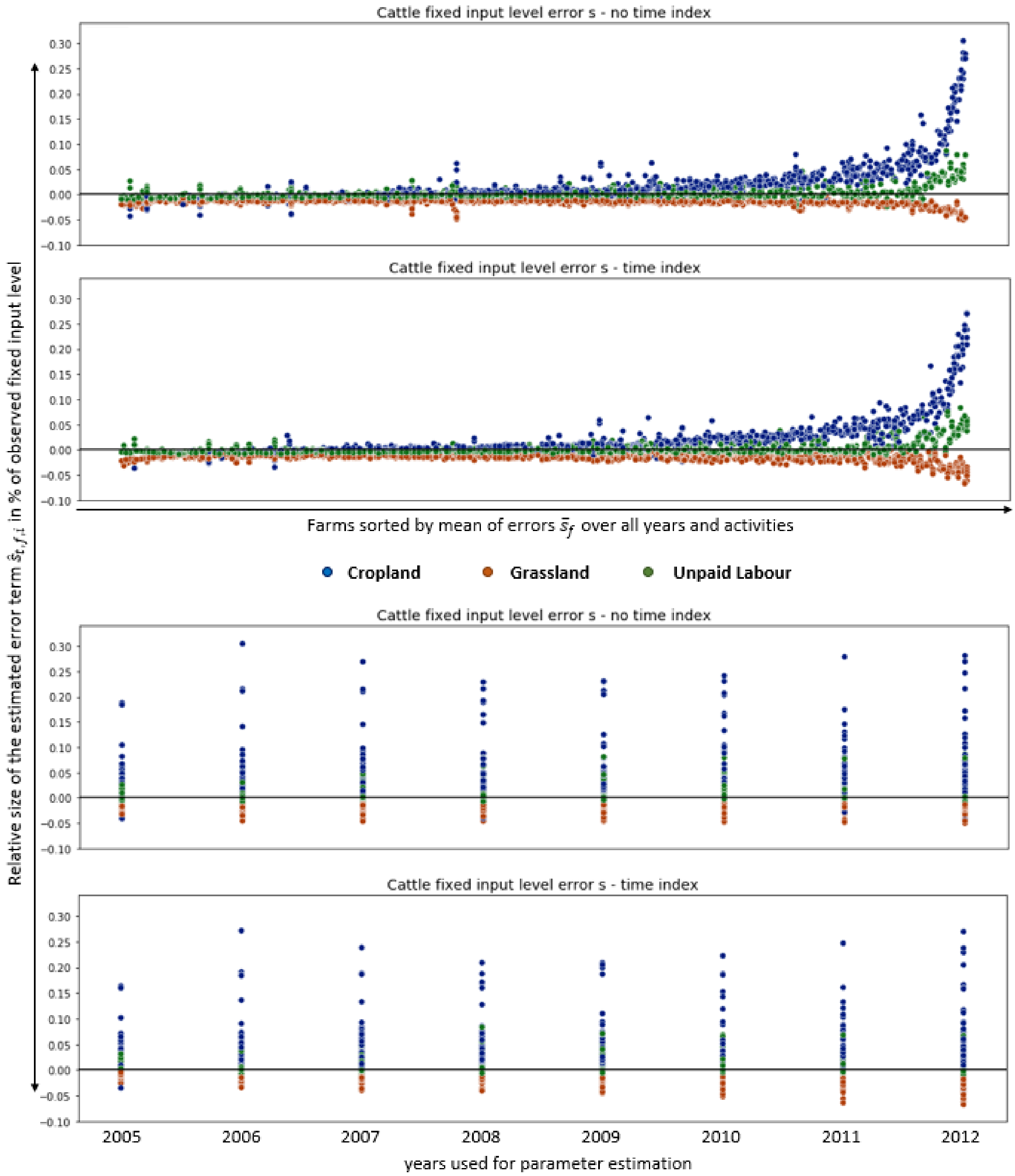

Figure A.16: Distribution of estimated fixed-input level related error term $\hat{s}$ for cattle farms in South-West France. *Relative size of error term, measured in % of observed value. Activity level and fixed-input level dependent cost function with q_min$_{ii}$ = 0.2. Source: own simulation and visualization.*

Annex IV – Application for a Greenhouse Gas Emission Cap Simulation

GHG emissions for all outputs in CO_2 equivalents

FADN variable	Source	CO_2 equivalents in kg per kg or ton per ton	CO_2 equivalents in tons per LU	description
K162	Agribalyse	1.18		Cow milk, national average, at farm gate/FR U
K163	Agribalyse	4.74		Average of all dairy products
N23	IPCC livestock emissions		3.68	Calves/fattening
N24	IPCC livestock emissions		5.4	Other cattle (<12m)
N25	IPCC livestock emissions		3.24	M cattle (12-24m)
N26	IPCC livestock emissions		3.24	F cattle (12-24m)
N27	IPCC livestock emissions		2.17	M cattle (>24m)
N28	IPCC livestock emissions		2.96	Breeding heifers
N29	IPCC livestock emissions		2.96	Heifers/ fattening
N30	IPCC livestock emissions		4.09	Dairy Cows
N31	IPCC livestock emissions		3.02	Cull dairy cows
N32	IPCC livestock emissions		2.77	Other Cows
N38	IPCC livestock emissions		1.44	Goat (breeding F)
n39	IPCC livestock emissions		1.44	Other goats
N40	IPCC livestock emissions		2.29	Ewes
N41	IPCC livestock emissions		2.29	Other Sheep
N43	IPCC livestock emissions		2.3	Piglets
N44	IPCC livestock emissions		0.76	Breeding sows
N45	IPCC livestock emissions		0.89	Pigs/fattening
N46	IPCC livestock emissions		1.07	Other pigs
N47	IPCC livestock emissions		0.08	Table chickens
N48	IPCC livestock emissions		0.06	Laying hens

N49	IPCC livestock emissions		0.05	Other poultry
N22	IPCC livestock emissions		0.68	horses
N33	IPCC livestock emissions		0	no data for bees
N34	IPCC livestock emissions		0	no data for rabbits
N50	IPCC livestock emissions		0	no data for other animals
K129	Agribalyse	0.29		average of the following categories in Agribalyse: French bean, conventional, national average, at farm gate/FR U; Spring faba bean, conventional, national average, at farm gate/FR U; Spring pea, conventional, 15% moisture, animal feed, at farm gate, production/FR U; Winter pea, conventional, 15% moisture, at farm gate/FR U
K132	Agribalyse	0.47		average of the following categories in Agribalyse: Sunflower grain, conventional, 9% moisture, national average, animal feed, at farm gate, production/FR U; Rapeseed, conventional, 9% moisture, national average, animal feed, at farm gate, production/FR U; Soybean, national average, animal feed, at farm gate/FR U
K120	Agribalyse	0.30		average of the following categories: Soft wheat grain, conventional, breadmaking quality, 15% moisture, at farm gate/FR U; Soft wheat grain, conventional, national average, animal feed, at farm gate, production/FR U
K121	Agribalyse	0.30		average of the following categories: Soft wheat grain, conventional, breadmaking quality, 15% moisture, at farm gate/FR U; Soft wheat grain, conventional, national average, animal feed, at farm gate, production/FR U
K126	Agribalyse	0.25		Maize grain, conventional, 28% moisture, national average, animal feed, at farm gate/FR U
K127	Agribalyse	0		no data for French rice
K122	Rajaniemi et al. 2011	0.85		rye

K123	Agribalyse	0.29		average of the following categories: Spring barley, conventional, downgraded quality, animal feed, at farm gate/FR U; Spring barley, conventional, malting quality, animal feed, at farm gate/FR U; Winter barley, conventional, malting quality, animal feed, at farm gate/FR U; Winter forage barley, conventional, animal feed, at farm gate/FR U
K124	Agribalyse	0.37		Oat grain, national average, animal feed, at farm gate/FR U
K125	Agribalyse	0.28		average of the following categories: Spring wheat, from intercrop, organic, system n°2, at farm gate/FR U; Sorghum grain, conventional, national average, animal feed, at farm gate/FR U; Triticale grain, conventional, national average, animal feed, at farm gate, production/FR U
K128	Agribalyse	0.28		average of the following categories: Spring wheat, from intercrop, organic, system n°2, at farm gate/FR U; Sorghum grain, conventional, national average, animal feed, at farm gate/FR U; Triticale grain, conventional, national average, animal feed, at farm gate, production/FR U
K130	Agribalyse	0.06		average of the following categories: Ware potato, conventional, variety mix, national average, at farm gate/FR U; Starch potato, conventional, national average, at farm gate/FR U
K131	Agribalyse	0.03		Sugar beet roots, conventional, national average, animal feed, at farm gate, production/FR U
K135	Agribalyse	0.05		average of the following categories: Sugar beet roots, conventional, national average, animal feed, at farm gate, production/FR U; Ware potato, conventional, variety mix, national average, at farm gate/FR U; Starch potato, conventional, national average, at farm gate/FR U
K133	Agribalyse	0		no data for French hop
K134	Agribalyse	0		no data for French tobacco

K136	Agribalyse	0.18		average of the following categories: Carrot, conventional, national average, at farm gate/FR U; Cauliflower, national average, at farm gate/FR U; Chicory witloof, conventional, national average at farm gate/FR U; Leek, national average, at plant/FR U; Lettuce, conventional, national average, at farm gate/FR U; Melon, open field, conventional, at farm gate/FR U; Onion, national average, at farm/FR U; Squash, open field, conventional, at farm gate/FR U; Strawberry, open field, conventional, at farm gate/FR U;
K137	Agribalyse	0.32		average of the values for pCO2otc_136 and pCO2otc_138
K138	Agribalyse	0.46		average of the following categories: Carrot, conventional, national average, at farm gate/FR U; Cauliflower, national average, at farm gate/FR U; Chicory witloof, conventional, national average at farm gate/FR U; Leek, national average, at plant/FR U; Lettuce, conventional, national average, at farm gate/FR U; Melon, national average, at farm gate/FR U; Onion, national average, at farm/FR U; Squash, conventional, national average, at farm gate/FR U; Strawberry, national average, at farm gate/FR U; Tomato, average basket, conventional, heated greenhouse, national average, at greenhouse/FR U; Tomato, average basket, conventional, soil based, non-heated greenhouse, at greenhouse/FR U;
K139	Agribalyse	0		no data for mushrooms
K140	Agribalyse	0.77		average of the following categories: Potted shrub, national average, at production site/FR U; Rose (cut flower), soilless, low-heated, conventional pest management, at greenhouse/FR U; Rose (cut flower), soilless, low-heated, integrated pest management, at greenhouse/FR U

K141	Agribalyse	0.93		average of the following categories: Potted shrub, national average, at production site/FR U; Rose (cut flower), production mix, national average, at greenhouse/FR U; Rose (cut flower), soilless, heated and enlighted, conventional pest management, at greenhouse/FR U; Rose (cut flower), soilless, heated and enlighted, integrated pest management, at greenhouse/FR U

Table A.17: GHG emissions of each activity in tons of CO_2equ/t or LU. *These values were used as the input values to calculate the emission factors for each aggregated activity and farm.*

Note: “Agribalyse” indicates Agribalyse (2020b) for conventional products and Agribalyse (2020a) for organic products. “IPCC livestock emissions” refers to IPCC (2006).

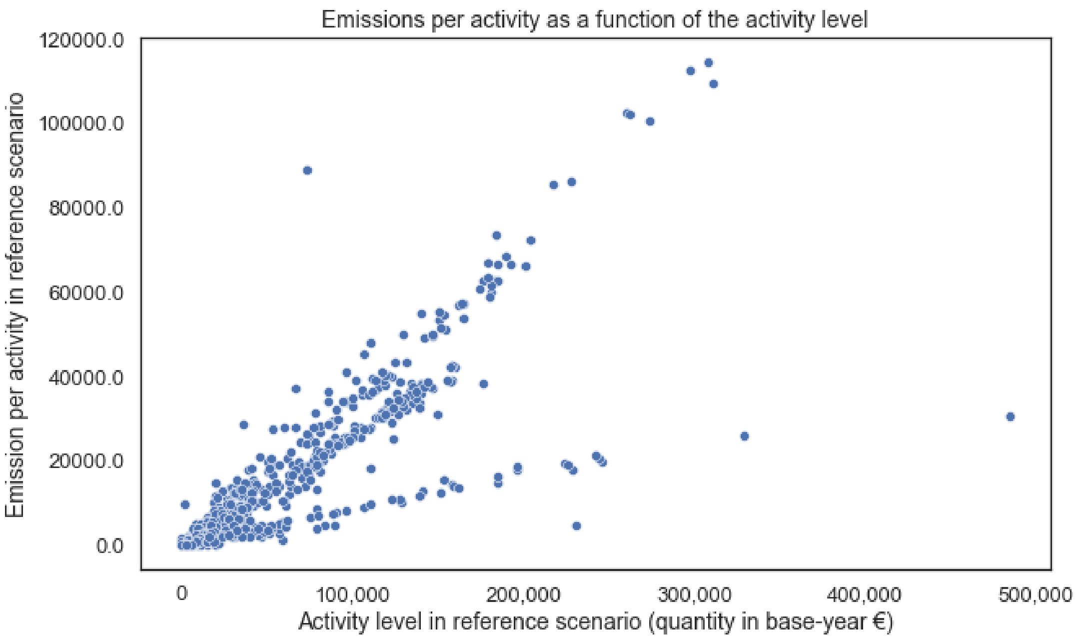

Figure A.17: Emissions per activity as a function of the activity level. *Activity emissions depend largely on the activity level. Source: own visualization and calculation.*

Emission Cap

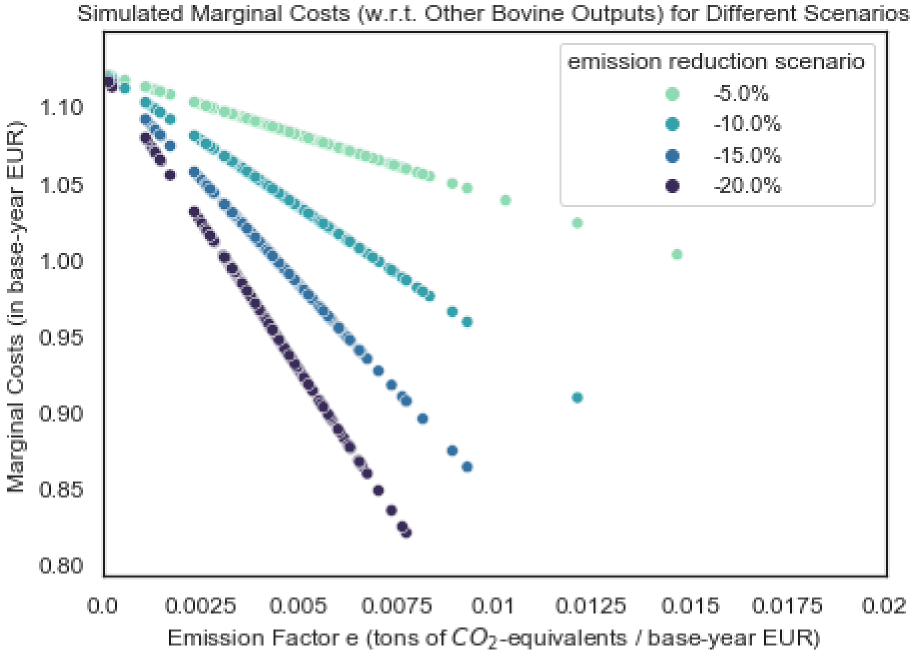

Figure A.18: Simulated marginal costs w.r.t. other bovine outputs for dairy farms, for four different emission reductions scenarios. *Source: own visualization and simulation.*

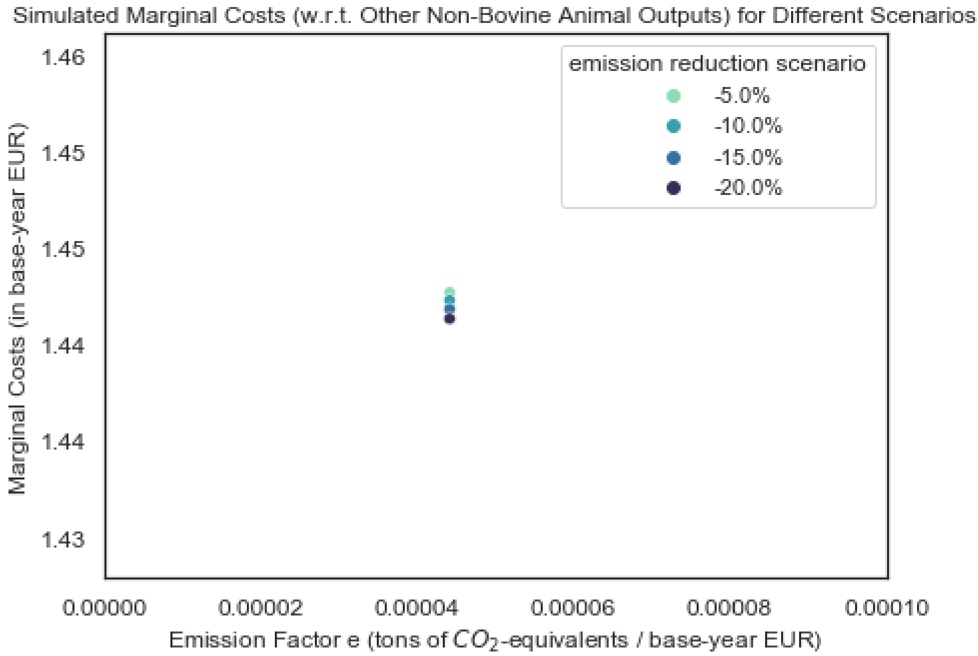

Figure A.19: Simulated marginal costs w.r.t. other non-bovine animal outputs for dairy farms, for four different emission reductions scenarios. *Source: own visualization and simulation.*

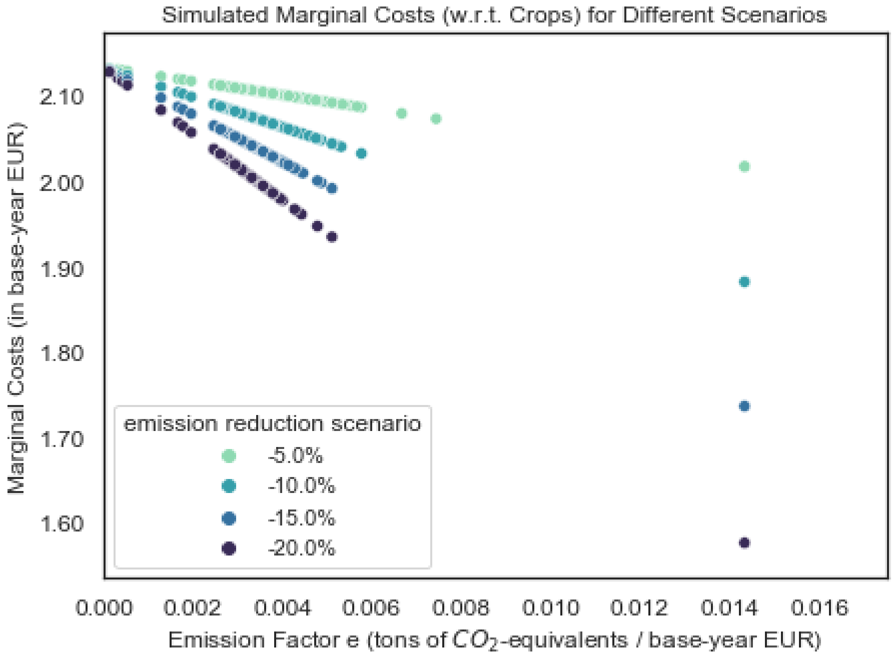

Figure A.20: Simulated marginal costs w.r.t. crop outputs for dairy farms, for four different emission reductions scenarios. *Source: own visualization and simulation.*

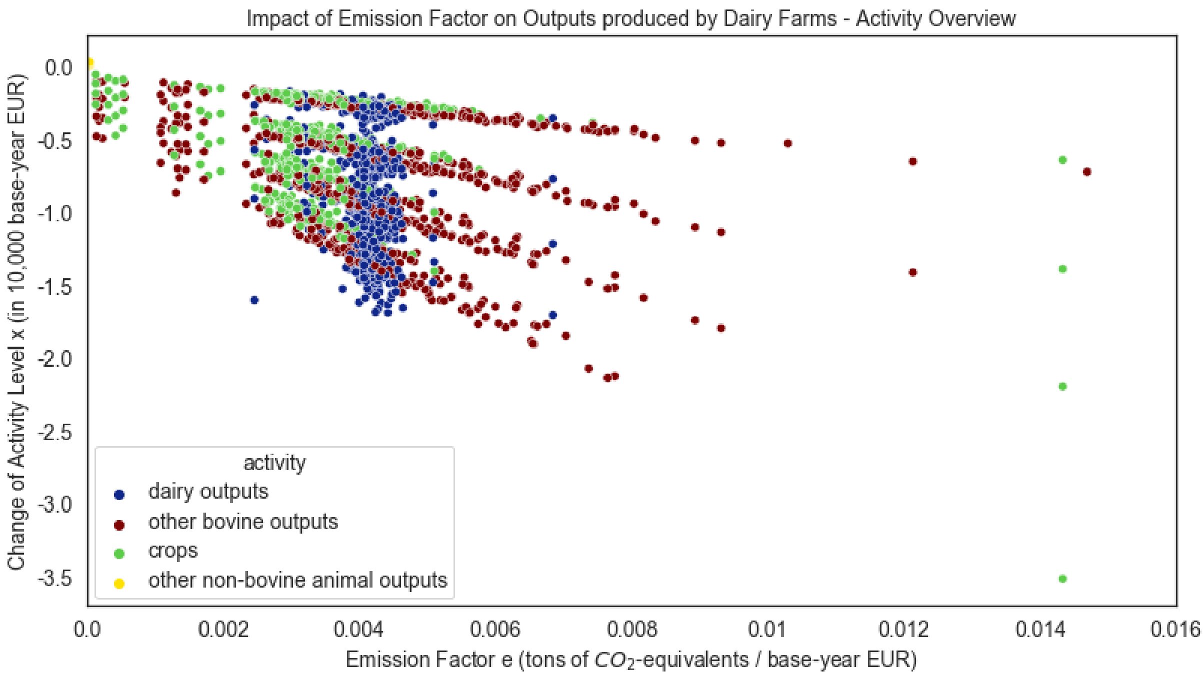

Figure A.21: Impact of emission factor on outputs by dairy farms. *Source: own visualization and simulation.*

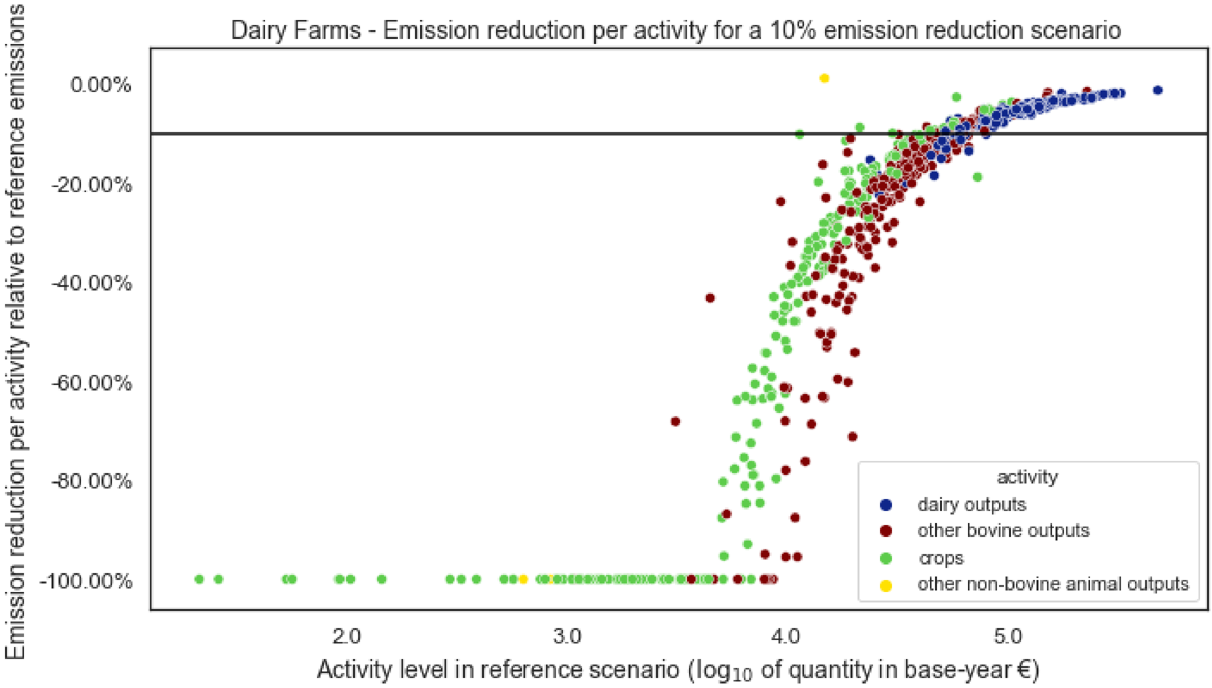

Figure A.22: impact of activity level in reference scenario on the emission reduction for a 10% emission cap. *Source: own visualization and simulation.*

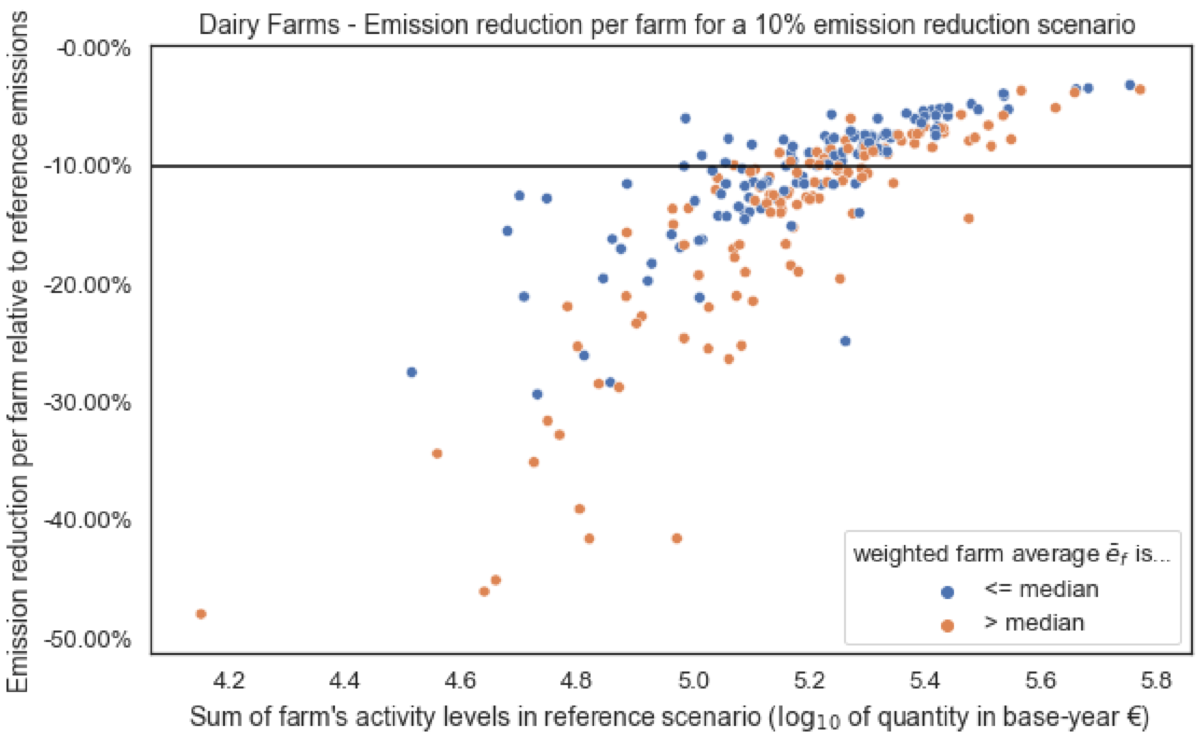

Figure A.23: Dairy farm's emission reduction for an emission cap of 10% compared to the reference scenario. *Farms that lie above the black line are net emission right buyers, whereas the farms below the line are net sellers of emission rights. Source: own visualization and simulation.*

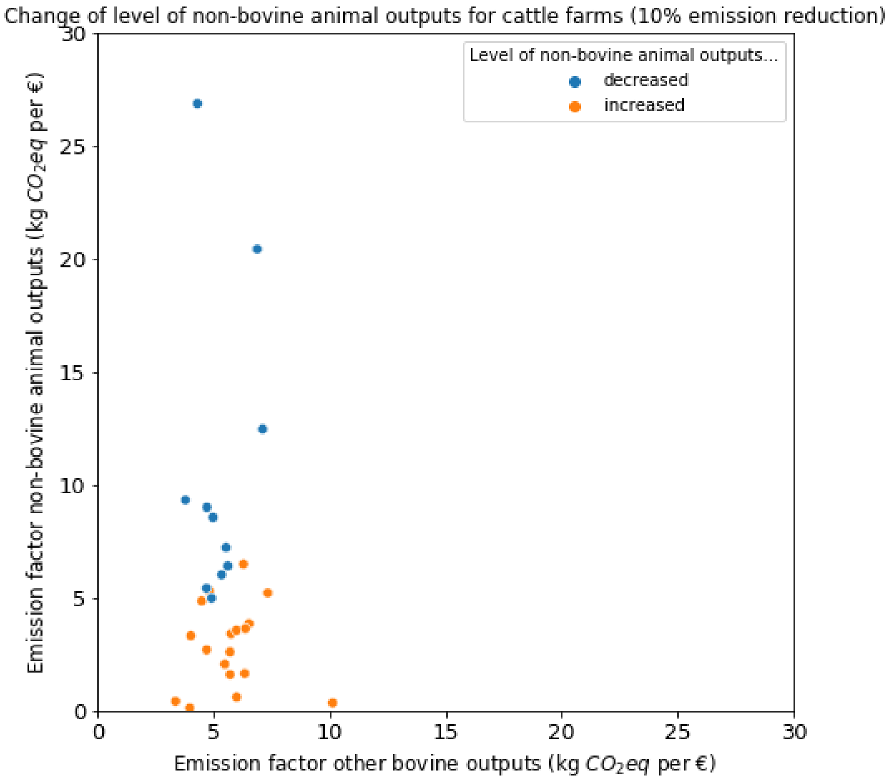

Figure A.24: Impact of the emission factors of other bovine outputs and non-bovine animal outputs of cattle farms on the activity level change of non-bovine animal outputs. *If the emission factor of non-bovine animal outputs is larger than the emission factor of other bovine outputs, cattle farms reduce the level of non-bovine animal outputs. If the emission factor of non-bovine animal outputs is comparably small, farms increase the level of non-bovine animal outputs. This relationship explains why for a few cattle farms the level of non-bovine animal outputs increases if an emission cap is imposed. Source: own visualization and simulation.*

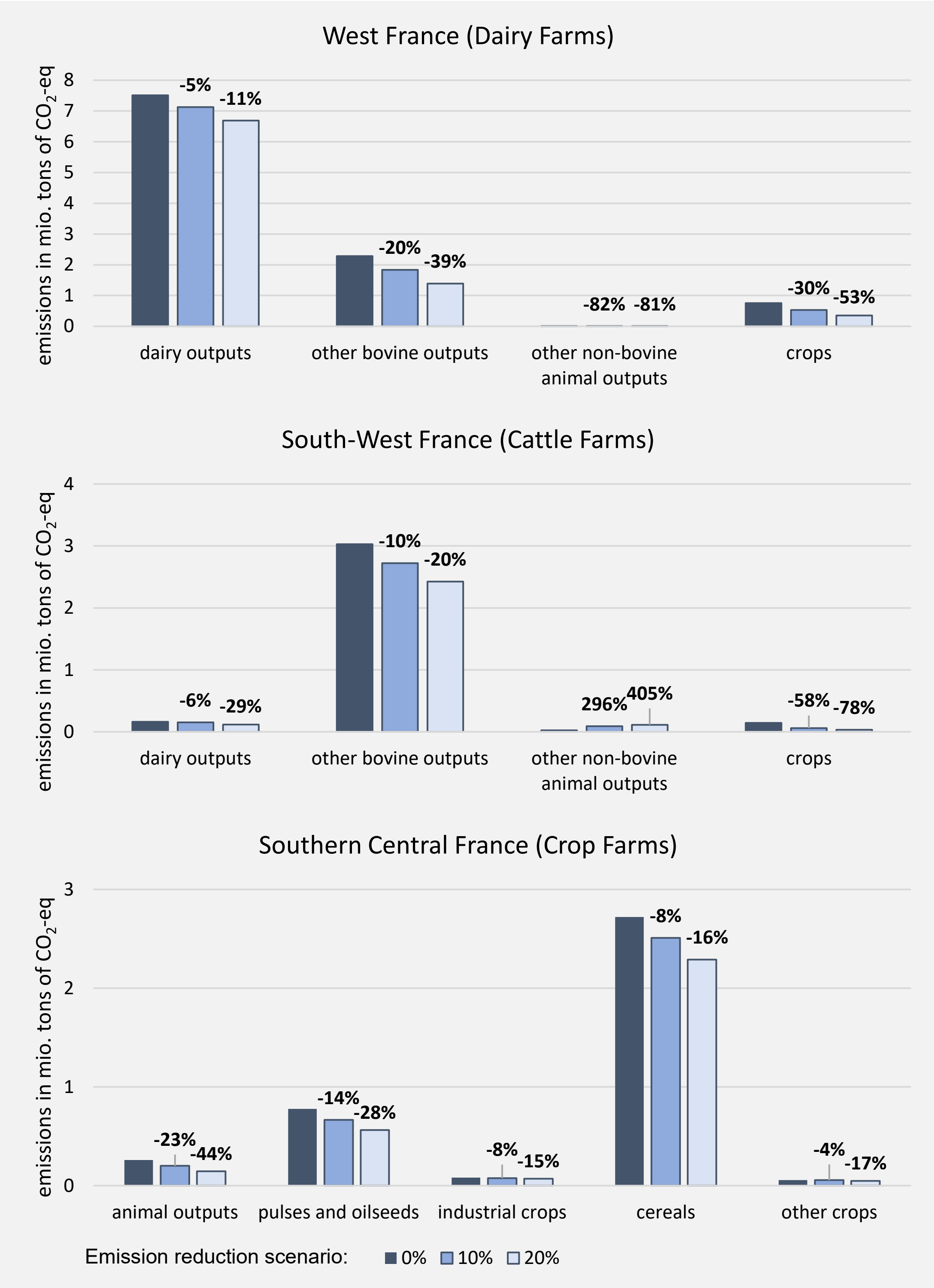

Figure A.25: Impact of different GHG emission caps on GHG emissions caused by each output category.
Source: own visualization and simulation.

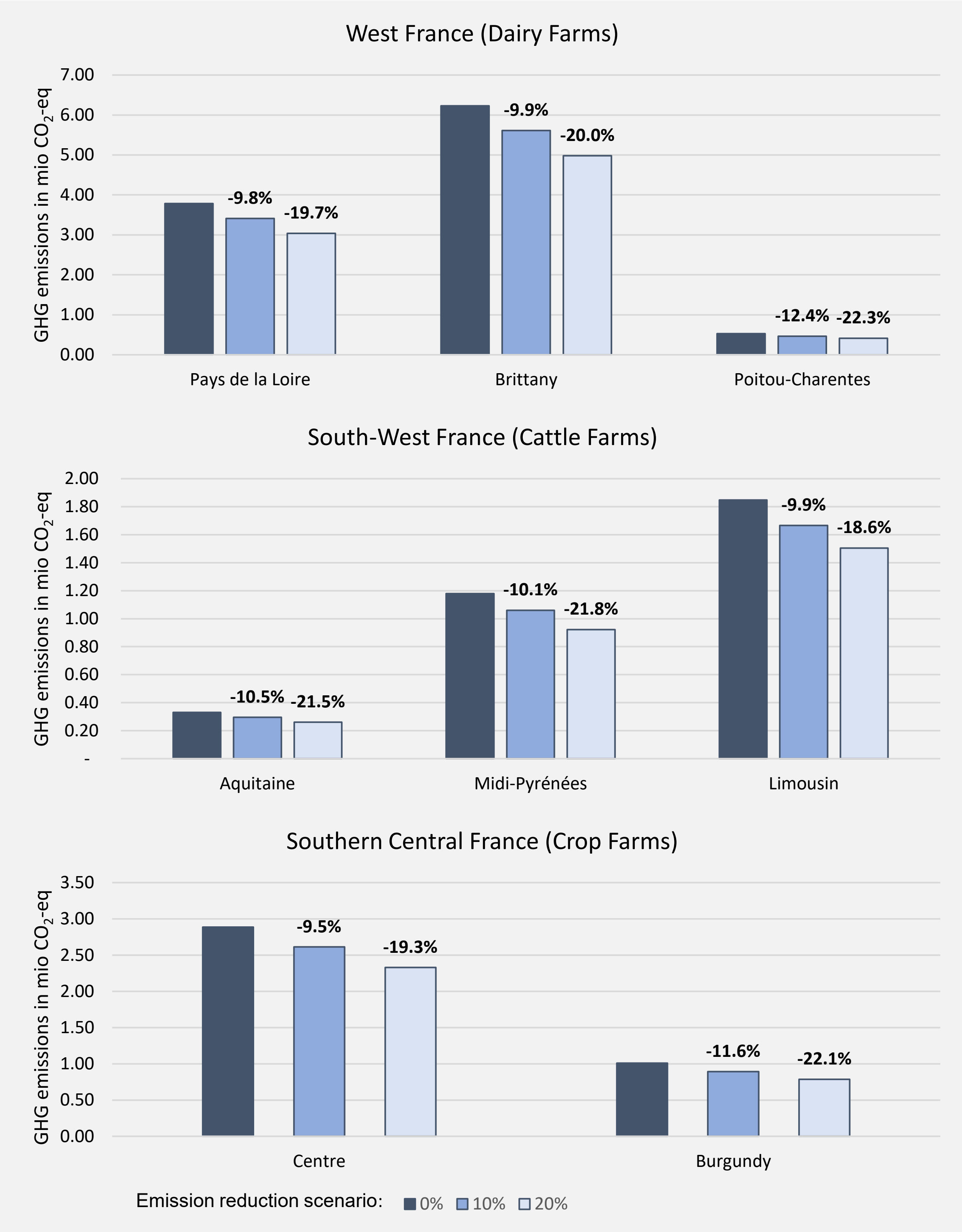

Figure A.26: Impact of emission cap on emissions from each subregion. *Source: own visualization and simulation.*