FACEPA

Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture

Methodology for the definition of case study farms and model structure for each case study

FACEPA Deliverable No. 6.2 – March 2010

INEA

Authors:

Filippo Arfini, Luca Cesaro, Quirino Paris

Michele Donati, Maria Giacinta Capelli, Sonia Marongiu

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under *grant agreement* n° 212292.

Executive Summary

The PMP model developed within the framework of the FACEPA project can be considered as a FADN specific accounting cost estimator open to policy and market assessment. Indeed, the model can recover the information about the specific costs per crop and use this new information to measure the impact of policy and/or market scenario changes on farm behaviour at territorial and sector level. In this deliverable, we discuss the first property of the model, the specific cost estimation.

The analysis has been developed recovering the specific costs per crop in three European case studies: farms belonging to the farm type "arable crops" in the Veneto-Lombardy-Piedmont regions in Italy, Belgium and Hungary. The data for the three Italian regions and Hungary were collected from the national FADN archive (year 2007), while for Belgium the model input came from European FADN information (year 2006). Despite the other available databases, the Italian FADN has allowed the estimates obtained for the three regions to be compared with the same information present in the national archives. This latter has represented the observed information against which the estimation procedure has been validated. The lack of observed specific accounting costs for Belgium and Hungary has prevented the possibility of extending the validation to the estimation for these two countries.

The estimation procedure uses the known information about acreage, prices, yields, other specific earnings, like coupled subsidies, per crop and the total variable costs at farm level. This exogenous information is used to estimate two types of costs: the specific marginal accounting cost and hidden marginal cost. The first type is directly related to the accounting information on total variable cost at farm level; the summation of specific accounting costs for the whole set of crops is equal to the total variable cost provided by the European FADN. The estimation of this latter cost component is the main aim of the present analysis. The second cost type is related to the part of the cost that eludes the farm accounting system, but is nevertheless considered within the farmer's decision process depicting the observed production plan.

The second cost type is the hidden cost, to be considered as a specific cost that is not registered by the farm accountancy but that influences the production choices. This is effectively an opportunity cost that each farmer takes in account during the

decision process and it is characterized by several factors, like the farmer's experience, risk attitude, market expectation and so on. These factors are not all explicitly considered by the PMP model, but implicitly assumed present within the observed production plan. Thanks to this cost component it is possible to recover the economic information that has led farmers to define the actual farm activity configuration and, thus, calibrate the observed situation.

The estimation procedure has been developed with respect to the three case studies, trying to identify homogenous groups of farms for improving the capacity of the model to estimate the observed accounting specific costs at farm and activity level. For this purpose, the analysis has adopted a multivariate analysis technique using principal component detection and the cluster analysis method (*k*-mean), which has contributed to reduce the variability of the information used in the estimation phase and to control the outliers. For the three Italian regions, the analysis has explored the estimation using farm information stratified according to region.

Just for the Italian results, the estimates have been submitted to a validation process using as a term of comparison the registered accounting costs available in the national FADN archive. The estimated accounting cost values have been compared with the observed accounting costs in order to verify that the average accounting costs per crop provided by the PMP model were not significantly different from the observed one. In respect to the estimate validation, the usual *t*-test has been implemented on paired groups of information (estimated and observed).

The results obtained demonstrate a strong influence of some factors in the estimation procedure, which can be summarized as follows:

- presence of outliers: the out-of-range values have without doubt an important effect on the estimation and a preventive check is fundamental for minimizing the interference of this kind of component in the estimation procedure; cluster analysis has also been adopted in order to reduce this risk further.
- variability in yield: the high internal variability in yield for some crops, like maize, has produced unreliable estimates in some cases even when there are a large number of observations. The case of maize in Veneto-Lombardy-Piedmont regions is highly emblematic of such a problem, because it had the highest number of observations among the considered crops but has generated estimates that are not statistically significant; in this specific case, the variability in yields is mainly due to the irrigation practices adopted, for

which the related costs are not considered within the observed specific accounting costs used as reference term for the validation phase.

- Level of internal sample homogeneity: the obtained estimates are much more significant the more homogeneous the sample is. This is evident for the three Italian regions, where the territorial stratification and the clustering have produced a marked improvement of the estimate significance.

Among the farm processes, soft wheat is the crop with the best significance in term of accounting cost estimation. For this crop, the hypothesis test has shown a good significance for all the estimations, with values not lower than 60%; only in the case of the sample formed using the cluster analysis the significance for the soft wheat accounting cost drops near 30%, but leading to an improvement for the other product estimates. This confirms the relevance of the grouping of farms in the estimation outcomes. For the three Italian regions, the territorial stratification has produced very good estimates for the most representative crops of the related farm type, i.e. soft wheat and barley, while in respect to the cluster analysis sample, the acceptable estimates are more distributed among the crops.

A cross comparison of the Belgian-Hungarian results with respect to the three Italian regions estimates has demonstrated that the estimation for Belgium presents the same scale value as the Italian one, while the Hungarian results are more distant. This comparison does not constitute a method for checking the estimation goodness, but rather, given the lack of observed accounting costs, provides a narrow judgement of the estimate scale and the degree of approximation to the three Italian regions validated estimates. It is quite clear that, for a deeper validation of the results achieved for Belgium and Hungary, observed accounting costs per crop are necessary. To date, the only information available for this scope is the estimation of soft wheat developed within WP5, which has reached estimates very close to the values obtained for the Italian and Belgian case studies; while, for Hungary the PMP estimates are much higher than the WP5 outcome.

In conclusion, the PMP model has demonstrated a good capacity to reproduce the observed accounting costs for cereals, apart from maize, and for the crops with a high level of homogeneity in prices and yields, like sugarbeet. To improve the level of estimation fitness it is important to reduce the variability as much as possible and, thus, the dispersion in the observations submitted to the estimation procedure, adopting an adequate method of farm grouping, like sector and territorial stratification and/or multivariate methods. The estimated accounting costs and the hidden marginal cost component will be used to evaluate possible productive reactions of farmers facing policy and market dynamics within a perspective of extensive use of the European FADN information.

Contents

Abbreviations and Acronyms

ACC_COST	Estimated marginal accounting costs
CA	Cluster Analysis
CAP	Common Agricultural Policy
D_WHEAT	Durum wheat
F_MAIZE	Fodder maize
FADN	Farm Accountancy Data Network
FT	Farm Type
FT1	Arable Crops Farm Type
FT4	Animal Production Farm Type
GAMS	General Algebraic Modeling System
GM	Gross Margin
GSP	Gross Saleable Production
HID_COST	Estimated marginal differential costs
INEA	Istituto Nazionale di Economia Agraria
OBS_COST	Observed marginal accounting costs
PCA	Principal Component Analysis
PMP	Positive Mathematical Programming
RICA	Réseau d'Information Comptable Agricole
S_WHEAT	Soft wheat
T_GRASS	Temporary grass
TVC	Total Variable Cost
UAA	Utilized Agricultural Area
VEGETABLE	Fresh vegetables grown in open field (FADN code: K136)
VEGETABLE2	Fresh vegetables grown in market garden (FADN code: K137)

List of Figures and Tables

Figures

Tables

1 PMP methodology for estimating specific production costs from FADN

The PMP methodology is widely used for evaluating the effects of policy and market changes on farm behaviour in order to give to policymakers some useful information for taking decisions about the CAP mechanisms. One of the major strengths of PMP is its capacity to recover the farm decision information using a relatively small amount of data. More specifically, PMP reconstructs, by the way of a calibration technique, the total variable cost function that farmers have taken into account in order to define the observed production plan. This information is used in the simulation phase for interpreting the farm responses to exogenous shocks. The PMP approach, adequately modified, is applied for the specific purpose of estimating the specific costs per product lacking in the FADN database.

The PMP in its classical approach, presented in the paper by Paris and Howitt (1998), is an articulated method consisting of three different phases, each of which is geared to obtaining additional information on the behaviour of each observed farm so as to be able to simulate its behaviour in conditions of maximization of the total gross margin (Howitt and Paris, 1998; Paris and Arfini, 2000). The PMP method has been widely used in the simulation of alternative policy and market scenarios, utilizing micro technical-economic data relative both to individual farms and to mean farms that are representative of a region or a sector (Arfini et al., 2005). The success of the method is to be largely attributed to the relatively low requirement for information on the business and, first and foremost, to the possibility of using databases, including the FADN database (Arfini et al., 2003, 2005, 2008).

Notwithstanding the numerous studies that adopt the PMP approach using the FADN data, the methodology nonetheless comes up against a limitation consisting of the lack of FADN data on specific production costs per process. The lack of this information poses a problem during the calibration phase of the model, when the estimation of the cost function requires a non negative marginal cost for all production processes activated by a single holding (Paris and Arfini, 2000).

This problem is dealt with in this analysis by resorting to an approach that utilizes dual optimality conditions directly in the estimation phase of the non linear function. The approach qualifies itself as an extension of the Heckelei proposal (2005), according to which the first phase of the classical PMP method can be avoided by imposing first order conditions directly in the second cost function estimation phase. Moreover, as a guide to the correct estimation of the explicit corporate costs, the model considers the information relative to the total corporate variable costs available in the European FADN archive. This "innovation" becomes particularly important as it enables us to perform analyses utilizing the European database without having to resort to parameters that are exogenous to the model.

According to this new approach, the PMP model divides into two phases: a) the aim of the first is to estimate specific accounting variable costs by activity through the reconstruction of a non linear function of the total variable cost that considers the exogenous observed information on the total variable costs for the individual farm; b) the aim of the second is the calibration of the observed production situation through the resolving of a farm gross margin maximization problem, in the objective function of which the cost function estimated in the previous phase is included.

The first phase is defined by an estimation model of a quadratic cost function in which the squares of errors are minimized:

$$\min_{u} LS = \frac{1}{2}\mathbf{u}'\mathbf{u} \tag{1}$$

subject to

$$\mathbf{c} + \boldsymbol{\lambda} = \mathbf{R}'\mathbf{R}\overline{\mathbf{x}} + \mathbf{u} \qquad \text{se } \overline{x} > 0 \tag{2}$$

$$\mathbf{c} + \boldsymbol{\lambda} \le \mathbf{R}'\mathbf{R}\overline{\mathbf{x}} + \mathbf{u} \qquad \text{se } \overline{x} = 0 \tag{3}$$

$$\mathbf{c}'\overline{\mathbf{x}} \le TC \tag{4}$$

$$\mathbf{u}'\overline{\mathbf{x}} + \frac{1}{2}\overline{\mathbf{x}}'\left(\mathbf{R}'\mathbf{R}\right)\overline{\mathbf{x}} \ge TC \tag{5}$$

$$\mathbf{c} + \boldsymbol{\lambda} + \mathbf{A}'\mathbf{y} \ge \mathbf{p} + \mathbf{A}'\mathbf{s} \tag{6}$$

$$\mathbf{b}'\mathbf{y} + \boldsymbol{\lambda}'\overline{\mathbf{x}} = \mathbf{p}'\overline{\mathbf{x}} + \mathbf{s}'\overline{\mathbf{h}} - \mathbf{c}\overline{\mathbf{x}} \tag{7}$$

$$\mathbf{R} = \mathbf{L}\mathbf{D}^{1/2} \tag{8}$$

$$\sum_{n=1}^{N} u_{n,j} = 0 \tag{9}$$

By means of the model (1)-(9) a non linear cost function can be estimated using the explicit information on the total farm variable costs (TVC) available in the FADN database. The restrictions (2) and (3) define the relationship between marginal costs derived from a linear function and marginal costs derived from a quadratic cost function. $\mathbf{c} + \boldsymbol{\lambda}$ defines the sum of the explicit process costs and the differential marginal costs, i.e. the costs that are implicit in the decision-making process of the entrepreneur and not accounted for in the holding's bookkeeping. Both components are variables that are endogenous to the minimization problem. To guarantee consistency between the estimate of the total specific costs and those effectively recorded by the corporate accounting system, the restriction (4) imposes that the total estimated explicit cost should not be more than the total variable cost observed in the FADN database. Restriction (5) defines a further restriction on the costs estimated by the model, where the non linear cost function must at least equal the value of the total variable cost (TVC) measured. In order to guarantee consistency between the estimation procedure and the optimal conditions, restriction (6) introduces the traditional condition of economic equilibrium, where total marginal costs must be greater or equal to marginal revenues.

The total marginal costs also consider the use cost of the production factors defined by the product of the technical coefficients matrix A' and the shadow price of the restricting factors y; while the marginal revenues are defined by the sum of the products' selling prices, p, and any existing public subsidies. The additional restriction (7) defines the optimal condition, where the value of the primary function must correspond exactly to the value of the objective function of the dual problem. In order to ensure that the matrix of the quadratic cost function is symmetrical, positive and semi-defined, the model adopts Cholesky's decomposition method, according to which a matrix that respects the conditions stated is the result of the product of a triangular matrix, a diagonal matrix and the transpose of the first triangular matrix (8). Last but not least, restriction (9) establishes that the sum of the errors, u, must be equivalent to zero.

The cost function estimated with the model (1)-(9) may be used in a model of maximization of the corporate gross margin, ignoring the calibration restrictions imposed during the first phase of the classical PMP approach. In this case, the dual relations entered in the preceding cost estimation model guarantee the reproduction of the situation observed. The model, therefore, appears as follows:

$$\max_{x\geq 0} ML = \mathbf{p'x} + \mathbf{s'h} - \left\{ \frac{1}{2}\mathbf{x'\hat{Q}x} + \mathbf{\hat{u}'x} \right\} \tag{10}$$

subject to

$$\mathbf{Ax} \leq \mathbf{b} \tag{11}$$

$$A_j x_j - h_j = 0 \qquad \forall j = 1,\dots,J \tag{12}$$

The model (10)-(12) precisely calibrates the farming system observed, thanks to the function of non linear cost entered in the objective function which preserves the (economic) information on the levels of production effectively attained. Restriction (11) represents the restriction on the structural capacity of the farm, while equation (12) enables us to obtain information on the hectares of land (or number of animals) associated with each process j. Once the initial situation has been calibrated through the maximization of the corporate gross margin, it is possible to introduce variations in the public aid mechanisms and/or in the market price levels in order to evaluate the reaction of the farm to the changed environmental conditions. The reaction of the farm business will take into account the information used during the estimation phase of the cost function, in which it is possible to identify a real, true matrix of the farm choices, i.e. Q. Within this framework, the PMP methodology described in this section will be implemented for recovering the specific production costs related to the process whose data are collected by the FADN.

2 The 3 Case Studies selected for the cost estimation and Model structure

2.1 FADN case studies

The specific cost estimation using FADN information and the PMP model described above, is developed with respect to national and regional FADN databases selected among the WP6 partners. Within this framework, the scope of the work is to develop and test the PMP-based methodology able to capture the information about variable costs per activity. For validating the methodology, the results obtained by the PMP are compared with observed information recovered from the same database when available. The partners providing FADN data have been selected considering the availability of observed specific costs in the national database. For this reason, the FADN databases selected for estimation are:

- Veneto, Piedmont, Lombardy regions (Italy),
- Belgium,
- Hungary.

The three national FADN databases contain the information on the specific variable costs, so permit the aforementioned comparison.

The Italian FADN liaison office, INEA, for example, collects the specific variable costs for each crop concerning seeds, fertilizers, pesticides and services provided by third parties. This information, obtained every year and not transferred to the European database, is the result of a process of accounting attribution starting from the farm invoice information collected by the RICA local interviewer. It is clear that the result of the process of cost distribution among activities leads to an imperfect evaluation of the farm specific costs, but it is the closest possible to the real information. For our purposes, it represents the benchmark, in respect of which we can validate the estimating methodology for the Italian specific costs.

In order to do the estimation, the quality check of the data is an important task that must be carried out to avoid results influenced by outliers. It is well known that FADN, despite a control on the statistical data goodness, is affected by “out of range” values that have to be adequately treated. This is why the estimation procedure is anticipated by an outliers check, so that the estimation can be applied reducing the influence of out of range values.

For the purpose of the present analysis, the cost estimation is developed for each farm belonging to a specific farm type stratification in order to keep a sufficient degree of homogeneity with respect to the farm technology. Not all farm types have been investigated, but only the most numerous in terms of observations, which are arable crops FT (11) and livestock production FT (41).

In the following sections, the estimates will be presented as well as a statistical description of the case study sample considered in the analysis.

2.2 Model architecture

The PMP model for the estimation of FADN specific costs is part of an articulated elaboration system developed using GAMS. This system is divided in different modules each one devoted to a specific task and interfaced with the others providing the input information. Four modules can be distinguished by which the PMP cost estimation model is composed: the data entry, the stratification, estimation and calibration, and output module. Figure 2.1 shows the flow chart where the different phases of the model are specified and linked.

2.2.1 Data entry

The basic information used is the FADN related to the countries specified above. The FADN database is preliminarily treated in order to fit the GAMS syntax requirements. More specifically, starting from a database composed of all the variables (fields) included in the FADN database, we have selected the group of variables relevant for the scope of the analysis. The complete list of variables is shown in Annex 1. The database obtained in the preliminary treatment has been organized in tables readable and manageable by GAMS. In this phase, GDX-routines have allowed the interface between GAMS and the common .csv and .xls files where the basic data is stored.

2.2.2 Stratification

This module is devoted to selecting the farm information used for the estimation procedure. Each database is stratified according two criteria: the specific farm type (8 groups) and the reference clusters obtained using the set of multivariate analysis tools described in the next sections. In this study, for each considered farm type, the reference cluster is the one that has the highest number of farms without outliers. As will be explained in section 3, the cluster analysis technique is adopted in order to group farms that are homogeneous in terms of production technology and market conditions. This procedure allows to not consider in the analysis farms that can be considered as outliers due to technology and market conditions or due to incomplete information or data entry errors.

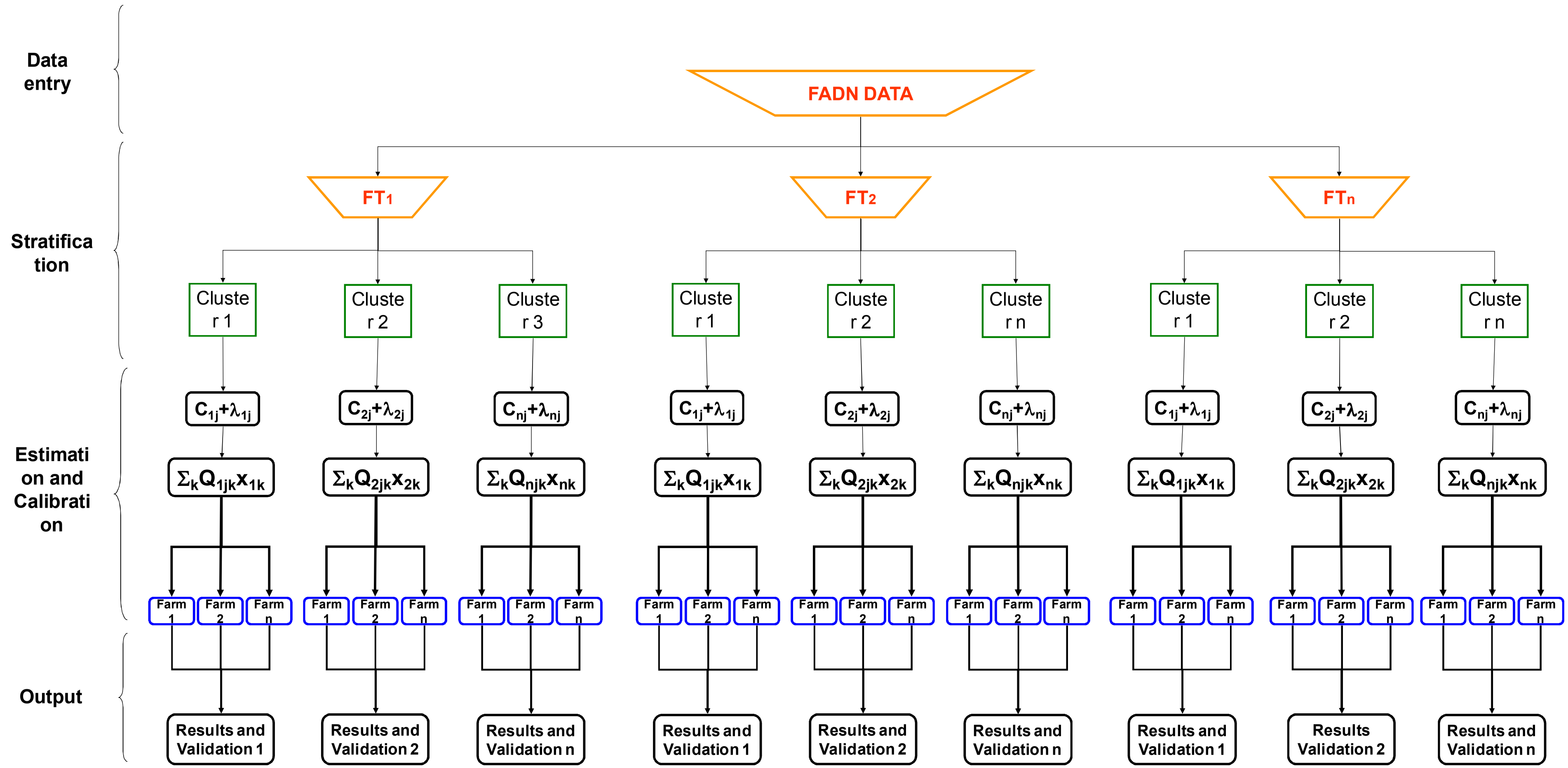
2.2.3 Estimation and Calibration

The estimation and calibration part of the model is connected with the two previous modules. Adopting the methodology presented in the first section, this module estimates the specific variable costs associated with each activity for each farm considered in the sample. The model foresees two calibration phases: the first is obtained during the estimation phase, while the second is achieved in a non-linear mathematical programming model, where the information related to the estimated costs is used within the objective function of this model; The latter is conducted to test the solution reached in the first estimation phase and to obtain a model for simulations.

2.2.4 Output

The results obtained are stored in specific output files readable by statistical and spreadsheet software, adopting GDX routines. The generated output is composed of calibrating checks, the matrices related to the accounting cost and latent cost estimation, c and λ, and the matrix of the specific variable cost per activity. For the Italian case study, the output also considers the observed specific variable costs that allow a direct comparison with the estimated variable cost. This last phase is carried out using Student's t-tests provided by statistical packages (SPSS) to evaluate the estimation goodness.

Fig. 2.1: PMP Model architecture



3 Italian regions - Veneto, Lombardy and Piedmont

3.1 Data entry description and quality control procedure

The Italian regions selected for the analysis are in Northern Italy (north of the Po River) and are characterized by highly specialized and intensive agricultural practices. The most important activities are livestock production (mainly dairy and beef cattle) and arable crops. According the 2009 Eurostat information, the Veneto-Lombardy-Piedmont area represents 50% of the entire livestock in Italy. The average size of each farm is 5 ha, against the national average of 2 ha (Eurostat, 2009).

The farm sample considered in this analysis is composed of 738 farms belonging to FT1 (arable crops). The average size of each farm in the sample is 50 ha. The RICA farms in Piedmont are the largest in terms of hectares. On average the incidence of cereals on the total UAA in the sample is 43%. The average GSP per hectare is 1,774 Euros, while the total variable cost per hectare is 600 Euros (Table 3.1).

Table 3.1: Statistical description of Italian FADN sample – Farm type 1

Area	N. of farms	Av. UAA (ha)	Cereals / tot (%)	GSP/ha (€)	Total Variable Costs /ha (€)
Veneto	220	44	62	1956	656
Lombardy	165	46	40	1763	370
Piedmont	353	56	36	1689	661
Total	738	50	43	1774	600

Considering the entire sample, rice covers 39% of the total land area, followed by maize with 25% and soft wheat with 15%. In Veneto maize is the main crop, while in Lombardy and Piedmont the most important crop in terms of area is rice. Another important crop is soya, that in the Veneto sample represents 17% of the entire acreage. Indeed, Veneto is specialized in producing maize and soya due to the presence of dairy and beef farms and important foodstuff industries.

Fig. 3.1.a: Crop distribution in Veneto Farm type 1 sample

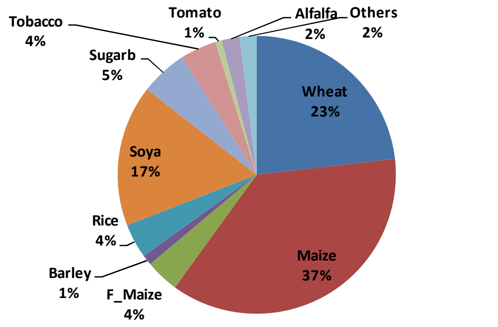


Fig. 3.1.b: Crop distribution in Lombardy Farm type 1 sample

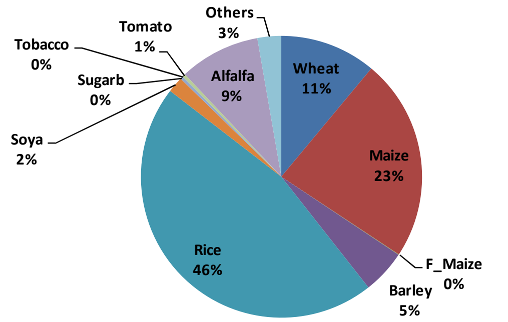


Fig. 3.1.c: Crop distribution in Piedmont Farm type 1 sample

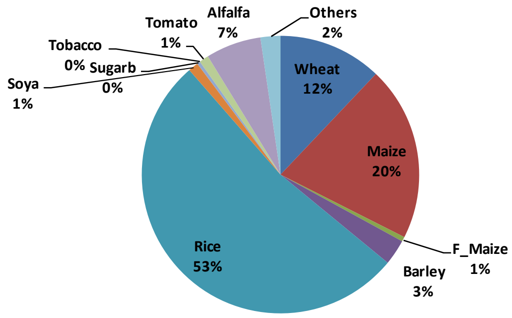
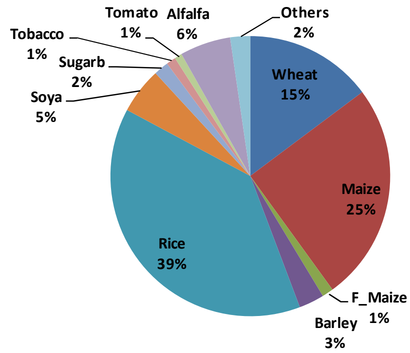


Fig. 3.1.d: Crop distribution in the entire Farm type 1 sample area



All the crops depicted above are considered in the PMP model analysis and specific variable production costs are estimated for each one. As described in the previous section, the estimation is made using the information about acreage, yields, prices for each crop at farm level and the total variable cost at farm level.

In order to achieve a good fitness of the estimation to the reality, it is important to avoid the presences of outliers, but it is also useful to utilize a homogeneous sample of farms with respect to the main variable that influences the production function and dynamics of production cost, like yields and output prices. Figure 3.2 and Table 3.2 present some descriptive information on prices and yields of four main crops included in the FADN sample.

As one can see at a first glance, the observations are less dispersed for some crops, like soft wheat and rice, while for others, like maize and soya, the dispersion is very high. The main factor that influences the observations' dispersion is the variation in yields. Indeed, for maize, the standard deviation is very high, equal to 31, which means a variation with respect the mean of 3.1 tons per hectare; while, for rice, the dispersion in yields is more restrained, 0.9 tons per hectare.

Fig. 3.2: Price and yield distribution in FT1 Sample for the Italian case study

Fig. 3.2.a: Soft wheat

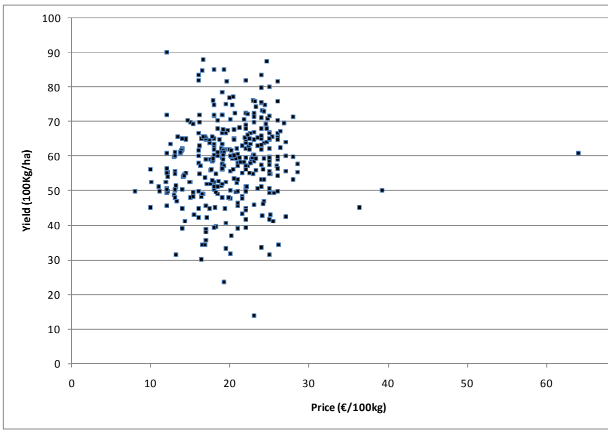


Fig. 3.2.b: Maize

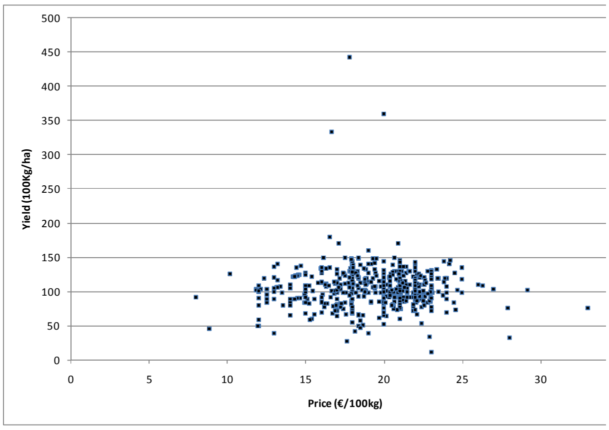


Fig. 3.2.c: Soya

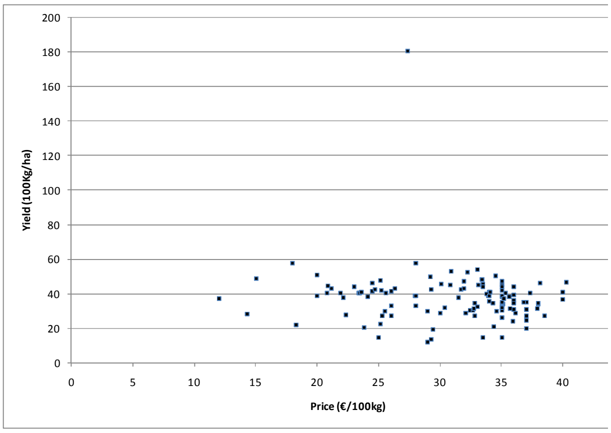


Fig. 3.2.d: Rice

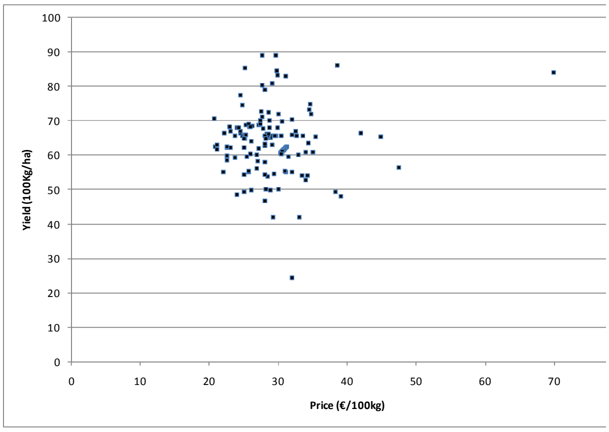


Table 3.2: Descriptives of some crops selected from FADN sample (Lombardy, Piedmont and Veneto): price in €/100 Kg; Yields in 100 kg/ha

Crop	Variable	N. of Obs.	MIN	MAX	Mean		Std. Deviation
					Statistic	Std. Error	
Soft wheat	Prices	335.00	8.00	64.00	20.01	0.27	5.09
	Yields	335.00	13.99	90.00	58.54	0.59	11.41
Maize	Prices	546.00	7.98	33.01	19.14	0.15	3.39
	Yields	548.00	12.00	442.48	106.02	1.32	30.83
Soya	Prices	127.00	12.00	40.26	30.85	0.53	5.98
	Yields	125.00	12.19	180.77	37.74	1.42	15.88
Rice	Prices	145.00	20.72	69.83	29.01	0.46	5.55
	Yields	144.00	24.36	89.05	63.80	0.77	9.23

The high level of dispersion also hides the presence of outliers that can strongly influence the estimation results for some crops. For example, maize is characterized by several observations that are out of range. Indeed, figure 3.2.b shows a cluster of points surrounded by several out of range observations. These points represent outliers that influence the capacity of the model to correctly estimate production cost and should thus be eliminated from the estimation procedure.

The disturbing information represented by the outliers can be appreciated at both process level and farm level. In regard to the farm information, one can see that the bad information is also present for the normalized variables concerning the gross saleable production (GSP) and farm total variable cost (TVC). Figure 3.3 shows the farms on a scatter plot considering the GSP per hectare and TVC per hectare on the axes. It is evident that some points are extremely far from the average observations and they can be considered as outliers. If we observe a detail of the same sample reducing the scale, the possibility to adopt statistical techniques aiming to detect a homogeneous set of observations is obvious.

Fig. 3.3: Farm distribution between GSP/ha and TVC/ha (Veneto-Lombardy-Piedmont)- standard scale

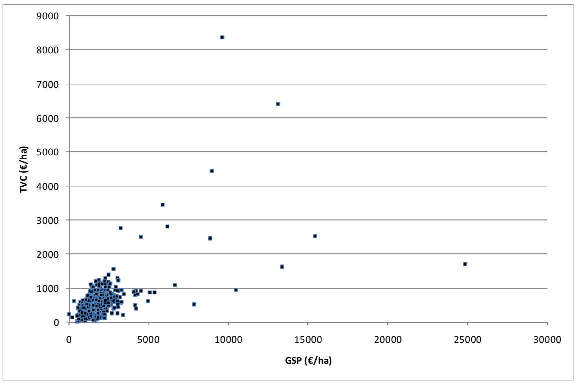
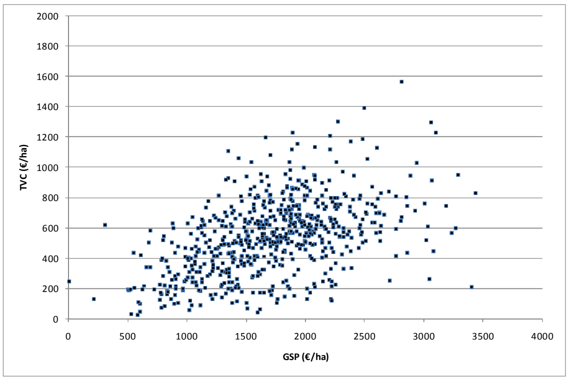


Fig. 3.4: Farm distribution between GSP/ha and TVC/ha (Veneto-Lombardy-Piedmont) - reduced scale



In conclusion, the strategy with respect to the outlier is to always consider all the activities present on the observed farm and to not consider in the cost estimations those farms that are outliers due to the fact that they present a single crop or many activities that are not homogeneous with respect to the characteristics of the sample. The homogeneity is evaluated by using Principal Component Analysis (PCA) and Cluster Analysis (CA). This latter is implemented using the K-mean methodology. Only the clusters with the highest number of homogeneous farms are used for the process of cost estimation by PMP model in order to guarantee sufficient numbers of observations for crops to submit to estimation.

3.2 Validation procedure

In order to validate the estimation procedure, the "estimated specific variable costs" are compared with the "observed specific variable costs" through *t*-test. The test allows to verify that the two means derive from a population with the same mean ($H_0 : \mu_1 = \mu_2$). When the probability is very low the hypothesis $\mu_1 = \mu_2$ is rejected.

This procedure is only possible for the Italian FADN that includes the information about specific costs per activity, while the other EU regional FADNs considered in this analysis (Belgium and Hungary) do not collect the information on production cost per activity at farm level.

3.3 Specific accounting cost estimation

In this section, we provide the estimation of variable cost per observed activity in the Italian FADN (Veneto, Lombardy and Piedmont) in different environments:

a) the entire area (the three regions together);
b) each region;
c) homogeneous farm belonging the Italian FADN detected by cluster analysis.

The objective is also to show how the criteria used in the definition of the set of data becomes crucial in order to obtain a good estimation of observed variable cost.

Before discussing the analysis of the results is useful to recall that PMP allows two types of specific variable costs for each activity to be estimated: the accounting cost (c) and the marginal cost (lambda). We should be aware that these costs are estimated under economic constraints because the dual property of a profit maximization problem is used, that is implicit in the model (1)-(9), where the shadow prices associated to production activities are exactly equal to the sum of the *estimated accounting cost* and the *estimated differential marginal costs*. The estimated accounting cost may be interpreted as the part of production shadow price that can be explained by the accounting values, while the estimated differential marginal cost might be considered as the opportunity cost associated to each activity. The sum of the estimated accounting cost and the estimated differential marginal cost provide the exact measure of the total variable (marginal) cost associated to each activity.

The estimated differential marginal costs are defined in this work as "*hidden costs*", to indicate the part of estimated marginal costs that are considered by farmers in defining their production plans but which are absent from the farm accounting sheets. These are the part of marginal costs related to the specific and individual opportunity costs that each farmer has considered for deciding to introduce a given crop in the production plan. We can consider this category of costs as “pure economic cost” due to the fact that it depends on the profit maximization logic (expressed by the observed price) and on characteristics of the production function (expressed by the observed yields).

On the contrary the *observed variable accounting cost* registered by FADN can, in theory, contain errors due to several reasons: gathering at farm level and/or imputation. In particular farmers may wrongly specify some costs related to a production technique. One example is provided by the irrigation costs where it is difficult for farmers to properly record such costs because they are often not explicit. For these reasons the comparison of estimated variable cost with the observed marginal accounting cost can fail when some types of cost are not explicit even for farmers.

3.3.1 The estimation for Veneto, Lombardy and Piedmont as homogenous area

Table 3.3 shows the results obtained using the information on the entire sample, where the observed accounting costs are compared with the estimated accounting. Certainly the comparison is only between observed and estimated accounting cost because the hidden cost, as opportunity cost, is not collected by FADN.

Table 3.3: Comparison between observed accounting cost and specific variable cost estimated from PMP model – Veneto, Lombardy and Piedmont – Farm type 1, Year 2007

Crop	Observed cost	Std. Error	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
D_wheat	**0.07575**	0.00598	**0.13428**	0.01738	**0.02680**	0.00205	**0.16108**	0.00244
S_wheat	**0.07016**	0.00170	**0.06602**	0.00332	**0.03275**	0.00289	**0.09878**	0.00309
Maize	**0.06232**	0.00161	**0.07439**	0.00172	**0.04685**	0.00243	**0.12124**	0.00206
Barley	**0.06052**	0.00329	**0.05130**	0.00543	**0.02099**	0.00167	**0.07229**	0.00206
Rice	**0.11425**	0.00313	**0.12368**	0.00470	**0.03833**	0.00363	**0.16201**	0.00575
Sorghum	**0.06466**	0.01705	**0.04719**	0.01233	**0.01949**	0.00200	**0.06669**	0.00200
Prot_crops	**0.08839**	0.00904	**0.08747**	0.01744	**0.01959**	0.00323	**0.10706**	0.00352
Soya	**0.11664**	0.00590	**0.09133**	0.00676	**0.02504**	0.00333	**0.11636**	0.00427
Suagarbeet	**0.01405**	0.00050	**0.01721**	0.00124	**0.00096**	0.00015	**0.01817**	0.00031
Potato	**0.05974**	0.01268	**0.12623**	0.02343	**0.03735**	0.00908	**0.16358**	0.01381
Rape	**0.18170**	0.04158	**0.11232**	0.02731	**0.02266**	0.00238	**0.13497**	0.00283
Sunflower	**0.11240**	0.02158	**0.11070**	0.03307	**0.02117**	0.00150	**0.13188**	0.00197
Tobacco	**0.97254**	0.10625	**1.03875**	0.08186	**0.02164**	0.01118	**1.06039**	0.02012
Melon	**0.11124**	0.02712	**0.12270**	0.03737	**0.01627**	0.00230	**0.13897**	0.00280
Tomato	**0.05094**	0.01876	**0.09376**	0.04093	**0.01844**	0.00624	**0.11219**	0.01364
F_maize	**0.02065**	0.00557	**0.00924**	0.00240	**0.00136**	0.00064	**0.01060**	0.00084
T_grass	**0.02434**	0.00287	**0.03165**	0.00779	**0.00224**	0.00022	**0.03389**	0.00038
Alfalfa	**0.01352**	0.00130	**0.02766**	0.00316	**0.00616**	0.00073	**0.03382**	0.00088
Meadow	**0.01403**	0.00086	**0.02986**	0.00336	**0.00789**	0.00058	**0.03775**	0.00086

The same consideration can be appreciated observing Figure 3.5, where the estimated total variable cost is split between the accounting costs (ACC_COST) and hidden costs

(HID_COST) for some relevant activities. Tobacco is the crop with highest accounting cost, justified by the high cost of production treatment.

Fig. 3.5: Total marginal cost distribution - Veneto, Lombardy and Piedmont – Farm type 1, Year 2007 (€/t).

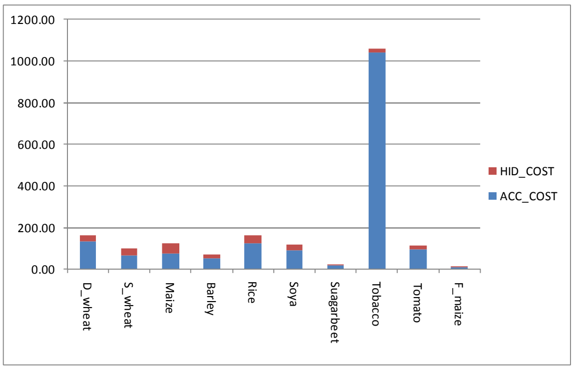


Fig. 3.6: Comparison between observed variable costs and estimated accounting costs - Veneto, Lombardy and Piedmont – Farm type 1, Year 2007 (€/t.)

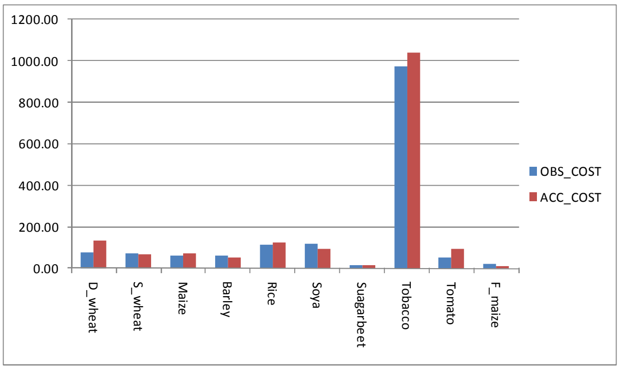


Figure 3.6 compares the observed accounting cost with the estimated one. For the most numerous crops, like soft wheat, maize, barley, rice and soya, the differences in absolute value remain within the range of 6% (soft wheat) to 20% (soya).

Nevertheless, the pure investigation of the differences says nothing about the statistical significance of the estimation from an inferential point of view. For this reason the *t*-test is introduced to verify the goodness of fit of the estimation through the comparison the mean of the estimated accounting cost with the mean of observed accounting costs.

The results obtained applying Student's *t*-test are presented in Table 3.4, where the most significant values are written in **bold**. For the entire Italian sample, the test of paired groups indicates a high significance for soft wheat, protein crops and sunflower, while for barley, rape and fodder maize the significance level is only good. For the other estimates, the null hypothesis has to be rejected for most of the crops, since the probability is lower than 1%.

Table 3.4: Student's *t*-test for estimated and observed accounting costs - Veneto, Lombardy and Piedmont – Farm type 1, Year 2007.

Crop	Paired Differences					t	Sig. (2-tailed)
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper		
D_wheat	0.0607708	0.0941463	0.0192175	0.0210163	0.1005253	3.162	0.004
S_wheat	-0.0013917	0.0557586	0.0035843	-0.0084523	0.0056688	-0.388	**0.698**
Maize	0.0139243	0.0340009	0.0016046	0.0107708	0.0170778	8.678	0.000
Barley	-0.0063194	0.0508202	0.0064542	-0.0192253	0.0065866	-0.979	**0.331**
Rice	0.0086862	0.0408357	0.0034762	0.0018123	0.0155601	2.499	0.014
Sorghum	-0.0273333	0.0231198	0.0133482	-0.0847660	0.0300993	-2.048	0.177
Prot. crops	-0.0096231	0.1022543	0.0283603	-0.0714148	0.0521686	-0.339	**0.740**
Soya	-0.0215915	0.0812341	0.0089708	-0.0394406	-0.0037424	-2.407	0.018
Sugarbeet	0.0024846	0.0074877	0.0011990	0.0000574	0.0049118	2.072	0.045
Potato	0.0659294	0.0646643	0.0156834	0.0326821	0.0991767	4.204	0.001
Rape	-0.0577750	0.1799401	0.0636184	-0.2082087	0.0926587	-0.908	**0.394**
Sunflower	-0.0011200	0.1504740	0.0475840	-0.1087626	0.1065226	-0.024	**0.982**
Tobacco	0.0662250	0.0761189	0.0380594	-0.0548971	0.1873471	1.740	0.180
Melon	0.0160250	0.1126294	0.0563147	-0.1631935	0.1952435	0.285	**0.794**
Tomato	0.0356211	0.2019461	0.0463296	-0.0617138	0.1329560	0.769	**0.452**
F_maize	-0.0105500	0.0256396	0.0090650	-0.0319852	0.0108852	-1.164	0.283
T_grass	0.0122818	0.0342730	0.0103337	-0.0107431	0.0353067	1.189	0.262
Alfalfa	0.0149816	0.0209442	0.0033976	0.0080974	0.0218658	4.409	0.000
Meadows	0.0175032	0.0275066	0.0034933	0.0105178	0.0244886	5.010	0.000

For instance, maize that presented a difference of the estimated mean with respect the observed mean of 19%, doesn't pass the test t at a level of probability equal to zero. In other words, it is not true that the estimated mean can explain the mean of the observed costs. According to the brief statistical description of maize observations previously presented, this results may be attributable to the strong dispersion in prices and yields and to the lack of gathering specific cost related to the irrigation (that strongly influences the yields).

3.3.2 The estimation of accounting costs for each region as homogenous area

In order to assess the capability of the model to capture the territorial specificities and, thus, improve the estimates, the entire Italian sample has been stratified in three groups of farms corresponding to the three regions considered for Italy. Also in this case, the PMP model performs the estimation using all the available information presents in the sample, that consists in the activity observations for each individual farm.

3.3.2.1 The case of Veneto Region

The table 3.5 shows the estimation outputs for the Veneto region. Observing and comparing the estimated accounting cost with the observed costs is evident a strong improvement of goodness, with respect to the previous analysis. The most part of the activities, like for

example soft wheat, barley, soya and sugarbeet, the difference between the estimated and the observed accounting costs is lower than 10%. On the contrary, for some crops the divergence on the observed values remains: durum wheat estimate is completely different with respect
For instance, maize, which presented a difference of the estimated mean with respect to the observed mean of 19%, doesn't pass the *t*-test at a level of probability equal to zero. In other words, it is not true that the estimated mean can explain the mean of the observed costs. According to the brief statistical description of maize observations given previously, this result may be attributable to the strong dispersion in prices and yields and to the lack of gathering the specific cost related to irrigation (that strongly influences yields).

3.3.2 The estimation of accounting costs for each region as a homogenous area

In order to assess the capacity of the model to capture the territorial specificities and, thus, improve the estimates, the entire Italian sample has been stratified in three groups of farms corresponding to the three regions considered. Also in this case, the PMP model performs the estimation using all the available information present in the sample, which consists of the activity observations for each individual farm.

3.3.2.1 The case of Veneto Region

Table 3.5 shows the estimation outputs for the Veneto region. Observing and comparing the estimated accounting cost with the observed costs, there is a strong improvement with respect to the previous analysis. For most of the crops, e.g. soft wheat, barley, soya and sugarbeet, the difference between the estimated and observed accounting costs is lower than 10%. On the contrary, for some crops the divergence from the observed values remains: the durum wheat estimate is completely different from the observed data, maize also presents a divergence of 30% with respect to the observed value.

Table 3.5: Specific cost estimates obtained from PMP model - Veneto sample

Crop	N_OBS	Observed cost	Std. Error	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
D_wheat	11	**0.06730**	0.00433	**0.22931**	0.02464	**0.02513**	0.00861	**0.25444**	0.02451
S_wheat	110	**0.07020**	0.00203	**0.06853**	0.00549	**0.04956**	0.00790	**0.11809**	0.00599
Maize	184	**0.06557**	0.00118	**0.08523**	0.00255	**0.03895**	0.00377	**0.12418**	0.00300
Barley	17	**0.07180**	0.00645	**0.06843**	0.02050	**0.04060**	0.01198	**0.10903**	0.01527
Rice	6	**0.12074**	0.00483	**0.14119**	0.04230	**0.18740**	0.04930	**0.32859**	0.06858
Sorghum	1	**0.07874**	0.00000	**0.00000**	0.00000	**0.04885**	0.02128	**0.04885**	0.02128
Prot_crops	4	**0.08886**	0.00755	**0.15265**	0.01891	**0.03100**	0.01266	**0.18365**	0.03276
Soya	82	**0.11336**	0.00437	**0.10317**	0.00927	**0.06615**	0.01234	**0.16933**	0.00913
Suagarbeet	42	**0.01426**	0.00045	**0.01538**	0.00145	**0.00330**	0.00050	**0.01868**	0.00069
Sunflower	2	**0.17323**	0.02564	**0.23890**	0.11892	**0.03637**	0.01555	**0.27527**	0.03879
Tobacco	4	**0.97254**	0.10625	**1.03875**	0.08186	**0.05621**	0.01332	**1.09496**	0.02355
Tomato	7	**0.03133**	0.00720	**0.03477**	0.00841	**0.00740**	0.00342	**0.04217**	0.00459
F_maize	8	**0.01120**	0.00043	**0.00601**	0.00234	**0.00940**	0.00359	**0.01541**	0.00367
Alfalfa	13	**0.00937**	0.00156	**0.02387**	0.00849	**0.00827**	0.00220	**0.03213**	0.00221
Meadows	7	**0.01966**	0.00506	**0.00872**	0.00000	**0.00231**	0.00038	**0.01103**	0.00038

If there is a problem of number of observations for durum wheat that may have influenced the estimation, for maize the problem is different. Indeed, the number observations for this crop is very high (184), but a strong dispersion of the information on prices and, more importantly, on yields, plays an important role in procuring a distortion in the estimation results.

The analysis of the estimated accounting and hidden marginal costs (Fig. 3.7) doesn't substantially change the considerations developed for the entire sample, in the sense that the hidden cost remains a residual cost component with respect the accounting cost. As stated previously, most of the estimations are in line with the observed values, but few other estimates amplify the divergence with respect to the previous estimation (Fig. 3.8).

Fig. 3.7: Total marginal cost distribution - Veneto sample (€/t)

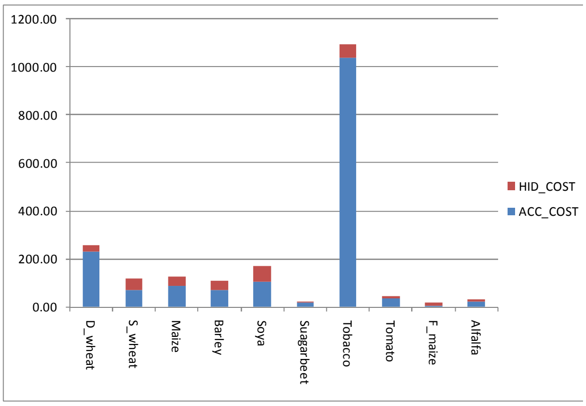


Fig. 3.8: Comparison between observed costs and estimated accounting costs - Veneto sample (€/t)

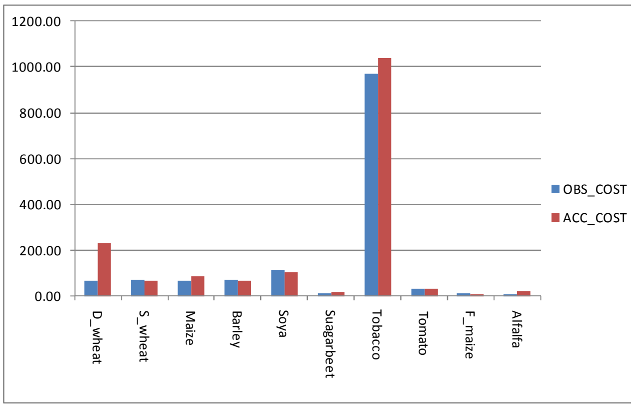


Table 3.6: *T*-test for estimated and observed accounting costs - Veneto sample

Crops	Paired Differences					t	Sig. (2-tailed)
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper		
D_wheat	0.1623200	0.0839499	0.0265473	0.1022659	0.2223741	6.114	0.000
S_wheat	-0.0004943	0.0554849	0.0059486	-0.0123197	0.0113312	-0.083	**0.934**
Maize	0.0194960	0.0304049	0.0022854	0.0149858	0.0240063	8.531	0.000
Barley	0.0009273	0.0647307	0.0195170	-0.0425594	0.0444139	0.048	**0.963**
Rice	0.0204333	0.1101644	0.0449744	-0.0951772	0.1360438	0.454	**0.669**
Protein crops	0.0637500	0.0545134	0.0272567	-0.0229930	0.1504930	2.339	0.101
Soya	-0.0132200	0.0873351	0.0104385	-0.0340443	0.0076043	-1.266	0.210
Sugarbeet	0.0011690	0.0086062	0.0015981	-0.0021047	0.0044426	0.731	**0.471**
Sunflower	0.0656500	0.2891360	0.2044500	-2.5321336	2.6634336	0.321	**0.802**
Tobacco	0.0662250	0.0761189	0.0380594	-0.0548971	0.1873471	1.740	0.180
Tomato	0.0019500	0.0239079	0.0097604	-0.0231398	0.0270398	0.200	**0.850**
F_maize	-0.0044667	0.0058586	0.0033825	-0.0190203	0.0100869	-1.321	**0.318**
Alfalfa	0.0152000	0.0142836	0.0101000	-0.1131327	0.1435327	1.505	**0.373**

The t-test shows a relevant improvement in the estimation significance for most crops. For soft wheat and barley the *t*-test indicates a higher than 90% probability that the estimated mean is equal to the observed mean, while for sunflower and tomato the significance is over 80%. Sugarbeet, fodder maize and alfalfa also present a very good significance of the mean estimation. The worst results correspond to durum wheat and maize, for which the level of probability that the two means are equal is null. The maize results confirm those obtained for the entire sample.

3.3.2.2 The case of Lombardy region

The estimate obtained for this region, presented in Table 3.7, provides a notable increase in the estimation fitness for durum wheat, barley and soya. For durum wheat the estimated accounting cost is 30% higher than the observed cost, barley +3.8% and soya +17%. For this subset, the estimated accounting cost for soft wheat worsens with respect to Veneto and the entire sample outcomes, with a difference of +14.5% from the observed data.

Table 3.7: Specific cost estimates obtained from PMP model - Lombardy sample

Crop	N_OBS	Observed cost	Std. Error	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
D_wheat	8	**0.05180**	0.01531	**0.06763**	0.01945	**0.01367**	0.00207	**0.08130**	0.00298
S_wheat	46	**0.04845**	0.00310	**0.05546**	0.00777	**0.03845**	0.01148	**0.09391**	0.01065
Maize	74	**0.04618**	0.00300	**0.05910**	0.00322	**0.03877**	0.00402	**0.09786**	0.00365
Barley	23	**0.04124**	0.00753	**0.04281**	0.00837	**0.01783**	0.00482	**0.06064**	0.00639
Rice	38	**0.09353**	0.00635	**0.11371**	0.01094	**0.11537**	0.01487	**0.22908**	0.01493
Sorghum	1	**0.03505**	0.00000	**0.00000**	0.00000	**0.03228**	0.00425	**0.03228**	0.00425
Protein crops	5	**0.07190**	0.01186	**0.15215**	0.03987	**0.02559**	0.01088	**0.17775**	0.01184
Soya	13	**0.07573**	0.01001	**0.08916**	0.01914	**0.04276**	0.01182	**0.13192**	0.01460
Rape	1	**0.03828**	0.00000	**0.00000**	0.00000	**0.02744**	0.00356	**0.02744**	0.00356
Melon	1	**0.01947**	0.00000	**0.08245**	0.02556	**0.26945**	0.12405	**0.35190**	0.14415
Tomato	2	**0.00853**	0.00053	**0.01041**	0.00000	**0.01232**	0.00159	**0.02273**	0.00160
Alfalfa	27	**0.01515**	0.00248	**0.03266**	0.00457	**0.04306**	0.00757	**0.07572**	0.00670
Meadows	24	**0.01526**	0.00158	**0.02669**	0.00612	**0.02157**	0.00575	**0.04826**	0.00680

Figure 3.9 shows that, for Lombardy, the hidden cost represents an important component of the farmer's decision process. In particular, this added marginal cost is important for cereals and alfalfa. Considering that the estimation deviations are all positive, this means that the outcomes overestimate the "real" accounting cost, so the production plan at regional level is strongly influenced by implicit costs that are not captured by the agricultural accounting systems.

An analysis of Figure 3.10 verifies that the estimations for cereals, soya and tomato are roughly near the target value represented by the observed accounting costs, while for protein crops and alfalfa the estimations are far from the target value. This is a simple comparison between two means that has to be verified using a statistical test.

Fig. 3.9: Total marginal cost distribution - Lombardy sample (€/t)

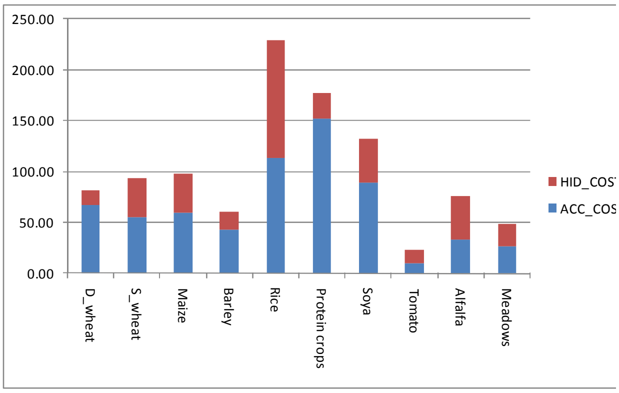
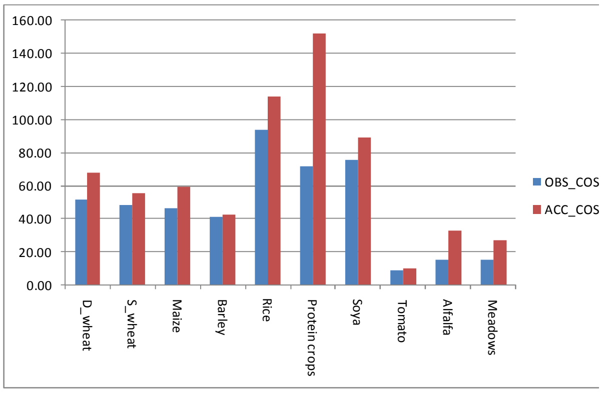


Fig. 3.10: Comparison between observed costs and estimated accounting costs - Lombardy sample (€/t)



In order to verify that the means obtained from the individual estimated accounting costs are representative of the mean originated from the observed values, this hypothesis is submitted to the *t*-test. In this way, it will be possible to accept or reject the estimate obtained by implementing the PMP model. Table 3.8 presents the level of probability associated to paired groups of values (estimated and observed) for each crop. The level of significance is high for durum wheat and soft wheat, indicating that it is not possible to reject the hypothesis that the two means are different, with a probability of 66% and 63% respectively. Barley also shows a high level of significance.

Table 3.8: *T*-test for estimated and observed accounting costs - Lombardy sample

Crops	Paired Differences					t	Sig. (2-tailed)
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper		
D_wheat	0.0086167	0.0448852	0.0183243	-0.0384875	0.0557208	0.470	**0.658**
S_wheat	0.0044280	0.0449068	0.0089814	-0.0141086	0.0229646	0.493	**0.626**
Maize	0.0038746	0.0259270	0.0031675	-0.0024495	0.0101987	1.223	0.226
Barley	-0.0068000	0.0339325	0.0102310	-0.0295961	0.0159961	-0.665	**0.521**
Rice	0.0164342	0.0495990	0.0080460	0.0001314	0.0327370	2.043	0.048
Protein crops	0.0823333	0.1128924	0.0651785	-0.1981069	0.3627736	1.263	**0.334**
Soya	0.0240000	0.0785200	0.0277610	-0.0416444	0.0896444	0.865	**0.416**
Alfalfa	0.0161696	0.0276196	0.0057591	0.0042260	0.0281132	2.808	0.010
Meadows	0.0095091	0.0273577	0.0082487	-0.0088701	0.0278882	1.153	0.276

Despite the t-test conducted for the previous samples, in this case the probability level for maize reveals a value different from zero equal to 22.6%. The significance is evidently low, but we have no reason to reject the hypothesis of equality between the two means for this crop. It is interesting to note that the significance level for maize is better where the cropping technique is almost homogenous in all the area.

3.3.2.3 The case of Piedmont region

The estimation results obtained for Piedmont are very similar to those described for the other two regions and for the entire sample taken as a whole. The crop with high degree of fitness to the observed data is soft wheat (-4.8% with respect to the observed accounting cost), which confirms the good estimates reached previously. Also for barley, this fourth estimate confirms the model stability providing an estimated accounting cost 7% lower than the observed data. The estimation for maize, which is also in this case the crop with the highest numbers observed, presents an estimation equal to +15% of the corresponding observed accounting cost. Among the main crops in the region, the most relevant divergence of the estimations produced and the exogenous information concerns soya (+30%), durum wheat (+57%), grassland (+60%) and tomato (+110%).

Table 3.9: Specific cost estimates obtained from PMP model - Piedmont sample

Crop	N_OBS	Observed cost	Std. Error	Estimated Accountig Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
D_wheat	16	**0.09354**	0.00794	**0.14723**	0.02763	**0.02327**	0.00328	**0.17050**	0.00404
S_wheat	178	**0.07594**	0.00268	**0.07230**	0.00508	**0.04728**	0.00404	**0.11959**	0.00441
Maize	243	**0.06484**	0.00301	**0.07498**	0.00286	**0.06277**	0.00416	**0.13775**	0.00328
Barley	75	**0.06388**	0.00398	**0.05893**	0.00791	**0.02663**	0.00262	**0.08556**	0.00311
Rice	98	**0.12188**	0.00347	**0.12831**	0.00501	**0.04299**	0.00458	**0.17130**	0.00805
Protein crops	14	**0.09415**	0.01373	**0.07256**	0.01998	**0.01450**	0.00171	**0.08706**	0.00216
Soya	23	**0.15148**	0.02335	**0.10574**	0.01940	**0.01785**	0.00311	**0.12359**	0.00517
Sugarbeet	2	**0.00951**	0.00456	**0.00435**	0.00187	**0.00080**	0.00015	**0.00515**	0.00016
Potato	18	**0.06075**	0.01406	**0.12942**	0.02649	**0.06079**	0.01541	**0.19022**	0.02385
Rape	6	**0.24419**	0.05470	**0.11220**	0.03693	**0.01281**	0.00359	**0.12501**	0.00438
Sunflower	7	**0.08773**	0.01833	**0.07497**	0.02138	**0.02376**	0.00212	**0.09873**	0.00248
Tomato	14	**0.06681**	0.02979	**0.13748**	0.05727	**0.01212**	0.00567	**0.14961**	0.01539
Maize	3	**0.04865**	0.01212	**0.01190**	0.00739	**0.00422**	0.00031	**0.01612**	0.00031
T_grass	23	**0.02434**	0.00287	**0.03912**	0.00809	**0.00161**	0.00026	**0.04072**	0.00101
Alfalfa	24	**0.01392**	0.00175	**0.02195**	0.00501	**0.02338**	0.00189	**0.04533**	0.00187
Pasture	1	**0.00571**	0.00000	**0.00000**	0.00000	**0.00275**	0.00049	**0.00275**	0.00049
Meadow	92	**0.01329**	0.00098	**0.03257**	0.00379	**0.01006**	0.00104	**0.04263**	0.00160

Fig. 3.11: Total marginal cost distribution - Piedmont sample (€/t)

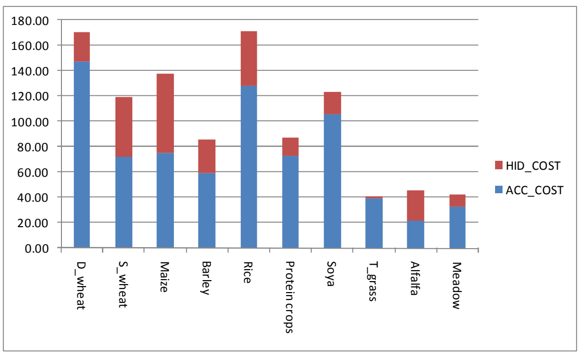
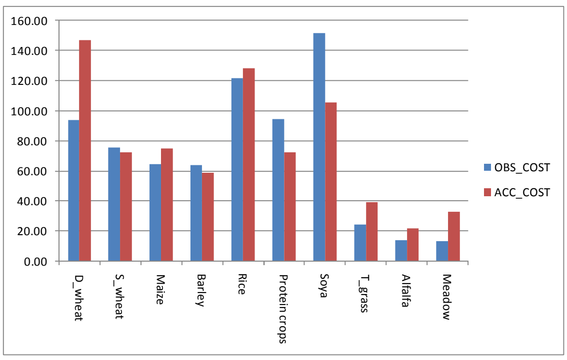


Fig. 3.12: Comparison between observed costs and estimated accounting costs - Piedmont sample (€/t)



Considering the *t*-test, Piedmont seems to be better than the estimation obtained for the entire sample, confirming that a greater degree of territorial homogeneity improves the fitness of the estimates to the observed data. Soft wheat, barley, protein crops, soya, sugarbeet and sunflower present a level of significance between 55% and 82%, which means a very high probability that the estimated mean is equal to the mean generated by the observed data. The mean derived from the estimated individual accounting cost for maize is associated to a very high probability that it is different from the observed mean.

Table 3.10: *T*-test for estimated and observed accounting costs - Piedmont sample

Crops	Paired Differences					t	Sig. (2-tailed)
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper		
D_wheat	0.0537818	0.0907642	0.0273664	-0.0071944	0.1147580	1.965	0.078
S_wheat	0.0022885	0.0605872	0.0053139	-0.0082251	0.0128020	0.431	**0.667**
Maize	0.0139077	0.0389144	0.0026918	0.0086010	0.0192143	5.167	0.000
Barley	-0.0061674	0.0704716	0.0103905	-0.0270949	0.0147601	-0.594	**0.556**
Rice	0.0062854	0.0404977	0.0041333	-0.0019202	0.0144910	1.521	0.132
Protein crops	-0.0194625	0.0911720	0.0322342	-0.0956842	0.0567592	-0.604	**0.565**
Soya	-0.0171571	0.1360563	0.0363626	-0.0957137	0.0613994	-0.472	**0.645**
Sugarbeet	-0.0052000	0.0128693	0.0091000	-0.1208265	0.1104265	-0.571	**0.670**
Potato	0.0678133	0.0687080	0.0177403	0.0297641	0.1058626	3.823	0.002
Rape	-0.1044000	0.2020111	0.0903421	-0.3552299	0.1464299	-1.156	0.312
Sunflower	-0.0076667	0.0787790	0.0321614	-0.0903402	0.0750069	-0.238	**0.821**
Tomato	0.0670846	0.2364237	0.0655721	-0.0757848	0.2099540	1.023	**0.326**
F_Maize	-0.0358500	0.0511238	0.0361500	-0.4951793	0.4234793	-0.992	**0.503**
T_grass	0.0197455	0.0337036	0.0101620	-0.0028969	0.0423878	1.943	0.081
Alfalfa	0.0079077	0.0222103	0.0061600	-0.0055139	0.0213293	1.284	0.223
Meadows	0.0191077	0.0275874	0.0038257	0.0114273	0.0267881	4.995	0.000

The significance level for durum wheat, potato, temporary grass and meadows causes the null hypothesis according to which the estimated and observed means are equal to be rejected.

3.3.3 Homogeneous group of farms identified through the cluster analysis

The previous analysis has demonstrated that there is a positive correlation between the level of homogeneity of the investigated group of farms and the estimation fitness on observed data. In order to improve the analysis and minimize the risk of retaining outliers in the estimation, it is necessary to increase the level of homogeneity of the groups to manage with the PMP model. The purpose of this further step of methodology testing is to evaluate the response of the model to a more homogenous group of farms. To meet this objective, a cluster analysis has been conducted on the entire set of information for the three Italian regions. The cluster analysis has been developed using the K-mean method, the best-known and applied partitioning method (for a review, see Atkinson et al., 2004). This procedure classifies the n units into k distinct clusters, with k chosen a priori by the analyst, according to a step-by-step iterative method that reaches the optimal distribution of observations into the defined groups.

In the first part of this section, the cluster analysis method is described with reference to the Italian dataset, in order to present the procedure generating the clusters and the statistical justification for the achieved results. The cluster analysis is preceded by a principal component analysis to identify the explanatory variables of the sample under investigation. Once the optimal number of clusters and the group of farms have been identified, the most numerous group will be submitted to the estimation procedure.

3.3.3.1 Multivariate Analysis

The goal of this analysis is the development of a process for the definition of homogeneous groups that can be used in the mathematical programming model.

As shown in section 3.2, farms belonging to the same FT in the same region can present a strong variability in terms of prices and yields. This variability can strongly influence the quality of the estimation of the accounting cost. The objective is therefore to use a group of farms homogeneous in respect to these two variables in the analysis.

In particular, the analysis is conducted through principal components analysis and cluster analysis starting from farm data on crop production of three different countries: Belgium, Italy and Hungary.

We can distinguish different steps in this phase:

1. Construction of the database for reference state;
2. Univariate and Bivariate Analysis;
3. Principal Components Analysis (PCA);
4. Cluster Analysis (CA).

1) Database construction for each region

All the data were arranged in a matrix where farms specialized in crops were analyzed for each crop by price and yield variables. In this way three datasets were made:

- The Belgian dataset (FADN 2006 data) is formed by 105 farms with 10 crops (soft wheat and spelt, barley, oats, grain maize, other cereals, potatoes, sugarbeet, other industrial crops, fresh vegetables in rotation, fresh vegetables grown in a market garden);
- The Italian dataset (2007 – from three Italian regions: Piedmont, Lombardy and Veneto) with 448 farms and 14 crops (soft wheat and spelt, durum wheat, grain maize, barley, rice, sorghum, soya, sugarbeet, potatoes, tomatoes, fodder maize, alfalfa, temporary grass, meadows);
- The Hungarian dataset (FADN 2007 data) is formed by 788 farms and 23 crops (soft wheat and spelt, durum wheat, rye, barley, oats, grain maize, rice, other cereals, potatoes, sugarbeet, rape, sunflower, soya, other oilseeds, fresh vegetables, grassland,

fodder maize, other silage, other fodder plants, temporary grass, pasture, rough grazing)

2) Univariate and Bivariate Analysis

The first insights into the data matrix were with univariate analysis, which provided a set of summary information with respect to the individual variables considered. This analysis emphasized the statistical units of the number of total cases, at a minimum, a maximum, the arithmetic mean and mean square deviation. This first analysis was very important because it provided a first indication on the data: the standard deviation, expressed in the same unit as the average, showing the dispersion from the mean. Correlation analysis has been conducted in order to assess the relationship between yields and prices of the different crops.

3) Principal Components Analysis

This phase identified a smaller number of synthetic indicators for yields and prices for each country. Many available variables have been summarized in a smaller number of not observable factors or latent variables. To achieve this, Principal Component Analysis (PCA) has been applied. The variables involved in this step are the same as those considered in the correlation matrices of each country. Indeed the quantitative and significance analysis showed good results among these groups of variables.

Annex 3 reports:

- the shares of variance of each variable, previously standardized;
- the variables are explained by the Principal Component (PC);
- the total variance explained by the PC.

The analysis can be summarized as follows:

- Belgium: PCA applied to the Belgian farm yields has identified four components; the total percentage of variance explained is equal to 62.525%. The analysis of Belgian farm prices has identified four components and the total percentage of variance explained is 59.257%
- Italy: PCA applied to the Italian farm yields has found seven components; the total percentage of variance explained is equal to 60.232%. The analysis of Italian farm prices has detected six components and the total percentage of variance explained is 53.639%.
- Hungary: PCA applied to the Hungarian farm yields has found eleven components; the total percentage of variance explained is equal to 58.214%. The analysis of Hungarian farm prices has identified ten components and the total percentage of variance explained is 57.495%.

Furthermore, the synthesis capacity of the components extracted in relation to each variable is shown in the communalities tables (Annex 1). As can be seen, almost all variables are reproduced quite well in the dimensional space and the number of Principal Components identified from the analysis is appropriate for three reasons:

- the percentage of variance explained by its principal component is greater than or close to 0.5987 (0.95^{10}) in the case of Belgium, and 0.4877 (0.95^{14}) and 0.3073 (0.95^{23}) for Italy and Hungary respectively;
- only the selected components have eigenvalue greater than one;
- the scree-plot of the eigenvalues obtained are decreasing.

The results obtained highlight that the extensive information provided by variable prices and yields for individual crops in different countries have been simplified in a satisfactory way by the PCA.

4) Cluster Analysis

As the number of cases exceed one hundred units for all three countries, the non-hierarchical type with the k-means method has been chosen. This classification methodology identifies, starting from the variables and units, one partition in a number of groups established a priori that satisfies a certain excellent condition.

The variables used in this analysis are the main components extracted by the PCA. These components summarize the variables of prices and yields of each country's crops. Given the difficulty that can be encountered in the objective definition of the number of groups, the analysis has been repeated several times in order to identify the best classification through the R^2 indicator.

As can be seen from the tables in Annex 2, increasing the number of groups R^2 also grows, thanks to the creation of more homogeneous groups. Moreover it can be highlighted that the classifications of each country can be satisfied only when the value of R^2 is greater than 0.5. Can thus be identified:

- 8 groups for Belgium with an R^2 value of 0.70.
- 10 groups for Italy with an R^2 value of 0.57.
- 10 groups for Hungary with an R^2 value 0.50.

ANOVA tables (Annex 2) can be derived from the values of the total deviance, the deviance among groups and the deviance within groups.

It can be highlighted that the partitions identified are quite satisfying, but the variability within groups can't be underestimated (30% of the total deviance for Belgium, 43% of the total deviance for Italy and 50% of the total deviance for Hungary).

For further research, the outputs on the initial clusters' centres, iteration history, final cluster centres and number of cases in each cluster are listed in Annex 2.

3.3.3.2 Estimation results

As described in the previous section, the number of clusters that responds better to the criteria of homogeneity for Veneto-Lombardy-Piedmont is 10, where it is possible to observe a strong concentration of farms within the sixth cluster (384 farms), while the others are dispersed among the remaining groups. These last groups are considered as the groups that may contain values disturbing the estimation of accounting costs when considered in a unique group for the PMP estimation.

The analysis of Table 3.11 highlights that some crops present in the previous evaluations are missing, like durum wheat and rice. These two crops, for instance, are present on farms that are not considered in the sixth cluster. The degree of homogeneity is thus reliant on the level of farm specialization, so that farms specialized in rice production with a technology quite different from the other farms are not captured by the most numerous group. The same consideration can be made for tomato production, which is also missing in the sixth cluster.

Comparing the observed with the estimated accounting cost, the percentage deviation is more smoothed than the results achieved with the other samples. Soft wheat and barley confirm the excellent estimation goodness with a deviation of 8.6% from the observed accounting costs. This is also a sign of the high uniformity in the technology for these two crops. All the estimates obtained for soft wheat and barley have given results close to the observed reality. For maize, the deviation is quite restrained, +16% on the observed information; for soya the variation is about 11%, while for sugarbeet and alfalfa the results are more satisfactory, with a deviation of 8% and 0.6% respectively. Only sunflower has a strong difference from the observed value, equal to -41% (see Figure 3.14).

Table 3.11: Specific cost estimates obtained from PMP model, Veneto-Lombardy-Piedmont, 10 groups, 6th cluster

Crop	N_OBS	Observed cost	Std. Error	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
S_wheat	197	**0.07113**	0.00231	**0.06501**	0.00383	**0.04752**	0.00718	**0.11254**	0.00511
Maize	311	**0.06106**	0.00133	**0.07100**	0.00175	**0.04258**	0.00267	**0.11357**	0.00230
Barley	62	**0.06208**	0.00504	**0.05673**	0.00699	**0.01929**	0.00367	**0.07602**	0.00612
Protein crops	11	**0.09320**	0.01502	**0.08331**	0.00980	**0.01383**	0.00897	**0.09714**	0.00948
Soya	74	**0.11812**	0.00659	**0.10452**	0.00804	**0.02489**	0.00608	**0.12941**	0.00784
Sugarbeet	17	**0.01452**	0.00079	**0.01569**	0.00207	**0.00179**	0.00043	**0.01747**	0.00065
Rape	6	**0.11845**	0.02814	**0.11202**	0.03581	**0.09389**	0.03235	**0.20591**	0.03229
Sunflower	8	**0.12248**	0.02720	**0.07227**	0.03161	**0.00601**	0.00245	**0.07828**	0.02434
T_grass	5	**0.03504**	0.00623	**0.03290**	0.01416	**0.00040**	0.00025	**0.03331**	0.00323
Alfalfa	6	**0.01432**	0.00296	**0.01423**	0.00168	**0.01149**	0.00760	**0.02572**	0.00750
Pasture	1	**0.00571**	0.00000	**0.00000**	0.00000	**0.00104**	0.00022	**0.00104**	0.00022
Meadows	77	**0.01404**	0.00113	**0.02553**	0.00407	**0.01834**	0.00230	**0.04387**	0.00331

Fig. 3.13: Total marginal cost distribution, Veneto-Lombardy-Piedmont - 10 groups, 6th cluster (€/t)

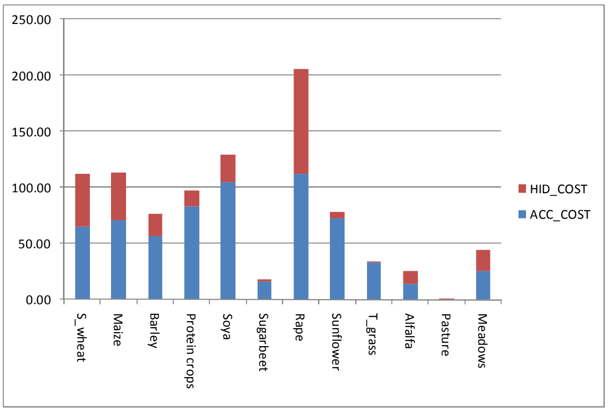
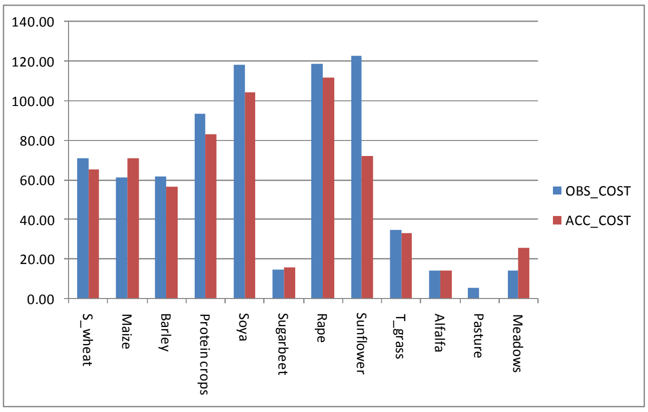


Fig. 3.14: Comparison between observed costs and estimated accounting costs, Veneto-Lombardy-Piedmont - 10 groups, 6th cluster (€/t)



The *t*-test for the estimates provides significance values that are higher on average than the results obtained for the previous samples. Estimated costs for some crops (like sugarbeet) that presented poor *t*-test significance in the other sample, improve their fitness in this sample, while for other crops, like soft wheat, the estimates worsen remaining in an acceptable range of significance. The case of soft wheat, for example, explains this kind of result: compared to the estimates obtained with the other sample stratification, the estimates made on clusters are worse, indicating a significance of 30% that is much lower than the significance of about 60% obtained in the other estimation procedures.

Table 3.12: T-test for estimated and observed accounting costs, Veneto-Lombardy-Piedmont, 10 groups, 6th cluster

Crops	Paired Differences					t	Sig. (2-tailed)
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper		
S_wheat	-0.0044612	0.0517577	0.0042689	-0.0128980	0.0039756	-1.045	**0.298**
Maize	0.0099957	0.0233423	0.0013499	0.0073391	0.0126522	7.405	0.000
Barley	-0.0058595	0.0642430	0.0099129	-0.0258790	0.0141600	-0.591	**0.558**
Protein crops	-0.0185571	0.0733837	0.0277364	-0.0864257	0.0493115	-0.669	**0.528**
Soya	-0.0104111	0.0820993	0.0103435	-0.0310876	0.0102653	-1.007	**0.318**
Sugarbeet	0.0012933	0.0082950	0.0021417	-0.0033003	0.0058869	0.604	**0.556**
Rape	-0.0375500	0.0942819	0.0471409	-0.1875735	0.1124735	-0.797	**0.484**
Sunflower	-0.0508286	0.1304534	0.0493068	-0.1714779	0.0698207	-1.031	**0.342**
T_grass	-0.0061500	0.0577706	0.0408500	-0.5251985	0.5128985	-0.151	**0.905**
Alfalfa	-0.0048000	0.0008485	0.0006000	-0.0124237	0.0028237	-8.000	0.079
Meadows	0.0111474	0.0277703	0.0045049	0.0020195	0.0202752	2.474	0.018

The cluster construction improves the average significance of the estimates, extending the number of processes with good fitness to observed data. Nevertheless, if this is true for most specific cost estimations, the accounting cost for maize continues to remain insignificant for the *t*-test. This is probably due to the wide variability of observed accounting costs with respect to the value of yields and prices for this particular crop.

4 Belgium

4.1 Data entry description and quality control procedure

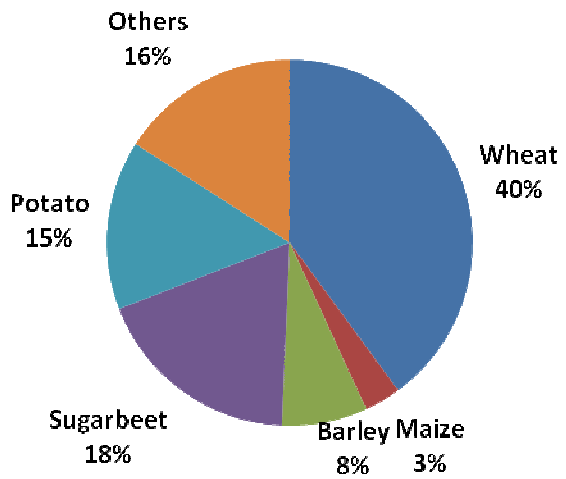
The farm sample considered in this study is composed of 105 farms belonging to FT1 (arable crops). The average size of each farm is 59 ha. The incidence of cereal acreage on the total UAA is 51% . The average GSP per hectare produced by Belgium is 2.206 Euros, while the total variable cost per hectare is 877 Euros (Table 4.1).

Table 4.1: Statistical description of Belgian Sample – Farm type 1

	N. of farms	Av. UAA (ha)	Cereals / tot (%)	GSP/ha (€)	Total Variable Costs /ha (€)
Total	105	59	51	2206	877

The most relevant crop, in terms of land area, is soft wheat that covers 40% of the total acreage, followed by sugarbeet with 18% and by other crops with 16%. The Belgian FT1 also has a significant presence of potatoes, which cover 15% of the total surface (Fig. 4.1).

Fig. 4.1: Crop distribution in the entire Farm Type 1 sample area - Belgium



Comparing this data with the information of the Italian sample, it is evident that the farm sample considered for Belgium consists of fewer units, and that the average size of each farm is roughly similar. The incidence of cereals on the total UAA is higher for Belgium, although in Veneto (Italy) the percentage of incidence exceeds 60%. Land productivity, measured by GSP/ha, is very high in Belgium, but this value is also accompanied by higher total variable costs.

Analyzing crop distribution, both countries have a relevant presence of cereals, but there is a wider variety of crops in Belgium. In addition to wheat, maize and barley, also sugarbeet and potatoes are significant in terms of land area.

Each crop is used in the PMP model in order to estimate the specific production cost. As known, the estimation is carried out using - for each crop at farm level - the information on acreage, yields and prices and - at farm level - the information on total variable cost. Furthermore, in order to obtain an appropriate level of estimation, it is important to verify the quality of the data in terms of the level of homogeneity and presence of outliers that can disturb the estimates.

Figure 4.2 and Table 4.2 present some descriptive information about prices and yields of soft wheat, barley, sugarbeet and potatoes. In particular, Figure 4.2 shows the degree of dispersion of observations for these crops with respect to prices and yields. For soft wheat the dispersion is low, although there are some observations out of the range. For this crop the standard deviation is irrelevant for the yields, while for the prices there is a variation with respect to the mean of 13.8 €/ton. For barley there is a high level of dispersion, mainly due to variation in prices, with a standard deviation equal to 14.6. For sugarbeet there is a good level of homogeneity, with few farms that deviate from the core of observations. In this case the main factor that influences the dispersion is the variation in yields, equal to 9.7 ton/ha with respect to the mean. Finally, the potato crop is characterized by a high level of dispersion, with a standard deviation equal to 84.9 for prices, while the variation in yields is more restrained: 10.7 ton/ha.

Fig. 4.2: Price and yield distribution in FT1 sample for Belgian case study

Fig. 4.2.a: Soft wheat

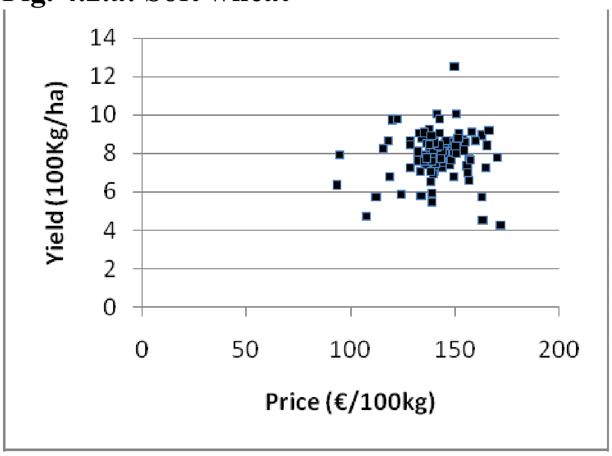


Fig. 4.2.b: Barley

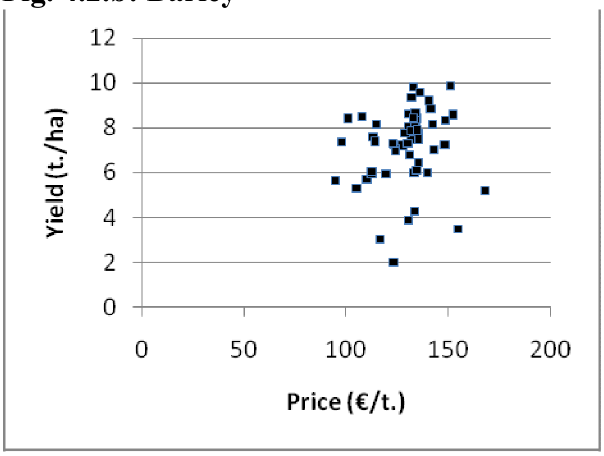


Fig. 4.2.c: Sugarbeet

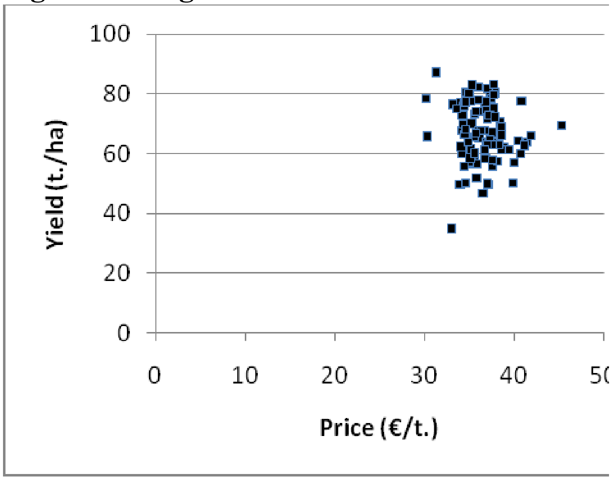


Fig. 4.2.d: Potatoes

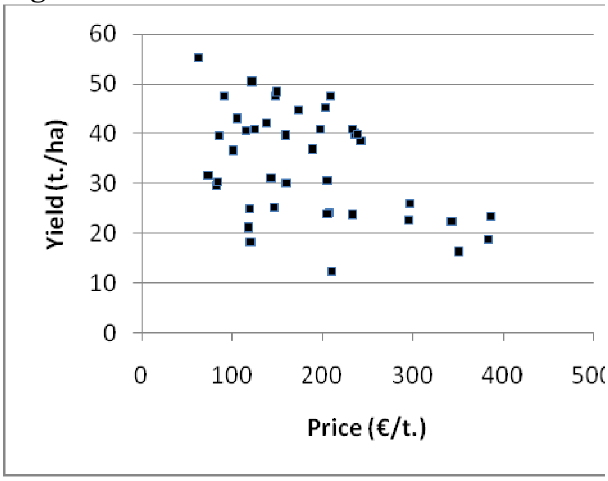


Table 4.2: Descriptives of some crops selected from Belgium sample: price in €/ton; yields in ton/ha

Crop	Variable	N. of Obs.	MIN	MAX	Mean		Std. Deviation
					Statistic	Std. Error	
Soft wheat	Prices	101.00	93.38	171.87	141.95	1.37	13.81
	Yields	101.00	4.27	12.50	7.94	0.12	1.20
Barley	Prices	54.00	94.63	168.27	129.04	1.99	14.63
	Yields	54.00	2.02	9.88	7.15	0.23	1.70
Sugar beet	Prices	95.00	30.21	45.38	36.52	0.25	2.40
	Yields	95.00	34.89	87.22	67.03	1.00	9.71
Potato	Prices	41.00	62.53	386.21	182.41	13.26	84.93
	Yields	41.00	12.38	55.33	33.97	1.68	10.78

The existence of outliers that can disturb the estimates is present not only at process level, but also at farm level. Indeed, considering the farm distribution between gross saleable production (GSP) and farm total variable cost (TVC) it is possible to note that there is a high level of dispersion, also at reduced scale.

Fig. 4.3: Farm distribution between GSP/ha and TVC/ha (Belgium) - standard scale

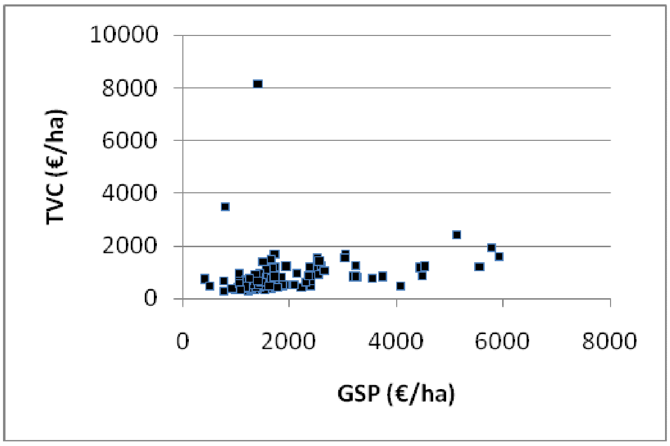
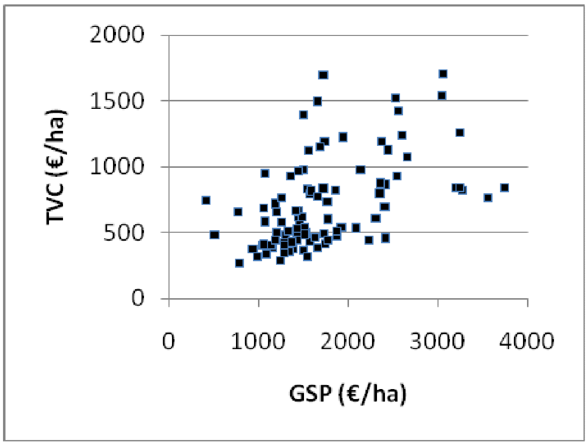


Fig. 4.4: Farm distribution between GSP/ha and TVC/ha (Belgium) - reduced scale



4.2 Estimation results

In comparison with the previous estimation on Italian FADN data, the cost assessment for Belgium cannot be made with the support of the observed specific cost information since, unfortunately, the Belgian national specific costs are not available for privacy policy issues. This is why the following results will be provided without the t-test and, thus, without the validation process adopted for the Italian estimates. This preliminary comment is essential for justifying the lack of a relevant phase of the estimation method test.

Table 4.3 shows the estimation achieved by the PMP method for the different Belgian crops. Also in this estimation, the hidden costs, which represent the part of the marginal cost not considered by the farm accounting system, are lower than the estimated accounting costs (see Figure 4.5). This means that the farm decision process has been largely driven by the registered costs rather than the cost related to the farmers experience, risk attitude and so on. In this perspective, the hidden cost may be viewed as a component of the transactional costs that each farmer takes into account for deciding which configuration to give to its own farm activity.

If one compares the results obtained for Belgium with the same outcomes obtained for Italy, it is possible to have a very narrow appreciation of the estimates. Soft wheat, for example, shows an estimated accounting cost of 48 €/t, while for the entire Italian sample estimation the accounting specific cost for this crop is 66 €/t. For the other important crops in Belgium, like barley, maize, potato and sugarbeet, the comparison is presented through table 4.4. This comparison obviously doesn't constitute a method for checking the goodness of estimation, but in any case it provides a reference point for judging the estimation scale and the degree of approximation to the Italian validated estimates.

Table 4.3: Specific cost estimates obtained from PMP model - Belgium

CROP	n. of observations	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
Soft wheat	89	**0.48046**	0.04293	**0.33426**	0.04211	**0.81472**	0.04376
Barley	46	**0.53470**	0.07100	**0.29442**	0.04667	**0.82912**	0.05754
Oats	5	**0.52034**	0.21729	**0.20539**	0.03489	**0.72573**	0.04798
Maize	24	**0.51728**	0.09020	**0.18926**	0.05328	**0.70654**	0.08664
Dry pulse	1	**0.11304**	0.00000	**0.02149**	0.01006	**0.13453**	0.01422
Potato	36	**0.89561**	0.12271	**0.36658**	0.05303	**1.26218**	0.09923
Sugarbeet	85	**0.21667**	0.00912	**0.10152**	0.01366	**0.31819**	0.00845
Industrial crops	32	**0.82557**	0.47571	**0.16823**	0.01947	**0.99380**	0.14274
Fresh vegetables	19	**0.70411**	0.11050	**0.36271**	0.08833	**1.06682**	0.12140
Grass seeds	3	**4.41615**	2.43979	**0.17420**	0.07206	**4.59035**	0.29726

Fig. 4.5: Total marginal cost distribution - Belgium - (€/t)

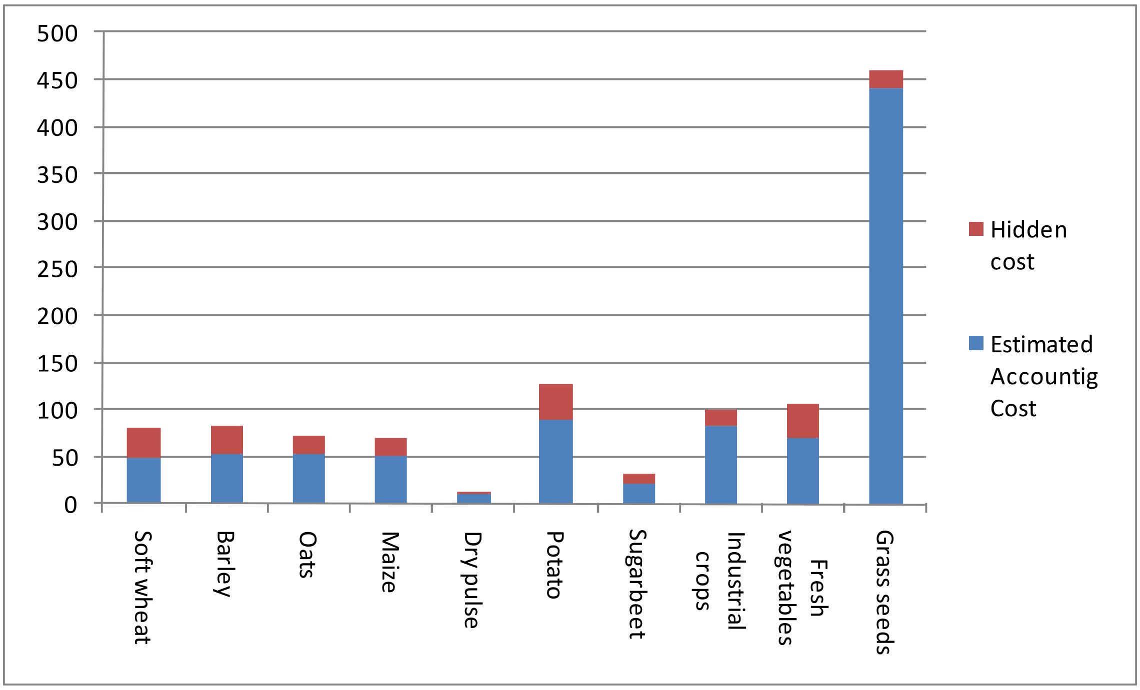


Table 4.4: Comparison between accounting cost estimates on Italian and Belgian sample -€/t

Crops	Italian FADN sample	Belgian FADN sample
Soft wheat	66	58
Barley	51	53
Maize	74	51
Potato	126	90
Sugarbeet	17	22

4.3 Estimation on cluster analysis results

As operated with the three Italian regions, the Belgian sample has been submitted to a cluster analysis for grouping farms on the basis of the specific production characteristics. The cluster analysis has identified 8 groups of farms similar for economic and production characteristics (mainly price and yield per crop), with a correlation index equal to 70%. A larger number of groups would increase the global R^2, but with an improvement lower than the preceding ones (see Table 4.5). Observing the distribution of farms among the eight clusters, it appears that the fifth group is the most numerous cluster. For this reason, the estimation procedure will be implemented with respect to the fifth cluster in order to reach outcomes based on more homogeneous information.

Table 4.5: Cluster analysis results for Belgium

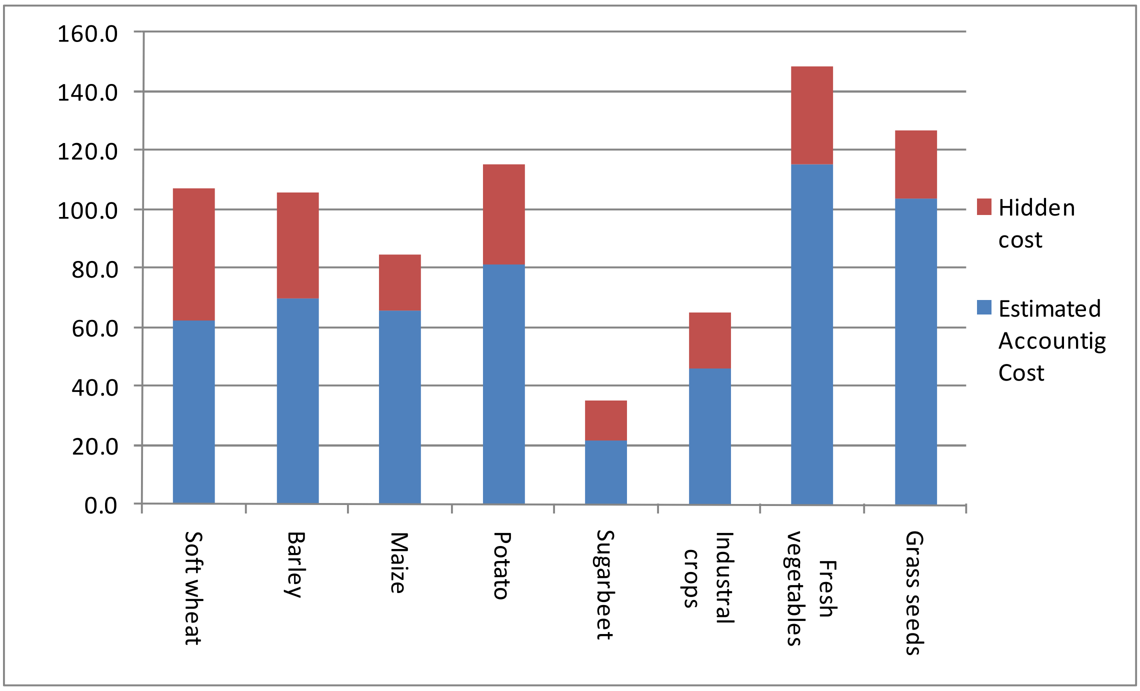
N. of clusters	Clusters										Global R2
	1	2	3	4	5	6	7	8	9	10	
2	8	97									0.18
3	94	3	8								0.34
4	61	33	8	3							0.47
5	9	3	2	2	89						0.49
6	10	2	34	2	54	3					0.61
7	9	2	1	83	6	3	1				0.60
8	**9**	**1**	**14**	**2**	**69**	**6**	**1**	**3**			**0.70**
9	7	1	12	4	3	2	1	3	72		0.71
10	5	33	8	2	4	1	1	3	47	1	0.76

According to the fifth cluster of the 8 identified groups, the PMP estimation has produced the results presented in Table 4.6. The stratification operated inside the entire sample has led to higher cost estimates than those previously obtained, indicating that the group of farms on average has more intensive input use than the entire sample considered as a whole. Observing the distribution of the marginal costs between the two components, accounting and hidden costs, it seems that the hidden component in this group of farms is more relevant. The same result was achieved for the Italian cluster outcomes, where the hidden marginal cost component was wider than in the case of the estimation for the entire group of farms. This result can be attributed to the presence of some farms out of scale that have perturbed the estimation, but also to the presence on average of higher opportunity costs on the specific crop. This means that the substitution of crops with higher hidden cost implies a greater sacrifice in economic terms.

Table 4.6: Specific cost estimates obtained from PMP model - Belgium, 8 groups, 5th cluster

Crop	No. of obervations	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
Soft wheat	62	**0.62145**	0.05590	**0.44848**	0.06059	**1.06993**	0.05311
Barley	35	**0.69956**	0.09701	**0.36031**	0.04760	**1.05987**	0.05792
Maize	9	**0.65595**	0.09826	**0.19377**	0.02505	**0.84972**	0.03416
Potato	16	**0.81329**	0.16252	**0.33673**	0.04855	**1.15002**	0.09060
Sugarbeet	62	**0.21687**	0.01237	**0.13403**	0.01698	**0.35090**	0.00736
Industrial crops	27	**0.45821**	0.10932	**0.19398**	0.02421	**0.65219**	0.04483
Fresh vegetables	9	**1.15084**	0.06589	**0.33529**	0.09024	**1.48613**	0.11352
Grass seeds	1	**1.04061**	0.00000	**0.22988**	0.03573	**1.27049**	0.04534

Fig. 4.6: Total marginal cost distribution - Belgium, 8 groups, 5th cluster - (€/t)



The results obtained for this group of farms can be compared with the same results obtained for Italy, as for the entire sample estimates. Table 4.7 presents the estimation results for most relevant Belgian FADN crops compared with the results obtained for Italy. From these values, it appears that the estimation obtained for the sub-sample given by the cluster analysis is in line with the data obtained for Italy. Only the estimation for barley seems to be quite far from the Italian results and Belgian entire sample outputs. This specific case may be attributed to the more specialized techniques for producing barley present in this sample that entail a more intensive input use. Sugarbeet estimation for both the Italian FADN sample and the Belgian sample remains substantially unchanged, since this is a crop characterized by a very uniform technology that doesn't change very much with respect to territory, farm type and farm size.

Table. 4.7: Comparison between accounting cost estimates on Italian and Belgian samples with cluster analysis (CA) - €/t

Crops	Italian FADN sample	Italian FADN sample with CA	Belgian FADN sample	Belgian FADN sample with CA
Soft wheat	66	65	58	62
Barley	51	57	53	70
Maize	74	71	51	66
Potato	126	-	90	81
Sugarbeet	17	16	22	22

5 Hungary

5.1 Data entry description and quality control procedure

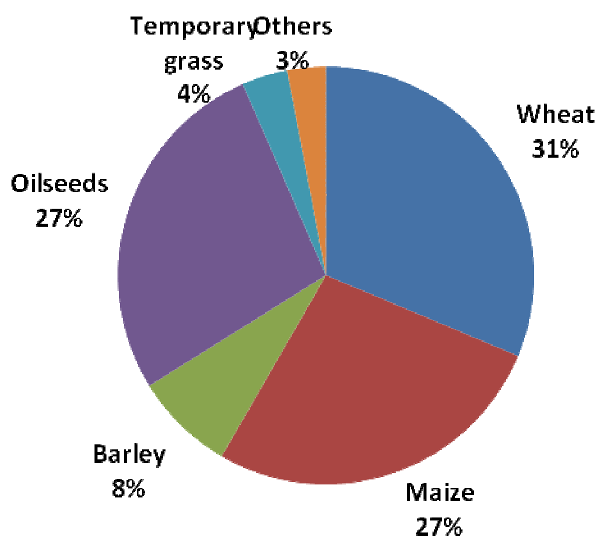
The farm sample considered for the Hungarian analysis is composed of 788 farms, belonging to Farm Type 1 (arable crops). As illustrated in the following table, the sample is characterized by a large farm size in term of hectares, with the average being equal to 228 ha. The incidence of cereals on the total UAA is also very high: about 68% on average in the entire sample. The average GSP per hectare is 734 Euros, while the total variable cost per hectare is 265 Euros (Table 5.1).

Table 5.1: Statistical description of Hungarian Sample – Farm Type 1

	N. of farms	Av. UAA (ha)	Cereals / tot (%)	GSP/ha (€)	Total Variable Costs /ha (€)
Total	788	228	68	734	265

The most important crop in terms of land area is wheat, which covers 31% of the total acreage, followed by maize and oilseed, both with 27%. Other important crops are barley, which represents 8% of the entire acreage, followed by temporary grass and the other crops, with 4% and 3% respectively (Figure 5.1).

Fig. 5.1: Crop distribution in the entire Farm Type 1 sample area - Hungary



The Hungarian sample consists of a good number of farms, as for the Italian analysis. Hungary shows the largest farm size in terms of hectares, about four times that of the other two countries. Cereals have a strong incidence on the total UAA, likewise Italy (43%) and Belgium (51%). The average GSP per hectare appears to be lower than the other two countries, as well as the total variable costs per hectare.

The Hungarian region is characterized by a high level of crop specialization, with a relevant presence of cereals in terms of occupied surface. As seen, this issue brings together the three case studies analyzed, although Belgium and Hungary present a higher level of diversification.

The PMP model has allowed the specific production cost for each crop to be estimated. This estimation is made using the information on acreage, yields, prices for each crop at farm level and the total variable cost at farm level. There are also outliers in this analysis that can disturb the estimates. Figure 5.2 shows the degree of dispersion of observations, in respect to prices and yields, of four main crops included in the sample. Instead, Table 5.2 presents some descriptive information about prices and yields for these crops.

On a first analysis, it is possible to note that the observations are less dispersed for some crops, like soft wheat and sunflowers, while for others, like barley and maize, the dispersion is very high. In all cases, the main factor that influences the dispersion is the variation in prices, although it is very high for maize - with a variation with respect to the mean of 123.6 €/ton - while for the other crops the dispersion in prices is more restrained. In the case of maize there are several observations out of range; these outliers should be corrected before the estimation of specific production cost.

Fig. 5.2: Price and yield distribution in FT1 sample for the Hungarian case study

Fig. 5.2.a: Soft wheat

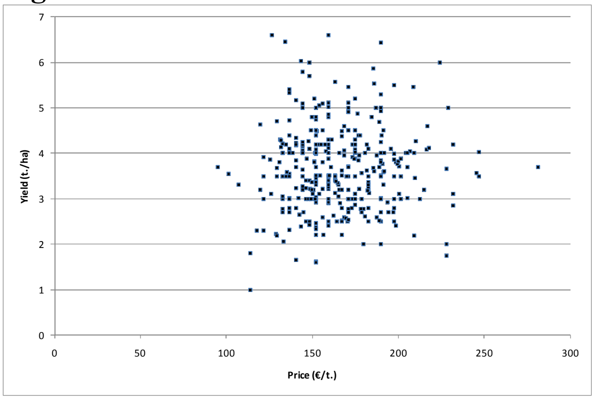


Fig. 5.2.b: Maize

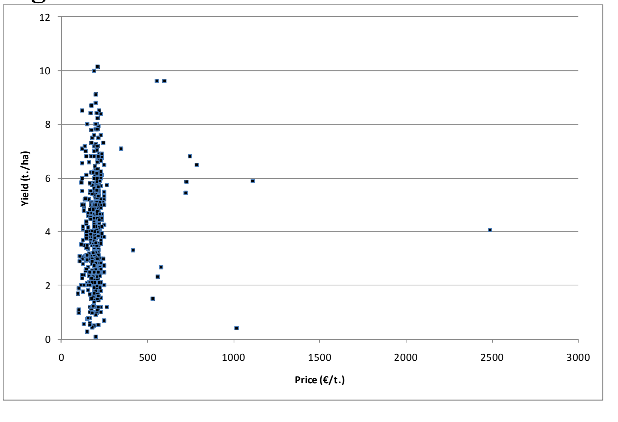


Fig. 5.2.c: Barley

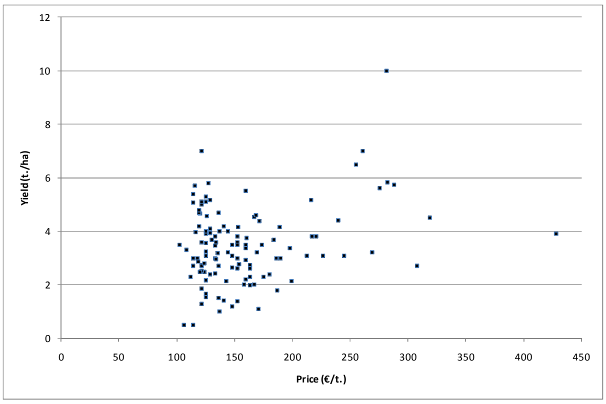


Fig. 5.2.d: Sunflowers

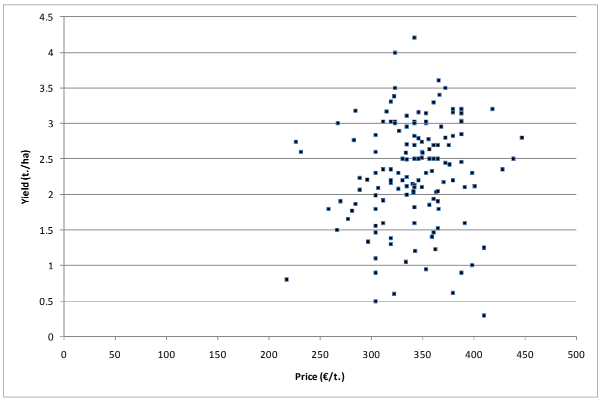


Table 5.2: Descriptives of some crops selected from Hungarian sample: price in €/ton; yields in ton/ha

Crop	Variable	N. of Obs.	MIN	MAX	Mean		Std. Deviation
					Statistic	Std. Error	
Soft wheat	Prices	690.00	76.00	287.13	164.80	1.08	28.44
	Yields	679.00	0.80	6.81	3.52	0.04	0.96
Maize	Prices	676.00	93.94	2489.00	204.52	4.75	123.61
	Yields	656.00	0.08	10.15	3.96	0.07	1.84
Barley	Prices	397.00	79.80	428.07	171.10	3.29	65.50
	Yields	371.00	0.14	10.57	3.61	0.09	1.78
Sunflower	Prices	498.00	190.00	1368.00	331.56	3.60	80.40
	Yields	493.00	0.06	4.33	2.20	0.04	0.78

Also in the Hungarian sample, the presence of outliers is evident not only at process level but also at farm level, in reference to gross saleable production (GSP) and farm total variable cost (TVC). As illustrated in the figure below, the outliers can be appreciated both at standard scale and at reduced scale.

Fig. 5.3: Farm distribution between GSP/ha and TVC/ha (Hungary) - standard scale

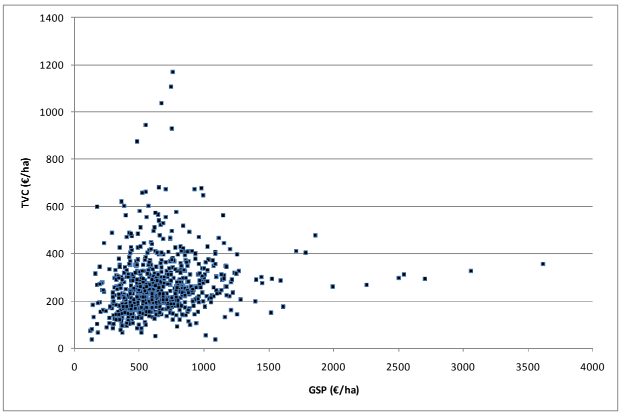
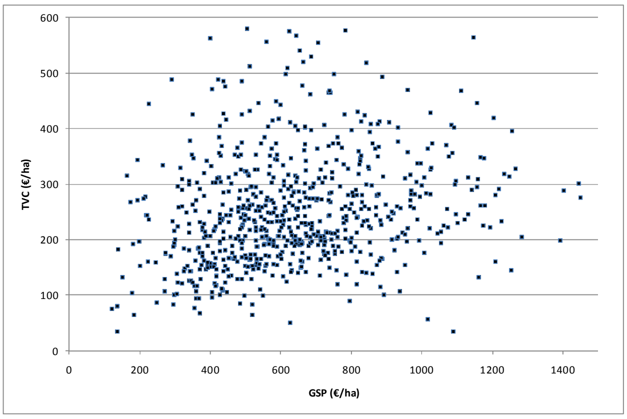


Fig. 5.4: Farm distribution between GSP/ha and TVC/ha (Hungary) - reduced scale



5.2 Estimation results

As described above, the Hungarian estimation is based on a sample of farms collected from the FADN 2007. The estimation has involved all the farms belonging to FT1 in Hungary, i.e. 778 farms. In this case study, as in the other ones, the PMP procedure is applied on the entire sample and then on a sub-group of farms identified using the cluster analysis technique. The estimation results cannot be compared with the observed marginal cost data, since privacy commitments prevent national information provision. T-tests are, thus, not allowed and the validation step cannot be carried out.

Table 5.3 presents the PMP estimation outputs. Hungary provides several observations for the main crops. For instance, the estimation developed for soft wheat is based on 679 observations, maize 656, sunflower 493 and barley 371. This rich amount of information could represent an important straightness for testing the goodness of fit of the model. Notwithstanding the lack of the observed accounting costs per crop, one can compare the results achieved with the same information obtained for the other two samples. Table 5.4 compares the estimates for the three case studies for the main crops resulting from the Hungarian FADN. Observing these estimates, a certain difference appears between the results obtained for Italy and Belgium, and the results obtained for Hungary. These latter are systematically lower than the estimation values related to the other two samples. In particular, the marginal accounting cost for soft wheat estimated for Hungary is roughly half of those estimated for Italy and Belgium. The difference is also quite high for maize and much more so for rape and oilseeds. These results have to be further investigated in order to evaluate the estimation correctness. The unique data available for this scope is the estimation developed within the framework of WP5, according to which the marginal variable cost for soft wheat in Hungary is assumed equal to 64 €/t. According to this estimation, it seems to reject the capacity of the model to correctly assess the cost for this kind of crop. However, if we compare the estimation for soft wheat obtained within the framework of WP5 with the estimation obtained for Italy and Belgium, we can note that the estimations are very similar. For Italy, the PMP method estimates a marginal accounting cost of 66 €/t., while the WP5 method 65 €/t; for Belgium, the marginal accounting cost for soft wheat according to the PMP model is equal to 58 €/t., while for the WP5 model 52 €/t. In this situation, it is very difficult to affirm that the estimation for Hungary is wrong: it requires a deeper investigation using the registered specific accounting costs at national level for the same farm type.

Table 5.3: Specific cost estimates obtained from PMP model - Hungary

CROP	No. of observations	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
Soft wheat	679	**0.30492**	0.01383	**0.63132**	0.02145	**0.93624**	0.02195
Durum wheat	8	**0.27556**	0.11053	**0.18593**	0.02772	**0.46148**	0.02928
Barley	371	**0.43864**	0.02903	**0.80392**	0.05134	**1.24256**	0.05137
Oats	95	**0.42917**	0.04391	**0.21249**	0.03163	**0.64166**	0.04748
Maize	656	**0.35739**	0.01650	**0.90416**	0.05715	**1.26154**	0.05821
Other cereals	87	**0.42766**	0.04205	**0.23380**	0.02888	**0.66146**	0.04003
Sugarbeet	25	**0.05900**	0.01472	**0.05883**	0.01063	**0.11783**	0.01139
Rape	321	**0.59215**	0.03683	**0.71090**	0.04066	**1.30305**	0.04898
Sunflower	493	**0.78590**	0.03041	**1.17138**	0.04190	**1.95728**	0.04702
Soya	36	**0.62591**	0.12765	**0.13972**	0.02483	**0.76563**	0.03673
Fodder maize	22	**0.04104**	0.00805	**0.04588**	0.01132	**0.08692**	0.01175
Fodder products	114	**0.15869**	0.02164	**0.20246**	0.02587	**0.36115**	0.02931
Pasture	72	**0.15268**	0.03439	**0.03602**	0.00896	**0.18870**	0.01283
Rough grazing	43	**0.13134**	0.03218	**0.05293**	0.00647	**0.18427**	0.00697

Fig. 5.5: Total marginal cost distribution - Hungary - (€/t)

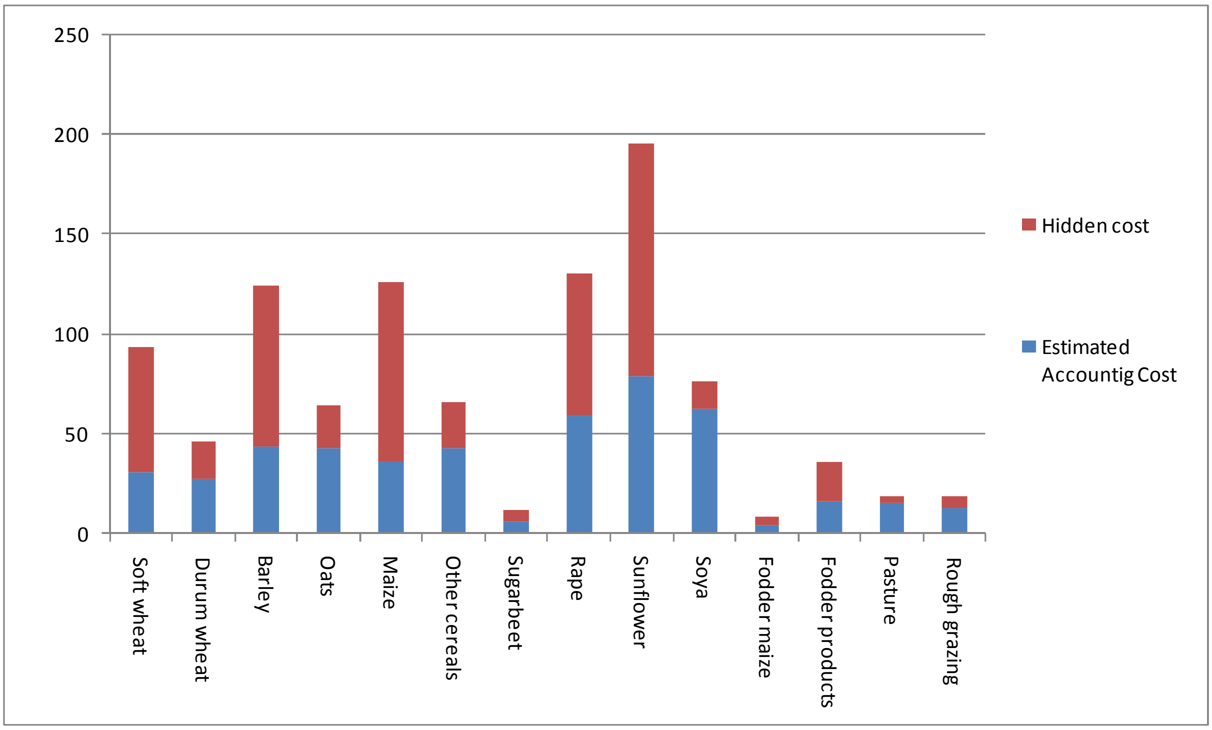


Table. 5.4: Comparison between accounting cost estimates in Italian and Belgian samples -€/t

Crops	Italian FADN sample	Belgian FADN sample	Hungarian FADN sample
Soft wheat	66	58	31
Barley	51	53	44
Maize	74	51	36
Rape	112	-	59
Sunflower	111	-	79

5.3 Estimation on cluster analysis results

The Hungarian sample has been submitted to a cluster analysis in order to group farms on the basis of characteristics of similarity identified in the productive and economic information provided by the FADN. The Hungarian sample has been tested for an increasing number of groups starting from 2 until reaching 10 groups. The results are shown in Table 5.5, where the distribution of farms among the groups is associated with the corresponding correlation index. According to the statistical tests developed for the cluster analysis, the 10 groups option seemed the best choice, even though the correlation index is equal to 50%. Among the 10 groups of farms the sixth group, which is the most numerous, has been submitted to the PMP estimation.

Table 5.5: Cluster analysis results for Hungary

N. clusters	Clusters										Global R2
	1	2	3	4	5	6	7	8	9	10	
2	769	19									0.06
3	1	786	1								0.14
4	1	1	784	2							0.19
5	1	778	2	1	6						0.25
6	1	443	2	335	6	1					0.30
7	1	5	125	2	637	1	17				0.35
8	11	615	2	1	31	5	1	122			0.41
9	11	5	1	476	5	286	1	2	1		0.44
10	**8**	**1**	**2**	**5**	**7**	**702**	**21**	**34**	**7**	**1**	**0.50**

As highlighted above, the estimates for Hungary cannot be put in relation with the real specific accounting costs. In this situation, Table 5.6 can be read on the basis of the considerations already given for the entire sample. The accounting cost estimation for the cluster reveals a very low modification with respect to the estimation carried out on the basic information. Comparing the results with the information provided by the PMP for the other two case studies, the estimates are much lower, perhaps indicating a less intensive use of variable inputs for the farm processes. However, this type of consideration should be further investigated in order to understand the statistical goodness of the outcomes and the relation with the farming system.

Table 5.6: Specific cost estimates obtained from PMP model - Hungary, 10 groups, 10 cluster

Crop	No. of observations	Estimated Accounting Cost	Std. Error	Hidden cost	Std. Error	Total Marginal Cost	Std. Error
Soft wheat	604	**0.32857**	0.01569	**0.60432**	0.02225	**0.93289**	0.02339
Durum wheat	8	**0.29425**	0.13266	**0.21409**	0.05071	**0.50834**	0.05426
Barley	333	**0.42934**	0.03013	**0.73954**	0.05137	**1.16888**	0.05277
Oats	84	**0.53430**	0.05756	**0.18925**	0.03122	**0.72355**	0.05429
Maize	581	**0.37097**	0.01751	**0.89794**	0.06366	**1.26892**	0.06434
Other cereals	79	**0.39653**	0.03942	**0.21570**	0.02710	**0.61223**	0.03629
Sugarbeet	25	**0.06411**	0.01526	**0.06308**	0.01150	**0.12719**	0.01232
Rape	288	**0.61140**	0.04010	**0.60049**	0.04106	**1.21189**	0.05214
Sunflower	443	**0.80307**	0.03315	**1.17744**	0.04520	**1.98051**	0.04995
Soya	34	**0.61474**	0.11788	**0.13614**	0.02174	**0.75088**	0.04103
Fodder maize	20	**0.04038**	0.00730	**0.03995**	0.00872	**0.08034**	0.00882
Fodder products	97	**0.17731**	0.02585	**0.18300**	0.02553	**0.36031**	0.03026
Pasture	64	**0.18172**	0.02510	**0.04689**	0.00894	**0.22862**	0.01508
Rough grazing	38	**0.13371**	0.07079	**0.04494**	0.00580	**0.17865**	0.00626

Fig. 5.6: Total marginal cost distribution - Hungary, 10 groups, 10 cluster (€/t)

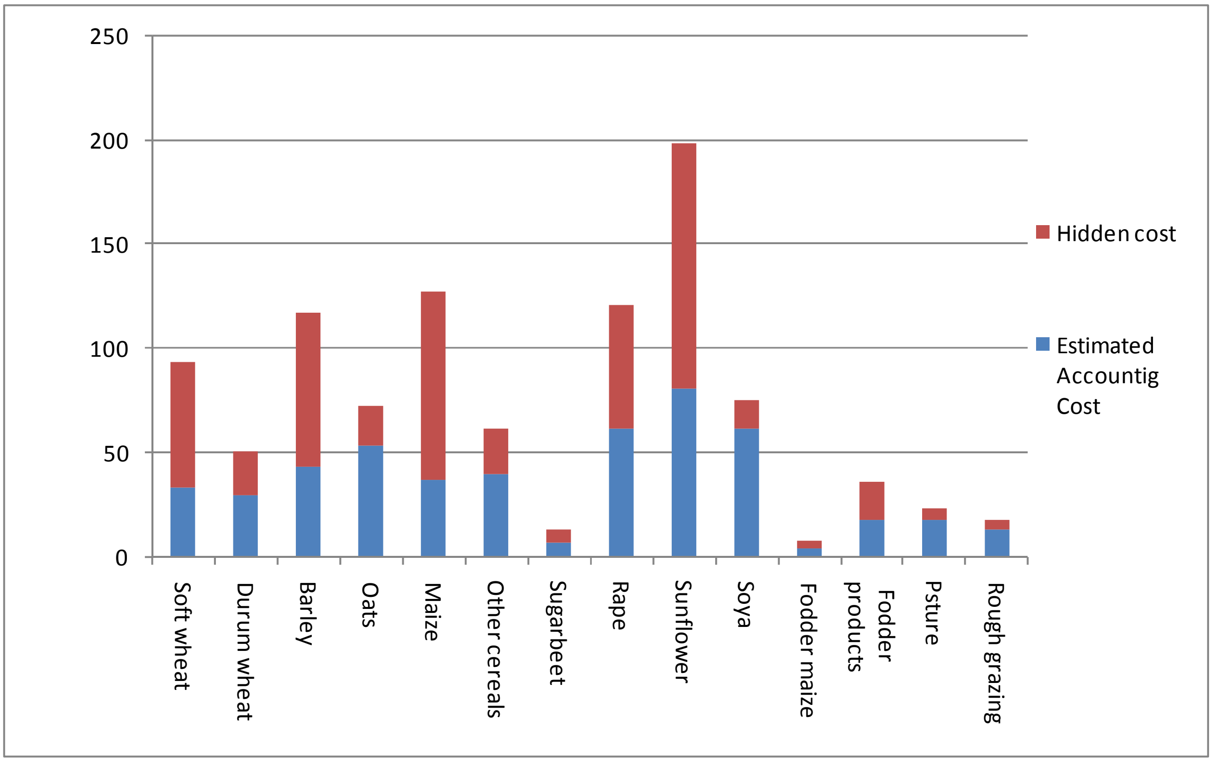


Table 5.7: Comparison between accounting cost estimates on Italian, Belgian and Hungarian samples with cluster analysis (CA) - €/t

Crops	Italian FADN sample	Italian FADN sample with CA	Belgian FADN sample	Belgian FADN sample with CA	Hungarian FADN sample	Hungarian FADN sample with CA
Soft wheat	66	65	58	62	31	33
Barley	51	57	53	70	44	43
Maize	74	71	51	66	36	37
Rape	112	112	-	-	59	61
Sunflower	111	72	-	-	79	80

6 Final Remarks

The discussion about the results developed in the previous sections has highlighted the capacity of the PMP model to recoer with a good degree of approximation the observed accounting costs for cereals, except for maize and some other crops that present information about price and yield with a low level of variability, i.e. sugarbeet. The number of available observations is important but not crucial for obtaining significant accounting cost estimates; an example of this is maize that, despite a high number of observations, has provided bad statistical test results for all the investigated samples submitted to the validation procedure. In this specific case, an important role has been played by the variability in yield mainly due to irrigation practices, for which the related costs are not considered within the observed specific accounting costs used as reference term for the validation phase. The lack of this information in the observed accounting costs has strongly contributed to the null hypothesis being rejected on the significance of the equality of estimated and observed means. Attempts to make the farm information more homogeneous through territorial stratification and cluster analysis have not improved the estimation for this crop. This example illustrates how the estimate validation may be influenced by lacks in the observed information that can produce distortions in statistical tests.

The results obtained also demonstrate a strong influence of two other factors in the estimation procedure:

- the presence of outliers: the out-of-range values have without doubt an important effect on the estimation and a preventive check is fundamental for minimizing the interference from this kind of component within the estimation procedure.
- the level of internal sample homogeneity: the obtained estimates are much more significant the more homogeneous the sample is. This is evident for the three Italian regions, where territorial stratification and clustering have produced a marked improvement in the estimate significance.

An attempt has been made to mitigate the problems encountered during the estimation procedure through cluster analysis, which has created homogenous groups of farms using their production (yields) and economic (prices) characteristics as variables. For Italy, where the presence of observed accounting costs permitted the validation, there has been a further grouping according the territorial area of each farm, with a notable improvement of the estimation fitness.

To minimize the information variability interference within the PMP estimation it is crucial to form farm samples where the information about prices and yields are more similar, otherwise there is a risk of obtaining unreliable accounting cost estimates. To make the sample to be estimated homogenous it is important first to consider the sector and territory relating to the farms and then try to eliminate the residual outliers and reduce information variability through a multivariate technique.

Table 6.1: Estimated and observed specific accounting costs for soft wheat per class of size - €/t.

Region	Class of size (ha)	Specific Accounting Cost			N. of obs.
		Estimated	Observed	Var. %	
Veneto	<10	71.44	72.63	*-1.6*	23
	10-20	62.53	77.13	*-18.9*	14
	20-50	67.81	71.55	*-5.2*	36
	50-100	82.01	67.48	*21.5*	20
	100-200	68.04	63.85	*6.6*	13
	>200	48.67	53.95	*-9.8*	4
	Total	**68.53**	**70.20**	***-2.4***	**110**
Lombardy	<10	28.94	31.17	*-7.2*	6
	10-20	45.55	61.71	*-26.2*	10
	20-50	79.03	48.34	*63.5*	15
	50-100	35.58	45.89	*-22.5*	8
	100-200	48.35	42.30	*14.3*	4
	>200	157.30	54.30	*189.7*	3
	Total	**55.46**	**48.45**	***14.5***	**46**
Piedmont	<10	68.04	63.07	*7.9*	50
	10-20	67.85	80.31	*-15.5*	39
	20-50	69.62	80.08	*-13.1*	50
	50-100	74.81	78.25	*-4.4*	22
	100-200	81.11	81.38	*-0.3*	11
	>200	207.00	101.72	*103.5*	6
	Total	**72.30**	**75.94**	***-4.8***	**178**
Veneto-Lombardy-Piedmont	<10	65.55	62.81	*4.4*	80
	10-20	58.30	76.84	*-24.1*	65
	20-50	63.18	71.72	*-11.9*	99
	50-100	64.26	69.50	*-7.5*	49
	100-200	74.45	68.07	*9.4*	29
	>200	146.76	77.20	*90.1*	13
	Total	**66.02**	**70.16**	***-5.9***	**335**

The presentation of case study results has been limited to the specific marginal costs at sample level, without considering a deeper level of information. As described in the first chapter, the developed PMP model is a micro-based model that uses individual farm information in such a way that the results are also obtained at individual level. The results can thus be aggregated in different ways according to the research objectives. The model can provide the specific marginal cost estimates for each crop from the farm level to a more aggregated level, like specific territorial area (e.g. altitude), economic size, physical size, and so on. Table 6.1 shows the results on specific accounting costs for soft wheat obtained for the three Italian regions aggregated according the physical size class. Moving from an aggregated result to a less aggregated one, the estimation variability increases. In particular, observing the results for the three regions considered as a whole, it is evident how the stratification leads to an amplification of the estimation errors for some size classes; for instance, the biggest class presents a very high divergence of the estimated specific accounting cost with respect to the observed one, while most of the other classes show differences with respect to the observed data higher than the average value calculated for the entire sample. This estimation behaviour is repeated for the three regions considered separately; the worst results generally correspond to the size classes where the number of observations is low, indicating that the estimation procedure tends to centre the specific cost estimation on the average information.

The PMP methodology properties allow, with the accounting costs, to recover the part of information that is hidden inside the production level and that each farmer has taken into account in the land allocation process. This kind of cost is important not only for the total marginal cost reconstruction but also for the calibration. The cost estimates obtained can be used to reproduce the basic production situation of each farm and the PMP methodology guarantees this result. This new information can be used in a model for evaluating the reaction of each farm included in the sample facing alternative policy and market scenarios. The farm response can be evaluated in terms of land allocation, variation in GSP, total variable costs and gross margin. This last information will be provided at individual and aggregated levels.

References

Arfini F., Donati M. (2008), Health check and farm efficiency: a comparative assessment of four European agricultural regions, contributed paper at 109th EAAE Seminar, November 20-21, 2008, Viterbo, Italy.

Arfini F. Donati M. and Paris Q. (2003), A national PMP model for policy evaluation in agriculture using micro data and administrative information, contributed paper to international conference: Agricultural policy reform and the WTO: Where are we heading?, Capri, June 2003.

Arfini F., Donati M. and Zuppiroli M. (2005), Agrisp: un modello di simulazione regionale per valutare gli effetti per l'Italia di modifiche delle politiche agricoli, in Anania G. (Ed.), La riforma delle politiche agricoli dell'UE ed il negoziato WTO, Franco Angeli Editore, Milano.

Eurostat (2009), http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/ data/database .

Heckelei T. and Britz W. (2005), Model based on positive mathematical programming: state of the art and further extension, in Arfini F. (Ed.), Modelling agricultural policies: state of the art and new challenges, MUP, Parma, pp. 48-74.

Paris Q. and Arfini F. (2000), Frontier Cost Functions, Self-Selection, Price Risk, PMP and Agenda 2000, Programme CT97-3403 "Eurotools", *Working Papers Series*, n.20.

Paris Q. and Howitt R.E. (1998), An analysis of ill-posed production problems using maximum entropy, *American Journal of Agricultural Economics*, n.80, pp. 124-138.

ANNEX 1 - Principal Component Analysis

Table A1.1: PCA Belgium (Yields)

Table A1.1.1: Communalities

	Initial	Extraction
Zscore(Y_WINTER_WHEAT)	1.000	0.520
Zscore(Y_BARLEY)	1.000	0.634
Zscore(Y_OATS)	1.000	0.599
Zscore(Y_MAIZE)	1.000	0.631
Zscore(Y_CEREALS)	1.000	0.645
Zscore(Y_POTATO)	1.000	0.658
Zscore(Y_SUGARBEET)	1.000	0.698
Zscore(Y_INDUSTRIAL)	1.000	0.445
Zscore(Y_VEGETABLES)	1.000	0.763
Zscore(Y_VEGETABLE2)	1.000	0.658

Extraction Method: Principal Component Analysis.

Table A1.1.2: Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.919	19.193	19.193	1.919	19.193	19.193
2	1.856	18.565	37.758	1.856	18.565	37.758
3	1.423	14.230	51.988	1.423	14.230	51.988
4	1.054	10.537	62.525	1.054	10.537	62.525
5	0.878	8.784	71.309			
6	0.795	7.952	79.261			
7	0.687	6.868	86.129			
8	0.535	5.351	91.480			
9	0.506	5.057	96.537			
10	0.346	3.463	100.000			

Figure A1.1: Scree plot of eigenvalues

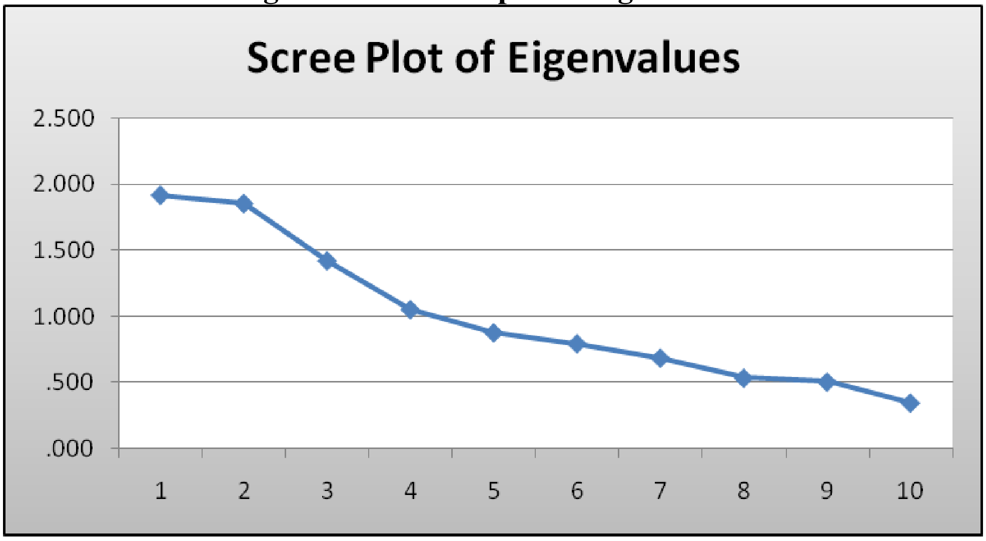


Table A1.1.3: Component Matrix[a]

	Component			
	1	2	3	4
Zscore(Y_WINTER_WHEAT)	0.520	0.498	0.020	-0.041
Zscore(Y_BARLEY)	0.658	-0.151	0.420	0.038
Zscore(Y_OATS)	0.226	-0.492	0.149	0.533
Zscore(Y_MAIZE)	-0.444	0.304	0.583	-0.032
Zscore(Y_CEREALS)	-0.198	-0.340	0.319	0.623
Zscore(Y_POTATO)	-0.321	0.674	0.166	0.269
Zscore(Y_SUGARBEET)	0.585	0.542	-0.153	0.196
Zscore(Y_INDUSTRIAL)	0.603	0.194	0.139	0.155
Zscore(Y_VEGETABLES)	-0.327	0.511	-0.421	0.466
Zscore(Y_VEGETABLE2)	-0.106	0.297	0.731	-0.157

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Table A1.2: PCA Belgium (Prices)

Table A1.2.1: Communalities

	Initial	Extraction
Zscore(P_WINTER_WHEAT)	1.000	0.796
Zscore(P_BARLEY)	1.000	0.587
Zscore(P_OATS)	1.000	0.646
Zscore(P_MAIZE)	1.000	0.658
Zscore(P_CEREALS)	1.000	0.480
Zscore(P_POTATO)	1.000	0.583
Zscore(P_SUGARBEET)	1.000	0.701
Zscore(P_INDUSTRIAL)	1.000	0.387
Zscore(P_VEGETABLES)	1.000	0.397
Zscore(P_VEGETABLE2)	1.000	0.690

Extraction Method: Principal Component Analysis.

Table A1.2.2: Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.819	18.192	18.192	1.819	18.192	18.192
2	1.622	16.225	34.417	1.622	16.225	34.417
3	1.360	13.602	48.020	1.360	13.602	48.020
4	1.124	11.237	59.257	1.124	11.237	59.257
5	0.976	9.756	69.013			
6	0.823	8.229	77.242			
7	0.725	7.251	84.493			
8	0.576	5.756	90.249			
9	0.517	5.165	95.414			
10	0.459	4.586	100.000			

Figure A1.2: Scree plot of eigenvalues

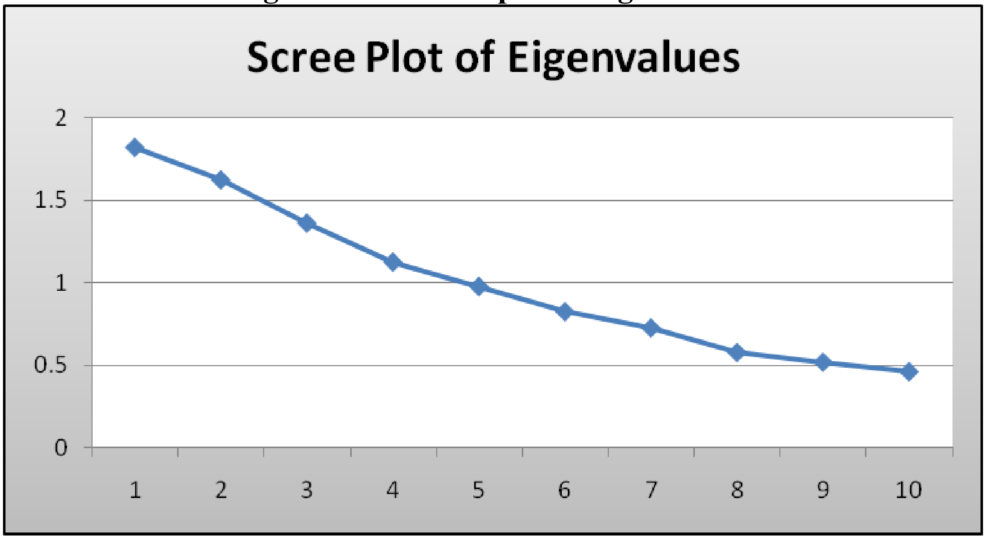


Table A1.2.3: Component Matrix[a]

	Component			
	1	2	3	4
Zscore(P_WINTER_WHEAT)	-0.287	-0.127	0.005	0.835
Zscore(P_BARLEY)	-0.551	0.350	0.400	0.029
Zscore(P_OATS)	-0.515	0.545	-0.209	0.200
Zscore(P_MAIZE)	0.588	0.229	0.487	0.149
Zscore(P_CEREALS)	-0.110	0.609	-0.218	0.223
Zscore(P_POTATO)	0.677	0.161	-0.055	0.308
Zscore(P_SUGARBEET)	-0.314	-0.646	0.344	0.260
Zscore(P_INDUSTRIAL)	0.456	0.227	-0.261	0.242
Zscore(P_VEGETABLES)	0.159	-0.473	-0.253	0.291
Zscore(P_VEGETABLE2)	0.137	0.212	0.787	0.084

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Table A1.3: PCA Italy (Yields)

Table A1.3.1: Communalities

	Initial	Extraction
Zscore(Y_DURUM_WHEAT)	1.000	0.566
Zscore(Y_WINTER_WHEAT)	1.000	0.576
Zscore(Y_MAIZE)	1.000	0.649
Zscore(Y_BARLEY)	1.000	0.545
Zscore(Y_RICE)	1.000	0.615
Zscore(Y_SORGHUM)	1.000	0.423
Zscore(Y_SOYA)	1.000	0.487
Zscore(Y_SUGARBEET)	1.000	0.578
Zscore(Y_POTATO)	1.000	0.839
Zscore(Y_TOMATO)	1.000	0.431
Zscore(Y_FODDER_MAIZE)	1.000	0.834
Zscore(Y_ALFALFA)	1.000	0.512
Zscore(Y_TEMPORARY_GRASS)	1.000	0.722
Zscore(Y_MEADOWS)	1.000	0.656

Extraction Method: Principal Component Analysis.

Table A1.3.2: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.588	11.345	11.345	1.588	11.345	11.345
2	1.463	10.447	21.791	1.463	10.447	21.791
3	1.205	8.610	30.401	1.205	8.610	30.401
4	1.082	7.729	38.131	1.082	7.729	38.131
5	1.059	7.566	45.697	1.059	7.566	45.697
6	1.021	7.290	52.987	1.021	7.290	52.987
7	1.014	7.245	60.232	1.014	7.245	60.232
8	0.923	6.593	66.825			
9	0.894	6.383	73.208			
10	0.852	6.089	79.297			
11	0.794	5.673	84.969			
12	0.751	5.367	90.336			
13	0.693	4.949	95.286			
14	0.660	4.714	100.000			

Figure A1.3: Scree plot of eigenvalues

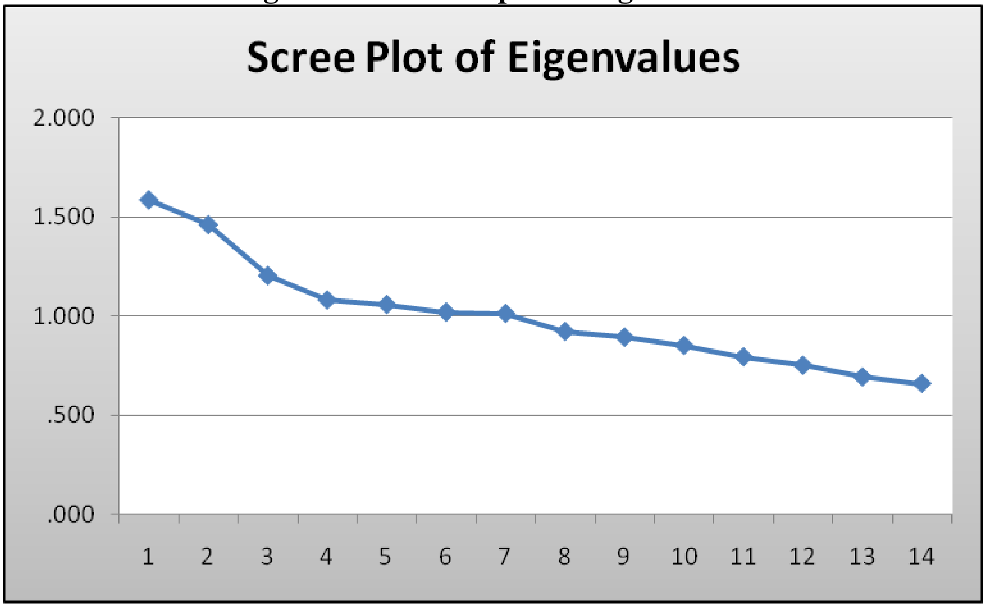


Table A1.3.2: Component Matrix[a]

	Component						
	1	2	3	4	5	6	7
Zscore(R_DURUM_WHEAT)	0.227	0.620	-0.262	0.138	0.094	0.077	-0.165
Zscore(R_WINTER_WHEAT)	0.366	0.110	0.609	0.129	0.148	0.073	0.121
Zscore(R_MAIZE)	-0.002	-0.319	-0.433	0.388	0.095	0.035	0.446
Zscore(R_BARLEY)	-0.285	0.473	0.088	-0.265	0.396	0.071	-0.012
Zscore(R_RICE)	-0.007	-0.204	0.261	0.653	0.183	0.166	0.130
Zscore(R_SORGHUM)	0.070	0.440	0.020	0.407	-0.050	-0.230	0.051
Zscore(R_SOYA)	0.682	-0.098	0.060	-0.065	0.008	0.049	-0.051
Zscore(R_SUGARBEET)	0.697	-0.138	-0.064	-0.089	0.231	-0.087	0.030
Zscore(R_POTATO)	-0.013	0.165	0.317	-0.350	-0.006	0.194	0.742
Zscore(R_TOMATO)	0.454	-0.035	-0.014	-0.083	0.084	0.379	-0.257
Zscore(R_FODDER_MAIZE)	0.137	-0.133	-0.084	-0.152	0.448	-0.743	0.121
Zscore(R_ALFALFA)	0.121	0.638	-0.033	0.220	-0.067	-0.146	0.120
Zscore(R_TEMPORARY_GRASS)	-0.112	0.047	-0.418	-0.023	0.608	0.395	0.081
Zscore(R_MEADOWS)	-0.338	-0.142	0.459	0.137	0.437	-0.045	-0.313

Extraction Method: Principal Component Analysis.

a. 7 components extracted.

Table A1.4: PCA Italy(Prices)

Table A1.4.1: Communalities

	Initial	Extraction
Zscore(P_DURUM_WHEAT)	1.000	0.489
Zscore(P_WINTER_WHEAT)	1.000	0.581
Zscore(P_MAIZE)	1.000	0.695
Zscore(P_BARLEY)	1.000	0.593
Zscore(P_RICE)	1.000	0.148
Zscore(P_SORGHUM)	1.000	0.435
Zscore(P_SOYA)	1.000	0.448
Zscore(P_SUGARBEET)	1.000	0.564
Zscore(P_POTATO)	1.000	0.703
Zscore(P_TOMATO)	1.000	0.448
Zscore(P_FODDER_MAIZE)	1.000	0.755
Zscore(P_ALFALFA)	1.000	0.539
Zscore(P_MEADOWS)	1.000	0.587
Zscore(P_TEMPORARY_GRASS)	1.000	0.525

Extraction Method: Principal Component Analysis.

Table A1.4.2: Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.668	11.912	11.912	1.668	11.912	11.912
2	1.394	9.957	21.869	1.394	9.957	21.869
3	1.169	8.347	30.216	1.169	8.347	30.216
4	1.130	8.071	38.287	1.130	8.071	38.287
5	1.082	7.728	46.015	1.082	7.728	46.015
6	1.067	7.625	53.639	1.067	7.625	53.639
7	0.991	7.079	60.719			
8	0.942	6.726	67.444			
9	0.893	6.377	73.821			
10	0.873	6.233	80.054			
11	0.802	5.726	85.780			
12	0.710	5.069	90.849			
13	0.650	4.644	95.493			
14	0.631	4.507	100.000			

Figure A1.4: Scree plot of eigenvalues

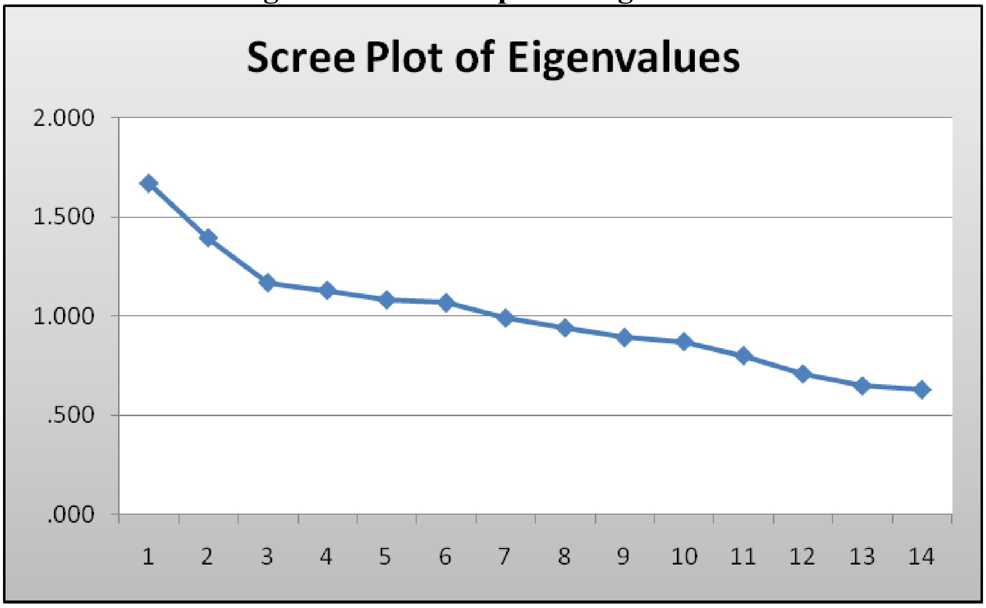


Table A1.4.2: Component Matrix[a]

	Component					
	1	2	3	4	5	6
Zscore(P_DURUM_WHEAT)	0.248	0.595	-0.202	0.173	0.031	-0.049
Zscore(P_WINTER_WHEAT)	0.334	-0.005	0.653	-0.140	-0.037	-0.146
Zscore(P_MAIZE)	0.345	-0.305	-0.175	-0.071	0.481	0.464
Zscore(P_BARLEY)	-0.285	0.536	0.307	0.147	0.194	-0.267
Zscore(P_RICE)	0.053	-0.212	0.095	-0.195	-0.136	0.186
Zscore(P_SORGHUM)	-0.009	0.392	-0.157	-0.258	-0.244	0.361
Zscore(P_SOYA)	0.626	0.022	0.195	-0.047	-0.119	0.031
Zscore(P_SUGARBEET)	0.712	0.010	0.032	0.236	-0.005	0.004
Zscore(P_POTATO)	0.033	0.131	0.416	-0.371	0.611	0.029
Zscore(P_TOMATO)	0.506	0.147	-0.034	0.383	-0.107	-0.106
Zscore(P_FODDER_MAIZE)	-0.176	0.087	0.201	0.516	0.138	0.625
Zscore(P_ALFALFA)	-0.030	0.609	0.039	-0.207	-0.080	0.341
Zscore(P_MEADOWS)	-0.319	-0.196	0.437	0.472	-0.142	0.118
Zscore(P_TEMPORARY_GRASS)	-0.023	0.072	-0.311	0.276	0.537	-0.240

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

Table A1.5: PCA Hungary (Yields)

Table A1.5.1: Communalities

	Initial	Extraction
Zscore(Y_WINTER_WHEAT)	1.000	0.594
Zscore(Y_DURUM_WHEAT)	1.000	0.670
Zscore(Y_RYE)	1.000	0.766
Zscore(Y_BARLEY)	1.000	0.453
Zscore(Y_OATS)	1.000	0.500
Zscore(Y_MAIZE)	1.000	0.580
Zscore(Y_RICE)	1.000	0.760
Zscore(Y_CEREALS)	1.000	0.521
Zscore(Y_POTATO)	1.000	0.360
Zscore(Y_SUGARBEET)	1.000	0.415
Zscore(Y_RAPE)	1.000	0.529
Zscore(Y_SUNFLOWER)	1.000	0.506
Zscore(Y_SOYA)	1.000	0.549
Zscore(Y_OTHER_OILSEEDS)	1.000	0.706
Zscore(Y_VEGETABALE)	1.000	0.591
Zscore(Y_VEGETABLE2)	1.000	0.526
Zscore(Y_GRASSLAND)	1.000	0.611
Zscore(Y_FODDER_MAIZE)	1.000	0.549
Zscore(Y_OTHER_SILAGE)	1.000	0.605
Zscore(Y_FODDER_PRODUCTS)	1.000	0.479
Zscore(Y_TEMPORARY_GRASS)	1.000	0.814
Zscore(Y_PASTURE)	1.000	0.545
Zscore(Y_ROUGH_GRAZING)	1.000	0.757

Extraction Method: Principal Component Analysis.

Table A1.5.2: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.758	7.642	7.642	1.758	7.642	7.642
2	1.598	6.946	14.588	1.598	6.946	14.588
3	1.290	5.611	20.199	1.290	5.611	20.199
4	1.245	5.413	25.612	1.245	5.413	25.612
5	1.190	5.174	30.786	1.190	5.174	30.786
6	1.113	4.837	35.623	1.113	4.837	35.623
7	1.090	4.739	40.363	1.090	4.739	40.363
8	1.057	4.596	44.958	1.057	4.596	44.958
9	1.026	4.463	49.421	1.026	4.463	49.421
10	1.020	4.434	53.856	1.020	4.434	53.856
11	1.002	4.359	58.214	1.002	4.359	58.214
12	0.983	4.275	62.489			
13	0.963	4.189	66.678			
14	0.928	4.034	70.712			
15	0.859	3.734	74.446			
16	0.849	3.690	78.135			
17	0.822	3.575	81.711			
18	0.779	3.385	85.096			
19	0.766	3.329	88.425			
20	0.740	3.216	91.641			
21	0.702	3.054	94.695			
22	0.654	2.843	97.538			
23	0.566	2.462	100.000			

Figure A1.5: Scree plot of eigenvalues

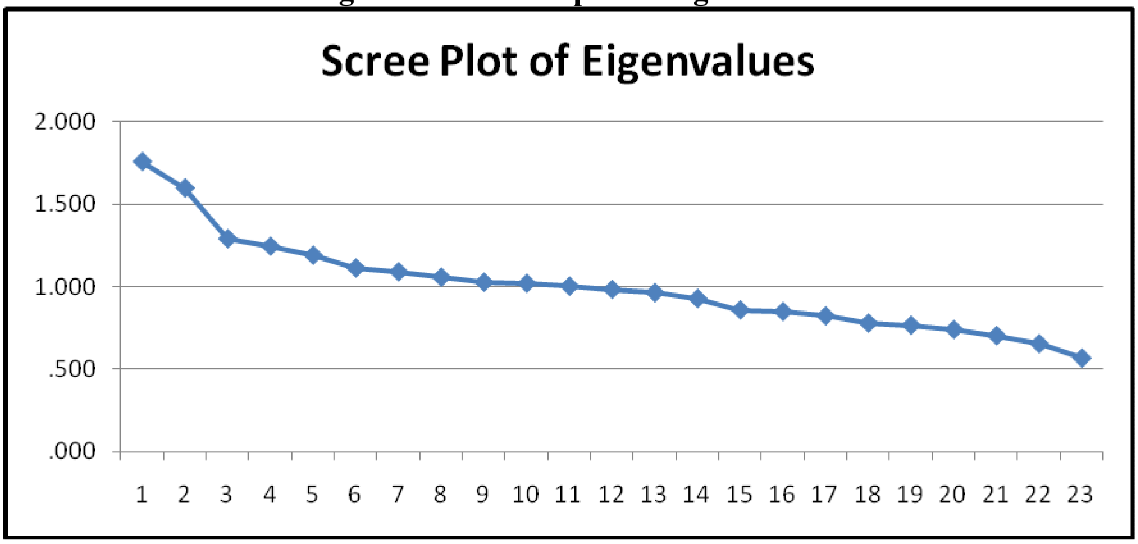


Table A1.5.2: Component Matrix[a]

	Component										
	1	2	3	4	5	6	7	8	9	10	11
Zscore(Y_WINTER_WHEAT)	0.704	0.052	0.033	0.059	0.164	0.018	0.164	-0.167	0.086	0.052	0.000
Zscore(Y_DURUM_WHEAT)	0.027	0.001	-0.410	0.372	0.207	0.215	-0.082	0.384	-0.270	0.017	-0.215
Zscore(Y_RYE)	-0.111	0.162	-0.345	0.275	-0.291	-0.168	0.087	-0.434	0.146	0.286	0.349
Zscore(Y_BARLEY)	0.451	0.309	-0.110	-0.203	-0.130	-0.173	-0.003	0.117	-0.161	-0.108	-0.049
Zscore(Y_OATS)	-0.116	0.652	0.157	0.034	-0.028	-0.091	0.065	0.070	0.019	-0.034	-0.128
Zscore(Y_MAIZE)	0.408	0.055	0.322	0.356	0.283	0.067	-0.138	0.147	-0.045	0.195	0.124
Zscore(Y_RICE)	0.002	-0.050	-0.178	0.137	0.118	0.003	0.187	-0.221	0.579	-0.350	-0.389
Zscore(Y_CEREALS)	-0.052	0.561	-0.279	-0.053	-0.044	0.061	-0.240	-0.043	-0.131	0.193	-0.049
Zscore(Y_POTATO)	0.044	0.141	0.211	0.076	0.270	-0.291	-0.133	-0.147	0.158	-0.109	-0.233
Zscore(Y_SUGARBEET)	0.461	0.052	-0.183	-0.275	-0.152	-0.127	0.208	0.034	-0.046	-0.068	0.001
Zscore(Y_RAPE)	0.649	0.115	-0.247	-0.045	-0.131	-0.010	0.000	-0.084	0.071	0.011	0.042
Zscore(Y_SUNFLOWER)	0.275	-0.031	0.237	0.025	-0.117	0.516	0.235	-0.045	-0.121	-0.084	-0.122
Zscore(Y_SOYA)	0.218	-0.077	0.304	0.267	0.338	-0.367	-0.155	-0.080	0.016	0.200	0.112
Zscore(Y_OTHER_OILSEED)	0.118	0.003	0.113	0.015	0.029	0.543	-0.356	-0.193	0.214	-0.234	0.345
Zscore(Y_VEGETABLE)	0.024	0.030	0.204	0.454	-0.571	0.005	-0.061	0.003	0.064	0.053	0.073
Zscore(Y_VEGETABLES2)	-0.061	-0.036	0.005	-0.087	-0.044	-0.005	0.025	0.417	0.499	0.297	0.002
Zscore(Y_GRASSLAND)	-0.018	0.095	-0.508	0.497	0.251	0.119	0.107	0.052	0.067	-0.028	0.015
Zscore(Y_FODDER_MAIZE)	0.065	0.189	0.031	-0.212	0.036	0.144	0.149	0.366	0.389	0.349	0.108
Zscore(Y_OTHER_SILAGE)	0.048	0.200	0.295	0.374	-0.447	-0.020	0.051	0.212	0.038	-0.158	-0.249
Zscore(Y_FODDER_PRODUCTS)	-0.188	0.540	0.180	0.032	0.182	-0.069	0.229	-0.088	-0.080	-0.112	0.043
Zscore(Y_TEMPORARY_GRASS)	-0.123	0.188	0.051	0.071	0.165	-0.039	0.437	0.234	-0.008	-0.396	0.569
Zscore(Y_PASTURE)	-0.082	0.513	0.025	-0.174	0.103	0.234	-0.387	-0.094	0.132	-0.044	-0.027
Zscore(Y_ROUGH_GRAZING)	-0.149	0.082	0.170	-0.009	0.131	0.316	0.481	-0.318	-0.129	0.442	-0.195

Extraction Method: Principal Component Analysis.

a. 11 components extracted.

Table A1.6: PCA Hungary (Prices)

Table A1.6.1: Communalities

	Initial	Extraction
Zscore(P_WINTER_WHEAT)	1.000	0.551
Zscore(P_DURUM_WHEAT)	1.000	0.660
Zscore(P_RYE)	1.000	0.636
Zscore(P_BARLEY)	1.000	0.556
Zscore(P_OATS)	1.000	0.510
Zscore(P_MAIZE)	1.000	0.362
Zscore(P_RICE)	1.000	0.767
Zscore(P_OTHER_CEREALS)	1.000	0.426
Zscore(P_POTATO)	1.000	0.707
Zscore(P_SUGARBEET)	1.000	0.468
Zscore(P_RAPE)	1.000	0.583
Zscore(P_SUNFLOWER)	1.000	0.482
Zscore(P_SOYA)	1.000	0.368
Zscore(P_OTHER_OILSEEDS)	1.000	0.620
Zscore(P_VEGETABLE)	1.000	0.608
Zscore(P_VEGETABLE2)	1.000	0.853
Zscore(P_GRASSLAND)	1.000	0.786
Zscore(P_FODDER_MAIZE)	1.000	0.383
Zscore(P_OTHER_SILAGE)	1.000	0.645
Zscore(P_FODDER_PRODUCTS)	1.000	0.523
Zscore(P_TEMPORARY_GRASS)	1.000	0.541
Zscore(P_PASTURE)	1.000	0.544
Zscore(P_ROUGH_GRAZING)	1.000	0.646

Extraction Method: Principal Component Analysis.

Table A1.6.2: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.030	8.826	8.826	2.030	8.826	8.826
2	1.892	8.228	17.054	1.892	8.228	17.054
3	1.455	6.325	23.378	1.455	6.325	23.378
4	1.360	5.912	29.290	1.360	5.912	29.290
5	1.192	5.181	34.472	1.192	5.181	34.472
6	1.141	4.960	39.431	1.141	4.960	39.431
7	1.072	4.663	44.094	1.072	4.663	44.094
8	1.054	4.583	48.677	1.054	4.583	48.677
9	1.021	4.439	53.117	1.021	4.439	53.117
10	1.007	4.378	57.495	1.007	4.378	57.495
11	0.978	4.252	61.747			
12	0.956	4.157	65.904			
13	0.916	3.985	69.888			
14	0.890	3.869	73.757			
15	0.848	3.686	77.443			
16	0.831	3.614	81.057			
17	0.754	3.279	84.336			
18	0.732	3.183	87.519			
19	0.697	3.031	90.550			
20	0.672	2.921	93.471			
21	0.649	2.820	96.291			
22	0.557	2.423	98.714			
23	0.296	1.286	100.000			

Figure A1.6: Scree plot of eigenvalues

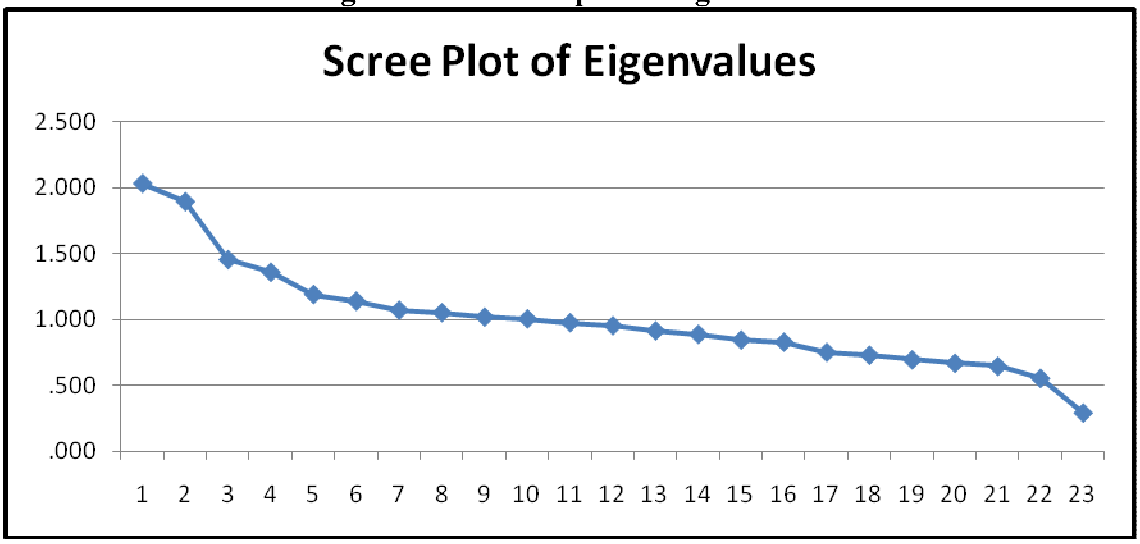


Table A1.6.2: Component Matrix[a]

	Component									
	1	2	3	4	5	6	7	8	9	10
Zscore(P_WINTER_WHEAT)	0.142	0.102	0.501	-0.327	0.220	-0.139	0.087	0.236	0.102	0.144
Zscore(P_DURUM_WHEAT)	0.035	0.122	0.000	-0.046	-0.033	0.237	0.737	0.166	-0.066	-0.098
Zscore(P_RYE)	0.138	-0.057	-0.028	-0.057	-0.068	0.757	-0.135	-2.450E-5	-0.077	0.088
Zscore(P_BARLEY)	0.418	-0.056	0.394	-0.039	-0.352	-0.082	-0.122	-0.204	-0.178	0.052
Zscore(P_OATS)	0.632	-0.237	-0.160	0.000	0.011	0.017	0.006	0.153	-0.041	0.062
Zscore(P_MAIZE)	0.219	0.396	0.003	0.234	0.105	-0.053	-0.151	-0.107	0.233	-0.017
Zscore(P_RICE)	0.187	0.830	-0.166	-0.095	0.043	0.027	-0.052	-0.043	-0.013	-0.014
Zscore(P_CEREALS)	0.474	-0.144	-0.137	-0.079	-0.302	0.077	0.099	-0.214	0.055	-0.005
Zscore(P_POTATO)	-0.037	0.002	-0.057	-0.006	-0.051	-0.096	-0.194	0.700	0.201	0.350
Zscore(P_SUGARBEET)	0.054	0.013	0.504	-0.169	-0.100	-0.033	-0.248	-0.168	-0.129	-0.254
Zscore(P_RAPE)	0.189	0.179	0.636	-0.160	-0.009	0.192	-0.068	0.176	0.112	0.001
Zscore(P_SUNFLOWER)	0.114	0.053	0.346	0.058	0.473	0.012	0.269	-0.160	-0.016	0.144
Zscore(P_SOYA)	0.034	0.382	-0.115	-0.024	-0.180	-0.267	-0.250	-0.035	-0.083	0.181
Zscore(P_OTHER_OILSEED)	-0.025	0.038	0.136	0.334	0.108	0.033	0.093	-0.277	0.607	0.146
Zscore(P_VEGETABLE)	0.104	0.046	0.129	0.745	0.066	0.053	-0.085	-0.042	-0.087	-0.012
Zscore(P_VEGETABLE2)	-0.030	-0.011	0.014	0.005	0.041	-0.038	0.119	-0.253	-0.377	0.793
Zscore(P_GRASSLAND)	0.247	0.812	-0.181	-0.094	-0.014	0.070	0.105	0.010	-0.078	-0.049
Zscore(P_FODDER_MAIZE)	0.372	-0.141	-0.113	-0.120	0.050	-0.220	0.006	-0.152	0.341	0.086
Zscore(P_OTHER_SILAGE)	0.230	0.046	0.146	0.604	0.118	-0.008	-0.039	0.293	-0.305	-0.097
Zscore(P_FODDER_PRODUCT)	0.629	-0.192	-0.137	-0.061	0.158	-0.106	-0.062	0.050	-0.142	-0.069
Zscore(P_TEMPORARY_GRASS)	0.242	-0.095	-0.163	-0.090	0.264	0.481	-0.307	-0.020	0.165	0.127
Zscore(P_PASTURE)	0.529	-0.101	-0.016	0.137	-0.320	-0.142	0.267	0.121	0.162	0.015
Zscore(P_ROUGH_GRAZING)	0.235	-0.131	-0.183	-0.183	0.640	-0.188	-0.053	-0.046	-0.163	-0.176

Extraction Method: Principal Component Analysis.

a. 10 components extracted.

ANNEX 2 - Cluster Analysis

Table A2.1: Belgium Cluster Analysis

Table A2.1.1: R^2

BELGIUM – R^2	
2 cluster	0.179081
3 cluster	0.344988
4 cluster	0.465605
5 cluster	0.494297
6 cluster	0.614953
7 cluster	0.604116
8 cluster	*0.704306*
9 cluster	0.708969
10 cluster	0.755578

Table A2.1.2: Initial Cluster Centres

	Cluster								
	1	2	3	4	5	6	7	8	9
REGR factor score 1 for analysis 1	1.5938	0.9331	1.0546	-2.1036	-1.8161	4.9017	-2.0166	0.6003	0.2608
REGR factor score 2 for analysis 1	1.3451	2.7237	-0.1922	1.3414	2.7518	2.8175	5.6106	1.4126	-1.9851
REGR factor score 3 for analysis 1	-0.5699	-0.3039	0.2324	-0.3564	-1.3014	-2.5176	-2.4876	5.2988	-1.2572
REGR factor score 4 for analysis 1	-4.3518	1.3019	0.5832	0.6532	0.5521	2.1534	1.8405	0.5875	0.9529
REGR factor score 1 for analysis 2	-2.5517	-1.6066	-1.7474	2.6464	-0.7712	-1.5523	-1.1980	-1.4138	0.5279
REGR factor score 2 for analysis 2	-1.9066	-1.5443	2.7101	-1.2562	-2.5425	-0.1825	-3.8173	2.0104	-0.0438
REGR factor score 3 for analysis 2	0.3616	2.7925	-1.4459	0.7816	0.2034	0.6380	1.6345	4.6129	-0.5404
REGR factor score 4 for analysis 2	-1.0639	4.6135	2.6369	2.7340	0.5179	-0.9040	4.4050	-0.9476	-0.2338

Table A2.1.3: Iteration History[a]

Iteration	Change in Cluster Centers								
	1	2	3	4	5	6	7	8	9
1	2.366	0.000	1.685	1.332	1.779	2.306	0.000	1.261	2.340
2	0.000	0.000	0.443	0.000	0.000	0.000	0.000	0.000	0.094
3	0.000	0.000	0.467	0.000	0.000	0.000	0.000	0.000	0.085
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A2.1.4: Final Cluster Centres

	Cluster								
	1	2	3	4	5	6	7	8	9
REGR factor score 1 for analysis 1	1.0625	0.9331	0.8576	-1.9473	-1.1617	3.3555	-2.0166	0.8302	-0.2024
REGR factor score 2 for analysis 1	0.9021	2.7237	-0.6811	1.1319	1.4922	2.0335	5.6106	1.2302	-0.3228
REGR factor score 3 for analysis 1	-0.4845	-0.3039	-0.2298	-0.2886	-0.9158	-1.5019	-2.4876	4.5359	0.0311
REGR factor score 4 for analysis 1	-2.6006	1.3019	0.9194	0.5476	0.2173	1.0660	1.8405	0.4775	-0.0330
REGR factor score 1 for analysis 2	-1.7078	-1.6066	-0.8742	1.7009	-0.6436	-1.8118	-1.1980	-0.5899	0.3579
REGR factor score 2 for analysis 2	-1.0644	-1.5443	1.4753	-1.3639	-1.7970	-0.1275	-3.8173	1.7173	0.0147
REGR factor score 3 for analysis 2	-0.1860	2.7925	-0.8345	0.4968	-0.1320	0.7484	1.6345	4.2340	-0.1236
REGR factor score 4 for analysis 2	-0.4934	4.6135	0.9838	1.8953	0.0659	-0.7749	4.4050	-0.9108	-0.2898

Table A2.1.5: ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
REGR factor score 1 for analysis 1	8.552	8	0.371	96	23.073	0.000
REGR factor score 2 for analysis 1	10.284	8	0.226	96	45.447	0.000
REGR factor score 3 for analysis 1	9.714	8	0.274	96	35.472	0.000
REGR factor score 4 for analysis 1	8.368	8	0.386	96	21.679	0.000
REGR factor score 1 for analysis 2	7.906	8	0.424	96	18.627	0.000
REGR factor score 2 for analysis 2	9.628	8	0.281	96	34.270	0.000
REGR factor score 3 for analysis 2	9.514	8	0.291	96	32.746	0.000
REGR factor score 4 for analysis 2	9.766	8	0.270	96	36.235	0.000

Table A2.1.6: Number of Cases in each Cluster

	Number of Cases in each Cluster									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Cluster 2	8	97								
Cluster 3	94	3	8							
Cluster 4	61	33	8	3						
Cluster 5	9	3	2	2	89					
Cluster 6	10	2	34	2	54	3				
Cluster 7	9	2	1	83	6	3	1			
Cluster 8	*9*	*1*	*14*	*2*	*69*	*6*	*1*	*3*		
Cluster 9	7	1	12	4	3	2	1	3	72	
Cluster 10	5	33	8	2	4	1	1	3	47	1

Table A2.2: Italy Cluster Analysis

Table A2.2.1: R^2

ITALY - R^2	
2 cluster	0.077707
3 cluster	0.174221
4 cluster	0.285238
5 cluster	0.357988
6 cluster	0.40856
7 cluster	0.491114
8 cluster	0.489014
9 cluster	0.608166
10 cluster	*0.569035*

Table A2.2.2: Initial Cluster Centres

	Cluster									
	1	2	3	4	5	6	7	8	9	10
REGR factor score 1 for analysis 1	6.445	0.309	-0.45	1.394	0.186	3.223	-0.63	-0.614	3.825	-1.81
REGR factor score 2 for analysis 1	1.022	1.196	-0.4	9.103	-1.96	-0.89	1.206	-1.033	-2.18	0.69
REGR factor score 3 for analysis 1	-0.91	-0.02	-0.79	-0.21	3.601	0.004	2.034	-0.471	-1.82	-5.13
REGR factor score 4 for analysis 1	-0.4	0.807	-0.1	5.7	6.359	-0.46	-3.16	-1.21	-2.36	-0.69
REGR factor score 5 for analysis 1	1.717	-1.06	-0.7	0.296	2.322	0.666	0.171	5.029	5.615	7.376
REGR factor score 6 for analysis 1	3.374	-1.07	-0.1	-2.79	1.668	-0.28	1.813	-6.591	-9.41	4.79
REGR factor score 7 for analysis 1	-2.69	-0.6	0.02	0.669	0.438	0.248	6.849	0.526	1.563	1.418
REGR factor score 1 for analysis 2	7.477	0.257	-1.4	-0.1	1.517	3.051	-0	-3.461	2.365	-0.8
REGR factor score 2 for analysis 2	3.765	4.03	1.736	7.744	-3.25	-0.43	1.806	1.152	-0.43	0.952
REGR factor score 3 for analysis 2	-0.88	-2.74	0.815	-1.38	2.009	0.746	5.499	4.373	-0.5	-3.04
REGR factor score 4 for analysis 2	5.368	-3.78	3.74	-3.6	-2.26	0.356	-4.2	10.35	0.504	2.554
REGR factor score 5 for analysis 2	-1.09	-3.68	1.269	-2.63	-1.26	-0.12	8.04	2.616	-0.04	5.707
REGR factor score 6 for analysis 2	-1.4	5.424	7.378	5.411	3.557	0.111	-0.27	11.01	0.412	-2.48

Table A2.2.3: Iteration History

Iteration	Change in Cluster Centres									
	1	2	3	4	5	6	7	8	9	10
1	1.967	5.610	0.000	4.366	3.747	4.749	2.696	0.000	2.011	3.211
2	3.074	1.660	0.000	0.000	0.000	0.112	0.000	0.000	0.000	1.045
3	0.000	1.191	0.000	0.000	0.000	0.131	0.000	0.000	0.000	0.000
4	0.000	0.292	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.000
5	0.000	0.076	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000
6	0.000	0.056	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A2.2.4: Final Cluster Centres

	Cluster									
	1	2	3	4	5	6	7	8	9	10
REGR factor score 1 for analysis 1	4.964	0.547	-0.45	1.102	-0.05	-0.11	-0.15	-0.614	2.576	-0.68
REGR factor score 2 for analysis 1	-0.31	1.927	-0.4	7.043	-1.75	-0.2	1.599	-1.033	-1.881	0.266
REGR factor score 3 for analysis 1	-0.19	-0.63	-0.79	0.225	2.248	0.061	2.927	-0.471	-1.275	-3.41
REGR factor score 4 for analysis 1	-0.85	0.52	-0.1	6.113	5.571	-0.11	-3.34	-1.21	-2.128	-0.23
REGR factor score 5 for analysis 1	0.962	-0.23	-0.7	-0.47	1.569	-0.13	-0.15	5.029	5.718	4.899
REGR factor score 6 for analysis 1	3.935	-0.22	-0.1	-3.31	1.41	-0.02	1.919	-6.591	-10.03	3.159
REGR factor score 7 for analysis 1	-2.7	-0.11	0.02	0.854	1.133	-0.08	7.64	0.526	1.806	0.667
REGR factor score 1 for analysis 2	5.027	0.253	-1.4	-0.53	0.419	-0.08	0.384	-3.461	1.262	-0.12
REGR factor score 2 for analysis 2	1.443	2.086	1.736	4.934	-1.68	-0.25	1.277	1.152	-0.469	0.457
REGR factor score 3 for analysis 2	-0.37	-0.38	0.815	-1.43	0.712	0.022	4.283	4.373	-0.187	-2.65
REGR factor score 4 for analysis 2	3.843	-0.32	3.74	-2.57	-1.62	-0	-3.85	10.35	0.028	2.328
REGR factor score 5 for analysis 2	-1.11	-0.26	1.269	-2.66	-1.05	-0.08	6.319	2.616	0.086	4.492
REGR factor score 6 for analysis 2	-1.09	0.514	7.378	3.329	1.422	-0.1	0.305	11.01	0.23	-1.96

Table A2.2.5: ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
REGR factor score 1 for analysis 1	14.853	9	0.715	438	20.763	0.000
REGR factor score 2 for analysis 1	32.656	9	0.350	438	93.428	0.000
REGR factor score 3 for analysis 1	17.243	9	0.666	438	25.880	0.000
REGR factor score 4 for analysis 1	37.122	9	0.258	438	144.009	0.000
REGR factor score 5 for analysis 1	29.140	9	0.422	438	69.090	0.000
REGR factor score 6 for analysis 1	46.335	9	0.068	438	676.825	0.000
REGR factor score 7 for analysis 1	31.600	9	0.371	438	85.122	0.000
REGR factor score 1 for analysis 2	13.911	9	0.735	438	18.935	0.000
REGR factor score 2 for analysis 2	30.574	9	0.392	438	77.935	0.000
REGR factor score 3 for analysis 2	16.523	9	0.681	438	24.262	0.000
REGR factor score 4 for analysis 2	33.887	9	0.324	438	104.515	0.000
REGR factor score 5 for analysis 2	35.549	9	0.290	438	122.541	0.000
REGR factor score 6 for analysis 2	28.014	9	0.445	438	62.962	0.000

Table A2.2.6: Number of Cases in each Cluster

	Number of Cases in each Cluster									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Cluster 2	3	445								
Cluster 3	3	42	403							
Cluster 4	37	3	4	404						
Cluster 5	379	35	2	28	4					
Cluster 6	2	4	377	31	2	32				
Cluster 7	4	2	377	2	6	32	25			
Cluster 8	4	3	2	1	6	6	4	422		
Cluster 9	3	6	6	1	2	139	28	4	259	
Cluster 10	*4*	*38*	*1*	*2*	*6*	*384*	*4*	*1*	*2*	*6*

Table A2.3: Hungary Cluster Analysis

Table A2.3.1: R^2

R^2	
2 cluster	0.064435
3 cluster	0.135428
4 cluster	0.192676
5 cluster	0.248481
6 cluster	0.300559
7 cluster	0.352136
8 cluster	0.406169
9 cluster	0.441386
10 cluster	*0.498562*

Table A2.3.2: Initial Cluster Centres

	Cluster									
	1	2	3	4	5	6	7	8	9	10
REGR factor score 1 for analysis 1	-0.28	-0.96	5.62	0.18	0.98	3.77	0.05	1.60	3.25	5.23
REGR factor score 2 for analysis 1	0.85	-0.26	0.17	0.08	-0.28	-1.24	0.07	-1.03	-1.33	23.26
REGR factor score 3 for analysis 1	0.41	0.29	3.62	1.33	0.07	-0.21	1.05	-1.32	-1.70	-4.65
REGR factor score 4 for analysis 1	-0.90	0.21	15.38	3.18	-0.19	-0.78	3.37	-1.30	-1.56	-2.66
REGR factor score 5 for analysis 1	-0.19	1.10	2.42	0.40	0.19	-2.33	0.35	4.39	3.85	1.19
REGR factor score 6 for analysis 1	3.34	-1.02	-0.23	0.24	-0.82	-1.47	2.98	-2.01	11.07	0.77
REGR factor score 7 for analysis 1	12.07	3.29	-0.72	2.19	-1.96	0.87	-0.78	0.22	-5.11	-1.45
REGR factor score 8 for analysis 1	2.63	-7.06	6.99	-2.41	7.11	-0.44	-0.38	-0.54	0.54	-1.21
REGR factor score 9 for analysis 1	-1.48	-10.58	-7.52	7.40	1.66	2.85	-0.93	0.59	1.75	-0.38
REGR factor score 10 for analysis 1	-1.42	21.99	-2.25	1.61	3.54	0.73	0.20	-0.37	2.53	-0.39
REGR factor score 1 for analysis 2	-0.08	-1.78	0.80	1.62	0.13	1.03	-1.21	-1.82	-0.86	0.05
REGR factor score 2 for analysis 2	-0.30	-1.27	4.02	0.81	3.91	3.08	1.06	0.23	0.69	-1.41
REGR factor score 3 for analysis 2	-3.01	0.04	6.30	1.58	3.93	-0.42	-3.66	2.51	-0.94	-4.98
REGR factor score 4 for analysis 2	1.69	-2.21	8.49	-0.38	1.09	-1.70	5.19	-0.54	1.37	3.85
REGR factor score 5 for analysis 2	0.95	-1.20	-10.56	0.68	4.27	0.40	-6.97	1.73	-0.66	3.30
REGR factor score 6 for analysis 2	1.10	-0.34	-0.46	10.53	-4.86	1.03	-2.19	5.86	-0.82	0.09
REGR factor score 7 for analysis 2	-0.51	0.47	0.72	-7.76	-0.94	0.05	1.06	8.65	1.70	5.26
REGR factor score 8 for analysis 2	3.03	11.08	4.22	-4.16	-2.63	3.45	-6.67	-5.65	-1.70	-6.20
REGR factor score 9 for analysis 2	-2.41	13.41	0.96	4.75	2.42	3.68	2.65	-2.13	0.61	16.23
REGR factor score 10 for analysis 2	-0.41	7.84	-3.14	-4.69	-2.69	4.19	4.44	7.66	0.28	-9.82
REGR factor score 11 for analysis 2	-2.03	0.00	-4.77	6.66	-4.68	1.01	5.80	-3.84	3.30	-10.89

Table A2.3.3: Iteration History

Iteration	Change in Cluster Centres									
	1	2	3	4	5	6	7	8	9	10
1	6.173	0.000	5.286	6.453	5.986	9.613	8.942	7.756	7.595	0.000
2	0.000	0.000	0.000	1.826	0.000	0.059	0.000	2.710	1.397	0.000
3	0.000	0.000	0.000	0.000	0.000	0.057	1.169	1.547	2.314	0.000
4	0.000	0.000	0.000	0.000	0.000	0.025	0.811	0.599	2.341	0.000
5	0.000	0.000	0.000	0.000	0.000	0.025	0.463	0.455	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.179	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.182	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.097	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.079	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A2.3.4: Final Cluster Centres

	Cluster									
	1	2	3	4	5	6	7	8	9	10
REGR factor score 1 for analysis 1	0.321	-0.955	4.406	-0.250	-0.334	-0.108	0.680	0.912	2.603	5.233
REGR factor score 2 for analysis 1	1.457	-0.256	1.092	0.491	0.010	-0.018	0.022	-0.625	-0.810	23.264
REGR factor score 3 for analysis 1	-0.024	0.286	2.776	1.786	-0.574	0.042	0.079	-0.730	-1.700	-4.652
REGR factor score 4 for analysis 1	-0.406	0.211	11.440	4.260	-0.054	-0.054	1.482	-0.748	-0.826	-2.665
REGR factor score 5 for analysis 1	-0.405	1.102	2.374	1.306	-0.548	-0.157	-0.333	2.733	2.587	1.194
REGR factor score 6 for analysis 1	2.059	-1.024	-0.135	0.421	-0.986	-0.111	3.308	-1.004	4.470	0.769
REGR factor score 7 for analysis 1	6.960	3.291	-0.804	1.140	-1.982	-0.012	-0.719	-0.080	-3.035	-1.454
REGR factor score 8 for analysis 1	1.534	-7.057	5.666	-3.452	7.054	-0.046	-0.318	-0.178	-0.363	-1.211
REGR factor score 9 for analysis 1	-0.479	-10.575	-5.834	7.533	2.005	-0.005	-0.566	-0.657	1.775	-0.378
REGR factor score 10 for analysis 1	-0.992	21.990	-1.878	1.820	3.551	-0.047	0.277	-0.750	1.229	-0.388
REGR factor score 1 for analysis 2	0.329	-1.776	0.964	1.825	0.453	0.041	-0.613	-0.697	-1.092	0.051
REGR factor score 2 for analysis 2	-0.011	-1.267	3.952	-0.023	1.101	-0.084	0.689	0.518	1.991	-1.413
REGR factor score 3 for analysis 2	-3.874	0.037	5.838	1.371	1.945	-0.013	-0.942	0.867	0.452	-4.978
REGR factor score 4 for analysis 2	3.503	-2.207	7.385	0.290	0.726	-0.143	2.262	-0.117	0.853	3.847
REGR factor score 5 for analysis 2	1.918	-1.202	-8.819	0.341	2.565	0.012	-2.742	0.558	1.548	3.304
REGR factor score 6 for analysis 2	2.101	-0.341	-0.385	6.124	-2.613	-0.065	-0.722	1.037	-0.368	0.087
REGR factor score 7 for analysis 2	-0.686	0.475	1.013	-3.906	-1.435	-0.089	0.046	1.710	4.344	5.256
REGR factor score 8 for analysis 2	3.556	11.078	4.199	-2.096	-1.341	0.043	-1.792	-0.841	1.995	-6.197
REGR factor score 9 for analysis 2	-2.536	13.411	0.751	2.285	1.514	-0.048	0.728	-0.414	-0.014	16.233
REGR factor score 10 for analysis 2	0.111	7.840	-3.134	-2.613	-0.893	-0.028	1.379	1.249	-3.595	-9.823
REGR factor score 11 for analysis 2	-2.030	-0.003	-4.945	3.945	-2.044	-0.029	1.642	-0.655	5.666	-10.893

Table A2.3.5: ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
REGR factor score 1 for analysis 1	18.066	9	0.803	778	22.509	0.000
REGR factor score 2 for analysis 1	64.438	9	0.266	778	242.113	0.000
REGR factor score 3 for analysis 1	10.564	9	0.889	778	11.878	0.000
REGR factor score 4 for analysis 1	48.099	9	0.455	778	105.679	0.000
REGR factor score 5 for analysis 1	38.496	9	0.566	778	67.984	0.000
REGR factor score 6 for analysis 1	50.652	9	0.426	778	119.006	0.000
REGR factor score 7 for analysis 1	56.830	9	0.354	778	160.471	0.000
REGR factor score 8 for analysis 1	60.865	9	0.307	778	197.947	0.000
REGR factor score 9 for analysis 1	59.688	9	0.321	778	185.893	0.000
REGR factor score 10 for analysis 1	70.700	9	0.194	778	364.981	0.000
REGR factor score 1 for analysis 2	6.439	9	0.937	778	6.871	0.000
REGR factor score 2 for analysis 2	10.564	9	0.889	778	11.878	0.000
REGR factor score 3 for analysis 2	32.741	9	0.633	778	51.740	0.000
REGR factor score 4 for analysis 2	39.823	9	0.551	778	72.288	0.000
REGR factor score 5 for analysis 2	47.705	9	0.460	778	103.772	0.000
REGR factor score 6 for analysis 2	35.825	9	0.597	778	59.995	0.000
REGR factor score 7 for analysis 2	40.157	9	0.547	778	73.410	0.000
REGR factor score 8 for analysis 2	50.307	9	0.430	778	117.098	0.000
REGR factor score 9 for analysis 2	61.850	9	0.296	778	208.899	0.000
REGR factor score 10 for analysis 2	44.595	9	0.496	778	89.968	0.000
REGR factor score 11 for analysis 2	67.123	9	0.235	778	285.529	0.000

Table A2.3.6: Number of Cases in each Cluster

	Number of Cases in each Cluster									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Cluster 2	769	19								
Cluster 3	1	786	1							
Cluster 4	1	1	784	2						
Cluster 5	1	778	2	1	6					
Cluster 6	1	443	2	335	6	1				
Cluster 7	1	5	125	2	637	1	17			
Cluster 8	11	615	2	1	31	5	1	122		
Cluster 9	11	5	1	476	5	286	1	2	1	
Cluster 10	*8*	*1*	*2*	*5*	*7*	*702*	*21*	*34*	*7*	*1*