**FACEPA**

*Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture*

# Methodology to assess the farm production costs using PMP farm models

FACEPA Deliverable No. D6.1 – April 2009

Luca Cesaro

Sonia Marongiu

Filippo Arfini

Michele Donati

Maria Giacinta Capelli

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 main objective of this deliverable is to present a positive mathematical programming based methodology for recovering the production specific costs for agricultural activities collected by the FADN database. It is well known that FADN, at European level, doesn't collect the information about the variable costs associated to the different farm activities, but only the total variable costs at farm level. In this respect, it is evident that all the analysis that aims to evaluate the production allocation decisions cannot be carried without costs derived from external sources (engineering information, literature, etc.), with the risk to not be able to differentiate the costs according to the farm specialization and the farm size (economic and physic). The policy and market evaluations that want use FADN are constrained to this important lack, because the specific profitability of the farm activities included in the sample is not explicit and it should be estimated.

In this context, the calibration process adopted inside the PMP is proposed for estimating this farm production decision component. The standard PMP calibration method is modified so that it is possible to generate the observed production plan using the dual structure of the problem proposed by Howitt and Paris (1998). This model allows to evidence the dual information linked to each farm production activity and estimate it using an approach based on the Heckelei's PMP proposal (2002). In the following sections, the standard approach will be presented with two PMP calibration variants for recovering the farm activity marginal costs. The first one aims to estimate the marginal costs using the technical farm information and the activity marginal revenue, according to the Heckelei's proposal; while, in the second proposal, the Heckelei' PMP approach is extended to the same estimation using the information about the farm total variable costs.

For this last methodology, the deliverable presents an application to a group of farms (35) collected from the Italian FADN archive, where the information about hectares, yields, prices, subsidies and farm total variable costs is used for estimating the specific costs associated to the farm production plan. The results achieved have been compared with the real specific accounting costs registered by the national database – instead of the European database, where this kind of information is, as known, not captured –, in order to test the estimation goodness and the related estimation errors. The estimation produces very interesting results, that fit with a very high approximation degree the observed levels. This results should be viewed as a preliminary results that should be tested with more observed data and on much more differentiated farm typologies.

# Contents

# Abbreviations and Acronyms

| | |
|---|---|
| CAP | Common Agricultural Policy |
| EU | European Union |
| FACEPA | Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture |
| FADN | Farm Accountancy Data Network |
| GM | Gross Margin |
| IACS | Information Accounting and Control System |
| LP | Linear Programming |
| PMP | Positive Mathematical Programming |
| REGIO | Regional Eurostat databank |

# List of Figures and Tables

# 1. Introduction

The main objective of this deliverable is to present a positive mathematical programming based methodology for recovering the production specific costs for agricultural activities collected by the FADN database. It is well known that FADN, at European level, doesn't collect the information about the variable costs associated to the different farm activities, but only the total variable costs at farm level. In this respect, it is evident that all the analysis that aims to evaluate the production allocation decisions cannot be carried without costs derived from external sources (engineering information, literature, etc.), with the risk to not be able to differentiate the costs according to the farm specialization and the farm size (economic and physic). The policy and market evaluations that want use FADN are constrained to this important lack, because the specific profitability of the farm activities included in the sample is not explicit and it should be estimated.

In this context, the calibration process adopted inside the PMP is proposed for estimating this farm production decision component. The standard PMP calibration method is modified so that it is possible to generate the observed production plan using the dual structure of the problem proposed by Howitt and Paris (1998). This model allows to evidence the dual information linked to each farm production activity and estimate it using an approach based on the Heckelei's PMP proposal (2002). In the following sections, the standard approach will be presented with two PMP calibration variants for recovering the farm activity marginal costs. The first one aims to estimate the marginal costs using the technical farm information and the activity marginal revenue, according to the Heckelei's proposal; while, in the second proposal, the Heckelei' PMP approach is extended to the same estimation using the information about the farm total variable costs.

# 2. Positive Mathematical Programming standard approach

Positive Mathematical Programming (PMP) is widely used for evaluating the effects of the CAP instruments on the dynamics of the agricultural processes and farm economic variables, both for ex-post and ex-ante analysis. The main contribution of this methodology to agricultural economics is due to its capacity to maximize the information contents in the agricultural datasets available at European level, as FADN, REGIO, IACS (Arfini et al., 2003; Paris and Howitt, 1998). Thanks to the recovering of farm decision variables, by the way of estimating the total variable cost function, PMP is capable to reproduce the exact observed farm allocation plan and the decision variables (total specific variable costs) that led farmers to decide for such a production plan.

Many papers have adopted the PMP methodology for developing models capable to assess the impact of proposed or already implemented CAP reforms. Also in European research projects, this approach is used with micro-based information, like FADN1. In most cases, the PMP is proposed in the so-called "classical" form, where the procedure is articulated in three phases: the differential costs recovering, the estimation of the non-linear cost function and, finally, the calibration by using a non constrained production model with non-linear objective function (Howitt, 1995). Applications of this basic version are the most diffused, e.g. for evaluating CAP's reform impacts (Arfini et al., 2005; Judez et al., 2002).

An attempt to introduce innovations in the basic approach is due to Heckelei and Wolff (2003) that proposed a methodology that overcomes the first phase for calibrating the observed situation by directly imposing the first order conditions in the cost function estimation phase. This approach was also used with cross-section data in order to enhance the consistency of the cost estimation (Heckelei and Britz, 2000). More advanced extensions of the PMP are due to Paris (2001) that generalizes the method adopting an equilibrium model in a static framework and in a dynamic price expectation approach.

The demand for an assessment of agricultural policy measures rose with force during this last decade and contributed to the development of a set of economic tools that would respond to such needs using all the available information. In this field, the PMP plays a first order role. This methodology can provide useful results to policy makers even in the presence of a limited set of information as it generally happens when European agricultural databases are adopted. PMP can responds with flexibility and in a consistent way to a large spectrum of policy issues, typically concerning the land use change, production dynamics, variation in gross margin and in other main economic variables (costs, subsidies, gross saleable

[1] Several European research projects have developed and applied models based on the Positive Mathematical Programming methodology, as CAPRI (Heckelei, 1997; Heckhelei and Britz, 2000) and EUROTOOLS (Paris and Arfini, 2000) in the V FP, GENEDEC (contract no. SSPE-CT-2004-502184 and CARERA (contract no. SSPE-CT-2005-022653) in the VI FP.

production, etc.). However, all these applications are developed exploring the supply side of the agricultural sector while avoiding to implement an evaluation of the demand side, by measuring the effects on the output market prices. Indeed, the literature about PMP models application seems to indicate that such a class of models was just developed for investigating the supply side of the agricultural sector, delegating the demand issues side to well-posed problems solved by econometric techniques.

## 2.1. The mathematical structure of the PMP model

The methodology described in Paris and Howitt (1998) and Paris and Arfini (2002), as a mathematical programming process to analyse farmer behaviour, that can recover the latent information driving the farmer decision process and using it in assessing the likely responses to market and policy scenarios. PMP consists of three steps.

The first is defined by N linear programming (LP) models, one for each macro-farm, and by an additional LP model for the entire sub-region. The n-th individual macro-farm model (n=1,…,N), uses all the available information pertaining to the n-th farm in order to derive the vector of shadow prices of the limiting allocable inputs, y, and the differential marginal cost vector corresponding to the vector of realized output levels, λ . The n-th farm LP model has the following structure:

(1) $$\max_{x_v^n} GM = \sum_{v=1}^{V}[x_v^n(pr_v^n - c_v^n)] + \sum_{v=1}^{V} xh_v^n sh_v^n$$

where $x_v^n$ is the production level for each process, v=(1,...,V), of each farm in the sample, n=(1,...,N), while $pr_v^n$ e $c_v^n$ are, respectively, the price and the cost associated with each product level. The objective function takes into consideration the amounts of farm aid — defined as the product of the growing area, $xh_v^n$, and the per hectare aid level, $sh_v^n$ — as part of the farm's gross margin (GM). The objective function specified in (1) is subject to a series of constraints that can be expressed as:

(2) $$\sum_{v=1}^{V}(a_v^n x_v^n) \le b^n$$

(3) $$x_v^n \le \overline{x}_v^n + \varepsilon$$

(4) $$x_v^n \ge 0$$

where $a_v^n$ is the element of the technical matrix of the different activities implemented by each of the n farms in the sample (the $n$-th matrix An of technical coefficients is defined as An=[anij], where anij=hRni /xRnJn), bn is the vector of availability of limiting allocable inputs. (4) presents the non-negativity constraint placed on the primal variables of the problem.

Constraints (2) are called structural constraints, while constraints (3) are called calibration constraints. The constraint in Equation (2) indicates the overall availability of scarce factors to be allocated among the various production processes V. In the present model, the only limiting factor is the land to be used for the various production processes. Constraint (3), on the other hand, concerns the production capacity of each activity on the farm, defined according to the levels of production observed. Constraint (3) reproduces the initial situation observed in terms of production levels for each farm activity. The term $\varepsilon$, a low positive number selected at will, serves to separate structural constraint (2) from calibrating constraint (3). In fact, if this term were omitted, the linear dependence between the two constraints would lead to dual positive values for all the calibration constraints while the shadow price for the structure constraint in (2) would remain at zero, making interpretation difficult and hardly reflecting reality (Paris and Arfini, 1995).

The problem of linear programming (1)-(4) uses calibration constraints to reconstruct the situation observed, restoring the dual values associated with the production capacity constraints in (3), $\lambda_v^n$ .

This initial phase, therefore, serves to derive the dual variables specific to the production processes used on the farm. This information incorporates the technical and economic elements the farmer considers in defining the farm production plan.

However, the lack of specific cost information at farm level means we cannot derive the cost function parameters for the marginal product, since its marginal cost value is null. So we have to implement an alternative first phase, different from the traditional PMP model formulation, where we derive the shadow prices associated with the binding and calibrating constraints, by the resolution of a problem in which the constraints are represented by the equilibrium conditions of the problem (1) – (4). This problem is solved by means of traditional econometric tools (Heckelei, 2003) and by an innovative methodology we propose below.

## 2.2. Deriving the cost function

The objective of the second phase of the PMP procedure is to estimate the farm cost function. Starting from the vector of the shadow prices associated with the calibration constraints, we can determine a new cost function that meets the criteria defined by both economic theory of production costs and farm reality. To meet the non-linearity condition for the objective function of the third phase, a quadratic functional shape is used (Howitt, 1995). Starting from the information on the problem of linear programming it is, therefore, possible to build a new quadratic cost function defined as follows:

(5) $$(\lambda + c)\overline{x} = \frac{1}{2}\overline{x}'Q\overline{x}$$

where $\lambda$ and $c$ are, respectively, the vector of the dual values that determine the first phase and the vector of the accounting costs, $\overline{x}$ is the vector of the known production levels and $Q$ the matrix of the non linear function of total costs. In (5) the elements for matrix $Q$ are still unknown and must be derived through suitable estimation methods. In the literature (Paris et al., 2000) estimation through application of the principle of maximum entropy is preferred. With this principle, the uncertainty regarding the realization of that event must be maximised in order to derive the probability of distribution for a given event. To clarify the concept, we introduce the general formula of the entropy for s possible occurrences of the same phenomenon:

(6) $$H(p_1, p_2, ..., p_s) = \sum_{i=1}^{s} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{s} p_i \log p_i$$

where $p_i$ is the i-th probability of a probability distribution made up of s elements. From (6) one can see that if the probability $p_i = \frac{1}{s}$ — that is the case of uniform distribution, where the degree of uncertainty is highest — the function  is maximised and is an increasing, monotone function of s. The case of uniform distribution corresponds to the case where some elements are available for a given phenomenon. However, when we know some distribution moments, following the above reasoning, we can maximize the entropy of the probability distribution by placing constraints on the moments used to derive it. In other words, we look for that probability distribution that comes closest to the uniform distribution (Jaynes, 1957).

Considering that entropy measures the degree of uncertainty regarding realization of a phenomenon, this approach can be applied to estimating a parameter, the value of which can be defined within an as-yet unknown probability distribution. On the basis of these concepts and the adaptations given by Paris and Howitt (1998), the parameters of matrix $Q$ can be recovered by maximizing the probability distribution associated with an interval of suitably specified support values.  The non linear programming problem of maximum entropy is applied to the estimation of the matrix Q decomposed according to the Cholesky factorization, where $Q = LDL' = TT'$, where L is a triangular matrix, D a diagonal matrix and $T = LD^{1/2}$. The problem can then be resolved by maximizing a probability distribution for which we know the expected value, which corresponds to the marginal cost $(\lambda + c)$ determined in the first phase. The objective function of the problem of maximum entropy is thus presented as follows:

$$\max_{p^l_{(\cdot)}, p^d_{(\cdot)}, p^u} \quad - \sum_{v=1}^{V}\sum_{v'=1}^{V}\sum_{w=1}^{W}\left(p^l_{vv'w}\log p^l_{vv'w}\right) - \sum_{v=1}^{V}\sum_{v'=1}^{V}\sum_{w=1}^{W}\left(p^d_{vv'w}\log p^d_{vv'w}\right) - \sum_{w=1}^{W}\left(p^u_w \log p^u_w\right) \tag{7}$$

where $p^l_{(\cdot)}$ and $p^d_{(\cdot)}$ are the probability of the distribution associated with elements of the triangular matrix L and of the diagonal matrix D while $p^u_{(\cdot)}$ are elements of the probability of errors, or differences, vs. the farm costs-sum. In fact, the cost matrix is estimated on the basis of the following equation:

$$\bar{\lambda}_v + \bar{c}_v = \sum_{v'=1}^{V}\left\{\sum_{v''=1}^{V}\left(T_{vv''}T_{v''v'}\right)\right\}\bar{x}_{v''} \tag{8}$$

In (8), $(\bar{\lambda}_{(\cdot)} + \bar{c}_{(\cdot)})$ is the average marginal cost of the production processes for the group of N farms considered in the model. $T_{(\cdot)}$ is an element of the matrix T obtained through Cholesky's * decomposition. In fact:

$$T_{vv''} = \sum_{v'=1}^{V}\left\{\sum_{w=1}^{W}\left(p^l_{vv'w}z^l_{vv'w}\right)\sum_{w=1}^{W}\left(p^d_{vv'w}z^d_{vv'w}\right)^{1/2}\right\} \tag{9}$$

The relationships inserted into (9) clarify the role of the support values in the process of estimating the cost matrix. The components $z^l_{(\cdot)}$ and $z^d_{(\cdot)}$ are the appropriately selected support values (Paris and Howitt, 1998). Associated with the distribution of probability, $p^l_{(\cdot)}$ and $p^d_{(\cdot)}$, they define the elements of the triangular matrix L and of the diagonal matrix D. It must be pointed out that matrix $Q$ is unique and is derived from the marginal costs of the farm-sum. In this context, the cost function specified according to the $Q$ matrix is also called the frontier cost function, indicating that the farm-sum cost function is the most efficient activity cost structure (Paris and Arfini, 2000).

To define the quadratic marginal cost associated with each form in the sample, the difference (or error) vs. the average marginal cost must be determined. Thus, for the processes implemented — that is for those which are strictly positive — the individual marginal cost function is:

(10) $$\lambda_v^n + c_v^n = \sum_{w=1}^{W}\left(p_{vw}^{un} z_{vw}^{un}\right) + \sum_{v'=1}^{V}\left\{\sum_{v''=1}^{V}\left(T_{vv''} T_{v'v''}\right)\right\}\bar{x}_{v''}^{n}$$

where $(\lambda_{(\cdot)}^n + c_{(\cdot)}^n)$ is the individual marginal cost of the n-th farm. The average errors are given by the product obtained, multiplying the specially identified support values $z_{(\cdot)}^{un}$ and the relative probabilities $p_{(\cdot)}^{un}$. Moreover, given that the cost function contains all production processes implemented by the sample of farms considered, we must also consider those farms that have not implemented the entire range of processes identified for the sample as a whole. For this reason, the model calls for the following relation for N farms:

(11) $$\bar{\lambda}_v^n + \bar{c}_v^n \leq \sum_{w=1}^{W}\left(p_{vw}^{un} z_{vw}^{un}\right) + \sum_{v'=1}^{V}\left\{\sum_{v''=1}^{V}\left(T_{vv''} T_{v'v''}\right)\right\}\bar{x}_{v''}^{n}$$

All the above probability distributions must meet the following condition:

(12) $$\begin{cases} \sum_{w=1}^{W} p_{(\cdot)}^{l} = 1 \\ \sum_{w=1}^{W} p_{(\cdot)}^{d} = 1 \\ \sum_{w=1}^{W} p_{(\cdot)}^{un} = 1 \end{cases}$$

Problem (7)-(12) provides the probability distribution values for the elements of the triangular matrix L, the diagonal matrix D and for the vector of the residual marginal variable costs for each farm in the sample. The reconstruction of the elements that make up matrix $Q$ is obtained from the following:

(13) $$q_{vv'} = \sum_{v''=1}^{V} \{T_{vv''} T_{v'v''}\}$$

where $q_{(\cdot)}$ is one of the parameters that make up the cost matrix $Q$. The cost function specified by the above method preserves the technical information regarding the calibration constraints.

If the cost function is inserted in a problem similar to the one identified in the first phase, it makes it possible to reproduce the situation observed, but without the calibration constraints. This last model exactly reproduces the base period allocation and output decision of the single $n$-th macro-farm and of the entire region.  That is, the primal and dual solutions of this quadratic programming models are exactly equal to the primal and dual solution of the initial LP model which, in turn, reproduces the results of the base period.  This is the meaning of calibration within the PMP methodology.  This model is analogous to the model specification and selection of econometric studies. The prediction step of PMP exploits the calibrated model to generate responses in the endogenous variables induced by the variation of some relevant parameters, assimilated to the exogenous variables of econometric models. It can be used to analyse various scenarios of agricultural policy with changes in output prices, and limiting resource availability.

# 3. An alternative to the traditional PMP model

One of problems of the FADN archive regarding the probable dynamics of farm production plans and farm revenue is the absence of any kind of specific activity costs. This means that, when specifying a PMP model, the cost function recovered in the second phase of the methodology cannot be correctly derived, because marginal costs associated with the calibrating constraint equal zero. This happens for the farm activity with the lowest marginal profit .

In other words, the lack of specific cost information at farm level means that we cannot derive the cost function parameters for the marginal product, since its marginal cost value is null. To solve this problem, the literature provides a number of contributions that modify PMP traditional formulation. In particular, Heckelei (2002) offers a wide range of instruments for assess the cost function starting from the observed production level.

In this case, one possible solution is an alternative first phase, different from the traditional PMP model, where one derives shadow prices associated with binding and calibrating constraints by using the equilibrium conditions of the problem (1) – (3). This implies that the first phase of the PMP methodology changes in such a way that all the marginal costs can be recovered from the observed production information values using the optimal conditions as the relevant information for calibrating the observed situation.

One of the most useful ways to overcome the absence of specific activity costs is to directly define the first order conditions of the problems (1)-(3) and to optimise the problem by deriving shadow prices and the production levels by minimizing the slackness variables associated with dual endogenous variables. In the following section we present an overview of the principles of the alternative method to traditional PMP.

In order to present the revisited PMP methodology without prior information about specific variable costs, we can say that that the first phase of PMP in the traditional approach as shown by Paris-Howitt (1998) is presented as a tautological procedure aiming to identify the marginal costs associated with different farm products. In fact, the level of activity variables is already known before the model resolution and does not require derivation by optimisation. So the first phase of PMP can be avoided, and substituted by a two step procedure. This is obtained starting from the model suggested by Heckelei (2002), in which first order optimality conditions are imposed in the first phase of PMP.

Heckelei's paper says that "the general alternative to PMP with respect to calibrating or estimating a programming model is a simple methodological principle: always to directly use the first order conditions of the very optimisation model that is assumed to represent or approximate producer behaviour and is suitable to the simulation needs of the analysts". This approach is a maximisation model in which one overcomes the problem of poor data transmission about farm behaviour by estimating shadow prices of resource constraints simultaneously with the other parameters of the model.

This methodology supplies a general and flexible tool for estimating parameters of duality based behavioural functions with explicit allocation of fixed factors. In this context, the only difference between programming and econometric models is the form of the simulation model (Heckelei, 2003). But if the solution of this problem is known at the beginning of the analysis, Heckelei's first phase can be omitted and optimality conditions imposed.

The approach proposed here as an alternative to traditional PMP in the absence of cost information on individual farm activities is mainly based on Heckelei's alternative solution. The difference is that it estimates marginal costs related to the binding inputs and the farm products.

In order to explain this, we can start with the Langrangian function associated with the problem (1)-(4), in which the variable costs in the objective function are supposed to be unknown :

(14) $$\boldsymbol{L} = (\boldsymbol{p}-\boldsymbol{c})'\boldsymbol{x} + \boldsymbol{sh}'\boldsymbol{h} + \boldsymbol{y}(\boldsymbol{b}-\boldsymbol{Ax}) + \boldsymbol{\lambda}(\overline{\boldsymbol{x}}+\boldsymbol{\varepsilon}-\boldsymbol{x})$$

From the Lagrangian function above we can derive, by Kuhn-Tucker conditions, the optimality relation for the problem, as follows:

(15) $$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{y}} = \boldsymbol{Ax}-\boldsymbol{b} = 0$$

and

(16) $$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{x}} = \boldsymbol{A}'\boldsymbol{y} + \boldsymbol{\lambda} - \boldsymbol{r} = 0$$

Optimality condition (15) states that for the maximum level of the objective function the level of the variable $\boldsymbol{x}$ should use all the quantities of input available, in such a way that the structural constraint is completed. Condition (16) establishes the economic condition on the basis of which the marginal cost must be equal or greater than the marginal revenue. In this case, marginal revenue is shown by vector $\boldsymbol{r}$. This vector is the result of all positive economic parameters considered by the objective function. More specifically, the elements of the vector are composed by the sum of the price of the product and its level of aid. The element v for farm n is specified as follows:

(17) $$r_v^n = p_v^n + [sh_v^n a_v^n]$$

where sh is multiplied by the inverse yields of the crops to obtain the value of aid associated with one unit of product quantity.

In the two optimality conditions above we can observe that for equation (15) we know perfectly all elements: the level of output that maximises the objective function of the PMP specification concerns the observed output. But equation (16) is not perfectly known at the beginning of the solution; we have no prior information on marginal costs of input or shadow prices for the different processes.

At this level, Heckelei provides two solutions:

1. The shadow prices from the first phase of the traditional PMP model can be derived by specific econometric tools (e.g. generalized least squares method) directly applied to Equations (15) and (16);
2. Stakeholders and land market information can supply the marginal value of the land, and this exogenous information can be used to derive shadow prices of outputs;

In this framework, we propose a third alternative that matches an endogenous estimation of all the dual values of the problem (1-4), but without the optimality condition (4), and using land value to calculate the dual values for each activity.

We suppose that the farmer prefers to rent land at a price not higher than the marginal internal process, or the product with the lowest contribution to farm total profit. In fact, the choosing process, in the short term, considers the marginal contribution of each crop to farm revenue. So land purchase is submitted to a comparison between the marginal cost of the land and the marginal productivity of this input for each production possibility. All processes in the production plan of the farm must show non-negative economic return, always taking into account the land rent costs. So the maximum value of the land entering the decision process should not be greater than the lowest contribution to farm revenue provided by a unit of output.

On the basis of this assumption, the value of the land in the decision process of each farmer can be calculated as:

(18) $$\bar{y} = \min\left\{ p_v \, 1/a_v + sh_v \right\}$$

where the $\bar{y}$ is the maximum value of land for a certain farm estimated as the minimum marginal profit contribution among all the farm processes. Prior information about the shadow price of the land, rather than being exogenous, is endogenously derived from observed information.

This mechanism is the same as that applied in the first phase of the traditional PMP. The resolution of the linear programming model together with the constraints associated with the production capacity of the different farm activities helps to define a shadow price for the land using the marginal profit of the least profitable crop. The calibrating constraints are generally defined adding to the right hand side of the inequality a small perturbation component in such a way that only one crop remains without a positive shadow price. In other words, the crop with the lowest contribution to total farm profit in fact defines the marginal value of the land. This is why the minimum value of marginal revenue among all the crops activated by the farms is introduced into the constraint.

Using the results of this condition to reconstruct the cost function of the farms, we can directly implement the maximum entropy problem in such a way that we can estimate a quadratic cost function to reproduce the observed production plan. In this case, like the traditional model, we reconstruct one Q matrix for the entire sample of farms, and we can estimate the deviations of marginal farm costs from the so-called frontier cost according to this matrix. This means that we state for two main relations: one on the sample, and the other for each individual for each farm.

Relation (18) represents the equivalence between the dual optimal condition and the marginal cost component of the quadratic cost function for all farms; equation (19) introduces farm information and thus also error component u, as the marginal cost deviation at farm level from the cost of the most efficient farm in the sample, that is the idealised farm from equation (18).

(18) $$p_v^{avg} + sh_v^{avg} a_v^{avg} - \overline{y}^{avg} a_v^{avg} = \sum_{v'} q_{vv'} \overline{x}_{v'}^{sum}$$

(19) $$p_v^n + sh_v^n a_v^n - \overline{y}^n a_v^n = \sum_{v'} q_{vv'} \overline{x}_{v'}^n + u_v^n$$

(The superscript “avg” means that the parameter of the relation is obtained as average of the sample values.)

The resolution of a maximum entropy problem is very similar to that in traditional PMP, with only the two above equations and without other equations or information obtained from optimization processes (i.e. first phase of PMP). A cost function able to reach the calibrated solution in the subsequent “simulation” phase can be derived.

## 3.1. Regional aggregation of PMP models for policy evaluation

The PMP model is therefore a regional model in which information on the farms is aggregated at sub-regional level and simulated by PMP procedure to provide predictions about agricultural policy change that are as representative as possible.

From a methodological point of view, the model allows aggregation of the single sub-regional models and maximisation during simulation. Each simulation is carried out simultaneously for all sub-regions, allowing for the introduction of constraints at regional level.

In many regional models based on the use of PMP for which literature is available, the simulation process involves the resolution of a problem of optimisation for each single sub-region, without considering the complex constraints at regional level and the profitability of other sub-regions within the same region. In this model, the simulation phase includes the maximisation of an objective function aggregated by group of sub-regions that comprise the region. For this reason, the model appears as a "concatenated" model. A model in which during the policy scenario simulation phase, the decisions taken by each sub-region are linked to the decisions taken by the bordering sub-regions through the definition of a problem of simultaneous optimisation.

In the phase concerning the effects of agricultural policy measures at regional level, the important aspects of the PMP model are therefore the aggregation of cost functions into a single regional model, and the construction of a suitable set of constraints to correctly simulate the policies for the whole region.

As illustrated in the previous section, the estimation of the cost function for each sub-region has the specific aim of estimating the parameters comprising the matrix $Q$, which incorporates all the information concerning the relations of substitution and complementarity between the processes, and represents the total cost function of the sub-region. Very briefly, we can express the $Q$ matrix as follows:

(20)
$$Q = \begin{bmatrix} q_{11} = \begin{bmatrix} w_{11}^1 \cdots w_{11}^s \end{bmatrix} \begin{bmatrix} p_{11}^1 \\ \vdots \\ p_{11}^s \end{bmatrix} & \cdots & q_{1n} = \begin{bmatrix} w_{1n}^1 \cdots w_{1n}^s \end{bmatrix} \begin{bmatrix} p_{1n}^1 \\ \vdots \\ p_{1n}^s \end{bmatrix} \\ \vdots & \ddots & \vdots \\ q_{n1} = \begin{bmatrix} w_{n1}^1 \cdots w_{n1}^s \end{bmatrix} \begin{bmatrix} p_{n1}^1 \\ \vdots \\ p_{n1}^s \end{bmatrix} & \cdots & q_{nn} = \begin{bmatrix} w_{nn}^1 \cdots w_{nn}^s \end{bmatrix} \begin{bmatrix} p_{nn}^1 \\ \vdots \\ p_{nn}^s \end{bmatrix} \end{bmatrix}$$

where, as shown by Paris and Howitt (1998), the parameters of the $Q$ matrix are estimated through the distribution of probability $p_{nn}^s$ (s=1,…,S) associated with the interval * * * * with the support weighting $w_{nn}^s$ .

In the second phase of the PMP procedure the model will therefore estimate the same number of cost functions as there are sub-regions in the referred region. Estimated in this way, the cost functions will represent the specific economic structure of each homogeneous area and will be used during the agricultural policy measure simulation phase.

The information on the estimated cost functions for the sub-region ($Q^*$) using Maximum Entropy are gathered in a parameter indicated by $\hat{Q}$, which links the single $Q^*$ matrices in one single vector. The same aggregation procedure is developed for other information needed for constructing the policy model; output price, yield and compensation payments.

The information reorganised into vectors is included in the regional model and allows for an efficient definition of the problem of maximisation, as the overall group of vectors joins all the components concerning the objective functions of the *n* sub-regions in the referred region in one single matrix. The aim of the regional model is therefore to reproduce the initial production situation for the entire referred region without any calibration constraints, then to calibrate the model for the entire region once more.

More precisely, the objective function of the regional model sums the single objective functions of each sub-region, maximising the overall gross income for the region.

(21)
$$\sum_{n=1}^{N} PROF^{n} = PROFT$$

where $PROFT$ represents the gross regional income, while $PROF^{n}$ represents the gross income of each sub-region.

Given that the structure of the constraints of each sub-region is independent from the others, the maximisation of the gross regional income is the result of the maximisation of the gross income of each sub-region, $PROF^{n}$, giving an optimal solution at regional level that is equally optimal at sub-regional level.

The maximisation of the regional objective function is subject to a series of constraints that fix the structural characteristics (the surface) and reproduce the agricultural policy scenarios for each sub-region. Specifically, the structural constraint (21) on the available resources obliges the land used for produce sold, $xh_{v}^{n}$, re-used, $xh_{r}^{n}$, the set-aside, $xhs^{n}$ and the non-productive cultivated land respecting good agronomic practice $xhb^{n}$, to be at least equal to the total land availability at sub-regional level, $b^{n}$.

(21)
$$\sum_{v=1}^{V} xh_{v}^{n} + \sum_{r=1}^{R} xh_{r}^{n} + xhs^{n} + xhb^{n} \leq b^{n}$$

The structure of (21) allows us to determine the set-aside due to the variable $xhs^n$, (set-aside surface), which is also present in the constraint (22) concerning the sub-region set-aside:

(22) $$\left(\sum_{Vcop} xh_v^n + xhs^n\right)\theta^n \le xhs^n$$

where $\theta^n$ is the set-aside rate for each sub-region that multiplies the surface used for COP crops.

In the same way constraints have been added for beet and tomato crops. Also in this case the production by sub-region has been divided into two quotas, the first relating to the quantity produced during the year of observation and the second to the excess production with respect to the quota on which the output price penalty is to be applied (23).

(23) $$\begin{aligned} x_{ORT}^n &= xq_{ORT}^n + xf_{ORT}^n \\ xq_{ORT}^n &\le \overline{x}_{ORT}^n \end{aligned}$$

where $x_{ORT}^n$ is the variable of production level of horticultural crops (ORT), and the other symbols have the same meaning as those used for the milk quota constraints.

For both the milk quota constraints and the horticultural crop constraints there is an objective function for each sub-region, $PROF^n$, introducing a negative income component for the part of the production that exceeds the quota.

This procedure can be replicated from regional to national level, replacing the sub-regional constraints with a single constraint that functions on national level. This would allow for the introduction of an agricultural policy system, such as the “Maximum Guaranteed Surface”, maintaining the production particularities of each single sub-region thanks to the gathered information. This last aspect could represent a highly useful element for agricultural policy analyses, which are increasingly having to consider the technological, structural, production and economic characteristics of the individual European sub-regions. Furthermore, the main agricultural intervention tools, the principle of decoupling farm aid and modulation, have been introduced into the PMP model.

# 4. PMP cost estimation approach using farm total variable costs

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 at obtaining additional information on the behaviour of the farm so as to be able to simulate its behaviour in conditions of maximization of the 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, utilising 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 to use data banks, among which also the FADN data bank (Arfini, 2005) .

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 utilises dual optimality conditions directly in the estimation phase of the non linear function. The approach qualifies itself as an extension of the Heckelei proposal (2002), 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 activity costs, the model considers the information relative to the total farm variable costs available in the European FADN archive. This "innovation" becomes particularly important as it enables us to perform analyses utilising the European data bank without having to resort to parameters that are exogenous to the model.

According to this new approach, the PMP model falls into two phases: a) the aim of the first is to estimate specific cultivation costs through the reconstruction of a non linear function of the total variable cost that considers the exogenous information on the total variable costs observed 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 entered.

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

(24) $$\min_{u} LS = \frac{1}{2}\mathbf{u}'\mathbf{u}$$

subject to

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

(26) $$\mathbf{c} + \boldsymbol{\lambda} \leq \mathbf{R}'\mathbf{R}\overline{\mathbf{x}} + \mathbf{u} \qquad \text{se } \overline{x} = 0$$

(27) $$\mathbf{c}'\overline{\mathbf{x}} \leq TC$$

(28) $$\mathbf{u}'\overline{\mathbf{x}} + \frac{1}{2}\overline{\mathbf{x}}'\left(\mathbf{R}'\mathbf{R}\right)\overline{\mathbf{x}} \geq TC$$

(29) $$\mathbf{c} + \boldsymbol{\lambda} + \mathbf{A}'\mathbf{y} \geq \mathbf{p} + \mathbf{A}'\mathbf{s}$$

(30) $$\mathbf{b}'\mathbf{y} + \boldsymbol{\lambda}'\overline{\mathbf{x}} = \mathbf{p}'\overline{\mathbf{x}} + \mathbf{s}'\overline{\mathbf{h}} - \mathbf{c}\overline{\mathbf{x}}$$

(31) $$\mathbf{R} = \mathbf{L}\mathbf{D}^{1/2}$$

(32) $$\sum_{n=1}^{N} u_{n,j} = 0$$

By means of the model (24)-(32) a non linear cost function can be estimated using the explicit information on the total farm variable costs (TC) available in the FADN data bank. The restrictions (25) and (26) 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 (27) imposes that the total estimated explicit cost should not be more than the total variable cost observed in the FADN data bank. Restriction (28) 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 cost (TC) measured. In order to guarantee consistency between the estimation process and the optimal conditions, restriction (29) 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 factors of production 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 (30) 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 (31). Last but not least, restriction (32) establishes that the sum of the errors, u, must be equivalent to zero.

The cost function estimated with the model (24)-(32) 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:

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

subject to

(34) $$\mathbf{Ax} \leq \mathbf{b}$$

(35) $$A_j x_j - h_j = 0 \qquad \forall j = 1,..., J$$

The model (33)-(35) 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. The matrix Q estimated is reconstructed using Cholesky's decomposition: $\mathbf{\hat{Q}} = \mathbf{\hat{R}}'\mathbf{\hat{R}} = \mathbf{\hat{L}\hat{D}\hat{L}}'$. Restriction (34) represents the restriction on the structural capacity of the farm, while the relation (35) 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 farm 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 firm choices, i.e. Q.

## 4.1 Empirical analysis

The methodology presented in this section is applied to a sample of farms belonging to the Emilia-Romagna region. The sample is composed by 35 farms placed in the region plain and specialized in arable crop productions. The sample is extracted by the Italian FADN for the year 2005. The national database contains more information than the European archives. Indeed, it has been possible to get information about specific costs used in this analysis as comparison term for the estimated information.

The information used concerns the crop area of each farm, the production level, the price information and the related subsidies, the total variable cost at farm level. The sample presents a production set of seven crops: sugarbeet, durum wheat, soft wheat, maize, processed tomato, rice and soya. The main characteristics are showed by the table 1.

**Table 1: Characteristics of the sample**

| Main information | |
|---|---|
| Number of farms | 35 |
| Incidence of sugarbeet (in %) | 11 |
| Incidence of durum wheat (in %) | 10 |
| Incidence of soft wheat (in %) | 21 |
| Incidence of maize (in %) | 18 |
| Incidence of processed tomato (in %) | 13 |
| Incidence of rice (in %) | 14 |
| Incidence of soya (in %) | 13 |
| Average AAU (ha) | 175 |
| Revenue by ha (in euros) | 1,565 |

The high level of the average land per farm reveals the presence of a number of farms with large dimension: two farm out of 35 show a total land over than 1.500 ha. These farms are placed in the area of Ferrara province, where the level of specialization in arable crops is particularly strong. The aims of this analysis is to estimate the level of specific variable costs for the crops associated to this group of farms and compare it with the real specific costs collected by the national archive.

The table 2 shows the results achieved by the model previously specified. These results should be view as preliminary results of a methodology that will be submitted to other tests using higher quantity of observations. For each activities, the average specific costs has been estimated and compared with the observed average cost. The estimation errors are very small for all the crops, but maize presents an estimation error high (31%) with respect the observed value. It is interesting to remark that the standards errors calculated for the estimated specific cost are very similar to those obtained for the observed values.

**Table 2: PMP estimation outcomes**

| Activities | Specific costs | | Standard errors | |
|---|---|---|---|---|
| | Observed | Estimated | In the observed information | In the estimated results |
| Sugarbeet | 0.01890 | 0.02130 | 0.03 | 0.04 |
| Durum wheat | 0.06531 | 0.07122 | 0.08 | 0.14 |
| Soft wheat | 0.06366 | 0.06290 | 0.08 | 0.13 |
| Maize | 0.08569 | 0.05895 | 0.11 | 0.07 |
| Processed Tomato | 0.03576 | 0.03726 | 0.02 | 0.02 |
| Rice | 0.19341 | 0.20527 | 0.56 | 0.58 |
| Soya | 0.09726 | 0.09224 | 0.20 | 0.20 |

It is important to highlight that the information presented below concerns the part of marginal costs that one can consider explicit, in an accounting sense. Indeed, this kind of information is related to the variable $c$ of the problem (24)-(32), that cost that is component of the total farm variable costs included inside the FADN data. The other marginal cost, $\lambda$, that in the same sense of the PMP methodology, represents the differential marginal cost. This latter added to the accounting cost, $c$, provides the total marginal cost. The model seems to reproduce quite precisely the observed accounting marginal cost of the sample's crops, with a little deviation from the real values for the great part of the information.

# References

Arfini F., Donati M. (2008), Health Check and Farm Efficiency: a Comparative Assessment of Four European Agricultural Regions, contributed paper of 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, contribute 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.

Golan A., Judge G. and Miller D. (1996), Maximum Entropy Econometrics: Robust Estimation With Limited Data, John Wiley, New York.

Hecklei T. (1997), Positive Mathematical Programming: Review of the standard approach, *CAPRI Working Paper*, University of Bonn.

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

Heckelei T. and Britz W. (2000), Positive mathematical programming with multiple data points: a cross-sectional estimation procedure, *Cahiers d'économie et sociologie rurales*, n. 57, 2000, pp. 27-50.

Heckelei T. and Wolff H. (2003), Estimation of constrained optimization models for agricultural supply analysis based on generalized maximum entropy, *European review of agricultural economics*, Vol. 30, pp. 27-50.

Howitt R.E. (1995), Positive Mathematical Programming, in *American Journal of Agriculture Economics* , Vol. 77, pp. 329-342.

Jaynes E.T. (1957), Information Theory and Statistical Mechanics II, *Phisics Review*, Vol.108, pp. 171-190.

Paris Q. (2001), Symmetric positive equilibrium problem: a framework for rationalizing the economic behaviour with limited information, *American journal of agricultural economics*, Vol. 83(4), pp.1049-1061.

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.

*Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture*
www.ekon.slu.se/facepa