

Multi-input multi-output cost function for
IFM-CAP model
Contract 153916-2013 A08-BE
Deliverable 8 : Final Report

Bruno Henry de Frahan, Jérémie Dong and Rembert De Blander
Earth and Life Institute, Université catholique de Louvain

Louvain-la-Neuve, 22 April 2015

Contents

I	Design and Development of a Method for Estimating Nested Cost and Input Demand Functions	5
	Introduction	6
1	Nested Cost and Input Demand Functions	7
1.1	Context	7
1.2	The Nested Cost Function and its Derived Input Demand Functions	9
1.3	The Nested Input Demand Functions	20
1.4	Conclusions	29
2	Data and Aggregation	31
2.1	Introduction	31
2.2	Data Preparation	32
2.3	Aggregation Scheme	33
II	Design and Development of a Method to Prepare and Update the Data from the EU-FADN	39
	Introduction	40
3	Törnqvist Index Construction	42
4	Imputation of Missing Prices	45
5	Output Data Preparation	48
5.1	Animal-specific Outputs	48

5.2	Crop-specific Outputs	52
5.3	Net Sale Value of Output Categories	52
6	Input Data Preparation	55
6.1	Variable Inputs	55
6.2	Fixed versus Variable Inputs	58
6.3	Land	60
6.4	Non-land Capital	62
6.5	Estimation	66
 III Estimation of the Nested Cost and Input Demand Functions for a Selection of Farm Types and EU Regions		 67
Introduction		68
7	Estimation Procedure	69
7.1	Preliminary Remark	69
7.2	Cost Function	69
7.3	System of Equations	70
7.4	Symmetry and Adding-up Restrictions	71
7.5	Monotonicity Conditions	71
7.6	Curvature Conditions	73
7.7	Estimation Method	75
7.8	Rescaling the Output and Fixed Input quantities	76
7.9	Further Remarks	77
8	Selected Estimation Results	79
8.1	Introduction	79
8.2	Estimated Cost Functions with Upper-level Inputs and Outputs for Crop Farms	81
8.3	Estimated Cost functions with Upper-level Inputs and Lower- level Outputs for Crop Farms	85
8.4	Estimated Expenditure Functions with Lower-level Inputs and Upper-level Outputs for Crop Farms	90
8.5	Estimated Expenditure Functions with Lower-level Inputs and Outputs for Crop Farms	93
8.6	Preliminary Conclusions	96

IV	Conclusions and Recommendations	98
	Introduction	99
9	Main Challenges	100
9.1	Data Issues	100
9.2	Aggregation of input and output categories	101
9.3	Robustness of the Results	101
9.4	Computational Requirements	102
10	Strengths and Limitations	103
10.1	Main Strenghts	103
10.2	Main Limitations	104
11	Improvements	105
11.1	Removing Outliers	105
11.2	Quasi-fixed Input Allocation to Output Categories	105
11.3	Alternative Functional Form	106
	Bibliography	107

Part I

Design and Development of a Method for Estimating Nested Cost and Input Demand Functions

Introduction

This first part gives the main concepts for estimating a nested multi-input and multi-output cost function, preparing farm accounting data into input and output categories at different levels of aggregation and proceeding to the econometric estimation. These concepts will be implemented in the following parts of this project. With respect to previous work in that empirical domain, in particular the work as reported in Henry de Frahan et al. (2011) and De Blander et al. (2011), this work adds the original feature of introducing and implementing a cost function that is nested in an upper and a lower level of both input and output categories that allows to consider a wider range of input and output categories that is generally not considered in the available empirical literature on cost functions. The authors are not aware of previous development of a nested cost function in the literature while the concept of nested functions is widely used for production and utility functions. With respect to this previous work, this work also calculates differently land and non-land capital inputs, using a better estimate of the opportunity cost of capital as suggested in Andersen et al. (2011).

This final report is organized in three parts, each one introducing the concept for estimating a nested multi-input and multi-output cost function, the concept for preparing farm accounting data into input and output categories at different levels of aggregation and the concept for proceeding to the econometric estimation.

Chapter 1

Concept for Estimating Nested Cost and Input Demand Functions

1.1 Context

The overall objective of the project consists in developing and applying a method for estimating a theoretically consistent and flexible multi-input multi-output cost function for a disaggregated set of input and output categories for individual FADN farms using EU-FADN data. The output categories need to be disaggregated at product level and the input categories at input level by farm type as reported in the EU-FADN data set.

The estimation of the theoretically consistent and flexible multi-input multi-output cost function reported in the FACEPA Deliverable 9.1 (De Blander et al., 2011) uses the Symmetric Generalized McFadden (SGM) functional form that is particularly ideal for applied work. Among the class of flexible quadratic cost functions, the multi-input multi-output SGM cost function is a function for which the global curvature properties of a cost function can be imposed if needed without destroying its second-order flexibility (see Diewert and Wales, 1987). It is expressed in terms of variable input prices, output quantities and quasi-fixed input quantities. The estimation also uses an augmentation of the SGM functional form to allow third-order terms in output quantities. This addition allows estimating cost functions for which marginal costs are downward sloping for some farms, a possible situation when outputs are limited by quotas. The estimation uses a medium-term and long-term versions of the SGM and augmented SGM functional forms. The estimates are obtained by a fixed-effects non-linear seemingly unrelated regression (SUR) of input demands using an unbalanced panel of FADN farms from 1990 to the latest available year and imposing the theoretical restrictions on parameters, i.e.,

the symmetry and adding up restrictions, the curvature conditions of a theoretically consistent cost function, and/or the monotonicity conditions (see Chambers, 1988, p. 52 and 102). This SGM functional form is also used for estimating disaggregated input demands because of its second-order flexibility.

Because of limitations in degrees of freedom, risk of multicollinearity and failure to converge, the specification of cost function includes a limited number (three to five) of variable input categories, a limited number (two to five) of output categories and a limited number (one to three) of quasi-fixed input categories by farm type. FACEPA Deliverable 9.2 (Bahta et al., 2011) reports a number of applications for crop, dairy and livestock farms for several representative EU regions and member states with convergence failures and unrealistic estimated marginal costs for some of the applications. IPTS would like to expand this theoretically consistent and flexible multi-input multi-output cost function to a wider set of input and output categories at a more disaggregated level.

For that purpose, we will test a method to disaggregate the cost and input demand functions into a greater number of input and output categories relying on the concept of hierarchical or nested functions that is widely used in consumer demand analysis (Deaton and Muellbauer, 1999, p. 117-147), production analysis (Sato, 1967) and often applied in computable general equilibrium models. This concept rests on the assumption that a function of many arguments could be separate into sub-functions (see Green, 1964). The application of this concept to a cost function and its derived input demand functions assumes then the functional separability of broad output and input categories.

This separability assumption is acceptable to the extent that outputs sharing a similar underlying technology are grouped together in the same broad output category such that the technology of producing these outputs in one particular broad output category is separate from the technology of producing outputs belonging to another broad output category. This implies that producing one output belonging to a broad output category cannot directly affect producing another output that belongs to another broad output category. It can only affect indirectly producing this another output through producing the broad output category to which it belongs. For instance, wheat and grain maize in one broad output category 'cereals' share the same technology while dry pulses and oil-seeds in another broad output category 'dry pulses & oilseeds' share another technology. Producing wheat cannot directly affect producing dry pulses through transformation effects, only indirectly if producing more wheat leads to producing more cereals and, hence, through transformation effects less dry pulses & oilseeds and, in turn, less dry pulses. If the

marginal rate of transformation between two outputs in one broad output category is independent of any other output outside of that broad output category, then the production possibility function is said to be weakly separable in partition (Berndt and Christensen, 1973).

Similarly, this separability assumption is acceptable to the extent that inputs having a strong substitution among them are grouped together in the same broad input category such that inputs are similar in technico-economic characteristics within the same broad input category (Sato, 1967). This implies that the use of one input belonging to a broad input category cannot directly affect the use of another input that belongs to another broad input category. It can only affect indirectly the use of this another input through the use of the broad input category to which it belongs. For instance, wages and contract work in one broad input category 'services' are more similar in technico-economic characteristics than the broad input categories 'services' and 'other intermediate inputs'. Using contract work cannot directly affect the use of inputs belonging to the broad input category 'intermediate inputs' through substitution effects, only indirectly if using more contract work leads to using more services and, hence, through substitution effects less intermediate inputs and, in turn, less inputs in that category. If the marginal rate of substitution between two inputs in one broad input category is independent of any other input outside of that broad input category, then the production function is said to be weakly separable in partition (Berndt and Christensen, 1973).

First, we describe the method for estimating a nested cost function and its derived input demand functions. Second, we describe the method for estimating nested conditional input demand functions. The description is done in generic terms to make it applicable to any specific situation of the FADN.

1.2 The Nested Cost Function and its Derived Input Demand Functions

The two-level cost function is proposed as an analogy to the two-level expenditure function proposed in the consumption theory (Deaton and Muellbauer, 1999, p. 117-147). The cost function has two levels: the lower level and the upper level. Assume that each broad output category m in the upper-level branch has an aggregate of output sub-categories n in its lower level as subsets. First, we present the concept for estimating the cost and input demand functions at the upper level of outputs. Second, we present the concept for estimating the cost and input demand functions at the lower level of outputs. These two concepts take inputs at their upper level.

In contrast to surveys on consumption expenditures that provide expenditures on specific consumption items, the EU-FADN data set does not provide disaggregated total costs for specific output categories, neither the use of input categories for specific output categories. This project proposes and tests a method able to retrieve these missing disaggregated farm data.

1.2.1 Cost Functions for Upper-Level Outputs

Let total variable cost for farm f at time t be represented by

$$TC_{ft} = TC(w_{ft}, y_{ft}, t; z_{ft}; \alpha) + \varepsilon_{0;ft}, \quad (1.1)$$

for $y \geq 0$, with the usual theoretical properties (Chambers, 1988, p.52), where $w_{ft} = (w_{1;ft}, \dots, w_{J;ft})$ represents the vector of broad input category prices, $y_{ft} = (y_{1;ft}, \dots, y_{M;ft})$ the vector of broad output category quantities, $z_{ft} = (z_{1;ft}, \dots, z_{K;ft})$ the vector of quasi-fixed broad input category quantities, and $\varepsilon_{0;ft}$ an error term normally distributed.

The dependent variable is obtained as

$$TC_{ft} = \sum_{i=1}^J w_{i;ft} \cdot x_{i;ft}.$$

where $x_{ft} = (x_{1;ft}, \dots, x_{J;ft})$ represents the vector of broad input category quantities.

Based on the cost function (1.1), cost minimization implies the following system of broad input demand equations

$$x_{i;ft} = x_i(w_{ft}, y_{ft}, t; z_{ft}; \alpha) + \varepsilon_{i;ft} \quad (1.2)$$

where $\varepsilon_{i;ft}$ represents an error term normally distributed.

By Shephard's lemma (Chambers, 1988, p.56),¹ it holds that

$$x_i(w, y, t; z; \alpha) = \frac{\partial TC(w, y, t; z; \alpha)}{\partial w_i},$$

for $x_i > 0$.

The system of broad input demand equations (1.2) is used to estimate the vector of parameters α . Estimated total cost $\widehat{TC}$ and estimated demands for a broad input

¹As an important result of the envelope theorem, the Shephard's lemma states that, if the cost function is differentiable in input prices, then there exists a unique vector of cost-minimizing input demands that is equal to the gradient of the cost function in input prices.

category $\hat{x}_i$ are generated as

$$\begin{aligned}\widehat{TC}_{ft} &= TC(w_{ft}, y_{ft}, t; z_{ft}; \hat{\alpha}) \\ \hat{x}_{i,ft} &= x_i(w_{ft}, y_{ft}, t; z_{ft}; \hat{\alpha})\end{aligned}\tag{1.3}$$

subject to

$$\begin{aligned}y &\geq 0 \\ \hat{x}_i &= x_i(w, y, t; z; \hat{\alpha}) \quad \text{if} \quad [(\hat{x}_i > 0) \quad \text{and} \quad (x_i > 0)] \\ \hat{x}_i &= 0 \quad \text{if} \quad \text{not} \quad [(\hat{x}_i > 0) \quad \text{and} \quad (x_i > 0)]\end{aligned}$$

implying that

$$\widehat{TC}_{ft} \geq \sum_{i=1}^J w_{i,ft} \cdot \hat{x}_{i,ft}.$$

Leaving aside indexes f and t for clarity, the marginal cost function for broad output category m is defined as

$$MC_m(w, y, t; z; \alpha) = \frac{\partial TC(w, y, t; z; \alpha)}{\partial y_m}.\tag{1.4}$$

Estimated marginal costs for a broad output category $\widehat{MC}_m$ are generated as

$$\widehat{MC}_{m;ft} = MC_{m;ft}(w_{ft}, y_{ft}, t; z_{ft}; \hat{\alpha}).$$

From the cost function (1.1), it is then possible to obtain pseudo-observations of the total and average variable cost of broad output category m as the following.

Leaving aside indexes f and t for clarity, let us now, *without loss of generality*, represent the estimated total cost function $\widehat{TC}$ as

$$\begin{aligned}\widehat{TC}(w, y, t; z; \hat{\alpha}) &= \sum_m f_m(y_m) + \sum_m \sum_{m' < m} f_{m,m'}(y_m, y_{m'}) \\ &\quad + \sum_m \sum_{m' < m} \sum_{m'' < m'} f_{m,m',m''}(y_m, y_{m'}, y_{m''}) \\ &\quad + \dots + f_{1,2,\dots,M}(y_1, \dots, y_M),\end{aligned}\tag{1.5}$$

i.e., the sum of additive components, such that $f_m(y_m)$ is only a function of one output, $f_{m,m'}(y_m, y_{m'})$ is a function of two outputs only, etc. Note that on the right hand side all arguments except broad output category quantities y are omitted for clarity.

We propose to define the estimated total variable cost of output category m at

time t , as²

$$\begin{aligned}\widehat{TC}_m(w, y, t; z; \hat{\alpha}) &= f_m(y_m) + \sum_{m' \neq m} \tau_{m;m,m'} \cdot f_{m,m'}(y_m, y_{m'}) \\ &+ \sum_{m' \neq m} \sum_{m'' < m'} \tau_{m;m,m',m''} \cdot f_{m,m',m''}(y_m, y_{m'}, y_{m''}) \\ &+ \dots + \tau_{m;1,2,\dots,M} \cdot f_{1,2,\dots,M}(y_1, \dots, y_M),\end{aligned}\quad (1.6)$$

where the following cross-equation restrictions apply

$$\begin{aligned}\sum_{k=m,m'} \tau_{k;m,m'} &= 1, \quad \forall m, m'; m' \neq m \\ \sum_{k=m,m',m''} \tau_{k;m,m',m''} &= 1, \quad \forall m, m', m''; m'' \neq m' \neq m \\ &\vdots \\ \sum_{k=1}^M \tau_{k;1,2,\dots,M} &= 1,\end{aligned}\quad (1.7)$$

which ensure that

$$\sum_{m=1}^M \widehat{TC}_m = \widehat{TC}.$$

The coefficients τ distribute the non-additive terms of the estimated cost function over the relevant broad output categories m . The above restrictions are fulfilled by the following weights

$$\begin{aligned}\tau_{k;m,m'} &= \frac{q_k}{\sum_{k=m,m'} q_k}, \quad k = m, m' \\ \tau_{k;m,m',m''} &= \frac{q_k}{\sum_{k=m,m',m''} q_k}, \quad k = m, m', m'' \\ &\vdots \\ \tau_{k;1,2,\dots,M} &= \frac{q_k}{\sum_{k=1}^M q_k}, \quad k = 1, \dots, M\end{aligned}\quad (1.8)$$

²As an alternative, we could have defined the estimated total variable cost of output category m at time t , keeping the other outputs at their observed level,

$$\widetilde{TC}_m(w, y, t; z; \hat{\alpha}) = \int_0^{y_m} \widehat{MC}_m(w, y, t; z; \hat{\alpha}) \Big|_{y_m=u} du.$$

However, since $\widehat{TC}$ is not additively separable in outputs y_m , we have that

$$\widehat{TC} \neq \sum_m \widetilde{TC}_m.$$

with natural candidates for q_m being 1, y_m or $f_m(y_m)$ if all such terms are present in the estimated cost function (1.5). We choose the former to preserve the symmetry restrictions on cost functions.³

The average cost function for broad output category m can now be derived straightforward

$$AC_m(w, y, t; z; \alpha) = \frac{TC_m(w, y, t; z; \alpha)}{y_m}. \quad (1.9)$$

Pseudo-observations for the total variable cost of broad category output m are now generated as

$$\widehat{TC}_{m;ft} = TC_m(w_{ft}, y_{ft}, t; z_{ft}; \hat{\alpha}). \quad (1.10)$$

Pseudo-observations for the average variable cost of broad category output m are now generated as

$$\widehat{AC}_{m;ft} = AC(w_{ft}, y_{ft}, t; z_{ft}; \hat{\alpha}). \quad (1.11)$$

An empirical verification on preliminary results from estimations of total cost functions over a panel of Belgian crop, dairy and livestock farms (1990-2008) shows that the sum of the estimated total variable cost of output category m , i.e., $\widehat{TC}_{m;ft}$, is equal to the estimated total variable cost, i.e., $\widehat{TC}_{ft}$

$$\widehat{TC}_{ft} = \sum_{m=1}^M \widehat{TC}_{m;ft}. \quad (1.12)$$

Note, however, that TC_m and AC_m from (1.10) and (1.11) respectively are fictitious total and average variable costs of broad output category m since TC_m and AC_m are valid given that the other broad output categories m are at their observed level.

Likewise, the marginal broad input demand function describes the amount of broad input category i that is allocated to broad output category m at the margin

³The symmetry restriction on cost functions imposes q_m be 1 to preserve such relationship (see Section 7.4)

$$\frac{\partial \widehat{MC}_m}{\partial y_{m'}} = \frac{\partial \widehat{MC}_{m'}}{\partial y_m} \quad \forall m, m'; m' \neq m.$$

of production (Beattie et al., 2009, p.132). It is given by

$$\begin{aligned}
 mx_{i;m}(w, y, t; z; \alpha) &= \frac{\partial x_i(w, y, t; z; \alpha)}{\partial y_m} \\
 &= \frac{\partial^2 TC(w, y, t; z; \alpha)}{\partial y_m \partial w_i} \\
 &= \frac{\partial MC_m(w, y, t; z; \alpha)}{\partial w_i}.
 \end{aligned} \tag{1.13}$$

Similarly as the estimated total cost function equation (1.5), leaving aside indexes f and t for clarity, represent the estimated input demand function $\hat{x}_i$ as a sum of additive components, such that $g_m(y_m)$ is only a function of one output, $g_{m,m'}(y_m, y_{m'})$ is a function of two outputs only, etc. Proceeding the same way, we obtain the estimated demand for broad input category i that can be allocated to broad output category m at time t , as expression (1.6)⁴

$$\begin{aligned}
 \hat{x}_{i;m}(w, y, t; z; \hat{\alpha}) &= g_m(y_m) + \sum_{m' \neq m} \tau_{m;m,m'} \cdot g_{m,m'}(y_m, y_{m'}) \\
 &+ \sum_{m' \neq m} \sum_{m'' < m'} \tau_{m;m,m',m''} \cdot g_{m,m',m''}(y_m, y_{m'}, y_{m''}) \\
 &+ \dots + \tau_{m;1,2,\dots,M} \cdot g_{1,2,\dots,M}(y_1, \dots, y_M),
 \end{aligned} \tag{1.14}$$

subject to

$$\begin{aligned}
 \hat{x}_{i;m} &= \hat{x}_{i;m}(w, y, t; z; \hat{\alpha}) \quad \text{if} \quad (\hat{x}_{i;m} > 0) \\
 \hat{x}_{i;m} &= 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i;m} > 0).
 \end{aligned}$$

The coefficients τ distribute the non-additive terms of the estimated input demand function over the relevant broad output categories m using the same weights as expressions (1.8) since it can be shown that $\hat{x}_{i;m}(w, y, t; z; \hat{\alpha})$ obtained from (1.14)

⁴As an alternative, we could have defined the estimated demand for broad input category i that can be allocated to broad output category m at time t , keeping the other outputs at their observed level,

$$\tilde{x}_{i;m}(w, y, t; z; \hat{\alpha}) = \int_0^{y_m} \widehat{m}x_{i;m}(w, y, t; z; \hat{\alpha})|_{y_m=u} du$$

However, since $\hat{x}_i$ is not additively separable in outputs y_m , we have that

$$\hat{x}_i \neq \sum_m \tilde{x}_{i;m}.$$

or derived from $\widehat{TC}_m(w, y, t; z; \hat{\alpha})$ (1.6) must have the same value.⁵

Pseudo-observations for the demand for broad input category i that can be allocated to broad output category m are generated as

$$\hat{x}_{i;m;ft} = x_{i,m}(w_{ft}, y_{ft}, t; z_{ft}; \hat{\alpha}). \quad (1.15)$$

As expected, mixes of broad input categories $x_{i;m}$ per broad output category m depend on relative prices of broad input category w_i and the levels of the broad output categories y_m . These input mixes can be different depending on the broad output category y_m .

An empirical verification on preliminary results from estimations of total cost functions over a panel of Belgian crop, dairy and livestock farms (1990-2008) shows that the sum of the calculated demands for a broad input category i for a broad output category m , i.e., $\hat{x}_{i;m;ft}$, is equal to the estimated total demand for that broad input i , i.e., $\hat{x}_{i;ft}$

$$\hat{x}_{i;ft} = \sum_{m=1}^M \hat{x}_{i;m;ft} \quad (1.16)$$

at the condition that the following restrictions are removed

$$\hat{x}_{i;m} = \hat{x}_{i;m}(w, y, t; z; \hat{\alpha}) \quad \text{if} \quad (\hat{x}_{i;m} > 0)$$

$$\hat{x}_{i;m} = 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i;m} > 0).$$

Again, note, however, that $x_{i;m}$ from (1.15) is a fictitious demand for a broad input category i for a broad output category m since $x_{i;m}$ is valid given that the other broad output categories m are at their observed level.

The cost for broad input category i that can be allocated to broad output category m is given by $(w_i \cdot x_{i;m})$, which implies a predicted unit cost of

$$\widehat{uc}_{i;m} = \frac{w_i \cdot \hat{x}_{i;m}}{y_m} \quad (1.17)$$

that varies according to y_m for a cost function that is non linear in y_m , and the total

⁵Note that the following comparative statics property of derived input demands is preserved (Chambers, 1988, p.262)

$$\frac{\partial x_{i;m}}{\partial y_m} = \frac{\partial MC_m}{\partial w_i} \quad \forall i, m.$$

variable cost of broad output category m can also be predicted as

$$\widehat{TC}_{m;ft} = \sum_{i=1}^J w_{i;ft} \cdot \hat{x}_{i;m;ft}. \quad (1.18)$$

Note that it is possible to estimate $\widehat{TC}_{m;ft}$ either through (1.10) or (1.18).

1.2.2 Cost Functions for Lower-Level Outputs

Now, consider the vector of output sub-category quantities $\tilde{y}_{m;ft} = (y_{m,1;ft}, \dots, y_{m,N_m;ft})$, for which it holds that

$$y_{m;ft} = \sum_{n=1}^{N_m} y_{m,n;ft}$$

and assume that the estimated total variable cost of broad output category m is a function of broad input category prices and of output sub-category quantities $y_{m,n}$ using the separability assumption that the total cost function for one particular broad output category m does not depend on the level of the other broad output categories $m' \neq m$

$$\widehat{TC}_{m;ft} = TC_m(w_{ft}, \tilde{y}_{m;ft}, t; z_{ft}; \beta_m) + \eta_{0;m;ft} \quad (1.19)$$

where $\eta_{0;m;ft}$ represents an error term normally distributed.

Then, for each broad output category m , we derive the system of broad input demand equations

$$\hat{x}_{i;m;ft} = x_{i;m}(w_{ft}, \tilde{y}_{m;ft}, t; z_{ft}; \beta_m) + \eta_{i;m;ft} \quad (1.20)$$

where $\eta_{i;m;ft}$ represents an error term normally distributed and where, by Shephard's lemma (Chambers, 1988, p.56), it holds that

$$x_{i;m}(w, \tilde{y}_m, t; z; \beta_m) = \frac{\partial TC_m(w, \tilde{y}_m, t; z; \beta_m)}{\partial w_i}$$

for $\hat{x}_{i;m} > 0$.

Estimation of β_m proceeds as above using the values of the pseudo-observations $\hat{x}_{i;m;ft}$ from (1.15). Estimated total cost of the broad output category $\widehat{TC}_m$ and estimated demands for a broad input category for that broad output category $\hat{x}_{i;m}$ are generated as

$$\begin{aligned}\widehat{\widehat{TC}}_{m;ft} &= TC_m(w_{ft}, y_{ft}, t; z_{ft}; \hat{\beta}_m) \\ \hat{x}_{i;m;ft} &= x_{i;m}(w_{ft}, y_{ft}, t; z_{ft}; \hat{\beta}_m)\end{aligned}\tag{1.21}$$

subject to

$$\begin{aligned}\hat{x}_{i;m} &= \hat{x}_{i,m}(w, y, t; z; \hat{\beta}_m) \quad \text{if} \quad ([\hat{x}_{i;m} > 0] \quad \text{and} \quad (\hat{x}_{i;m} > 0)) \\ \hat{x}_{i;m} &= 0 \quad \text{if} \quad \text{not} \quad [(\hat{x}_{i;m} > 0) \quad \text{and} \quad (\hat{x}_{i;m} > 0)]\end{aligned}$$

implying that

$$\widehat{\widehat{TC}}_{m;ft} \geq \sum_{i=1}^J w_{i;ft} \cdot \hat{x}_{i;m;ft}.$$

If needed, the marginal cost $MC_{m,n}$ as in expression (1.4), the total variable cost $\widehat{TC}_{m,n}$ as in expression (1.6) and the average cost $AC_{m,n}$ as in expression (1.9) functions for output sub-category m, n of broad output category m can be defined as above. The same verifications on consistency between the aggregate and disaggregate quantities apply also here

$$\widehat{\widehat{TC}}_{m;f,t} = \sum_{n=1}^{N_m} \widehat{TC}_{m,n;ft}$$

where $\widehat{\widehat{TC}}_{m;f,t} = TC_m(w_{ft}, \tilde{y}_{m;ft}, t; z_{ft}; \hat{\beta}_m)$.

Again, the marginal broad input demand function describes the amount of broad input category i that is allocated to output sub-category m, n of broad output category m at the margin of production. It is given by

$$\begin{aligned}mx_{i;m,n}(w, \tilde{y}_m, t; z; \beta_m) &= \frac{\partial x_{i;m}(w, \tilde{y}_m, t; z; \beta_m)}{\partial y_{m,n}} \\ &= \frac{\partial^2 TC_m(w, \tilde{y}_m, t; z; \beta_m)}{\partial y_{m,n} \partial w_i} \\ &= \frac{\partial MC_{m,n}(w, \tilde{y}_m, t; z; \beta_m)}{\partial w_i}.\end{aligned}\tag{1.22}$$

Similarly as the estimated total cost function equation (1.5), represent the estimated input demand function $\hat{x}_{i;m}$ as a sum of additive components, such that $g_{m,n}(y_{m,n})$ is only a function of one output, $g_{mn,mn'}(y_{m,n}, y_{m,n'})$ is a function of two outputs only, etc. Proceeding the same way, we obtain the estimated demand for broad input category i that can be allocated to output sub-category m, n at time t ,

as expression (1.6)⁶

$$\begin{aligned}\hat{x}_{i;m,n} \left(w, \tilde{y}_m, t; z; \hat{\beta}_m \right) &= g_{m,n} (y_{m,n}) + \sum_{n' \neq n} \tau_{m,n;mn,mn'} \cdot g_{m,n;m,n'} (y_{m,n}, y_{m,n'}) \\ &+ \sum_{n' \neq n} \sum_{n'' < n'} \tau_{m,n;mn,mn',mn''} \cdot g_{m,n;m,n';m,n''} (y_{m,n}, y_{m,n'}, y_{m,n''}) \\ &+ \dots + \tau_{m,n;m1,m2,\dots,mN} \cdot g_{m,1;m,2,\dots,m,N} (y_{m,1}, \dots, y_{m,N})\end{aligned}\quad (1.23)$$

subject to

$$\begin{aligned}\hat{x}_{i;m,n} &= \hat{x}_{i;m,n} \left(w, \tilde{y}_m, t; z; \hat{\beta}_m \right) \quad \text{if} \quad (\hat{x}_{i;m,n} > 0) \\ \hat{x}_{i;m,n} &= 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i;m,n} > 0).\end{aligned}$$

The coefficients τ distribute the non-additive terms of the estimated input demand function over the relevant output sub-categories m, n using the same weights as expressions (1.8) but $q_{m,n}$ being 1, $y_{m,n}$ or $g_{m,n}(y_{m,n})$ since it can be shown that $\hat{x}_{i;m,n}(w, \tilde{y}_m, t; z; \hat{\beta}_m)$ obtained from 1.23 or derived from $\widehat{TC}_{m,n}(w, \tilde{y}_m, t; z; \hat{\beta}_m)$ must have the same value.⁷

Pseudo-observations for the demand for broad input category i that can be allocated to output sub-category m, n are generated as

$$\hat{x}_{i;m,n;ft} = x_{i;m,n} \left(w_{ft}, \tilde{y}_{m;ft}, t; z_{ft}; \hat{\beta}_m \right). \quad (1.24)$$

As expected, mixes of estimated broad input categories $x_{i;m,n}$ per output sub-category m, n depend on relative prices of broad input category w_i and the levels

⁶As an alternative, we could have defined the estimated demand for broad input category i that can be allocated to output sub-category m, n at time t , keeping the other outputs at their observed level,

$$\tilde{x}_{i;m,n} \left(w, \tilde{y}_m, t; z; \hat{\beta}_m \right) = \int_0^{y_{m,n}} \widehat{m}x_{i;m,n} \left(w, \tilde{y}_m, t; z; \hat{\beta}_m \right) \Big|_{y_{m,n}=u} du$$

However, since $\hat{x}_{i,m}$ is not additively separable in outputs $y_{m,n}$, we have that

$$\hat{x}_{i;m} \neq \sum_n \tilde{x}_{i;m,n}.$$

⁷Note that the following comparative statics property of derived input demands is preserved (Chambers, 1988, p.262)

$$\frac{\partial x_{i;m,n}}{\partial y_{m,n}} = \frac{\partial MC_{m,n}}{\partial w_i} \quad \forall i, m, n.$$

of the output sub-categories $y_{m,n}$. These input mixes can be different depending on the output sub-category $y_{m,n}$.

The same verifications on consistency between the aggregate and disaggregate quantities apply also here

$$\hat{x}_{i,m;ft} = \sum_{n=1}^{N_m} \hat{x}_{i,m,n;ft}$$

at the condition that the following restrictions are removed

$$\hat{x}_{i,m,n} = \hat{x}_{i,m,n}(w, \tilde{y}_m, t; z; \hat{\beta}_m) \quad \text{if} \quad (\hat{x}_{i,m,n} > 0)$$

$$\hat{x}_{i,m,n} = 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i,m,n} > 0).$$

Again, note, however, that $\hat{x}_{i,m,n}$ from (1.24) is a fictitious demand for a broad input category i for an output sub-category m, n since $\hat{x}_{i,m,n}$ is valid given that the other output sub-categories m, n are at their observed level.

The cost for broad input category i that can be allocated to output sub-category m, n is given by $(w_i \cdot x_{i,m,n})$, which implies a predicted unit cost of

$$\widehat{u}_{i,m,n} = \frac{w_i \cdot \hat{x}_{i,m,n}}{y_{m,n}} \quad (1.25)$$

that varies according to $y_{m,n}$ for a cost function that is non linear in $y_{m,n}$, and the total variable cost of output sub-category m, n can also be predicted as

$$\widehat{TC}_{m,n;ft} = \sum_{i=1}^J w_{i;ft} \cdot \hat{x}_{i,m,n;ft}. \quad (1.26)$$

1.2.3 A Remark about Theoretical Consistency between Cost Functions

Define $\mathbf{y}_{ft} = (\tilde{y}_{1;ft}, \dots, \tilde{y}_{M;ft})$, then total variable cost can be written as⁸

$$TC_{ft} = \mathbf{TC}(w_{ft}, \mathbf{y}_{ft}, t; z_{ft}; \boldsymbol{\alpha}). \quad (1.27)$$

Then, which restrictions does our procedure impose on $\mathbf{TC}(w, \mathbf{y}, t; z; \boldsymbol{\alpha})$? Are the marginal cost functions $\mathbf{MC}_{m,n}(w, \mathbf{y}, t; z; \hat{\boldsymbol{\alpha}})$ of this total cost identical to the ones obtained by our nested procedure? In other words, given $TC(w, y, t; z; \alpha)$ and $TC_m(w, \tilde{y}_m, t; z; \beta_m)$, what does $\mathbf{TC}(w, \mathbf{y}, t; z; \boldsymbol{\alpha})$ looks like and is it an ac-

⁸Bold letters are used to distinguish the function, variables and parameteres of expression (1.27) from those of expression (1.1).

ceptable form for a cost function? The difference is that $TC(w, y, t; z; \alpha)$ and $TC_m(w, \tilde{y}_m, t; z; \beta_m)$ are more restrictive than $\mathbf{TC}(w, \mathbf{y}, t; z; \boldsymbol{\alpha})$ considering the functional separability of the broad output categories that is embedded.

1.3 The Nested Input Demand Functions

The two-level production function was proposed by Sato (1967). The production function has two levels: the lower level and the upper level. Assume that each broad input category i in the upper-level branch has an aggregate of input sub-categories i, j in its lower level as subsets. First, we present the concept for estimating the input demand functions in its lower level conditional to the upper level of input. It is then possible to calculate the demands for input sub-categories i, j and broad output category m and, similarly, the demands for input sub-categories i, j and output sub-category m, n . Note that the estimation of the nested conditional input demand functions has no implication on the estimation of the nested cost functions.

1.3.1 Demand Functions for Lower-Level Inputs

Consider the vector of input sub-category quantities $\tilde{x}_{i,ft} = (x_{i,1,ft}, \dots, x_{i,J_i,ft})$ and the vector of input sub-category prices $\tilde{w}_{i,ft} = (w_{i,1,ft}, \dots, w_{i,J_i,ft})$, for which it holds that

$$\begin{aligned} x_{i,ft} &= \sum_{j=1}^{J_i} x_{i,j,ft} \\ E_{i,ft} &= \sum_{j=1}^{J_i} w_{i,j,ft} \cdot x_{i,j,ft} \end{aligned}$$

Expenditure minimization E_i on the use of input sub-category quantities $(x_{i,1,ft}, \dots, x_{i,J_i,ft})$ subject to input sub-category price levels $(w_{i,1,ft}, \dots, w_{i,J_i,ft})$ and an broad input category quantity level $x_{i,ft}$ that is function of $x_{i,ft}(x_{i,1,ft}, \dots, x_{i,J_i,ft}, t)$ as in Sato (1967)⁹ implies the following system of input sub-category equations for each broad input category i

$$x_{i,j,ft} = x_{i,j}(\tilde{w}_{i,ft}, x_{i,ft}, t; \gamma_i) + \mu_{i,j,ft} \quad (1.28)$$

where $\mu_{i,j,ft}$ represents an error term normally distributed, and the corresponding indirect expenditure function on broad input category $x_{i,ft}$

$$E_{i,ft} = E_i(\tilde{w}_{i,ft}, x_{i,ft}, t; \gamma_i) + \mu_{0,i,ft} \quad (1.29)$$

⁹Sato (1967) uses a constant elasticity substitution (CES) form.

for $x_{i,ft} \geq 0$, where $\mu_{0;i;ft}$ represents an error term normally distributed.

By Shephard's lemma, it holds that

$$x_{i,j}(\tilde{w}_i, x_i, t; \gamma_i) = \frac{\partial E_i(\tilde{w}_i, x_i, t; \gamma_i)}{\partial w_{i,j}}$$

for $x_{i,j} > 0$.

The system of conditional input sub-category demand equations (1.28) is used to estimate the vector of parameters γ_i . Estimated expenditure for a broad input category $\hat{E}_i$ and estimated demands for input sub-category $\hat{x}_{i,j}$ are generated as

$$\begin{aligned}\hat{E}_{i,ft} &= E_i(\tilde{w}_{i,ft}, x_{i,ft}, t; \hat{\gamma}_i) \\ \hat{x}_{i,j;ft} &= x_{i,j}(\tilde{w}_{i,ft}, x_{i,ft}, t; \hat{\gamma}_i)\end{aligned}\tag{1.30}$$

subject to

$$x_i \geq 0$$

$$\hat{x}_{i,j} = x_{i,j}(\tilde{w}_i, x_i, t; \hat{\gamma}_i) \quad \text{if} \quad [(\hat{x}_{i,j} > 0) \quad \text{and} \quad (x_{i,j} > 0)]$$

$$\hat{x}_{i,j} = 0 \quad \text{if} \quad \text{not} \quad [(\hat{x}_{i,j} > 0) \quad \text{and} \quad (x_{i,j} > 0)]$$

implying that

$$\hat{E}_{i,ft} \geq \sum_{j=1}^J w_{i,j;ft} \cdot \hat{x}_{i,j;ft}.$$

As expected, mixes of estimated sub-category inputs $\hat{x}_{i,j}$ depend on relative sub-category input prices $w_{i,j}$ and the level of the broad input category x_i , which depends in turn on the levels of broad output categories y_m .

Note that

$$\sum_{j=1}^{J_i} \hat{x}_{i,j;ft} \neq \hat{x}_{i,ft}$$

but

$$\sum_{j=1}^{J_i} (\hat{x}_{i,j;ft} + \mu_{i,j;ft}) = \hat{x}_{i,ft} + \varepsilon_{i,ft}$$

since

$$\sum_{j=1}^{J_i} x_{i,j;ft} = x_{i,ft}.$$

Leaving aside indexes f and t for clarity, the marginal expenditure function for

broad input category i is defined here as

$$ME_i(\tilde{w}_i, x_i, t; \gamma_i) = \frac{\partial E_i(\tilde{w}_i, x_i, t; \gamma_i)}{\partial x_i}. \quad (1.31)$$

Estimated marginal expenditures for a broad input category $\widehat{ME}_i$ are generated as

$$\widehat{ME}_{i,ft} = ME_{i,ft}(\tilde{w}_{i,ft}, x_{i,ft}, t; \hat{\gamma}_i).$$

The average expenditure function for broad input category i is defined as

$$AE_i(\tilde{w}_i, x_i, t; \gamma_i) = \frac{E_i(\tilde{w}_i, x_i, t; \gamma_i)}{x_i}. \quad (1.32)$$

Estimated average expenditures for a broad input category $\widehat{AE}_i$ are generated as

$$\widehat{AE}_{i,ft} = AE(\tilde{w}_{i,ft}, x_{i,ft}, t; \hat{\gamma}_i).$$

1.3.2 Demand Functions of Lower-Level Inputs for Upper-Level Outputs

From the estimated input demand function (1.30), it is then possible to obtain pseudo-observations for the demand for input sub-category i, j that can be allocated to broad output category m as the following. But, first define this new predicted estimated demand function for input sub-category i, j as¹⁰

¹⁰Note that $\hat{\hat{x}}_{i,j,ft} \neq \hat{x}_{i,j,ft}$ since

$$\hat{\hat{x}}_{i,j,ft} = x_{i,j}(\tilde{w}_{i,ft}, \hat{x}_{i,ft}, t; \hat{\gamma}_i)$$

$$\hat{x}_{i,j,ft} = x_{i,j}(\tilde{w}_{i,ft}, x_{i,ft}, t; \hat{\gamma}_i)$$

where $x_{i,ft} = \hat{x}_{i,ft} + \epsilon_{i,ft}$.

Therefore,

$$\begin{aligned} \hat{\hat{x}}_{i,j} &= x_{i,j}(\tilde{w}_i, (x_i - \epsilon_i), t; \hat{\gamma}_i) \\ &= x_{i,j}(\tilde{w}_i, x_i, t; \hat{\gamma}_i) \\ &\quad + x_{i,j}(\tilde{w}_i, -\epsilon_i, t; \hat{\gamma}_i) \\ &\quad + x_{i,j}(x_i, \epsilon_i, t; \hat{\gamma}_i) \end{aligned}$$

and

$$\hat{\hat{x}}_{i,j} = \hat{x}_{i,j} + x_{i,j}(\tilde{w}_i, -\epsilon_i, t; \hat{\gamma}_i) + x_{i,j}(x_i, \epsilon_i, t; \hat{\gamma}_i).$$

$$\hat{x}_{i,j;ft} = x_{i,j}(\tilde{w}_{i,ft}, \hat{x}_i, t; \hat{\gamma}_i).$$

Leaving aside indexes f and t for clarity, let us now, *without loss of generality*, represent this new predicted demand function for input sub-category i, j as

$$\begin{aligned} \hat{x}_{i,j}(\tilde{w}_i, \hat{x}_i, t; \hat{\gamma}_i) &= \sum_m g_{i;m}(\hat{x}_{i;m}) + \sum_m \sum_{m' < m} g_{i;m,m'}(\hat{x}_{i;m}, \hat{x}_{i;m'}) \\ &+ \sum_m \sum_{m' < m} \sum_{m'' < m'} g_{i;m,m',m''}(\hat{x}_{i;m}, \hat{x}_{i;m'}, \hat{x}_{i;m''}) \\ &+ \dots + g_{i;1,2,\dots,M}(\hat{x}_{i;1}, \dots, \hat{x}_{i;M}), \end{aligned} \quad (1.33)$$

i.e., the sum of additive components, such that $g_{i;m}(\hat{x}_{i;m})$ is only a function of one broad input i allocated to output m , $g_{i;m,m'}(\hat{x}_{i;m}, \hat{x}_{i;m'})$ is a function of two broad inputs i allocated to outputs m and m' only, etc. since

$$\hat{x}_i = \sum_{m=1}^M \hat{x}_{i;m}$$

Note that on the right hand side all arguments except broad input category quantities $\hat{x}_i$ are omitted for clarity.

We propose to define the estimated demand for input sub-category i, j that can be allocated to output category m at time t , as ¹¹

$$\begin{aligned} \hat{x}_{i,j;m} &= g_{i;m}(\hat{x}_{i;m}) + \sum_{m' \neq m} v_{i,j;m;m,m'} \cdot g_{i;m,m'}(\hat{x}_{i;m}, \hat{x}_{i;m'}) \\ &+ \sum_{m' \neq m} \sum_{m'' < m'} v_{i,j;m;m,m',m''} \cdot g_{i;m,m',m''}(\hat{x}_{i;m}, \hat{x}_{i;m'}, \hat{x}_{i;m''}) \\ &+ \dots + v_{i,j;m;1,2,\dots,M} \cdot g_{i;1,2,\dots,M}(\hat{x}_{i;1}, \dots, \hat{x}_{i;M}) \end{aligned} \quad (1.34)$$

¹¹ As an alternative, we could have defined the estimated demand for input sub-category i, j that can be allocated to broad output category m at time t , keeping the other outputs at their observed level,

$$\begin{aligned} \tilde{x}_{i,j;m}(\tilde{w}_i, \hat{x}_{i,m}, t; \hat{\gamma}_i) &= \int_0^{y_m} \frac{\partial \hat{x}_{i,j}(\tilde{w}_i, \hat{x}_i, t; \hat{\gamma}_i)}{\partial \hat{x}_i} \frac{\partial \hat{x}_i(w, y, t; z; \hat{\alpha})}{\partial y_m} \Big|_{y_m=v} dv \\ &= \int_0^{\hat{x}_{i;m}} \hat{x}_{i,j}(\tilde{w}_i, \hat{x}_i, t; \hat{\gamma}_i) \Big|_{\hat{x}_i=v} dv \end{aligned}$$

However, since $\hat{x}_{i,j}$ is not additively separable in inputs $\hat{x}_{i;m}$, we have that

$$\hat{x}_{i,j} \neq \sum \tilde{x}_{i,j;m}.$$

for $\hat{x}_{i;m} > 0$, where the following cross-equation restrictions apply

$$\begin{aligned}
\sum_{k=m,m'} v_{i,j;k;m,m'} &= 1, \quad \forall m, m'; m' \neq m \\
\sum_{k=m,m',m''} v_{i,j;k;m,m',m''} &= 1, \quad \forall m, m', m''; m'' \neq m' \neq m \\
&\vdots \\
\sum_{k=1}^M v_{i,j;k;1,2,\dots,M} &= 1,
\end{aligned} \tag{1.35}$$

subject to

$$\begin{aligned}
\hat{x}_{i,j;m} &= \hat{x}_{i,j;m}(\tilde{w}_i, \hat{x}_i, t; \hat{\gamma}_i) \quad \text{if} \quad (\hat{x}_{i,j;m} > 0) \\
\hat{x}_{i,j;m} &= 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i,j;m} > 0).
\end{aligned}$$

The coefficients v distribute the non-additive terms of the estimated demand function for broad input category i for broad output categories m over the relevant broad input categories i, m . The above restrictions are fulfilled by the following weights

$$\begin{aligned}
v_{i,j;k;m,m'} &= \frac{u_k}{\sum_{k=m,m'} u_k}, \quad k = m, m' \\
v_{i,j;k;m,m',m''} &= \frac{u_k}{\sum_{k=m,m',m''} u_k}, \quad k = m, m', m'' \\
&\vdots \\
v_{i,j;k;1,2,\dots,M} &= \frac{u_k}{\sum_{k=1}^M u_k}, \quad k = 1, \dots, M
\end{aligned} \tag{1.36}$$

with natural candidates for u_m being 1, $\hat{x}_{i;m}$ or $g_{i;m}(\hat{x}_{i;m})$ if all such terms are present in the estimated demand function for input sub-category i, j (1.33). We choose the former to preserve the symmetry restrictions on expenditure functions (see footnote 3 of Chapter 1).¹²

Pseudo-observations for the demand for input sub-category i, j that can be allocated to broad output category m are generated as

$$\hat{x}_{i,j;m;ft} = x_{i,j;m}(\tilde{w}_{i;ft}, \tilde{\hat{x}}_{i;ft}, t; \hat{\gamma}_i) \tag{1.37}$$

¹²Note that the symmetry restriction on derived input demand functions is preserved due to the restrictions on the estimated coefficients shown in Section 7.4

$$\frac{\partial x_{i,j;m}}{\partial w_{i,j'}} = \frac{\partial x_{i,j';m}}{\partial w_{i,j'}} \quad \forall i, j, j', m; j' \neq j.$$

where the vector of broad input quantities $\tilde{x}_{i;ft} = (\hat{x}_{i;1;ft}, \dots, \hat{x}_{i;M;ft})$.

As expected, mixes of estimated sub-category inputs $\hat{x}_{i,j;m}$ per broad output category m depend on relative sub-category input prices $w_{i,j}$ and the level of the estimated broad input categories $\hat{x}_{i;m}$ allocated to the broad output category y_m . These input mixes can be different depending on the broad output category y_m .

The same verifications on consistency between the aggregate and disaggregate input quantities apply also here

$$\hat{x}_{i,j;ft} = \sum_{m=1}^M \hat{x}_{i,j;m;ft}$$

at the condition that the following restrictions are removed

$$\hat{x}_{i,j;m} = \hat{x}_{i,j;m}(\tilde{w}_i, \hat{x}_i, t; \hat{\gamma}_i) \quad \text{if} \quad (\hat{x}_{i,j;m} > 0)$$

$$\hat{x}_{i,j;m} = 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i,j;m} > 0).$$

Note, however, that $\hat{x}_{i,j;m}$ from (1.37) is a fictitious demand for an input sub-category i, j for a broad output category m since $\hat{x}_{i,j;m}$ is valid given that the other input sub-categories i, j are at their observed level.

The cost for input sub-category i, j that can be allocated to broad output category m is given by $(w_{i,j} \cdot x_{i,j;m})$, which implies a predicted unit cost of

$$\widehat{uc}_{i,j;m} = \frac{w_{i,j} \cdot \hat{x}_{i,j;m}}{y_m} \quad (1.38)$$

that varies according to y_m for a cost function that is non linear in y_m , and the variable expenditure on broad input category i for output category m can also be predicted as

$$\widehat{E}_{i;m;ft} = \sum_{j=1}^{J_i} w_{i,j;ft} \cdot \hat{x}_{i,j;m;ft}. \quad (1.39)$$

1.3.3 Demand Functions of Lower-Level Inputs for Lower-Level Outputs

Now, consider the vector of broad input category quantities for output sub-category quantities $\tilde{x}_{i;m;ft} = (\hat{x}_{i;m,1;ft}, \dots, \hat{x}_{i;m,N_m;ft})$ in addition to the vector of input sub-category prices $\tilde{w}_{i;ft} = (w_{i,1;ft}, \dots, w_{i,J_i;ft})$, for which it holds that

$$\begin{aligned}\hat{x}_{i;m;ft} &= \sum_{n=1}^{N_m} \hat{x}_{i;m;n;ft} \\ \hat{E}_{i;m;ft} &= \sum_{j=1}^{J_i} w_{i,j;ft} \cdot \hat{x}_{i,j;m;ft}\end{aligned}$$

for $\hat{x}_{i;m} > 0$

and assume that in the following system of input sub-category equations the estimated input sub-categories i, j for each broad input category i for each broad output category m is a function of input sub-category prices and of broad input category quantities $\hat{x}_{i;m,n}$

$$\hat{x}_{i,j;m;ft} = x_{i,j;m}(\tilde{w}_{i;ft}, \tilde{x}_{i;m;ft}, t; \delta_{i;m}) + \nu_{i,j;m;ft} \quad (1.40)$$

where $\nu_{i,j;m;ft}$ represents an error term normally distributed, and the corresponding indirect expenditure function on broad input category $x_{i;m;ft}$ for each broad output category m

$$\hat{E}_{i;m;ft} = E_{i;m}(\tilde{w}_{i;ft}, \tilde{x}_{i;m;ft}, t; \delta_{i;m}) + \nu_{0;i;m;ft} \quad (1.41)$$

or $\tilde{x}_{i;m} \geq 0$, where $\nu_{0;i;m;ft}$ represents an error term normally distributed.

By Shephard's lemma, it holds that

$$\hat{x}_{i,j;m}(\tilde{w}_i, \tilde{x}_{i;m}, t; \delta_{i;m}) = \frac{\partial \hat{E}_{i;m}(\tilde{w}_i, \tilde{x}_{i;m}, t; \delta_{i;m})}{\partial w_{i,j}}$$

for $\hat{x}_{i,j;m} > 0$.

Estimation of $\delta_{i;m}$ proceeds as above using the pseudo-observations of $\hat{x}_{i,j;m;ft}$ from (1.37). Estimated expenditure for a broad input category $\hat{E}_{i;m}$ and estimated demands for input sub-category $\hat{x}_{i,j;m}$ are generated as

$$\begin{aligned}\hat{E}_{i;m;ft} &= E_{i;m}(\tilde{w}_{i;ft}, \tilde{x}_{i;m;ft}, t; \hat{\delta}_{i;m}) \\ \hat{x}_{i,j;m;ft} &= x_{i,j;m}(\tilde{w}_{i;ft}, \tilde{x}_{i;m;ft}, t; \hat{\delta}_{i;m})\end{aligned} \quad (1.42)$$

subject to

$$\tilde{x}_{i;m} \geq 0$$

$$\hat{x}_{i,j;m} = x_{i,j;m}(\tilde{w}_i, \tilde{x}_{i;m}, t; \delta_{i;m}) \quad \text{if} \quad (\hat{x}_{i,j;m} > 0)$$

$$\hat{x}_{i,j;m} = 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i,j;m} > 0)$$

implying that

$$\hat{E}_{i;m;ft} \geq \sum_{i=1}^J w_{i,j;ft} \cdot \hat{x}_{i,j;m;ft}.$$

From the estimated input demand function (1.42), it is then possible to obtain pseudo-observations for the demand for input sub-category i, j that can be allocated to output sub-category m, n as the following.

Similarly as the estimated demand function equation (1.33), leaving aside indexes f and t for clarity, represent the estimated demand function $\hat{x}_{i,j;m}$ as a sum of additive components, such that $g_{i;m,n}(\hat{x}_{i;m,n})$ is only a function of one input, $g_{i;m,n;m,n'}(\hat{x}_{i;m,n}, \hat{x}_{i;m,n'})$ is a function of two inputs only, etc. Proceeding the same way, we obtain the estimated demand for input sub-category i, j that can be allocated to output sub-category m, n at time t , as expression (1.34)¹³

$$\begin{aligned} \hat{x}_{i,j;m,n} &= g_{i;m,n}(\hat{x}_{i;m,n}) + \sum_{m' \neq m} v_{i,j;m,n;mn,mn'} \cdot g_{i;mn,mn'}(\hat{x}_{i;m,n}, \hat{x}_{i;m,n'}) \\ &+ \sum_{n' \neq n} \sum_{n'' < n'} v_{i,j;m,n;mn,mn',mn''} \cdot g_{i;mn,mn',mn''}(\hat{x}_{i;m,n}, \hat{x}_{i;m,n'}, \hat{x}_{i;m,n''}) \\ &+ \dots + v_{i,j;m,n;m1,m2,\dots,mN} \cdot g_{i;m1,m2,\dots,mN}(\hat{x}_{i;m,1}, \dots, \hat{x}_{i;m,N}) \end{aligned} \quad (1.43)$$

for $\hat{x}_{i;m,n} > 0$, where the cross-equations restrictions (1.35) apply, subject to

$$\hat{x}_{i,j;m,n} = \hat{x}_{i,j;m,n}(\tilde{w}_i, \tilde{x}_{i;m}, t; \delta_{i;m}) \quad \text{if} \quad (\hat{x}_{i,j;m,n} > 0)$$

$$\hat{x}_{i,j;m,n} = 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i,j;m,n} > 0).$$

The coefficients v distribute the non-additive terms of the estimated input demand function over the relevant output sub-categories m, n using the same weights as expressions (1.36) but $u_{m,n}$ being 1, $\hat{x}_{i;m,n}$ or $g_{i;m,n}(\hat{x}_{i;m,n})$. We choose the former to preserve the symmetry restrictions on expenditure functions (see footnote 3 of Chapter 1).¹⁴

¹³As an alternative, we could have defined the estimated demand for input sub-category i, j that can be allocated to output sub-category m, n at time t , keeping the other outputs at their observed level,

$$\begin{aligned} \tilde{x}_{i,j;m,n}(\tilde{w}_i, \hat{x}_{i;m,n}, t; \hat{\gamma}_i) &= \int_0^{y_{m,n}} \frac{\partial \hat{x}_{i,j;m}(\tilde{w}_i, \hat{x}_{i;m}, t; \hat{\gamma}_i)}{\partial \hat{x}_{i;m}} \frac{\partial \hat{x}_{i;m}(w, \tilde{y}_m, t; z; \hat{\beta}_m)}{\partial y_{m,n}} \bigg|_{y_{m,n}=v} dv \\ &= \int_0^{\hat{x}_{i;m,n}} \hat{x}_{i,j;m}(\tilde{w}_i, \hat{x}_{i;m}, t; \hat{\gamma}_i) \big|_{\hat{x}_{i;m}=v} dv \end{aligned}$$

However, since $\hat{x}_{i,j;m}$ is not additively separable in inputs $\hat{x}_{i;m,n}$, we have that

$$\hat{x}_{i,j;m} \neq \sum_{n=1}^{N_m} \tilde{x}_{i,j;m,n}.$$

¹⁴Again, note that the symmetry restriction on derived input demand functions is preserved due to the restrictions on the estimated coefficients shown in Section 7.4

Pseudo-observations for the demand for input sub-category i, j that can be allocated to output sub-category m, n are generated as

$$\hat{x}_{i,j;m,n;ft} = x_{i,j;m,n}(\tilde{w}_i, \tilde{x}_{i;m}, t; \hat{\delta}_{i;m}). \quad (1.44)$$

As expected, mixes of estimated sub-category inputs $x_{i,j;m,n}$ per output sub-category m, n depend on relative sub-category input prices $w_{i,j}$ and the level of the estimated broad input categories $\hat{x}_{i;m,n}$ allocated to the output sub-category $y_{m,n}$. These input mixes can be different depending on the output sub-category $y_{m,n}$.

The same verifications on consistency between the aggregate and disaggregate quantities apply also here

$$\hat{x}_{i,j;m,ft} = \sum_{n=1}^{N_m} \hat{x}_{i,j;m,n;ft}$$

at the condition that the following restrictions are removed

$$\hat{x}_{i,j;m,n} = \hat{x}_{i,j;m,n}(\tilde{w}_i, \tilde{x}_{i;m}, t; \delta_{i;m}) \quad \text{if} \quad (\hat{x}_{i,j;m,n} > 0)$$

$$\hat{x}_{i,j;m,n} = 0 \quad \text{if} \quad \text{not} \quad (\hat{x}_{i,j;m,n} > 0).$$

Again, note, however, that $\hat{x}_{i,j;m,n}$ from (1.44) is a fictitious demand for an input sub-category i, j for an output sub-category m, n since $\hat{x}_{i,j;m,n}$ is valid given that the other input sub-categories i, j are at their observed level.

The cost for input sub-category i, j that can be allocated to output sub-category m, n is given by $(w_{i,j} \cdot x_{i,j;m,n})$, which implies a predicted unit cost of

$$\widehat{uc}_{i,j;m,n} = \frac{w_{i,j} \cdot \hat{x}_{i,j;m,n}}{y_{m,n}} \quad (1.45)$$

that varies according to $y_{m,n}$ for a cost function that is non linear in $y_{m,n}$, and the variable expenditure on broad input category i for output sub-category m, n can also be predicted as

$$\hat{E}_{i;m,n;ft} = \sum_{j=1}^{J_i} w_{i,j;ft} \cdot \hat{x}_{i,j;m,n;ft}. \quad (1.46)$$

$$\frac{\partial x_{i,j;m,n}}{\partial w_{i,j'}} = \frac{\partial x_{i,j';m,n}}{\partial w_{i,j'}} \quad \forall i, j, j', m, n; j' \neq j.$$

1.3.4 A Remark about Theoretical Consistency between Input Demand Functions

Define $\mathbf{x}_{ft} = (\tilde{x}_{1;ft}, \dots, \tilde{x}_{J;ft})$ and $\mathbf{w}_{ft} = (\tilde{w}_{1;ft}, \dots, \tilde{w}_{J;ft})$, for which it holds that

$$\mathbf{x}_{ft} = \sum_{i=1}^J \sum_{j=1}^{J_i} x_{i,j;ft},$$

then total expenditure on inputs can be written as¹⁵

$$\mathbf{E}_{ft} = \mathbf{E}(\mathbf{w}_{ft}, \mathbf{x}_{ft}, t; \boldsymbol{\gamma}). \quad (1.47)$$

Then, which restrictions does our procedure impose on $\mathbf{E}(\mathbf{w}, \mathbf{x}, t; \boldsymbol{\gamma})$? Are the input demand functions $\mathbf{x}_{i,j}(\mathbf{w}, \mathbf{x}, t; \hat{\boldsymbol{\gamma}})$ of this total expenditure identical to the ones obtained by our nested procedure? In other words, given $E_i(\tilde{w}_i, x_i, t; \gamma_i)$ and $x_{i,j}(\tilde{w}_i, x_i, t; \gamma_i)$, what does $\mathbf{E}(\mathbf{w}, \mathbf{x}, t; \boldsymbol{\gamma})$ looks like and is it an acceptable form for an expenditure function? The difference is that $E_i(\tilde{w}_i, x_i, t; \gamma_i)$ is more restrictive than $\mathbf{E}(\mathbf{w}, \mathbf{x}, t; \boldsymbol{\gamma})$ considering the functional separability of the broad input categories that is embedded.

1.3.5 A Remark about Functional Form Consistency between the Cost and Expenditure Functions

To what extent the choice of the functional form of the indirect input expenditure functions needs to be consistent with the choice of the functional form of the cost function? For instance, if the SGM functional form is used to specify the cost function, what would be a consistent functional form for specifying the indirect input expenditure functions? It is not necessary to select the same functional form for specifying both the cost function and the indirect input expenditure functions to the extent that both forms meet the theoretical restrictions. But for facilitating the coding we use the same functional form applying the same theoretical restrictions.

1.4 Conclusions

This section provides a concept for estimating nested cost and input demand functions. Thanks to this concept it is possible to estimate total variable cost and input demand functions at different levels of aggregation:

¹⁵Bold letters are used to distinguish the function, variables and parameters of expression (1.47) from those of expression (1.29).

1. at the level of output and input broad categories,
2. at the level of output sub-categories and input broad categories,
3. at the level of output broad categories and input sub-categories,
4. at the level of output and input sub-categories.

From these functions, it is possible to derive:

1. the marginal and average cost functions at the level of output broad and sub-categories,
2. the marginal input demand functions at the level of output broad and sub-categories,
3. the unit cost per input at the level of output and input broad and sub-categories.

It is, however, necessary to use a procedure that distributes non-additive terms over the relevant input and output categories to obtain consistency between aggregate and disaggregate input and output quantities.

Chapter 2

Data Preparation and Aggregation Scheme

2.1 Introduction

In this part, we present the organization of the Stata programme. The routine is divided in three sections, each section being dedicated to a specific task.

The first section consists of preparing the data in order to have all needed variables available and ready for the next sections. This includes the verification of data sets integrity (i.e., detecting missing variables and verifying data coherence), the generation of necessary input and output variables, the imputation of missing data, the importation of prices and the construction of indices using both EU-FADN and international data bases¹

The second section is dedicated to the settings and options necessary to obtain a given estimation. In particular, the user is invited to choose the type of farm (TF), the time specification (TM) and the aggregation schemes for inputs and outputs. The routine then generates automatically an adapted data set for further estimations.

Using the previously generated data set, the third and last section is dedicated to estimations using the standard and the augmented Symmetric Generalized McFadden (SGM) cost function. Figure 2.1 provides a summary of the global Stata routine structure, where the resulting output for each section is shown as a typical *.dta* file. The next sections give more details about these three specific tasks.

¹We resort to international data bases when prices and interest rates are missing in the EU-FADN dataset. First, we use Eurostat for missing prices and interest rates. If prices and interest rates are still not available in Eurostat, we use OECD, if not the Penn World Table.

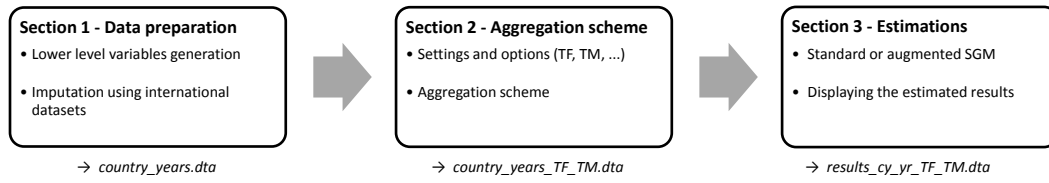


Figure 2.1: Schematic view of the Stata routine.

2.2 Data Preparation

This first section of the Stata routine is pre-processing the data in order to have the EU-FADN data sets ready for further manipulations. No specific options or settings are available to the user in this part, except for the choice of the member state or region and the year range to include in the data set.

The routine is here divided into two main operations. The first operation consists in constructing every needed variables at the lowest level of aggregation using the EU-FADN data set. Variables are classified into three categories: fixed inputs, variable inputs and outputs. It is important to notice that the choice of the aggregation at the lowest level of aggregation of the variables is an *ex ante* decision, i.e., once the variable aggregation at the lowest level is decided, the present section of the Stata routine generates the needed aggregated variables at that level, while any other variable at a more disaggregated level is not accessible any more for further estimations. For all these variables, we also construct the farm prices and, when missing, we impute then with the regional prices.² The second operation consists in importing data from Eurostat, the OECD and the Penn World Table (PWT) in order to impute the remaining missing prices and interest rates. During the two described operations, a special attention is given to the construction of variables adapted to each time specification (medium- and long-term), i.e., input variables that are either fixed inputs or variable inputs depending on the time specification.

At the end of this routine, a complete data set including all input and output values and related prices and interest rates is stored in a data file.

²When imputing with the regional prices, the most disaggregated available regional level is chosen. The scheme for regional levels depends on the availability of regional variables in the EU-FADN dataset.

2.3 Aggregation Scheme

This second section of the routine allows the user to choose between different options to make the last data adjustments before estimations. Important options concern the choice of the farm type, the time specification and the resulting aggregation scheme for fixed inputs, variable inputs and outputs. The farm type choice allows to select the most relevant farms in the data set, while the choice between the medium- and long-term specifications has an influence on the input aggregation scheme. After choosing the farm type, the user has two possibilities to determine the aggregation scheme. Either, she lets the aggregation scheme to be automatically built by the Stata routine using aggregation schemes that are implemented for each farm type by default, or she chooses the farm type and customizes the aggregation scheme. In that latter case, the user should care to design the scheme in a way that is relevant for the chosen farm type. Once the aggregation scheme is decided either automatically or customized, the routine builds the upper level of variable aggregations following that scheme, i.e., it builds new aggregated variables.

We want to emphasize the fact that flexibility resides in the customization of the aggregation scheme. However, as we already mentioned, the only variables that can be used in this aggregation scheme are the ones that have been chosen *ex ante* and prepared in the first section of the programme.

Figures 2.2, 2.3, 2.4 and 2.5 respectively show the aggregation schemes adopted for fixed inputs for the medium- and long-term specifications, variable inputs for the medium-term specification, variable inputs for the long-term specification and outputs for sale. In each figure, the dotted line marks the limit of disaggregation. Any variable that is below the line is not accessible any more for estimations, while the variables just above the line represent the most disaggregated variables, i.e., those chosen *ex ante* and constructed in the section for data preparation. The first line of variables above the dotted line include the variable that are considered at the lower lever of aggregation while the second line of variables above the dotted line include the variables that are considered at the upper level of aggregation.

As reported in Figure 2.2 for the medium term, we consider three fixed inputs as being agricultural area, non-land capital and unpaid labour input while, for the long term, we only consider one fixed input as being unpaid labour input. Accordingly, as reported in Figure 2.4 for the long term, agricultural area and non-land capital become variable inputs (see Chapter 6).

Note carefully that there is no need to single out explicitly on-farm forage crops that are used to feed on-farm animals. Inputs to produce those on-farm forage crops are already counted in the different input categories. On-farm forage crops that are

used to feed on-farm animals and, hence, not for sale are intermediate farm inputs. Note in parallel that outputs are only those sold outside the farm. There is no need then to figure out how much inputs are used to grow those on-farm forage crops and how much on-farm crops are produced. Note finally that animal products are denominated in terms of either live animals or dairy products since farms sell those products, nothing else.

At the end of this routine, a data set with only the needed variables for estimations is generated, accounting for the chosen type of farm, the time specification and the resulting aggregation scheme.

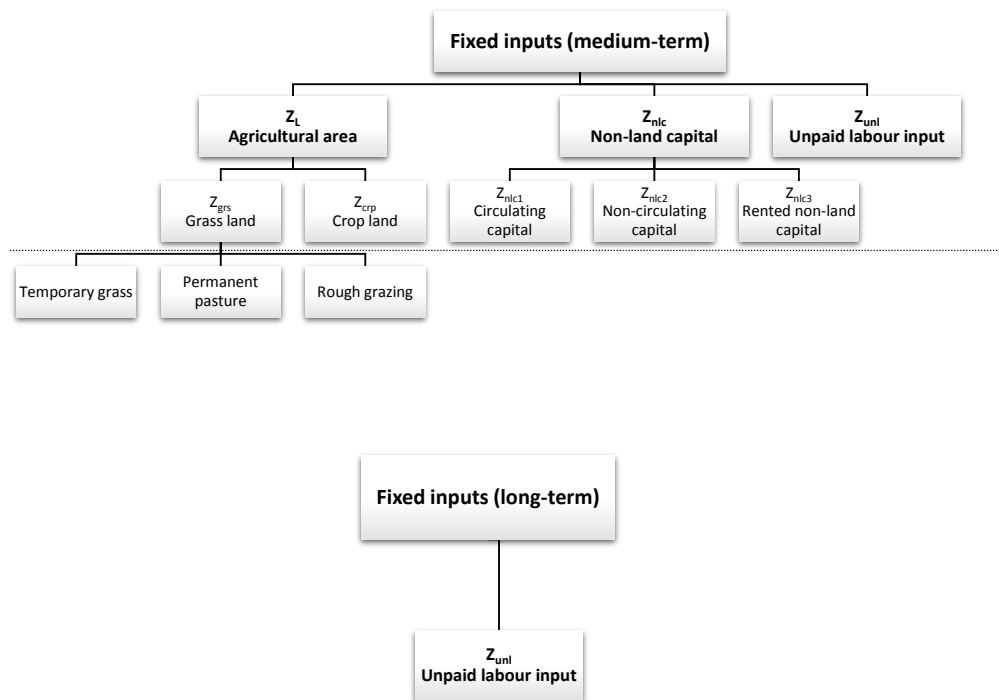


Figure 2.2: Aggregation scheme for fixed inputs (medium- and long-term specifications).

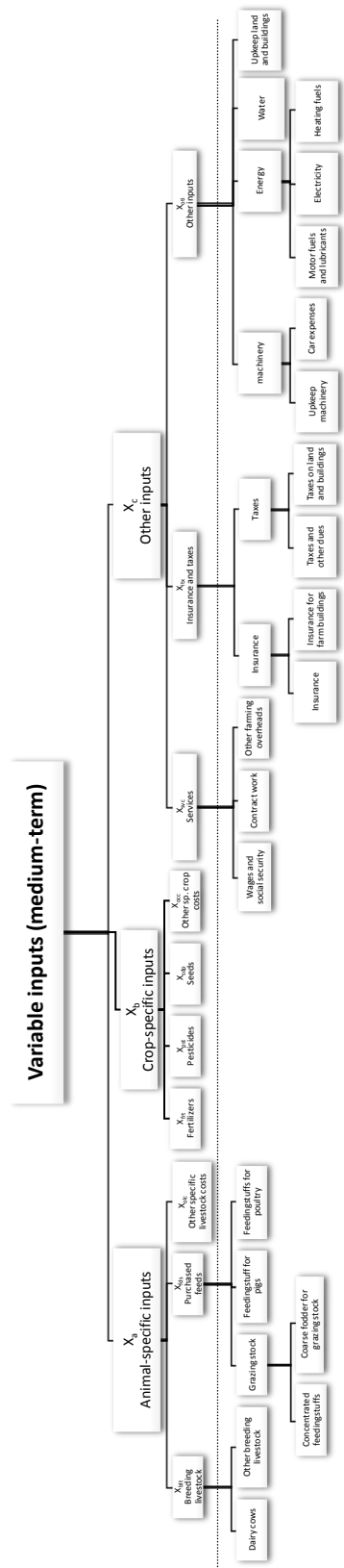


Figure 2.3: Aggregation scheme for variable inputs (medium term).

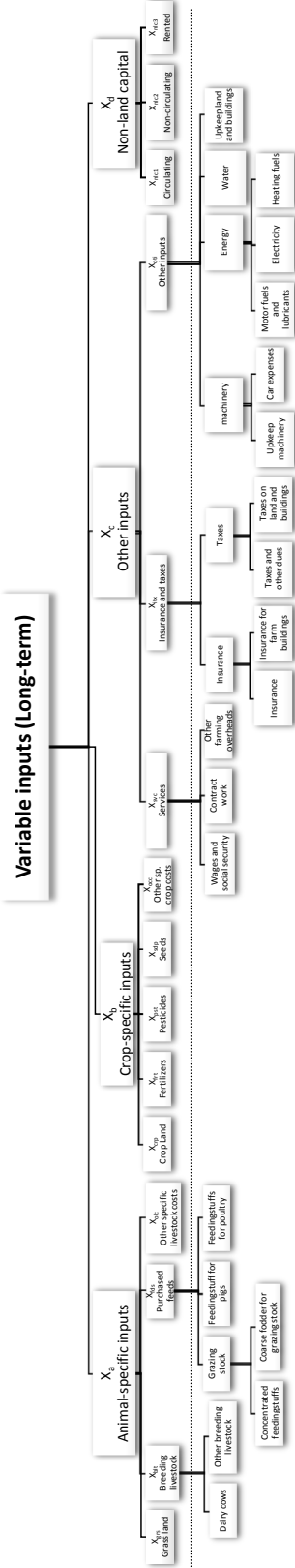


Figure 2.4: Aggregation scheme for variable inputs (long term).



Figure 2.5: Aggregation scheme for all outputs.

Part II

Design and Development of a Method to Prepare and Update the Data from the EU-FADN

Introduction

This report on Part 2 focusses on the design and development of the method to prepare and update the data from the EU-FADN for econometric estimations. The data preparation is a two steps process. First, every lower-level variables and their price are generated. This is done for variable inputs, fixed inputs and outputs for the medium- and long-term specifications. Missing prices are imputed using data from EUROSTAT. Second, following the choices of the user, including farm type and time specification, the data set is more precisely shaped in order to have only the relevant and needed data left for further estimations. In this second step, variables are aggregated in order to have upper-level variables available too. Two default aggregation schemes are defined, depending on the chosen type of farm: one for bovine farms (livestock and dairy) and one for crop farms. The user may also choose to define its own aggregation scheme. After data preparation, the resulting dataset contains then both aggregated and disaggregated input and output values (X , respectively Y), input and output prices (W , respectively P) and input and output quantities (x , respectively y), the level of aggregation being distinguished using indices.

In general, value and price information for the component items allows the construction of a Törnqvist price index for the aggregate. The quotient of aggregate value and Törnqvist index gives the quantity measure “value at base-year prices”. The construction of these variables is provided in greater detail in this section.

To ensure numerical stability, output quantities and fixed input quantities are rescaled such that their mean squared error lies between 1 and 10. For example y_m is transformed into

$$\begin{aligned} y_m^{(s)} &= y_m \cdot s_m \\ s_m &= 10^{\left(-\text{int}\left[\log_{10} \sqrt{\mathbb{E}[y_m]^2 + \text{Var}[y_m]}\right]\right)}, \end{aligned}$$

where $\text{int}[\cdot]$ denotes the nearest integer.

In the next chapters, we describe more precisely this overall data preparation

process. This includes the construction of the Törnqvist index, the imputation of prices, the construction of lower-level and the definition of the defaults upper-level aggregation schemes.

Chapter 3

Törnqvist Index Construction

Since farm-level aggregate price indices proved too erratic, Törnqvist price indices are constructed at the regional level. They are expressed with respect to base year $t_0 = 2005$. Törnqvist price indices are generated for every variable at the lower and upper level.¹

Supposing all inputs $h = 1, \dots, \sum_{i=1}^I N_i$ are grouped into $i = 1, \dots, I$ categories, the Törnqvist index w_{irt} is defined for each input aggregate i , each geographical unit r and each period t as

$$w_{irt} = \prod_{j=1}^{N_i} \left(\frac{w_{jrt}}{w_{jrt_0}} \right)^{\frac{g_{jrt} + g_{jrt_0}}{2}}$$
$$g_{jrt} = \frac{V_{jrt}}{\sum_{k=1}^{N_i} V_{krt}},$$

where N_i denotes the number of input-components encompassed by the aggregate input i , w_{jrt} represents the average of farm-gate prices of input-component j in geographical unit r in period t , and V_{jrt} represents the total value spent on input j in geographical unit r in period t . We tried different geographical subdivisions, among which individual farms, but finally decided to supply Törnqvist price indices at the regional level, because farm-level aggregate price indices proved too erratic.

Farm-gate prices w_{jft} for each input-component j at time t for farm f are obtained by dividing the value of total purchases of the farm (V_{jft}) by the farm-total volume purchased (N_{jft}). An average regional price of input j in region $r = 1, \dots, R$ in period t is estimated by dividing total purchases within region r : $V_{jrt} = \sum_{f=1}^{F_r} V_{jft}$ by total volume purchased $N_{jrt} = \sum_{f=1}^{F_r} N_{jft}$. A country-wide price average is obtained similarly. If needed, country-wide average prices are also provided by EURO-

¹It is relevant to also generate Törnqvist price indices for lower-level variables since these are aggregates of variables too.

STAT. Chapter 4 explains how farm-gate input prices are imputed when missing.

Similarly, the Törnqvist index τ_{mrt} for output aggregate m in geographical unit r in period t is given by

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

$$g_{nrt} = \frac{V_{nrt}}{\sum_{o=1}^{N_o} V_{ort}},$$

where N_m denotes the number of output-components n encompassed by the output aggregate m , p_{nrt} represents the average farm-gate price of output-component n produced in geographical unit r in period t , and V_{nrt} represents the total revenue generated by output-component n in geographical unit r in period t . Average regional and national prices of product n in period t are again obtained by dividing the value of total production by total number of units sold. Chapter 4 explains how farm-gate output prices are imputed when missing.

Note that average regional or country prices of inputs and outputs are calculated by removing extreme farm-gate prices of inputs and outputs. Farm-gate prices that have a probability of occurrence that is smaller than one over twice the sample size are disregarded from that calculation. For input prices, extreme values of w_{jrt} are defined as

$$| (w_{jft} - w_{jrt}) | > \Phi^{-1} (1 - 1/2F_r) * SD_{w_{jft}}$$

where F_r is the sample size, w_{jrt} the regional or country average of w_{jft} , and $SD_{w_{jft}}$ the standard deviation of w_{jft} .

Similarly for output prices, extreme values of p_{nrt} are defined as

$$| (p_{nft} - p_{nrt}) | > \Phi^{-1} (1 - 1/2F_r) * SD_{p_{nft}}$$

where F_r is the sample size, p_{nrt} the regional or country average of p_{nft} , and $SD_{p_{nft}}$ the standard deviation of p_{nft} . Extreme and missing values of farm gate prices are replaced by their regional or country price average.

To preserve the relationship

$$\tau_{mft} = \frac{p_{mft}}{p_{mft;b}},$$

We use the following definitions for prices of aggregates

$$p_{mft} = \prod_{n=1}^{N_m} (p_{nft})^{\frac{g_{nft}+g_{nft_0}}{2}}$$

$$p_{mft;b} = \prod_{n=1}^{N_m} (p_{nft_0})^{\frac{g_{nft}+g_{nft_0}}{2}} .$$

The price in the base year of an aggregate, $p_{mft;b}$, can thus vary from year to year, depending on the varying exponent $(g_{nrt}+g_{nrt_0})/2$.

Chapter 4

Imputation of Missing Prices

Output prices p_{nft} are in general computed from sale values and volumes in the EU-FADN database, if not from production values and volumes. If the necessary information is missing it is supplemented with prices from EUROSTAT. Information on input prices w_{jft} is mainly obtained from EUROSTAT, although for some inputs it also stems from the EU-FADN database. Missing prices are imputed according to the following scheme, in which we take dry pulses as an example. From all data sources, we obtain several price variables pertaining to dry pulses (see Table 4.1).

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	$\frac{\sum_{region}(K129tp)}{\sum_{region}(K129qq)}$
$p_{country;t}$	FADN	Dry pulses / country average	$\frac{\sum_{country}(K129tp)}{\sum_{country}(K129qq)}$
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 4.1: Illustration of the price imputation scheme using dry pulses as an example

We now proceed as follows:

- For each observation, i.e., each farm f at each observed period t , we compute p_{ft} , $p_{region;t}$ and $p_{country;t}$. If p_{ft} is missing, we assign it $p_{region;t}$ or, failing that, $p_{country;t}$.
- We retrieve price indices from the dataset `apri_pi00_outa` from EUROSTAT.¹ The EUROSTAT description of this dataset as well the description of the other datasets that are used in this Part II are shown in Table 4.2. This Table also shows related datasets that are used to impute price variables when they are

¹The dataset `apri_pi00_outa` corresponds to the *Price indices of agricultural products, output [2000 = 100]* (Eurostat, 2013a).

missing from the original dataset. From `apri_pi00_outa`, we take the most detailed index covering the whole crop category (in this example p41980 - pulses), we take the index for the smallest super-category (here “Other fresh vegetables”) and we take the index covering all crop outputs. Now for each country, and for each $p_{imp} = p_{pls}, p_{oth}, p_{crp}, cpi$:

- if p_{ft} is completely missing, it is replaced by p_{imp} ;
 - if p_{ft} has missing observations, it is regressed on p_{imp} and missing observations are replaced with the predicted values of this regression.
- We now end up with a price variable for pulses that contains no missing values.

Eurostat file	Eurostat description	Range	Processing
<i>apri_pi95_proda.dta</i>	EC index of producer prices of agricultural products (1995 = 100) - annual data	1988-2004	Using variables from <i>apri_pi00_outa.dta</i> . For missing years from 1988 to 1993, predict with <i>apri_pi95_proda.dta</i> . For missing years from 2009 to 2012, predict with <i>apri_pi05_outa.dta</i> .
<i>apri_pi00_outa.dta</i>	Price indices of agricultural products, output (2000 = 100) - annual data	1994-2008	
<i>apri_pi05_outa.dta</i>	Price indices of agricultural products, output (2005 = 100) - annual data	2000-2012	
<i>apri_ap_haouta.dta</i>	Selling prices of animal products (absolute prices) - annual - old codes	1969 - 2005	Using variables from <i>apri_ap_haouta.dta</i> . For missing years from 2006 to 2013, predict with prices from the dataset <i>apri_ap_anouta.dta</i> .
<i>apri_ap_anouta.dta</i>	Selling prices of animal products (absolute prices) - annual price	2000-2013	
<i>apri_pi95_purcha.dta</i>	EC indices of purchase prices of the means of agricultural production (1995 = 100) - annual data	1988-2004	Using values from <i>apri_pi00_ina.dta</i> . For missing years from 1988 to 1994, predict with <i>apri_pi95_purcha.dta</i> . For missing years from 2009 to 2012, predict with <i>apri_pi05_ina.dta</i> .
<i>apri_pi00_ina.dta</i>	Price indices of the means of agricultural production, input (2000 = 100) - annual data	1995-2008	
<i>apri_pi05_ina.dta</i>	Price indices of the means of agricultural production, input (2005 = 100) - annual data	2000-2012	
<i>aact_eaa04.dta</i>	Economic accounts for agriculture - values at real prices	1973-2014	Using production value at basic price for "compensation of employees" and "other taxes on production".
<i>apri_ap_aland.dta</i>	Land prices and rents - annual data	1985-2009	Using prices for "arable land", for "grassland" and for "agricultural land". Using also rental rates for "arable land", for "grassland" and for "agricultural land".

Table 4.2: Summary of the Eurostat datasets used in the Stata routine.

Chapter 5

Output Data Preparation

For convenience, outputs are described considering two main categories: animal-specific outputs and crops-specific outputs. However, in the Stata routine, all outputs are actually part of a common aggregation scheme, as shown by Figure 2.5.

5.1 Animal-specific Outputs

In this section, we first consider two subcategories of variables, i.e., dairy outputs and other animal outputs. We then summarize all animal-specific outputs in a table giving an overview of the data structure for this category.

The construction of these aggregate variables and their corresponding price indices are now described in more detail.

1. The dairy output Y_a consists of both milk ($K162$) and milk products ($K163$). Based on prices computed from total production values V_s and volumes Q_s (measured in quintals, i.e. units of 100kg, for butter and in “liquid milk equivalent in quintals” for milk) in the EU-FADN database as

$$\tilde{p}_q = V_q/Q_q,$$

a region-specific Törnqvist index for this product group is constructed. If the necessary information for computing said prices is missing, we use information from the dataset `apri_ap_haouta` from EUROSTAT.¹ The EUROSTAT description of this dataset is shown in Table 4.2. The connection between products and prices is described in Table 5.1. How the EUROSTAT information is used to obtain imputed price variables is described in Chapter 4.

¹The dataset `apri_ap_haouta` corresponds to the *Selling prices of animal products (absolute prices)* (Eurostat, 2012).

Value of sales			Price		
Var	EU-FADN description	unit	Var	EUROSTAT description	unit
K162sa	Milk sales	€	p5411	Whole cows milk for human consumption	€/100l
K163sa	Milk products sold	€	p5435	Butter	€/100kg

Table 5.1: Aggregate output Y_a : milk products and their corresponding prices

2. The aggregate of other animal outputs consists of the products $N22 - N50$. The bovine outputs Y_b ($N23 - N32$) are described in Table 5.2, while the non-bovine outputs Y_c ($N22$ and $N33 - N50$) are given in Table 5.4. Prices are obtained from the FADN by dividing the total sale value V_s by the number of animals sold N_s (measured in 10 LU)

$$\bar{p}_s = V_s / N_s$$

or, failing that, from the EUROSTAT datasets `apri_ap_haouta` and `apri_pi00_outa`, the former containing country-specific prices of different animal prices and the latter country-specific price indices of several animals. In Table 5.2, prices used for each category of bovine are listed in columns three and four. The last four columns of this table list the variables used for imputation of missing values (see Chapter 4). Their description is given in Table 5.3.

Value of sales (€)		Price (€/100kg)		Variables for imputation			
Var	Description	Var	Description	€/head		(base = 2005)	
EU-FADN		EUROSTAT: apri_ap_haouta				EUROSTAT: apri_pi00_outa	
N23sa	Calves/fattening	p4150	Calves	p4232	p4233	p111200	p111000
N24sa	Oth.cattle(<12m)	p4150	Calves	p4232	p4233	p111200	p111000
N25sa	M cattle(12-24m)	p4161	Young cattle	p4234		p111100	p111000
N26sa	F cattle(12-24m)	p4161	Young cattle	p4234		p111100	p111000
N27sa	M cattle(>24m)	p4180	Bullocks			p111100	p111000
N28sa	Breeding heifers	p4162	Heifers	p4235		p111100	p111000
N29sa	Heifers/fattening	p4162	Heifers	p4235		p111100	p111000
N30sa	Dairy cows	p4171	Cows A (1st quality)			p111100	p111000
N31sa	Cull dairy cows	p4173	Cows C (3rd quality)			p111100	p111000
N32sa	Other cows	p4172	Cows B (2nd quality)			p111100	p111000

Table 5.2: Aggregate output Y_b : bovine components and their corresponding prices and price indices from EUROSTAT

Var	Prices	Unit	Source
p4232	Calves (of a few days) - prices per head	€/head	apri_ap_haouta
p4233	Calves (of a few weeks) - prices per head	€/head	apri_ap_haouta
p4234	Young cattle (store) - prices per head	€/head	apri_ap_haouta
p4235	Heifers (store) - prices per head	€/head	apri_ap_haouta
Var	Price index	Base year	Source
p111000	Cattle	2005	apri_pi00_outa
p111100	Cattle excluding calves	2005	apri_pi00_outa
p111200	Calves	2005	apri_pi00_outa

Table 5.3: Description of EUROSTAT bovine prices and price indices used for imputation in Table 5.2

In Table 5.4, non-bovine outputs are described. While the first two columns list the EU-FADN variable names and their description, columns three and four give the corresponding EUROSTAT price index. The last two columns list the variables used for imputation of missing values (see Chapter 4).

Value of sales (€)		EUROSTAT price indices (base = 2000)		Variables for imputation (base = 2005)	
Var	FADN description	Var	EUROSTAT description	Var	EUROSTAT description
N22sa	Horses	p113000	Equines	p110000	Animals
N33sa	Bees	p119000	Other animals	p110000	Animals
N34sa	Rabbits (breed.F)	p119000	Other animals	p110000	Animals
N38sa	Goat (breeding F)	p114000	Sheep and goats	p110000	Animals
N39sa	Other goats	p114000	Sheep and goats	p110000	Animals
N40sa	Ewes	p114000	Sheep and goats	p110000	Animals
N41sa	Other sheep	p114000	Sheep and goats	p110000	Animals
N43sa	Piglets	p112000	Pigs	p110000	Animals
N44sa	Breeding sows	p112000	Pigs	p110000	Animals
N45sa	Pigs/fattening	p112000	Pigs	p110000	Animals
N46sa	Other pigs	p112000	Pigs	p110000	Animals
N47sa	Table chickens	p115100	Chickens	p115000	Poultry
N48sa	Laying hens	p115100	Chickens	p115000	Poultry
N49sa	Other poultry	p115900	Other poultry	p115000	Poultry
N50sa	Other animals	p119000	Other animals	p110000	Animals

Table 5.4: Aggregate output Y_c : non-bovine components and corresponding price indices from EUROSTAT

Using all these variables, we finally define the lower-level variables and default upper-level variables for animal-specific outputs. Table 5.5 shows an overview of the aggregation scheme for these variables at both levels.

Value of sales (€)			
Var	EU-FADN description	lower level	upper level
K162sa	Milk sales	Y_{mlk}	Y_a
K163sa	Milk products sold	Y_{mlp}	Y_a
N23sa	Calves/fattening	Y_{clv}	Y_b
N24sa	Oth.cattle(<12m)	Y_{ctl}	Y_b
N25sa	M cattle(12-24m)	Y_{ctl}	Y_b
N26sa	F cattle(12-24m)	Y_{ctl}	Y_b
N27sa	M cattle(>24m)	Y_{ctl}	Y_b
N28sa	Breeding heifers	Y_{hfr}	Y_b
N29sa	Heifers/fattening	Y_{hfr}	Y_b
N30sa	Dairy cows	Y_{cow}	Y_b
N31sa	Cull dairy cows	Y_{cow}	Y_b
N32sa	Other cows	Y_{cow}	Y_b
N38sa	Goat (breeding F)	Y_{got}	Y_c
N39sa	Other goats	Y_{got}	Y_c
N40sa	Ewes	Y_{shp}	Y_c
N41sa	Other sheep	Y_{shp}	Y_c
N43sa	Piglets	Y_{pig}	Y_c
N44sa	Breeding sows	Y_{pig}	Y_c
N45sa	Pigs/fattening	Y_{pig}	Y_c
N46sa	Other pigs	Y_{pig}	Y_c
N47sa	Table chickens	Y_{plt}	Y_c
N48sa	Laying hens	Y_{plt}	Y_c
N49sa	Other poultry	Y_{plt}	Y_c
N22sa	Horses	Y_{ota}	Y_c
N33sa	Bees	Y_{ota}	Y_c
N34sa	Rabbits (breed.F)	Y_{ota}	Y_c
N50sa	Other animals	Y_{ota}	Y_c

Table 5.5: Animal-specific outputs: lower- and upper-level variables overview

5.2 Crop-specific Outputs

The crop-specific outputs aggregates Y_d to Y_g are constructed using EU-FADN variables $K120 - K141$. Again, their prices are computed from total production values V_q and production volumes Q_q (measured in 100 kg) in the EU-FADN database as

$$\tilde{p}_q = V_q/Q_q.$$

The description of the variables is provided in column two of Table 5.6. Columns three and four contain the corresponding price variable and its description from EUROSTAT's `apri_pi00_outa`. These are the main variables used to impute missing prices. The last two columns give price variables used for imputation of missing values (see subsection 4).

Using all these variables, we finally define the lower-level variables and default upper-level variables for crops-specific outputs. Table 5.7 shows an overview of the aggregation scheme for these variables at both levels.

5.3 Net Sale Value of Output Categories

The sale values of the output categories are corrected for the variation in output stock values in the data preparation stage so that net sale values reflect better the output production for sales made during the year. These net sale values are calculated as the following:

$$NY_{m,n} = KxxxSA + (KxxxCV - KxxxBV)$$

for m, n corresponding to crop and milk output categories,

and

$$NY_{m,n} = NxxSV + (NxxCV - NxxBV)$$

for m, n corresponding to the other animal output categories,

where NY corresponds to sale values net of yearly stock variation, xxx takes the value from 120 to 375, xx takes the value from 22 to 58, SA and SV correspond to sale value, CV to closing value and BV to beginning value.

Note that farm-gate output unit values are still calculated as before:

$$\tilde{p}_q = V_q/Q_q$$

Value of sales (€)		Price index (base = 2005)		Variables for imputation (base = 2005)	
Var	FADN description	Var	EUROSTAT description	Var	Description
K120sa	Common wheat	p11100	Soft wheat and spelt	p11000	Wheat and spelt
K121sa	Durum wheat	p11200	Durum wheat	p11000	Wheat and spelt
K122sa	Rye	p12000	Rye and meslin	p10000	Cereals (including seeds)
K123sa	Barley	p13000	Barley	p10000	Cereals (including seeds)
K124sa	Oats	p14000	Oats and summer cereal mixtures	p10000	Cereals (including seeds)
K125sa	Summer cer_mix.	p14000	Oats and summer cereal mixtures	p10000	Cereals (including seeds)
K126sa	Grain maize	p15000	Grain maize	p10000	Cereals (including seeds)
K127sa	Rice	p16000	Rice	p10000	Cereals (including seeds)
K128sa	Oth.cereals	p19000	Other cereals	p10000	Cereals (including seeds)
K129sa	Dry pulses	p41980	Pulses	p41900	Other fresh vegetables
K130sa	Potatoes	p50000	Potatoes (including seeds)	p101000	Crop output, excluding fruits and vegetables
K131sa	Sugar beet	p24000	Sugar beet	p20000	Industrial crops
K132sa	Oil seed	p21000	Oil seeds and oleaginous fruits (including seeds)	p20000	Industrial crops
K133sa	Hops	p29200	Hops	p29000	Other industrial crops
K134sa	Tobacco	p23000	Raw tobacco	p20000	Industrial crops
K135sa	Oth. indust._crp	p29900	Other industrial crops: others	p29000	Other industrial crops
K136sa	Vegetables	p41000	Fresh vegetables	p40000	Vegetables and horticultural products
K137sa	Vegetables	p41000	Fresh vegetables	p40000	Vegetables and horticultural products
K138sa	Vegetables	p41000	Fresh vegetables	p40000	Vegetables and horticultural products
K139sa	Mushrooms	p41999	Other fresh vegetables: other	p40000	Vegetables and horticultural products
K140sa	Flowers	p42000	Plants and flowers	p40000	Vegetables and horticultural products
K141sa	Flowers	p42000	Plants and flowers	p40000	Vegetables and horticultural products

Table 5.6: Aggregate outputs Y_d to Y_g : crop categories and corresponding price indices from EUROSTAT

for crop and milk outputs,
and

$$\bar{p}_s = V_s/N_s$$

for the other animal outputs.

Value of sales (€)			
Variable	EU-FADN description	Lower level	Upper level
K129sa	Dry pulses	Y_{drp}	Y_d
K132sa	Oil seeds	Y_{oil}	Y_d
K120sa	Common wheat	Y_{wht}	Y_e
K121sa	Durum wheat	Y_{wht}	Y_e
K126sa	Grain maize	Y_{mze}	Y_e
K127sa	Rice	Y_{wht}	Y_e
K122sa	Rye	Y_{otc}	Y_e
K123sa	Barley	Y_{otc}	Y_e
K124sa	Oats	Y_{otc}	Y_e
K125sa	Summer cereral mix.	Y_{otc}	Y_e
K128sa	Other cereals	Y_{ocr}	Y_e
K130sa	Potatoes	Y_{ptt}	Y_f
K131sa	Sugar beet	Y_{sbt}	Y_f
K135sa	Other industrial crops	Y_{oin}	Y_f
K133sa	Hops	Y_{hps}	Y_g
K134sa	Tobacco	Y_{tbc}	Y_g
K136sa	Vegetables	Y_{otc}	Y_g
K137sa	Vegetables	Y_{otc}	Y_g
K138sa	Vegetables	Y_{otc}	Y_g
K139sa	Mushrooms	Y_{otc}	Y_g
K140sa	Flowers	Y_{otc}	Y_g
K141sa	Flowers	Y_{otc}	Y_g

Table 5.7: Crop-specific outputs : lower- and upper-level variables overview

Chapter 6

Input Data Preparation

For convenience, inputs are described considering two main categories: variable inputs and fixed inputs.

6.1 Variable Inputs

Table 6.1 gives an overview of the different lower-level variable inputs categories in the medium term, their prices and the way they are obtained. The table also shows the default aggregation schemes for these variable, i.e. the upper-level variables X_a to X_c .¹

The lower-level variable inputs for the medium term are :

- X_{blt} : breeding livestock
- X_{fds} : purchased feeds
- X_{olc} : other specific-livestock costs
- X_{sds} : purchased seeds
- X_{ftr} : fertilizers
- X_{pst} : pesticides
- X_{occ} : other specific-crop costs
- X_{svc} : services
- X_{itx} : insurance and taxes

¹The *default* aggregation schemes refers to the aggregation scheme automatically constructed by the Stata routine when the user decides not to define its own aggregation schemes.

- X_{oti} : other inputs

The default upper-level aggregate variable inputs for the medium-term are :

- X_a : animal-specific inputs (regrouping X_{blt} to X_{olc})
- X_b : crop-specific inputs (regrouping X_{sds} to X_{occ})
- X_c : other inputs (regrouping X_{svc} to X_{oti})

For each variable obtained from the EU-FADN database, the columns “Variable” give the names of the variables as they appear in their source dataset. The columns “lower-level” and “upper-level” denote the aggregated input variable in which it is represented, at the corresponding level. The column “Source” gives the data source from where the corresponding price is taken, with E denoting EUROSTAT and F the EU-FADN database. All EUROSTAT variables are obtained from the dataset `apri_pi00_ina` from EUROSTAT,² except the indices on “Compensation of employees” and “Other taxes on production” which are obtained from the EUROSTAT’s dataset `aact_eaa04`.³ Descriptions for `apri_pi00_ina` and `aact_eaa04` are available in Table 4.2.

The price for “Interest and financial charges” is computed by dividing all financial expenses by the value of outstanding debt and the rate of “Depreciation” by dividing it by the value of capital (Ivaldi et al., 1996).

The computation of prices and quantities for X_{blt} is now explained in greater detail. Define the following quantities:

- $N_{o;j}, V_{o;j}$: number of breeding animals of type j in stock at the beginning of the year and their total value (from FADN table D)
- $N_{c;j}, V_{c;j}$: number of breeding animals of type j in stock at the end of the year and their total value (from FADN table D)
- $N_{p;j}, V_{p;j}$: number of breeding animals of type j purchased and their total value (from FADN table N)
- $N_{s;j}, V_{s;j}$: number of breeding animals of type j sold and their total value (from FADN table N)

²The dataset `apri_pi00_ina` corresponds to the *Price indices of the means of agricultural production, input (2000 = 100)* Eurostat (2013b).

³The dataset `aact_eaa04` corresponds to the *Economic accounts for agriculture - values at real prices* Eurostat (2014).

The types of breeding animals counted as inputs are breeding heifers, dairy cows, other or suckler cows, breeding goats, ewes and breeding sows, the other or suckler cows making up the largest part of the FADN category “other cows”. For each sub-category j = “breeding heifers”, “dairy cows”, “other cows”, “breeding goats”, “ewes”, “breeding sows” of X_{blt} , the value of the input is computed as the value of the average stock of animals

$$X_{blt;j} = (V_{c;j} + V_{o;j})/2.$$

For each period, the average price of purchases (in €/LU) for each farm can be obtained by dividing the total value of purchases by the number of purchases

$$\bar{w}_{X_{blt;j}} = V_{p;j}/N_{p;j}.$$

If the price variable is missing, it is calculated from the FADN table D as the average of the price used for the opening and closing valuations (Community Committee for the Farm Accountancy Data Network, 2005, p.13)

$$\hat{w}_{X_{blt;j}} = (V_o/N_o + V_c/N_c)/2.$$

For the observations with missing price variable, it is imputed by the regional average or failing that, by the country average. If this whole procedure results in a missing price value, prices are obtained from EUROSTAT, as described in section 5.1.

Value of costs (€)					Prices / Indices		
Variable	EU-FADN description	Lower level	Upper level	Source	Variable	EUROSTAT description	Unit /base
$\sum_{j=28,30,32,38,40,44} D_j$	Breeding livestock	X_{blt}	X_a	F	$\sum_{j=28,30,32,38,40,44} N_j$	/	€/10L U
F64	Concentrated feedingstuffs for grazing stock	X_{fds}	X_a	E	p206200	Compound feeding stuffs	2005
F65	Coarse fodder for grazing stock	X_{fds}	X_a	E	p206100	Straight feeding stuffs	2005
F66	Feedingstuffs for pigs	X_{fds}	X_a	E	p206230	Compound feeding stuffs for pigs	2005
F67	Feedingstuffs for poultry and other small animals	X_{fds}	X_a	E	p206240	Compound feeding stuffs for poultry	2005
F71/SE330	Other specific livestock costs	X_{olc}	X_a	E	p205000	Veterinary expenses	2005
F72	Seeds and seedlings purchased	X_{sdp}	X_b	E	p201000	Seeds and planting stock	2005
F74/SE295	Fertilizers and soil improvers	X_{frt}	X_b	E	p203000	Fertilizers and soil improvers	2005
F75/SE300	Crop protection products	X_{pst}	X_b	E	p204000	Plant protection products and pesticides	2005
F76/SE305	Other specific crop costs	X_{occ}	X_b	E	p211141	Farm machinery and installations for crop production	2005
F59/SE370	Wages and social security	X_{svc}	X_c	E	23000	Compensation of employees	2005
F60/SE350	Contract work	X_{svc}	X_c	E	23000	Compensation of employees	2005
F84	Other farming overheads	X_{svc}	X_c	E	p209000	Other goods and services	2005
F82	Insurance	X_{itx}	X_c	E	p209000	Other goods and services	2005
F87	Insurance for farm buildings	X_{itx}	X_c	E	p209000	Other goods and services	2005
F83	Taxes and other dues	X_{itx}	X_c	E	24000	Other taxes on production	2005
F88	Taxes on land and buildings	X_{itx}	X_c	E	24000	Other taxes on production	2005
F61	Current upkeep of machinery and equipment	X_{oti}	X_c	E	p207000	Maintenance of materials	2005
F62	Motor fuels and lubricants	X_{oti}	X_c	E	p202300	Motor fuels	2005
				E	p202400	Lubricants	2005
F63	Car expenses	X_{oti}	X_c	E	p211200	Transport equipment	2005
F78	Upkeep of land improvements and buildings	X_{oti}	X_c	E	p208000	Maintenance of buildings	2005
F79	Electricity	X_{oti}	X_c	E	p202100	Electricity	2005
F80	Heating fuels	X_{oti}	X_c	E	p202200	Fuels for heating	2005
F81	Water	X_{oti}	X_c	E	p209000	Other goods and services	2005

Table 6.1: Variable inputs (medium-term)

6.2 Fixed versus Variable Inputs

Table 6.2 provides an overview of the input variables not yet considered in section 6. For each of those inputs, we indicate whether we consider them to be fixed or variable, and to which aggregate input they are allocated. The symbols indicated

as quantity (Q), value (V) and price (p) are explained in subsections 6.3 and 6.4 below.

MT (lower level)	MT (upper level)	LT (lower level)	LT (upper level)	Description	Quantity	Value	Price
Z_{grs}, Z_{crp}	Z_L	X_{grs}, X_{crp}	X_a, X_b	Land in full ownership without debt	S_{LF}	$R_{LF} = S_{LF} \cdot r_{rL}$	r_{rL}
Z_{grs}, Z_{crp}	Z_L	X_{grs}, X_{crp}	X_a, X_b	Land with outstanding debt	$S_{LD} = D_L$	$R_{LD} = S_{LD} \cdot (r_{rL} + r_{iL})$	$r_{rL} + r_{iL}$
Z_{grs}, Z_{crp}	Z_L	X_{grs}, X_{crp}	X_a, X_b	Rented land	S_{LR}	$R_{LR} = S_{LR} \cdot r_{rL}$	r_{rL}
Z_{unl}	Z_{unl}	Z_{unl}	Z_{unl}	Unpaid labour input (AWU)	SE015		/
Z_{nlc1}	Z_{nlc}	X_{nlc1}	X_d	circulating debt-free non-land assests	F_C	$F_C \cdot r_{oK}$	r_{oK}
Z_{nlc2}	Z_{nlc}	X_{nlc2}	X_d	noncircul. debt-free non-land assests	F_N	$F_N \cdot (r_{oK} + r_{dK})$	$r_{oK} + r_{dK}$
Z_{nlc1}	Z_{nlc}	X_{nlc1d}	X_d	circulating non-land assests with debt	D_C	$D_C \cdot (r_{oK} + r_{iK})$	$r_{oK} + r_{iK}$
Z_{nlc2}	Z_{nlc}	X_{nlc2}	X_d	noncircul. non-land assests with debt	D_N	$D_N \cdot (r_{oK} + r_{iK} + r_{dK})$	$r_{oK} + r_{iK} + r_{dK}$
Z_{nlc3}	Z_{nlc}	X_{nlc3}	X_d	Rented (non-land) capital	$R_K / (r_{oK} + r_{dK})$	R_K	$r_{oK} + r_{dK}$

Table 6.2: Medium- versus long-term perspective

6.2.1 Medium-term Specification

Medium-term fixed inputs are:

- Z_L : agricultural area
 - Z_{grs} : grass land
 - Z_{crp} : crop land
- $Z_{nlc} = X_d$: non-land capital
 - Z_{nlc1} : circulating non-land capital
 - * Z_{nlc1f} : circulating debt-free non-land capital
 - * Z_{nlc1d} : circulating indebted non-land capital
 - Z_{nlc2} : non-circulating non-land capital
 - * Z_{nlc2f} : non-circulating debt-free non-land capital
 - * Z_{nlc2d} : non-circulating indebted non-land capital
 - Z_{nlc3} : rented non-land capital
- Z_{unl} : unpaid labour on farm

6.2.2 Long-term Specification

In the long-term specification, land and non-land capital are both treated as variable inputs. Therefore, in the long-term specification, the only fixed-input is unpaid labour (Z_{unl}).

6.3 Land

When grass land Z_{grs} and crop land Z_{crp} are treated separately, the former is the sum of three types of grass land

$$Z_{grs} = K147aa + K150aa + K151aa,$$

where K147aa is the acreage of temporary grass, K150aa the acreage of meadow and pasture and K151aa the acreage of rough grazing,

while the latter is being computed as

$$Z_{crp} = \max(SE025 - Z_{grs} - K149AA, 0).$$

removing land that is leased to others (K149AA) since this leased land is not anymore considered as an input in the related production function, from the total utilised agricultural area (UAA) (SE025) (Community Committee for the Farm Accountancy Data Network, 2007, p.15) and grass land Z_{grs} .

However, when insufficient information is available to split up land into two separate types

$$Z_L = \max(SE025 - K149AA, 0).$$

Three types of land are considered: rented land, land on which there is still an outstanding debt, and land in full ownership. Rented land is valued at its rental rate. Non-rented land is valued as its opportunity rate. Non-rented land on which there is still an outstanding debt has however an extra price-component.

6.3.1 Rental Rate of Land

The sale value of rented land is given by

$$S_{LR} = SE030 \cdot p_{sL}$$

and the rental value of rented land is given by⁴

$$R_{LR} = SE030 \cdot p_{rL} \tag{6.1}$$

⁴In other words, the amount of rent paid to use the land.

where p_{sL} the sale price and p_{rL} the rental price of land are obtained from the dataset `apri_ap_aland` from EUROSTAT,⁵ and where SE030 is the rented UAA (Community Committee for the Farm Accountancy Data Network, 2007, p.15). A description of `apri_ap_aland` is available in Table 4.2.

Consequently, the rental rate of land is defined by

$$r_{rL} = \frac{R_{LR}}{S_{LR}} = \frac{p_{rL}}{p_{sL}},$$

where the rental price of land p_{rL} and the sales price of land p_{sL} are both obtained from the dataset `apri_ap_aland` from EUROSTAT. For non-rented, but owned land r_{rL} represents an opportunity rate.

6.3.2 Interest Rate of Land

Land on which there is an outstanding debt has an extra price-component necessary to acquire it. Interest paid on land, is given in Table F (Community Committee for the Farm Accountancy Data Network, 2005, p.15) by $I_L = \#291 = \text{F91}$. The interest rate can be calculated by dividing the interest paid by the outstanding debt on land, which is given by $D_L = \#392 = \text{H106c}$.

$$r_{iL} = \frac{I_L}{D_L} = \frac{\#291}{\#392} = \frac{\text{F91}}{\text{H106c}}.$$

If the total debt on land is missing, only the long-term debt on land might be considered ($D_L = \#376 = \text{H104c}$).

Dubious observations are treated as follows. Regress

$$I_{L;ft} = \rho D_{L;ft} + \varepsilon_{ft}, \quad (6.2)$$

with f denoting a farm index and t meaning time, in order to obtain $\hat{\rho}_1$. Remove observations for which $r_{iL;ft} > 5\hat{\rho}_1$ and re-regress (6.2), to obtain $\hat{\rho}_2$. Assign this value to the removed observations.

6.3.3 Opportunity Cost of Land

Regarding the rental, sale and opportunity values of rented land, we have the following:

⁵The dataset `apri_ap_aland` corresponds to the *Land prices and rents - annual data* Eurostat (2009).

1. The sale value of land on which there is an outstanding debt, is equal to the outstanding debt on land D_L . In order to determine the yearly recurring value for the farm of this asset as an input, we need to attribute it $r_{iL} + r_{rL}$ as a price, so

$$R_{LD} = D_L (r_{iL} + r_{rL}) = S_{LD} (r_{iL} + r_{rL}).$$

2. Opportunity value of utilised owned land R_{LF} is taken to be equal to the rental value of this land, since the farm faces the choice between farming or renting out the land it owns. In other words,

$$R_{LF} = \max(\text{SE025-SE030- K149aa} - D_L/p_{sL}, 0) \cdot p_{rL},$$

where, as before, SE025 is the total UAA, SE030 the rented UAA, K149aa the land that is leased to others, D_L/p_{sL} the indebted acreage, and p_{rL} the rental price. The sales value of utilised owned land S_{LF} is given by

$$S_{LF} = \max(\text{SE025-SE030- K149aa} - D_L/p_{sL}, 0) \cdot p_{sL},$$

where p_{sL} is the sale price of land is obtained from `apri_ap_aland`. The opportunity rate of utilised owned land is thus equal to the rental rate of rented land

$$r_{oL} = r_{rL} = \frac{p_{rL}}{p_{sL}}.$$

6.4 Non-land Capital

6.4.1 Stock of Non-land Capital

At opening valuation, the stock of capital, excluding land, is given by⁶

$$Z_{nlc} = S_K = \max[\text{G103bv} - \text{G95bv} - \text{G100bv}, 0].$$

⁶An alternative way of computing this quantity is

$$\begin{aligned} S_K &= \#312 + \#320 + \#328 + \#336 + \#352 + \#360 \\ &= \text{G96bv} + \text{G97bv} + \text{G98bv} + \text{G99bv} + \text{G101bv} + \text{G102bv}. \end{aligned}$$

It is divided into circulating capital

$$Z_{nlc1} = S_C = \max[G102bv, 0],$$

non-circulating capital

$$Z_{nlc2} = S_N = \max[S_K - S_C, 0],$$

and the rented non-land capital Z_{nlc3} determined at the end of this Section 6.4.

6.4.2 Depreciation Rate of Non-land Capital

Depreciation of non-circulating capital is given by⁷

$$d_N = SE360,$$

from which we calculate the depreciation rate as

$$r_{dK} = \frac{d_N}{S_N}.$$

To treat outliers, a similar procedure is used as with the interest rate on land. Regress

$$d_{N;ft} = \xi S_{N;ft} + u_{ft}, \quad (6.3)$$

with f denoting a farm index and t meaning time, in order to obtain $\hat{\xi}_1$. Remove observations for which $r_{dK;ft} > 5\hat{\xi}_1$ and re-regress (6.3), to obtain $\hat{\xi}_2$. Assign this value to the removed observations.

⁷Alternative ways of computing this quantity

$$\begin{aligned} SE360 &= \#300 + \#348 + \#356 \\ &= \#300 + \#356 \\ &= \#316 + \#324 + \#332 + \#340 + \#356 \\ &= G94dp + G101dp \\ &= G96dp + G97dp + G98dp + G99dp + G101dp, \end{aligned}$$

since #308 and #348 are not in use (Community Committee for the Farm Accountancy Data Network, 2005, 2007, p.22). Depreciation should not be applied to quotas (Community Committee for the Farm Accountancy Data Network, 2005, p.22, G99). [...] However any depreciation of quotas and other rights must not be applied in Table G (position 340) (Community Committee for the Farm Accountancy Data Network, 2005, p.43, L col7).

6.4.3 Interest Rate of Non-land Capital

From outstanding debt on capital (D_K), which is calculated as total debt (H106a) minus debt on land (H106c)

$$D_K = \max [H106a - H106c, 0],$$

and interests paid on capital excluding land (I_K), calculated as total interests paid (F89) minus interests paid on land (F91)

$$I_K = \max [F89 - F91, 0],$$

we calculate the interest rate on non-land capital as

$$r_{iK} = \frac{I_K}{D_K}.$$

To treat outliers, a similar procedure is used as with the depreciation rate of non-circulating capital (see expression 6.3) and the interest rate on land (see expression 6.2).

6.4.4 Opportunity Cost of Non-land Capital

As opportunity cost for capital, we take

$$r_{oK} = \text{irt_lt_gby10_a}$$

for circulating capital, and $\bar{r}_{oK}$, i.e., the country average of interest rates r_{oK} on non-land capital over the period of observations, for non-circulating capital (see Andersen et al., 2011, p.732).

The outstanding debt on capital D_K is allocated *pro rata* to circulating and non-circulating capital as

$$\begin{aligned} D_C &= S_C \frac{D_K}{S_K} \\ D_N &= S_N \frac{D_K}{S_K}. \end{aligned}$$

By the same logic, debt-free own non-land circulating and non-circulating capital are given respectively by

$$F_C = S_C^{(S_K - D_K)} / S_K$$

$$F_N = S_N^{(S_K - D_K)} / S_K.$$

To compute the rent paid on non-land assets, we subtract from the total rent paid (SE375: Rent paid for farm land and buildings and rental charges (Community Committee for the Farm Accountancy Data Network, 2007, p.23)),⁸ the amount of rent paid on rented land, which is given by R_{LR} (see expression 6.1), so

$$R_K = \max[SE375 - R_{LR}, 0].$$

The rate corresponding to rented assets consists of the sum of the opportunity cost-rate and depreciation rate: $r_{oK} + r_{dK}$.

This can be justified as follows. For rented non-land capital we observe the amount of rent paid, but not the value of the non-land assets, nor the rental rate.

In an attempt to find a proxy for the actual rental rate, consider the renting company (of equipment, for example). The annual rental rate it would charge covers:

1. their opportunity cost, which we take to be identical to the opportunity cost of farm-owned capital,
2. wear and tear, i.e., depreciation of the rented equipment,
3. a risk premium for mis-treatment,
4. overhead to run the rental company,
5. profits for stockholders.

Without any information on items 3 to 5, the rental rate is taken as the sum of the first two components. Note that in doing so, we assume that the rented capital:

1. closely resembles the non-rented capital since its depreciation rate is taken to be the depreciation rate of fully owned capital,
2. is rented the whole year round.

Finally, the rented non-land capital is given by

$$Z_{nlc3} = R_K / (r_{oK} + r_{dK}).$$

⁸Remark that there remains some ambiguity here, since SE375 (Rent paid for farm land and buildings and rental charges) refers to #285 = F85 (rent paid, but under the header "Land charges").

6.5 Estimation

The last section of the Stata routine takes care of estimations. The user is here able to choose estimation options such as the constraints to impose on the coefficients of the cost and input demand functions, in order to make it theoretically consistent. Estimations follow the procedure described in *FACEPA Deliverable 9-1* (De Blander et al., 2011), i.e., using the SGM cost function. This is outlined in Chapter 7.

Part III

Estimation of the Nested Cost and Input Demand Functions for a Selection of Farm Types and EU Regions

Introduction

This report on Part 3 details the estimation procedure and presents its applications on EU-FADN samples. First, the estimation procedure is given in details. Second, preliminary estimation results are reported and discussed for the standard SGM long-term specification at upper level of output aggregation for three Belgian EU-FADN samples: crop, dairy and livestock farms, and at lower level of output aggregation for the Belgian EU-FADN sample of crop farms.

Chapter 7

Estimation Procedure

7.1 Preliminary Remark

Sections 7.2 to 7.6 generically describe the properties of both cost functions (1.1) at the upper level and (1.6) at the lower level of Part I.

7.2 Cost Function

The specification of the cost function is based upon the standard Symmetric Generalized McFadden (SGM) cost function (Diewert and Wales, 1987; Wieck and Heckelei, 2007; Henry de Frahan et al., 2011). It is a second order Taylor approximation to the unknown total variable cost function. In that sense, the SGM specification is said to be flexible in all its arguments. Under some regularity conditions, flexible cost functions that are twice continuously differentiable in all their arguments are consistent with theory and well-behaved. This is the reason why the following sections detail how these regularity conditions can be imposed.

This cost function is to produce L_y goods, using L_x variable inputs and L_z quasi-fixed inputs

$$\begin{aligned} TC = & (\theta'W) a'Yt + (\phi'Y) b'Wt + Y'CW + Z'DW (\phi'Y) \\ & + \frac{1}{2} (\theta'W)^{-1} W'EW (\phi'Y) \\ & + (\theta'W) \left\{ Z'FZ (\phi'Y) + Y'GY + Z'HY + \sum_{l_y} \mathbf{y}_{l_y} (\mathbf{Y}'\mathbf{Q}_{l_y}\mathbf{Y}) \right\}, \quad (7.1) \end{aligned}$$

with the vector of output quantities $Y = (y_1, \dots, y_{l_y}, \dots, y_{L_y})'$, the vector of input prices $W = (w_1, \dots, w_{l_x}, \dots, w_{L_x})'$ and the vector of fixed inputs $Z = (z_1, \dots, z_{l_z}, \dots, z_{L_z})'$.

For readability, the time index $t = 1, \dots, T$ and farm index $f = 1, \dots, F$ are omitted.

The product $(\theta'W)$ can be interpreted as a fixed-weight input price index, with

$$\theta_{l_x} = T^{-1} \sum_{t=1}^T \frac{\sum_{f=1}^F x_{l_x;ft}}{\sum_{i=1}^{L_x} \sum_{f=1}^F x_{i;ft}},$$

where the vector $X = (x_1, \dots, x_{l_x}, \dots, x_{L_x})'$ denotes the vector of input quantities. The input price index is inserted to ensure first-order homogeneity in input prices (Chambers, 1988, p.52, property 2B-4).

Similarly, the product $(\phi'Y)$ can be interpreted as a fixed-weight output quantity index, with

$$\phi_{l_y} = T^{-1} \sum_{t=1}^T \frac{\sum_{f=1}^F p_{l_y;ft}}{\sum_{i=1}^{L_y} \sum_{f=1}^F p_{i;ft}},$$

where the vector $P = (p_1, \dots, p_{l_y}, \dots, p_{L_y})'$ denotes the vector of output prices. The output quantity index is inserted to ensure regularity condition $TC(Y=0, W, Z) = 0$ (Chambers, 1988, p.52, property 2B-6).

The dependent variable TC is constructed as $X'W = \sum_{l_x=1}^{L_x} x_{l_x} \cdot w_{l_x}$, with the vectors X and W the vector of input quantities, respectively prices.

Third-order terms in outputs to the cost function (bold term in equation (7.1)):

- are optional,
- can estimate cost functions for which the marginal costs are downward sloping at some of the observations (dairy and sugar quotas),
- are not considered in the rest of this text. If interested, the reader is oriented to *FACEPA Deliverable 9-1* (De Blander et al., 2011).

7.3 System of Equations

Shephard's (1970) lemma yields a set of input demand equations

$$\begin{aligned} x_{l_x} &= \frac{\partial TC}{\partial w_{l_x}} \\ &= \theta_{l_x} a' Y t + (\phi' Y) b_{l_x} t + Y' C_{l_x} + Z' D_{l_x} (\phi' Y) \\ &\quad + (\theta' W)^{-1} \left\{ W' E_{l_x} - \frac{1}{2} \theta_{l_x} (\theta' W)^{-1} W' E W \right\} (\phi' Y) \\ &\quad + \theta_{l_x} \{ Y' G Y + (\phi' Y) Z' F Z + Z' H Y \}, \end{aligned} \tag{7.2}$$

where the observed input quantities x_i are equated with the optimal input quantities $\frac{\partial TC}{\partial w_{lx}}$, i.e., those that minimize total costs. Observed input quantities x_i must be strictly positive since Shephard's lemma does not hold in a corner solution. Note that a sub-scripted matrix denotes its corresponding column.

7.4 Symmetry and Adding-up Restrictions

The obvious symmetry restrictions apply, i.e., we impose coefficients with permuted indices to be identical

$$q_{mno} \equiv q_{mon} \equiv q_{nmo} \equiv q_{omn} \equiv q_{nom} \equiv q_{onm}, \quad m, n, o = 1, \dots, L_y$$

$$e_{ij} \equiv e_{ji}, \quad i, j = 1, \dots, L_x$$

$$f_{kl} \equiv f_{lk}, \quad l, k = 1, \dots, L_z$$

$$g_{mn} \equiv g_{nm}, \quad m, n = 1, \dots, L_y$$

The adding up constraint is ensured by

$$\sum_{j=1}^{L_x} e_{ij} = 0, \quad i = 1, \dots, L_x$$

See Diewert and Wales (1987, p. 54).

These restrictions are always imposed.

7.5 Monotonicity Conditions

A well-behaved cost function should be:

- non-decreasing in output quantities (Chambers, 1988, p.52, property 2B-5),
- non-decreasing in input prices (Chambers, 1988, p.52, property 2B-2), i.e., the input demands can not be negative,
- non-increasing in fixed inputs (Chambers, 1988, p.102).

We allow the possibility to impose all monotonicity conditions. But, we do *not* recommend to impose *both* monotonicity and curvature of input prices or *both* monotonicity and curvature of fixed inputs (see below).

7.5.1 Non-negativity of every MC_{l_y} for the Standard SGM

The marginal cost for output l_y can be written as

$$\begin{aligned} MC_{l_y} &= (\theta'W) a_{l_y} t + b'W t \phi_{l_y} + C'_{l_y} W + Z' D W \phi_{l_y} \\ &\quad + \frac{1}{2} (\theta'W)^{-1} W' E W \phi_{l_y} + (\theta'W) \{ Z' F Z \phi_{l_y} \} \\ &\quad + (\theta'W) \{ 2Y' G_{l_y} + Z' H_{l_y} \}, \end{aligned} \quad (7.3)$$

where the vector $C_{l_y} = (c_{l_y1}, c_{l_y2}, \dots, c_{l_yL_x})'$, the vector $G_{l_y} = (g_{l_y1}, g_{l_y2}, \dots, g_{l_yL_y})'$ and the vector $H_{l_y} = (h_{l_y1}, h_{l_y2}, \dots, h_{l_yL_z})'$.

The l_y restrictions $MC_{l_y} \geq 0$ can be implemented as

$$\begin{aligned} c_{l_y1} \geq & -\min_{obs} \left[\frac{(\theta'W) a_{l_y} t + b'W t \phi_{l_y} + C'_{l_y(-1)} W_{(-1)}}{w_1} \right. \\ & + \frac{Z' D W \phi_{l_y} + \frac{1}{2} (\theta'W)^{-1} W' E W \phi_{l_y}}{w_1} \\ & \left. + \frac{(\theta'W) \{ Z' F Z \phi_{l_y} + 2Y' G_{l_y} + Z' H_{l_y} \}}{w_1} \right], \end{aligned}$$

where the symbol $W_{(-1)}$ denotes the vector W , with the first element removed.

In general, a “bound $c_{l_y1} \geq 0$ can be removed by defining a new parameter $\tilde{c}_{l_y1}$ which replaces c_{l_y1} , such that

$$c_{l_y1} = \tilde{c}_{l_y1}^2.$$

Then, for any $\tilde{c}_{l_y1} \in (-\infty, \infty)$ it follows that $c_{l_y1} \geq 0$, so the bound does not need to be explicitly enforced. [...] For strict constraints $c_{l_y1} > 0$ it is possible to use $c_{l_y1} = \exp(\tilde{c}_{l_y1})$. The advantage of these transformations is that they do extend the range of problems which can be handled by an unconstrained minimization routine” (Fletcher, 1993, p. 147). When the bound takes on the form $c_{l_y1} \geq \kappa$, with $\kappa \in (-\infty, \infty)$ a constant, it can be removed by defining a new parameter $\tilde{c}_{l_y1}$ which replaces c_{l_y1} , such that

$$c_{l_y1} = \tilde{c}_{l_y1}^2 + \kappa.$$

Then, for any possible value the new parameter $\tilde{c}_{l_y1}$ takes, the inequality $c_{l_y1} \geq \kappa$ is automatically fulfilled. We thus write the parameter c_{l_y1} as the sum of the constant right-hand side, κ , plus some positive amount $\tilde{c}_{l_y1}^2$, and optimize the objective function over $\tilde{c}_{l_y1}$ resulting in the estimator with respect to the new parameter $\tilde{c}_{l_y1}$. The

old parameter c_{l_y1} can be recovered by

$$c_{l_y1} = \tilde{c}_{l_y1}^2 - \min_{obs} \left[\frac{(\theta'W) a_{l_y} t + b'W t \phi_{l_y} + C'_{l_y(-1)} W_{(-1)}}{w_1} + \frac{Z'DW \phi_{l_y} + \frac{1}{2} (\theta'W)^{-1} W'EW \phi_{l_y}}{w_1} + \frac{(\theta'W) \{Z'FZ \phi_{l_y} + 2Y'G_{l_y} + Z'H_{l_y}\}}{w_1} \right]$$

and is thus guaranteed to have such a value that $MC_{l_y} \geq 0$ at all observed data points.

7.5.2 Non-negativity of every Input Demand x_{l_x}

The demand for input l_x is given in expression (7.2). Following the same reasoning as above, imposing a positive input demand can be obtained by the re-parametrization

$$c_{1l_x} = \tilde{c}_{1l_x}^2 - \min_{obs} \left[\frac{\theta_{l_x} a'Y t + (\phi'Y) b_{l_x} t + \sum_{l_y=2}^{L_y} y_{l_y} c_{l_y l_x} + Z'D_{l_x} (\phi'Y)}{y_1} + \frac{(\theta'W)^{-1} \{W'E_{l_x} - \frac{1}{2} \theta_{l_x} (\theta'W)^{-1} W'EW\} (\phi'Y)}{y_1} + \frac{\theta_{l_x} \{Y'GY + Z'F (\phi'Y) Z + Z'HY\}}{y_1} \right].$$

7.5.3 Non-increase of TC in Fixed Inputs

The derivative of (7.1) with respect to fixed input l_z is given by

$$\frac{\partial TC}{\partial z_{l_z}} = D_{l_z} W (\phi'Y) + 2 (\phi'Y) F_{l_z} Z (\phi'Y) + (\theta'W) H_{l_z} Y. \quad (7.4)$$

The restriction that it should be negative can be imposed by the re-parametrization

$$d_{l_z1} = -\tilde{d}_{l_z1}^2 - \max_{obs} \left[\frac{D'_{l_z(-1)} W_{(-1)} (\phi'Y) + 2 (\phi'Y) F_{l_z} Z (\phi'Y) + (\theta'W) H_{l_z} Y}{w_1 (\phi'Y)} \right].$$

7.6 Curvature Conditions

A well-behaved cost function fulfills the following requirements:

- concavity of TC in input prices (Chambers, 1988, p.52, property 2B-3),
- convexity of TC in fixed inputs (Chambers, 1988, p.109),

- convexity of TC in output quantities (Chambers, 1988, p.139).

We allow the possibility to impose all curvature conditions. But, we do *not* recommend to impose *both* monotonicity and curvature of input prices or *both* monotonicity and curvature of fixed inputs.¹

7.6.1 Concavity of TC in Input Prices

Concavity of the cost function in input prices (?, p. 52, property 2B-3) requires that the Hessian matrix $\frac{\partial^2 TC}{\partial w \partial w'}$ be *negative* semi-definite, a condition that holds by requiring the matrix E to be **negative semi-definite**. Now this restriction needs to be combined with the adding up constraint (see Diewert and Wales, 1987, p. 54). For example, writing a 4×4 negative semi-definite matrix E as the product of its Cholesky factors $\mathcal{L}_E$ and $\mathcal{L}'_E$

$$E = -\mathcal{L}_E \cdot \mathcal{L}'_E = \begin{pmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} & l_{11}l_{41} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} & l_{21}l_{41} + l_{22}l_{42} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 & l_{31}l_{41} + l_{32}l_{42} + l_{33}l_{43} \\ l_{11}l_{41} & l_{21}l_{41} + l_{22}l_{42} & l_{31}l_{41} + l_{32}l_{42} + l_{33}l_{43} & l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 \end{pmatrix},$$

we have that the adding up constraint results in following restrictions on the elements of the Cholesky factors

$$\begin{aligned} l_{41} &= -(l_{11} + l_{21} + l_{31}) \\ l_{42} &= -(l_{22} + l_{32}) \\ l_{43} &= -l_{33} \\ l_{44} &= 0. \end{aligned}$$

In other words, the columns of L sum to zero or, in general, $\sum_{i=1}^I l_{ij} = 0$.

7.6.2 Convexity of TC in Fixed Inputs

Convexity of the cost function in fixed inputs (Chambers, 1988, p.109) requires that the Hessian matrix $\frac{\partial^2 TC}{\partial z \partial z'}$ be *positive* semi-definite, a condition that holds by requiring the matrix F to be **positive semi-definite**, which is ensured by writing it as the

¹Since the SGM is truncated Taylor series in w and z , it is simply impossible to impose that the cost function is *globally* non-decreasing in input prices, while at the same time being concave in input prices. Wolff et al. (2004) make a similar observation.

product of its Cholesky factors:

$$F = \mathcal{L}_F \cdot \mathcal{L}'_F.$$

7.6.3 Convexity of TC in Output Quantities for the Standard SGM

Convexity of the standard SGM cost function in output quantities (Chambers, 1988, p.139) requires that the Hessian matrix $\frac{\partial^2 TC}{\partial y \partial y'}$ be *positive* semi-definite, a condition that holds by requiring the matrix G to be **positive semi-definite**, which is ensured by writing it as the product of its Cholesky factors:

$$G = \mathcal{L}_G \cdot \mathcal{L}'_G.$$

7.7 Estimation Method

The system of input demands (7.2) is estimated by a fixed effects non-linear seemingly unrelated regression (NLSUR) estimator.

Note that only observed and estimated input demands that are strictly positive are automatically used to estimate the systems of input demands 1.2, 1.20 and 1.28 of Part I, since zero input demands represent a corner solution of the farms cost minimisation problem. Consequently, Shephard's lemma does not hold and demand equations with input demands that are not strictly positive are dropped from the system of equations *for that particular farm*.

Also, note that when output categories are not present in the observations, e.g., rice in Belgium, then these output categories are automatically dropped from the estimation.

When estimated coefficients have both an infinite standard error and a value identical to its initial value, then these coefficients can be successively and automatically set to zero in an iterative process. The first estimated coefficient that appears in the list to have these two features, is first set to zero. The system of input demands is then estimated. The next first estimated coefficient that appears in the list to have these two features, is set to zero. The system of input demands is then estimated. This process is repeated until no estimated coefficient has these two features in the list. Then, if there are still coefficients with an infinite standard error, the first estimated coefficient that appears in the list to have an infinite standard error, is set to zero. The system of input demands is then estimated again. The next first estimated coefficient that appears in the list to have an infinite standard

error, is set to zero. The system of input demands is then estimated again. This process is repeated until no estimated coefficient has an infinite standard error in the list.

7.8 Rescaling the Output and Fixed Input quantities

The absolute marginal cost of output category m at time t is given by expression 7.3. As mentioned in the introduction of Part II, we estimate a cost function with output quantity variables expressed as rescaled values in the base year $y_{mt}^{(s)} = y_{mt} \cdot s_m$, where

$$y_{mt} = \frac{p_{mt} \cdot q_{mt}}{\tau_{mt}},$$

with q_{mt} the quantity of output category m at time t and the Törnqvist index $\tau_{mt} = \frac{p_{mt}}{p_{mt0}}$,

and $s_m = 10^{\left(-\text{int} \left[\log_{10} \sqrt{E[y_{mt}]^2 + \text{Var}[y_{mt}]} \right] \right)}$. The marginal cost of output category m at time t , expressed in nominal currency units per output quantity, is retrieved as

$$\begin{aligned} \frac{\partial TC}{\partial q_{mt}} &= \frac{\partial TC}{\partial y_{mt}} \cdot \frac{\partial y_{mt}}{\partial y_{mt}^{(s)}} \cdot \frac{\partial y_{mt}^{(s)}}{\partial q_{mt}} \\ &= \frac{\partial TC}{\partial y_{mt}} \cdot s_m \cdot \frac{p_{mt}}{\tau_{mt}} \\ &= \frac{\partial TC}{\partial y_{mt}} \cdot s_m \cdot p_{mt0}, \end{aligned}$$

where $\frac{\partial TC}{\partial y_{mt}}$ represents the derivative of the actually estimated cost function, i.e., using rescaled values in the base-year for each y_m . The marginal cost relative to output price p_{mt} is given by

$$MCp_{mt} = \frac{MC_{mt}}{p_{mt}}.$$

This rescaling applies for the derivation of total cost with respect to output quantity to obtain the marginal cost per output quantity as shown above, but also for the derivation of marginal cost with respect to output quantity, the calculation of the average cost per unit of output quantity, the derivation of input demand with respect to output quantity to obtain the marginal input demand per output quantity, and the calculation of the unit cost of input per output quantity. The rescaling applies for the derivation of total cost with respect to fixed in quantity since fixed input quantities are rescaled as output quantities (see the introduction

of Part II).

7.9 Further Remarks

In order to clarify the link between this Chapter and Chapter 1, notice that equation (7.1) corresponds with equation (1.1) of Part I, when we consider estimation at the upper level of output. However, when estimating at the lower level of output, it corresponds with equation (1.19) of Part I. At both levels, different sets of assumptions can be imposed on the cost function.

Similarly, the indirect input expenditure function (1.29) of Part I can be specified using the standard or augmented SGM functional form and applying the same regularity conditions as exposed in Sections (7.4) to (7.6). This has the twofold advantages of using (i) a flexible function for describing the indirect expenditure functions and their corresponding input demands at the lower level of input aggregation and (ii) the same Stata codes for implementation, in particular for imposing these regularity conditions. Notice that using a flexible functional form in this case implies that elasticities of substitution among inputs within the same input nest can differ.

In that case, the indirect expenditure function for upper-level input i using L_x^i lower-level variable inputs takes this form

$$\begin{aligned} E_i = & (\theta'_i W_i) a'_i X_i t + (\phi'_i X_i) b'_i W_i t + X'_i C_i W_i \\ & + \frac{1}{2} (\theta'_i W_i)^{-1} W'_i E_i W_i (\phi'_i X_i) \\ & + (\theta'_i W_i) \{X'_i G_i X_i\} \quad \forall i, \end{aligned} \quad (7.5)$$

with the vector of input quantities $X_i = (x_i)'$ and the vector of input prices $W_i = (w_{i,1}, \dots, w_{i,l_x}, \dots, w_{i,L_x^i})'$. For readability, the time index $t = 1, \dots, T$ and farm index $f = 1, \dots, F$ are omitted.

The product $(\theta'_i W_i)$ can be interpreted as a fixed-weight input price index, with

$$\theta_{i,l_x} = T^{-1} \sum_{t=1}^T \frac{\sum_{f=1}^F x_{i,l_x;ft}}{\sum_{j=1}^{L_x^i} \sum_{f=1}^F x_{i,j;ft}}, \quad (7.6)$$

where the vector $X_i = (x_{i,1}, \dots, x_{i,l_x}, \dots, x_{i,L_x^i})'$ denotes the vector of lower-level input quantities. The input price index is inserted to ensure first-order homogeneity in input prices.

Similarly, the product $(\phi'_i X_i)$ can be interpreted as a fixed-weight input quantity index, with

$$\phi_i = T^{-1} \sum_{t=1}^T \frac{\sum_{f=1}^F w_{i;ft}}{\sum_{i=i}^i \sum_{f=1}^F w_{i;ft}} = 1, \quad (7.7)$$

where the vector $W_i = (w_i)'$ denotes the vector of input sub-category prices. The input quantity index is inserted to ensure regularity condition $E_i(X_i = 0, W_i) = 0$.

The dependent variable E_i is constructed as $X'_i W_i = \sum_{l_x=1}^{L_x^J} x_{i,l_x} \cdot w_{i,l_x}$, with the vectors X_i and W_i the vector of input sub-category quantities, respectively prices.

Shephard's (1970) lemma yields a set of lower-level input demand equations

$$\begin{aligned} x_{i,l_x} &= \frac{\partial E_i}{\partial w_{i,l_x}} \\ &= \theta_{il_x} a'_i X_i t + (\phi'_i X_i) b_{il_x} t + X'_i C_{il_x} \\ &\quad + (\theta'_i W_i)^{-1} \left\{ W'_i E_{il_x} - \frac{1}{2} \theta_{il_x} (\theta'_i W_i)^{-1} W'_i E_i W_i \right\} (\phi'_i X_i) \\ &\quad + \theta_{il_x} \{ X'_i G_i X_i \}, \end{aligned} \quad (7.8)$$

where the observed input quantities x_{i,l_x} are equated with the optimal input quantities $\frac{\partial E_i}{\partial w_{i,l_x}}$, i.e., those that minimize total expenditure. Observed input quantities x_{i,l_x} must be strictly positive since Shephard's lemma does not hold in a corner solution. Note that a sub-scripted matrix denotes its corresponding column.

Note that for the indirect input expenditure function $E_{i;m}$ of Part I for upper-level input i and output m , the fixed-weight input price index $\theta_{i,l_x;m}$

$$\theta_{i,l_x;m} = T^{-1} \sum_{t=1}^T \frac{\sum_{f=1}^F \hat{x}_{i,l_x;m;ft}}{\sum_{j=1}^{L_x^i} \sum_{f=1}^F \hat{x}_{i,j;m;ft}}, \quad (7.9)$$

and the fixed-weight quantity index $\phi_{i;m,n}$

$$\phi_{i;m,l_y} = T^{-1} \sum_{t=1}^T \frac{\sum_{f=1}^F w_{i;m,l_y;ft}}{\sum_{n=1}^{L_y^m} \sum_{f=1}^F w_{i;m,n;ft}} = \frac{1}{L_y^m}. \quad (7.10)$$

Chapter 8

Selected Estimation Results

8.1 Introduction

This estimation procedure is tested using different EU-FADN samples covering a large panel of EU NUTS-1 regions or member states over the period from 1990 to 2008. In general, the selected regions or member states as reported in Table 8.1 are those that produce the largest amounts of cereals under the crop farm category or dairy products under the dairy farm category, or have the largest herds of non-dairy cows under the cattle farm category according to Eurostat in 2002.¹ The EU-FADN data set for Belgium is first used to test the whole procedure because of the empirical knowledge that the authors have with the agricultural sector of that particularly member state, so that knowledge should help interpret the preliminary results. In total 24 regions or members states are used.

With the EU-FADN data set for Belgium, both the standard and the augmented SGM specifications of the cost function are tested in both their medium- and long-term specifications on the three EU-FADN samples of crop, dairy and cattle farms. This leads to $(2 \times 2 \times 3)$ 12 preliminary estimations in total. Since it is preferable that simulations use long-term specifications, we decide to continue the test using the long-term specification to have a more manageable number of tests to carry. The standard or augmented SGM long-term specifications are not systematically better from each other in terms of the number of significant coefficients, except for the sample with cattle farms where the standard SGM long-term specification is better. The cubic terms of the augmented SGM long-term specifications are,

¹The crop farm sample includes types of farms (TF) 1110, 1120, 1130, 1210, 1220, 1243 and 1244 up to 1993 and TF 1310, 1320, 1330, 1410, 1420 or 1443 from 1994. The dairy farm sample includes types of farms (TF) 4110, 4120 or 4310. The cattle farm sample includes types of farms (TF) 4210, 4220 or 4320. These samples contain farms for which at least two observations are present from 1990 to 2008.

Selected NUTS-1 Regions and Member States (by decreasing order of production importance)		
Crop Farm	Dairy Farm	Cattle Farm
FR2 (Centre)	FR5 (Ouest)	FR6 (Sud-Ouest)
FR5 (Ouest)	DE2 (Bayern)	FR2 (Centre)
ES4 (Centro)	FR2 (Centre)	ES4 (Centro)
FR6 (Sud-Ouest)	DE9 (Niedersachsen)	FR5 (Ouest)
DE2 (Bayern)	FR3 (Nord-Ouest)	FR7 (Centre-Est)
FR4 (Est)	NL2 (Oost)	ES2 (Noroeste)
FR3 (Nord-Ouest)	ES2 (Noroeste)	BEL (Belgium)
DE9 (Niedersachsen)	BEL (Belgium)	
BEL (Belgium)		
9	8	7

Table 8.1: Selected NUTS-1 Regions and Member States

however, generally significant for the three farm samples. The number of significant coefficient is generally greater for the dairy farm sample, than the livestock farm sample and then the crop farm sample probably because of a decreasing number of observations. We, therefore, decide to continue the test using the standard SGM specification to reduce further the number of tests. Thereafter, these tests are carried on the 24 regions and member states reported in Table 8.1 using the standard SGM long-term specification. To make it short, we however report below only key estimation results for the EU-FADN crop farm samples from the three most important cereal regions of the EU: FR2 (Centre), FR5 (Ouest) and ES4 (Centro). We report first key estimations results for the cost function with upper-level input and output categories, second for the cost function with upper-level input categories and lower-level output categories, third for the expenditure function with lower-level input categories and upper -level output categories, and fourth for the expenditure function with lower-level input and output categories. The full set of estimation results for these three regions as well as for the other 21 regions or member states are available from the authors upon request.

8.2 Estimated Cost Functions with Upper-level Inputs and Outputs for Crop Farms

For the long-term specification for the crop farm sample, the upper-level output categories y include:

- a for the animal outputs (milk, dairy products, fattening calves, cattle, heifers cow, cows, goats, sheep, pigs, poultry, other animals),
- b for dry pulses & oil seeds (dry pulses, oil seeds),
- c for the industrial crops (potatoes, sugar beet, other industrial crops),
- d for the cereals (wheat, maize, rice, other cereals),
- e for other crops (hops, tobacco, other crops).

The upper-level input categories x include:

- a for the animal-specific inputs (breeding livestock, purchased feeds, other specific livestock costs, grass land),
- b for the crop-specific inputs (seeds, fertilizers, pesticides, other specific crop costs, crop land),

- c for the other inputs (services, insurance & taxes, other inputs),
- d for the non-land capital inputs (capital as intermediate input).

The upper-level input category z includes:

- a for the unpaid labour input.

We now present some key results from the estimations of the cost and input demand functions at upper levels.

8.2.1 Monotonicity Restrictions

All the theoretical restrictions are imposed on the the standard SGM long-term specification except the monotonicity of the cost and expenditure functions with respect to variable input prices and fixed inputs.

Table 8.2 shows to what extent these monotonicity restrictions are violated across the three crop farm samples. Violation of the monotonicity of the cost function with respect to upper-level input prices imply negative input demands, an impossible situation. Observations with negative estimated upper-level input demands are disregarded in the following calculations.

In general, across the 24 samples, violation of the monotonicity of the cost function with respect to the upper-level fixed input is more frequent than violation of the monotonicity of the cost function with respect to upper-level input prices. Violation of the monotonicity of the cost function with respect to the upper-level fixed input implies a negative shadow value of the fixed input, an unexpected situation that calls for further investigation.

At the upper level of output aggregation, the sum of the estimated input demands by upper-level output category is well equal to the estimated aggregated upper-level input demand. The sum of the estimated costs by upper-level output category is well equal to the estimated aggregated total cost.

Monotonicity violation at upper level of inputs and outputs (%)			
Function	FR2	FR5	ES4
Xa animal-specific, incl. grass land	1	6	0
Xb crop-specific, incl. crop land	0	0.2	0
Xc intermediate	15	25	3
Xd non-land capital	2	26	60
Z non-paid labour	0	45	55

Table 8.2: Monotonicity violation for upper-level input and output estimation

8.2.2 Significance Level and Goodness-of-fit for Input Demand and Cost Functions

Table 8.3 shows the percentage of significant coefficients and the goodness-of-fit of regressing the estimated upper-level input demands with respect to the observed upper-level input uses as well as the goodness-of-fit of regressing the calculated total cost from those estimated input demands with respect to the observed total costs. The sample size over the whole period is also indicated.

Significance and goodness-of-fit for upper-level input demand and cost functions for all upper-level outputs (adj-R ²)			
Function	FR2	FR5	ES4
% of significant coefficients	45	42	25
Xa animal-specific, incl. grass land	0.36	0.34	0.27
Xb crop-specific, incl. crop land	0.72	0.79	0.65
Xc intermediate	0.01	0.41	0.66
Xd non-land capital	0.62	0.35	0.09
TC total cost	0.71	0.71	0.76
Sample size over the period	16095	2934	20609

Table 8.3: Significance and goodness-of-fit for input demand and cost functions from upper-level input and output estimation

8.2.3 Own Elasticities of Input Demands and Marginal Costs

Table 8.4 reports the medians of the elasticities of upper-level input demands with respect to their own price as well as the medians of the elasticities of upper-level marginal costs with respect to their output quantity. Medians are given here and subsequently because of the non normality of the distribution of the estimations results. The standard deviations can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results as a consequence of using a second-order flexible cost and demand functions and the heterogeneity in the data. Own price elasticities of input demand are plausible while own output elasticities of marginal cost are particularly low implying very large price elasticities of supply. The full matrices of elasticities of upper-level input demands and marginal costs are available upon request.

Own elasticities at upper level of inputs and outputs (median)			
Input demand for X / Marginal cost for Y	FR2	FR5	ES4
Xa animal-specific, incl. grass land	-0.97	-2.52	-0.42
Xb crop-specific, incl. crop land	-0.02	-0.08	-0.00
Xc intermediate	-0.26	-0.14	-0.29
Xd non-land capital	-0.42	-0.66	-0.34
Ya animal products	0.01	0.00	0.04
Yb pulse & oil-seeds	0.00	0.02	0.00
Yc roots	0.00	0.00	0.17
Yd cereals	0.01	0.05	0.00
Ye other crops	0.07	0.06	0.01

Table 8.4: Own elasticities from upper-level input and output estimation

8.2.4 Marginal Costs over Output Price

Table 8.5 reports the medians of the ratio of the marginal cost of one upper-level output over the price of the same output. A ratio inferior to one indicates a profit margin for the unpaid labour or some other farm assets while a ratio superior to one indicates a loss for at least the last unit of output that can possibly be compensated from the profit margins realised on other outputs. Again, medians are given here because of the non normality of the distribution of the estimations results. The standard deviation can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results reflecting the heterogeneity in the data. Average costs over price are close to the marginal costs over prices.

Marginal costs over output price for upper-level outputs (median)			
Marginal cost / price for Y	FR2	FR5	ES4
Ya animal products	0.25	0.18	0.21
Yb pulse & oil-seeds	0.55	0.62	0.94
Yc roots	0.21	0.32	0.24
Yd cereals	0.59	0.45	0.49
Ye other crops	0.47	0.48	0.36

Table 8.5: Marginal costs over output price from upper-level input and output estimation

8.2.5 Unit Costs for Cereals

Table 8.6 reports the medians of the unit costs for the cereal upper-level output category. As expected, the unit costs of crop-specific inputs, including crop land, for

cereals are larger than those of the other input categories. A median of zero unit cost only implies that the unit cost can still be positive for some farms. Again, median is given here because of the non normality of the distribution of the estimations results. The standard deviation can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results reflecting the heterogeneity in the data. Unit costs for other upper-level output categories are available upon request.

Unit costs for cereals in €/ton (median)			
Function	FR2	FR5	ES4
Xa animal-specific, incl. grass land	7	38	10
Xb crop-specific, incl. crop land	34	91	99
Xc intermediate	10	0	31
Xd non-land capital	18	9	0

Table 8.6: Unit costs for cereals from upper-level input and output estimation

8.3 Estimated Cost functions with Upper-level Inputs and Lower-level Outputs for Crop Farms

For all the long-term costs functions with lower-level output categories for crop farms, the upper-level input categories x include:

- a for the animal-specific inputs (breeding livestock, purchased feeds, other specific livestock costs, grass land),
- b for the crop-specific inputs (seeds, fertilizers, pesticides, other specific crop costs, crop land),
- c for the other inputs (services, insurance & taxes, other inputs),
- d for the non-land capital inputs (capital as intermediate input).

The upper-level fixed-input category z includes:

- a for the unpaid labour input.

The lower-level output categories y for the dry pulse & oil seed output category include:

- drp for dry pulses,
- oil for oil seeds.

The lower-level output categories y for the industrial crop output category include:

- *ptt* for potatoes,
- *sbt* for sugar beet,
- *oin* for other industrial crops.

The lower-level output categories y for the cereal output category include:

- *wht* for wheat,
- *mze* for maize,
- *rce* for rice (not always available),
- *ocr* for other cereals.

The lower-level output categories y for the other crop output category include:

- *hps* for hops,
- *tbc* for tobacco,
- *otc* for other crops.

The lower-level output categories y for the animal outputs also include:

- *a* for the bovine milk & dairy product outputs (milk & dairy products),
- *b* for other bovine outputs (fattening calves, cattle, heifers cow, cows),
- *c* for the other non-bovine outputs (goats, sheep, pigs, poultry, other animals).

We now report some key results from the estimations of the cost and input demand functions for the cereal upper-level output category only.

8.3.1 Monotonicity Restrictions

Again, all the theoretical restrictions are imposed except the monotonicity of the cost function with respect to variable input prices and fixed inputs.

Table 8.7 shows to what extent these monotonicity restrictions are violated across the three crop farm samples for the cereal upper-level output category. Violation of the monotonicity of the cost function with respect to upper-level input prices imply negative input demands, an impossible situation. Observations with negative estimated upper-level input demands are here also disregarded in the following calculations.

In general, across the 24 samples, violation of the monotonicity of the cost function with respect to the upper-level fixed input is more frequent than violation of the monotonicity of the cost function with respect to upper-level input prices. Violation of the monotonicity of the cost function with respect to the upper-level fixed input implies a negative shadow value of the fixed input, an unexpected situation that calls for further investigation.

At the lower level of output aggregation, the sum of the estimated upper-level input demands by lower-level output category is well equal to the estimated aggregated upper-level input demand by upper-level output category. The sum of the estimated costs by lower-level output category is well equal to the estimated aggregated upper-level cost.

Monotonicity violation at upper level of inputs and lower level of outputs (%)			
Function	FR2	FR5	ES4
Xa animal-specific, incl. grass land	1	0	0
Xb crop-specific, incl. crop land	0	1	0
Xc intermediate	100	4	0
Xd non-land capital	4	35	4
Z non-paid labour	34	40	98

Table 8.7: Monotonicity violation for upper-level input and output estimation

8.3.2 Significance Level and Goodness-of-fit for Input Demand and Cost Functions

Table 8.8 shows the percentage of significant coefficients and the goodness-of-fit of regressing the estimated upper-level input demands by upper-level output category with respect to the pseudo-observed input uses by upper-level output category as well as the goodness-of-fit of regressing the calculated total cost by upper-level output category from those estimated input demands with respect to the pseudo-observed total costs by upper-level output category. The sample size over the whole period is also indicated.

8.3.3 Own Elasticities of Input Demands and Marginal Costs

Table 8.9 reports the medians of the elasticities of upper-level input demands with respect to their own price for the cereal upper-level output category as well as the medians of the elasticities of lower-level marginal costs with respect to their output quantity. Medians are given because of the non normality of the distribution of the

Significance and goodness-of-fit for upper-level input demand and cost functions for cereals (adj-R ²)			
Function	FR2	FR5	ES4
% of significant coefficients	82	72	55
Xa animal-specific, incl. grass land	0.56	0.79	0.99
Xb crop-specific, incl. crop land	0.93	0.97	0.99
Xc intermediate	0.00	0.61	0.95
Xd non-land capital	0.95	0.40	0.04
TC total cost for cereals	0.94	0.94	0.78
Sample size over the period	16095	2934	20609

Table 8.8: Significance and goodness-of-fit for input demand and cost functions from upper-level input and lower-level output estimation

estimations results. The standard deviations can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results as a consequence of using a second-order flexible cost and demand functions and the heterogeneity in the data. Own price elasticities of input demand are plausible while own output elasticities of marginal cost are again particularly low implying very large price elasticities of supply. The full matrices of elasticities of upper-level input demands by upper-level output and lower-level marginal costs are available upon request.

Own elasticities at upper level for input demands and lower level for marginal costs (median)			
Input demand for X / Marginal cost for Y	FR2	FR5	ES4
Xa animal-specific, incl. grass land	-3.20	-1.90	-1.10
Xb crop-specific, incl. crop land	-0.04	-0.21	-0.03
Xc intermediate	-0.34	-0.17	-0.01
Xd non-land capital	-0.91	-1.06	-0.60
Ywheat	0.02	0.01	0.01
Ymaize	0.05	0.02	0.01
Yother cereals	0.12	0.06	0.00

Table 8.9: Own elasticities from upper-level input and lower-level output estimation

8.3.4 Marginal Costs over Output Price

Table 8.10 reports the medians of the ratio of the marginal cost of one lower-level output over the price of the same output. A ratio inferior to one indicates a profit margin for the unpaid labour or some other farm assets while a ratio superior to one indicates a loss for at least the last unit of output that can possibly be compensated from the profit margins realised on other outputs. Again, medians are given here

because of the non normality of the distribution of the estimations results. The standard deviation can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results reflecting the heterogeneity in the data. Average costs over price are close to the marginal costs over prices.

Marginal costs over output price for lower-level outputs of the cereal upper-level output category (median)			
Marginal cost / price for Y	FR2	FR5	ES4
Ywheat	0.66	0.42	0.47
Ymaize	0.26	0.43	0.44
Yother cereals	0.33	0.55	0.47

Table 8.10: Marginal costs over output price from upper-level input and lower-level output estimation

8.3.5 Unit Costs for Wheat

Table 8.11 reports the medians of the unit costs for the wheat lower-level output category. As expected, the unit costs of crop-specific inputs, including crop land, for wheat are generally larger than those of the other input categories. A median of zero unit cost only implies that the unit cost can still be positive for some farms. Again, median is given here because of the non normality of the distribution of the estimations results. The standard deviation can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results reflecting the heterogeneity in the data. Unit costs for other lower-level output categories are available upon request.

Unit costs for wheat in €/ton (median)			
Function	FR2	FR5	ES4
Xa animal-specific, incl. grass land	23	11	5
Xb crop-specific, incl. crop land	27	33	48
Xc intermediate	0	4	14
Xd non-land capital	31	3	5

Table 8.11: Unit costs for wheat from upper-level input and lower-level output estimation

8.4 Estimated Expenditure Functions with Lower-level Inputs and Upper-level Outputs for Crop Farms

For the long-term specification for the crop farm sample, the upper-level output categories y include:

- a for the animal outputs (milk, dairy products, fattening calves, cattle, heifers cow, cows, goats, sheep, pigs, poultry, other animals),
- b for dry pulses & oil seeds (dry pulses, oil seeds),
- c for the industrial crops (potatoes, sugar beet, other industrial crops),
- d for the cereals (wheat, maize, rice, other cereals),
- e for other crops (hops, tobacco, other crops).

The lower-level input categories x for the the animal-specific upper-level input category include:

- 1: breeding livestock,
- 2: purchased feeds,
- 3 : other specific-livestock costs,
- 12 : grass land.

The lower-level input categories x for the the crop-specific upper-level input category include:

- 4: purchased seeds,
- 5: fertilizers,
- 6: pesticides,
- 7: other specific-crop costs,
- 11 : crop land.

The lower-level input categories x for the other input upper-level input category include:

- 8 : services,
- 9: insurance and taxes,
- 10: other inputs.

The lower-level input categories x for the non-land capital upper-level input category include:

- 13: non-land capital

No fixed input is considered in the expenditure function.

We now report some key results from the estimations of the expenditure and input demand functions for the crop-specific lower-level input category only.

8.4.1 Monotonicity Restrictions

Again, all the theoretical restrictions are imposed except the monotonicity of the expenditure function with respect to variable lower-level input prices.

Table 8.12 shows to what extent this monotonicity restriction is violated across the three crop farm samples for the expenditure function on the crop-specific inputs. Violation of the monotonicity of the expenditure function with respect to lower-level input prices imply negative input demands, an imposible situation. Observations with negative estimated lower-level input demands are here also disregarded in the following calculations.

At the lower level of input aggregation, the sum of the estimated lower-level input demands by upper-level output category is well equal to the pseudo-observed aggregated upper-level input demand by upper-level output category.

Monotonicity violation at lower level of inputs and upper level of outputs (%)			
Function	FR2	FR5	ES4
X4 seeds	93	94	0
X5 fertilizers	4	1	41
X6 pesticides	0	6	4
X7 other crop specific	2	1	42
X11 crop land	0	0	0

Table 8.12: Monotonicity violation for lower-level input and upper-level output estimation

8.4.2 Significance Level and Goodness-of-fit for Input Demand and Expenditure Functions

Table 8.13 shows the percentage of significant coefficients and the goodness-of-fit of regressing the estimated lower-level input demands by upper-level output category with respect to the pseudo-observed lower-level input uses by upper-level output category as well as the goodness-of-fit of regressing the calculated total expenditure on the upper-level input category by upper-level output category from those estimated lower-level input demands with respect to the pseudo-observed total expenditure on the upper-level input category by upper-level output category. The sample size over the whole period is again provided.

Significance and goodness-of-fit for lower-level input demand and expenditure functions for all upper-level output categories (adj-R ²)			
Function	FR2	FR5	ES4
% of significant coefficients	32	32	32
X4 seeds	0.04	0.00	0.43
X5 fertilizers	0.39	0.41	0.21
X6 pesticides	0.48	0.39	0.31
X7 other crop specific	0.13	0.46	0.06
X11 crop land	0.39	0.42	0.49
TC total expenditure on Xb	0.84	0.76	0.83
Sample size over the period	16095	2934	20609

Table 8.13: Significance and goodness-of-fit for input demand and cost functions from lower-level input and upper-level output estimation

8.4.3 Own Elasticities of Input Demands

Table 8.14 reports the medians of the elasticities of lower-level input demands with respect to their own price for the cereal upper-level output category. Medians are given because of the non normality of the distribution of the estimations results. The standard deviations can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results as a consequence of using a second-order flexible expenditure and demand functions and the heterogeneity in the data. Own price elasticities of lower-level input demand are plausible. The full matrices of elasticities of lower-level input demands by upper-level output are available upon request.

Own elasticities at lower level of input demands for cereals (median)			
Input demand for X	FR2	FR5	ES4
X4 seeds	-0.00	-0.04	-0.00
X5 fertilizers	-0.11	-0.10	-0.09
X6 pesticides	-0.10	-0.20	-0.00
X7 other crop specific	-0.11	-2.30	-0.83
X11 crop land	-0.01	-0.00	-0.12

Table 8.14: Own elasticities from lower-level input and upper-level output estimation

8.5 Estimated Expenditure Functions with Lower-level Inputs and Outputs for Crop Farms

For the long-term specification for the crop farm sample, the lower-level output categories y for the dry pulse & oil seed output category include:

- *drp* for dry pulses,
- *oil* for oil seeds.

The lower-level output categories y for the industrial crop output category include:

- *ptt* for potatoes,
- *sbt* for sugar beet,
- *oin* for other industrial crops.

The lower-level output categories y for the cereal output category include:

- *wht* for wheat,
- *mze* for maize,
- *rce* for rice (not always available),
- *ocr* for other cereals.

The lower-level output categories y for the other crop output category include:

- *hps* for hops,
- *tbc* for tobacco,
- *otc* for other crops.

The lower-level output categories y for the animal outputs also include:

- a for the bovine milk & dairy product outputs (milk & dairy products),
- b for other bovine outputs (fattening calves, cattle, heifers cow, cows),
- c for the other non-bovine outputs (goats, sheep, pigs, poultry, other animals).

The lower-level input categories x for the the animal-specific upper-level input category include:

- 1: breeding livestock,
- 2: purchased feeds,
- 3 : other specific-livestock costs,
- 12 : grass land.

The lower-level input categories x for the the crop-specific upper-level input category include:

- 4: purchased seeds,
- 5: fertilizers,
- 6: pesticides,
- 7: other specific-crop costs,
- 11 : crop land.

The lower-level input categories x for the other input upper-level input category include:

- 8 : services,
- 9: insurance and taxes,
- 10: other inputs.

The lower-level input categories x for the non-land capital upper-level input category include:

- 13: non-land capital

No fixed input is considered in the expenditure function.

We now report some key results from the estimations of the expenditure and input demand functions for the crop-specific lower-level input category spent on the wheat lower-level output category only.

8.5.1 Monotonicity Restrictions

Again, all the theoretical restrictions are imposed except the monotonicity of the expenditure function with respect to variable lower-level input prices.

Table 8.15 shows to what extent this monotonicity restriction is violated across the three crop farm samples for the expenditure function of wheat lower-level output category on the crop-specific inputs. Violation of the monotonicity of the expenditure function with respect to lower-level input prices imply negative input demands, an impossible situation. Observations with negative estimated lower-level input demands are here also disregarded in the following calculations.

At the lower level of input and output aggregation, the sum of the estimated lower-level input demands by lower-level output category is well equal to the pseudo-observed aggregated upper-level input demand by lower-level output category.

Monotonicity violation at lower level of inputs and outputs (%)			
Function	FR2	FR5	ES4
X4 seeds	0	24	0
X5 fertilizers	0	0	67
X6 pesticides	0	6	0
X7 other crop specific	2	0	2
X11 crop land	0	0	0

Table 8.15: Monotonicity violation for lower-level input and output estimation

8.5.2 Significance Level and Goodness-of-fit for Input Demand and Expenditure Functions

Table 8.16 shows the percentage of significant coefficients and the goodness-of-fit of regressing the estimated lower-level input demands by lower-level output category with respect to the pseudo-observed lower-level input uses by lower-level output category as well as the goodness-of-fit of regressing the calculated total expenditure on the upper-level input category by lower-level output category from those estimated lower-level input demands with respect to the pseudo-observed total expenditure on the upper-level input category by lower-level output category. The sample size over the whole period is again provided.

Significance and goodness-of-fit for lower-level input demand and expenditure functions for cereal upper-level output (adj-R ²)			
Function	FR2	FR5	ES4
% of significant coefficients	64	69	71
X4 seeds	0.41	0.36	0.98
X5 fertilizers	0.86	0.90	0.06
X6 pesticides	0.91	0.90	0.99
X7 other crop specific	0.92	0.70	0.99
X11 crop land	0.90	0.90	0.99
TC total expenditure on Xb for cereals	0.92	0.85	0.99
Sample size over the period	16095	2934	20609

Table 8.16: Significance and goodness-of-fit for input demand and cost functions from lower-level input and output estimation

8.5.3 Own Elasticities of Input Demands

Table 8.17 reports the medians of the elasticities of lower-level input demands with respect to their own price for the wheat lower-level output category. Medians are given because of the non normality of the distribution of the estimations results. The standard deviations can range from 20 to 50 percent of the mean, implying a wide heterogeneity of the results as a consequence of using a second-order flexible expenditure and demand functions and the heterogeneity in the data. Own price elasticities of lower-level input demand for wheat are plausible. The full matrices of elasticities of lower-level input demands by lower-level output are available upon request.

Own elasticities at lower level of input demands for wheat (median)			
Input demand for X	FR2	FR5	ES4
X4 seeds	-0.01	-0.01	-0.00
X5 fertilizers	-0.04	-0.47	-1.11
X6 pesticides	-0.68	-0.49	-0.03
X7 other crop specific	-0.11	-2.25	-2.40
X11 crop land	-0.16	-0.83	-0.96

Table 8.17: Own elasticities from lower-level input and output estimation

8.6 Preliminary Conclusions

These tests show that it is possible to estimate standard and augmented SGM specifications of cost and expenditure functions for both the medium- and long-

term forms at upper level of output and input aggregation for three Belgian EU-FADN samples: crop, dairy and cattle farms from 1990 to 2008. They also show that it is possible to estimate standard SGM specifications of cost and expenditure functions for the long-term form at upper and lower levels of both input and output aggregations for 24 NUTS1 regions and member states for crop, dairy and cattle farm samples of the EU-FADN from 1990 to 2008.

Estimation results show that:

- the monotonicity restrictions of the cost and expenditure functions are not consistently fully met, which leads to negative input demands,
- the number of significant coefficients and goodness-of-fit are acceptable for most estimations,
- there is a large heterogeneity in the calculated results as a consequence of the second-order flexibility of the functions and the heterogeneity in the data,
- the input demand elasticities are plausible but marginal cost elasticities are low,
- the unit costs per output calculated from the estimations at upper and lower-level output aggregation are also plausible.

Part IV

Conclusions and Recommendations

Introduction

This last Part discusses the following four issues: (i) main challenges encountered in this project with respect to estimation of the multi-input multi-output cost functions (e.g., data issues, aggregation of inputs and outputs, robustness of the results, and computational requirements), (ii) the strengths and limitations of extending the proposed multi-input multi-output cost function approach for the whole EU-28 member states (e.g., data issues, sample sizes, feasibility to automate the regression process and the solving for problematic coefficients), (iii) and improvements and alternative approaches that may be considered to address the perceived limitations.

Chapter 9

Main Challenges

9.1 Data Issues

Most of farm data to estimate the multi-input multi-output cost functions and its related input demands are found in the EU-FADN data set. Most farm-gate prices for outputs can be calculated as unit values (i.e., total value divided by quantity) using farm data from the EU-FADN data set as discussed in Chapter 5. Since these farm-gate prices for outputs show a large variability across farms within the same year, we prefer to use output price indices calculated at the regional or national level from EUROSTAT statistics as discussed in Chapter 3. In contrast, most farm-gate prices for inputs cannot be calculated as unit values except for breeding animals (i.e., total value divided by livestock units of animals) as discussed in Chapter 6. We use input price indices calculated at the regional or national level from EUROSTAT statistics as discussed in Chapter 3.¹ As a drawback of using price indices defined at the regional or national level, there is no more variability in prices across farms within the same region or member state and year.

Valuing farm land and capital at farm gate is possible under several hypothesis as discussed in Sections 6.3 and 6.4. However, valuing farm unpaid labour input as reported in the EU-FADN is more controversial. As a result, we prefer to leave that farm input as fixed in the long-term specification.

Outliers of farm-gate input and output prices as defined in Chapter 3 are removed from the calculation of average regional or national input and output prices.

¹Some of those input price indices are missing for some member states, e.g., for the UK.

9.2 Aggregation of input and output categories

Of course, given the number of observations (i.e., the sample size) available for the estimation, the number of input and output categories that can be considered in a multi-input and multi-output cost function for econometric estimation depends on the number of parameters of this function to estimate and, hence, the order of flexibility of the function. For a second-order flexibility as the standard SGM functional form, it is reasonable to consider fewer than four to five variable input categories and four to five output categories. For a third-order flexibility as the augmented SGM functional form, it becomes reasonable to reduce the number of variable input and output categories to three to four categories each. This, however, all depends on the sample size. Note that the number of input and output categories can be extended with a higher number of lower levels of aggregation.

The nested approach that is used to circumvent the lack of degrees of freedom if it is for estimating many parameters for many input and output categories for a flexible form, rests on key separability assumptions that have been discussed in Section 1.1.

9.3 Robustness of the Results

To test for robustness of the estimated results, we need to use different economic models and compare the estimated results, which we did not.

As reported in Section 7.7, estimated coefficients can have values that are identical to their initial values and/or show infinite standard error. If this is the case, the model exhibits insufficient curvature in those coefficients to allow meaningful estimations. This may result from the many theoretical restrictions that are imposed on the estimation process. In particular, the monotonicity restrictions introduce non-linearity in the estimation process. Removing them does not, however, alter the estimation results. Insufficient curvature in coefficients may also result from multicollinearity between regressors. At the end of the iteration process that removes successively estimated coefficients with values that are identical to their initial values and/or show infinite standard error, we notice that coefficients for quadratic terms are removed.

Without the monotonicity restrictions of the cost function on input prices, output quantities, and fixed inputs, we notice that monotonicity on fixed inputs is largely violated which implies many negative estimated shadow values for fixed inputs, monotonicity on input prices moderately violated which implies some negative estimated input demands, and monotonicity on output quantities marginally vio-

lated which implies few negative estimated marginal costs. The Shephard's lemma, however, applies only for strictly positive fixed-input shadow values, input demands and marginal costs given the properties of cost functions (Chambers, 1988, p. 57).

We also notice that marginal costs often depart from their output prices, an indication that a profit margin is available for unpaid labour and other farm assets. Some farms also may not at their maximum profit. There is also a large variability in marginal costs across farms, an indication that farms are economically heterogenous in the sample.

9.4 Computational Requirements

The estimation of the multi-input multi-output cost functions necessitates three stages outlined in Chapter (2): data preparation, input and output aggregation and estimations. Once the first stage is performed for a specific period, a specific base year and a specific member state, different input and output aggregation schemes can be selected and implemented for a specific member state, a specific farm type and a specific time horizon in the second stage. Once the second stage is performed for a specific aggregation scheme, then different estimations can be selected and implemented for a specific member state or region, a specific farm type, a specific time horizon, a specific functional form, and specific theoretical restrictions for fixed effects or not in the third stage. The third stage concerning the econometric estimations is in turn decomposed into four steps: (1) estimation of the specified cost function for upper-level outputs (see Section 1.2.1), (2) estimation of the specified cost function for lower-level outputs (see Section 1.2.2), (3) estimation of demand functions for lower-level inputs at upper-level outputs, and (4) estimation of demand functions for lower-level inputs at lower-level outputs (see Section 1.3). This decomposition allows a great flexibility for the user in data preparation, input and output aggregation and estimations.

Since econometric estimations can use some computer time for running the Stata codes (from 10 to 30 minutes) but also user time for analysing the estimations results, it is recommended to perform the estimation work on selected regions or member states for selected farm types and time horizons taken then as case studies.

Chapter 10

Strengths and Limitations

10.1 Main Strengths

The approach is generic for the whole EU-FADN data set and, therefore, applicable to any region and member state with enough farm observations for dealing with the requirement in degree of freedom. It is recommended to use EU-FADN panel data that show a high degree of balance, i.e., a high degree of repetition of the same farm through the time period.

The approach is flexible in selecting the time period, the base year, the input and output aggregation scheme, the region, the member state, the farm type (but also many other farm characteristics such as location, altitude, size, organic or not, etc. as far as these are reported in the EU-FADN data set), the time horizon, the functional form, the theoretical restrictions and estimation through fixed effects or not.

The approach is theoretically sound avoiding risks of implosion or explosion of simulations when farm models embed these estimated flexible cost and input demand functions.

The approach brings competition among outputs and substitution among inputs at a high disaggregation level conserving the full second-order flexibility. As a result, unit costs and yields are endogeneously determined.

The approach is consistent. The sum of the estimated lower-level costs and input demands are equal to their respective estimated upper-level costs and input demands.

The approach is user-friendly since this selection is performed by yes-or-no type of statements in the Stata codes. Routines are devised to take care of missing input and output quantities and prices, negative estimated input demands and marginal costs and estimated coefficients with values that are identical to their initial values

and/or show infinite standard error.

10.2 Main Limitations

The approach rests on separability assumptions in both output and input categories and, thus, on the underlying technology embedded into the cost functions.

The theoretical restrictions, in particular on monotonicity of the cost and expenditure functions in input prices and fixed inputs, impose highly nonlinear restrictions on parameters during the econometric estimation phase. When the monotonicity restrictions are not imposed *ex-ante* as it is recommended when imposing *ex-ante* the curvature restrictions, then those monotonicity restrictions are not necessarily respected *ex-post* on some farms leading to negative input demands for those farms. It is, however, possible to impose all together the theoretical restrictions as simple parameter restrictions with an alternative functional form that is introduced in the following section.

The non-additive terms of the estimated functions for input subcategories are distributed over output and input categories or sub-categories in a *ad hoc* way. We, however, recommend to rely on the symmetry restrictions of cost and expenditure functions (see footnotes 3 and 12 of Chapter 1) to select the distributive weights and proceed accordingly.

The approach requires an intermediate background in microeconomics and econometrics to understand it and interpret its estimation results. It also requires an introduction to EU-FADN and Stata.

The approach requires computer time and user attention. The approach should therefore be considered as an econometric tool for analysing the economic behaviour of EU-FADN farms in specific regions or sub-sectors. It is not a press-the-button type of estimation tool. It involves econometrics that rests on skills in implementation, interpretation and analysis.

Chapter 11

Improvements

One adjustment, one correction and one alternative functional form are proposed as improvements.

11.1 Removing Outliers

Although outliers in terms of input and output quantities are removed *ex-post* after the estimations for the following calculations, it would be wiser to remove them *ex-ante* before the estimations or to remove them *ex-post* after the a round of estimations but before a second round of estimations for the following calculations. We could apply the same exclusion rule as the one applied for removing extreme farm-gate prices of inputs and outputs but for farms that show extreme values of input or output quantities (see Chapter 3) .

11.2 Quasi-fixed Input Allocation to Output Categories

To improve the estimation results, one possible correction would consist in allocating quasi-fixed inputs to the different upper-level outputs. Chambers and Just (1989) use rules to allocate fixed inputs among outputs for input-nonjoint (expression 4 in their paper) and input-joint (expression 21 in their paper) technologies that basically implies the equalization of the shadow values of one particular fixed input across outputs in either a profit maximising problem as in their paper or a cost minimizing problem as in our study. However, it is not clear how to apply these rules when the fixed input allocations across outputs are not observed in the data.

11.3 Alternative Functional Form

To overcome the problem of imposing monotonicity of the cost and expenditure functions in input prices and fixed input, we propose to use another functional form than the standard or augmented SGM that would have the advantage of operationalizing all theoretical restrictions as simple parameter restrictions, instead of implying a highly nonlinear restriction on parameters given the data, as does the SGM functional form at least for the monotonicity restrictions.

Following Chambers (1988), we identify the criteria for a *neo-classically* well-behaved cost function as follows.

A well-behaved cost function should be:

- regular (i.e., $TC(Y = 0, W, Z) = 0$) (Chambers, 1988, p.52, property 2B-6)
- non-decreasing Chambers (1988, p.52, property 2B-5),
- *convex* (Chambers, 1988, p.139, property 4B*-7)

in output quantities,

- non-decreasing (Chambers, 1988, p.52, property 2B-2)
- *concave* (Chambers, 1988, p.52, property 2B-3)
- first-order homogenous¹ (Chambers, 1988, p.52, property 2B-4)

in input prices, and

- non-increasing (Chambers, 1988, p.102)
- *convex*² (Chambers, 1988, p.109)

in fixed inputs.

One well-behaved cost function could be

$$TC = f(Y) \cdot g(W) \cdot h(Z),$$

with output quantities $Y = (y_1, \dots, y_M)'$, input prices $W = (w_1, \dots, w_I)'$ and fixed inputs $Z = (z_1, \dots, z_K)'$. In order to be well-behaved, the multiplicatively separable components $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ could look like

$$f(Y) = \alpha_F \prod_{m=1}^M y_m^{\alpha_{E;m}},$$

¹Hence the adding up constraint (Diewert and Wales, 1987, p.54).

²This is needed for consistency with long-run cost-minimization.

with

$$\begin{aligned}\alpha_F &\geq 0 \\ \alpha_{E;m} &\geq 1,\end{aligned}$$

$$g(W) = \beta_F \prod_{i=1}^J w_i^{\beta_{E;i}},$$

for which

$$\begin{aligned}\beta_F &\geq 0 \\ \beta_{E;i} &\geq 0 \\ \sum_{i=1}^J \beta_{E;i} &= 1\end{aligned}$$

and

$$h(Z) = \gamma_F \prod_{k=1}^K (z_k - \gamma_{C;k})^{\gamma_{E;k}},$$

with

$$\begin{aligned}\gamma_F &\geq 0 \\ \gamma_{C;k} &> 0 \\ \gamma_{E;k} &< 0.\end{aligned}$$

The use of such a functional form would have the advantage of operationalizing all theoretical restrictions as simple parameter restrictions, instead of implying a highly non-linear restriction on parameters given the data, as does the SGM functional form, at least for the monotonicity restrictions.

Bibliography

- Andersen, M. A., J. M. Alston, and P. G. Pardey (2011). Capital services in U.S. agriculture: Concepts, comparisons, and the treatment of interest rates. *American Journal of Agricultural Economics* 93(3), 718–738.
- Bahta, S., F. Offermann, R. De Blander, and B. Henry de Frahan (2011). Comparative ex-post analysis of past CAP reforms across selected member states and regions with FADN data. Technical report, von Thünen Institute, Braunschweig, Germany.
- Beattie, B. R., C. R. Taylor, and M. J. Watts (2009). *The Economics of Production*. Krieger Publishing Company.
- Berndt, E. R. and L. R. Christensen (1973). The internal structure of functional relationships: Separability, substitution, and aggregation. *The Review of Economic Studies* 40, 403–410.
- Chambers, R. G. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge, UK: Cambridge University Press.
- Chambers, R. G. and R. E. Just (1989). Estimating multioutput technologies. *American Journal of Agricultural Economics*, 980–995.
- Community Committee for the Farm Accountancy Data Network (2005). Farm return data definitions. accounting years 2004, 2005. Technical report, Community Committee for the Farm Accountancy Data Network.
- Community Committee for the Farm Accountancy Data Network (2007). Definitions of variables used in FADN standard results. Technical report, Community Committee for the Farm Accountancy Data Network.
- De Blander, R., B. Henry de Frahan, and F. Offermann (2011, July). Ex-post evaluations of agricultural and environmental policies in the EU with FADN data:

- Methods and results. European FACEPA project - wp9 working paper, Université catholique de Louvain, Louvain-la-Neuve, Belgium: Université catholique de Louvain.
- Deaton, A. and J. Muellbauer (1999). *Economics and Consumer Behavior*. Cambridge University Press.
- Diewert, W. E. and T. J. Wales (1987). Flexible functional forms and global curvature conditions. *Econometrica* 55(1), 43–68.
- Eurostat (2009). Statistics : land prices and rents - annual data (1985-2009).
- Eurostat (2012). Statistics : selling prices of animal products (absolute prices) - annual data (1969-2005).
- Eurostat (2013a). Statistics : price indices of agricultural products, output (2000 = 100) - annual data (1994-2008).
- Eurostat (2013b). Statistics : price indices of the means of agricultural production, input (2000 = 100) - annual data (1995-2008).
- Eurostat (2014). Statistics : economic accounts for agriculture - values at real prices (1973-2014).
- Fletcher, R. (1993). *Practical Methods of Optimization*. Chichester, UK: Wiley.
- Green, H. A. J. (1964). *Aggregation in Economic Analysis: An Introductory Survey*. Princeton University Press.
- Henry de Frahan, B., A. Baudry, R. De Blander, P. Polomé, and R. Howitt (2011). Dairy farms without quotas in Belgium: Estimations and simulations with a flexible cost function. *European Review of Agricultural Economics* 38(4), 469–495.
- Ivaldi, M., N. Ladoux, H. Ossard, and M. Simioni (1996). Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers. *Journal of Applied Econometrics* 11(6), 649–667.
- Sato, K. (1967). A two-level constant-elasticity-of-substitution production function. *Review of Economic Studies* 34(2), 201–218.
- Shephard, R. W. (1970). *Theory of Cost and Production Functions*, Volume 4 of *Princeton Studies in Mathematical Economics*. Princeton University Press.

Wieck, C. and T. Heckelei (2007). Determinants, differentiation, and development of short-term marginal costs in dairy production: An empirical analysis for selected regions of the EU. *Agricultural Economics* 36, 203–220.

Wolff, H., T. Heckelei, and R. C. Mittelhammer (2004). Imposing monotonicity and curvature on flexible functional forms. paper presented at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 1-4.