Multi-input Multi-output Farm-level Cost Function: A Comparison of Least Squares and Entropy Estimators

P. Polomé, B. Fernagut, O. Harmignie, and B. Henry de Frahan

Corresponding author: Philippe Polomé, Université catholique de Louvain, ECRU, Place Croix du Sud, 2 - bte 15, 1348 Louvain-La-Neuve, Belgium, polome@ecru.ucl.ac.be

P. Polomé, O. Harmignie and B. Henry de Frahan are with the Rural Economics Unit (ECRU) from the Université catholique de Louvain, Belgium

B. Fernagut is with the Centre for Agricultural Economics, Brussels, Belgium

Paper prepared for presentation at the 11th Congress of the EAAE
(European Association of Agricultural Economists),
The Future of Rural Europe in the Global Agri-Food System,
Copenhagen, Denmark, August 24-27, 2005

Multi-input Multi-output Farm-level Cost Function: A Comparison of Least Squares and Entropy Estimators

Draft version last modified 14/07/2005

Abstract

We introduce a modification of the quadratic-Leontieff multi-output cost function that is particularly suitable for the data of the Farm Accountancy Data Network. We present least squares and entropy estimates of that function and compare their results for a sample of crop farms. Our results are encouraging for the use of entropy estimators in cases in which farms are not assumed to share the same technology. Our approach can be seen as an extension of the Positive Mathematical Programming approach (Howitt, 1995). The extension consists in an explicit specification of inputs in the cost function and in the possibility of modeling several farms simultaneously.

Keywords: Cost function, Least squares estimator, Entropy estimator, Heterogeneity

JEL classification:

C3 Econometric Methods: Multiple/Simultaneous Equation Models

C61 Mathematical Methods and Programming: Optimization Techniques • Programming Models • Dynamic Analysis

Q12 Micro Analysis of Farm Firms, Farm Households, and Farm Input Markets

Multi-input Multi-output Farm-level Cost Functions: A Comparison of Least Squares and Entropy Estimators

1. Introduction

In a recent issue of the *American Journal of Agricultural Economics*, Paris (2001) suggested the Symmetric Positive Equilibrium Problem (SPEP) generalization of the well-known Positive Mathematical Programming (PMP) developed by Howitt (1995) and Paris and Howitt (1998). The main advantages of SPEP with respect to PMP are to treat inputs explicitly and to model a sample of farms instead of the typical implicit-input one-farm PMP. SPEP is criticized by Britz et al. (2003). These critiques are detailed later on. A first objective of this paper is to address some of them and suggest an alternative extension of PMP to several farms and to an explicit specification of the inputs.

PMP is essentially a method to calibrate a multi-output cost function with implicit inputs. Generalized Maximum Entropy (GME) is an econometric technique that is used in conjunction with PMP to estimate the parameters of the cost function on the basis of very few observations – even as little as one farm can be used to estimate the cost function. The second objective of this paper is to compare the results of the GME estimation technique with the more conventional multi-equations Ordinary Least Squares (OLS) estimator.

The paper is intended for the same type of data as SPEP is: output prices and quantities, technical coefficients (input use per unit of output) and input prices data, either at a detailed farm-level sample, or an aggregated sample (for example at regional level). These data correspond to the European Union Farm Accountancy Data Network (FADN) data. The third objective of this paper is to present a multi-output cost function that fully accommodates for that type of data. It is a modification of a quadratic cost function, similar to Paris and Howitt's (2001) quadratic-Leontieff cost function, but modified to fully satisfy the regularity properties of a cost function. It also has an additional "decomposition" property that is desirable with FADN data. The framework that is presented in this paper can be used with as little as one farm, but can accommodate any number of farms.

We apply the method to the year 2000 FADN sample of 37 farms from the crops sector of the Belgian Walloon Loam region. The cost function is estimated imposing a common technology to the whole sample by OLS and GME. These estimates are then compared to an average of farm-specific estimates using GME, for which the common technology restriction is relaxed.

The next section reviews the PMP and SPEP methodologies and underlines their advantages and shortcomings. Section 3 specifies the cost function and its properties. Section 4 introduces the OLS and GME estimators. Section 5 presents the data and section 6 the results. Section 7 discusses the model and the results; only limited conclusions are presented at this stage of development of the paper as the estimation process can be improved in several clearly marked ways.

2. The Positive Mathematical Programming and alternative methodologies

PMP (Howitt, 1995) is primarily devised as a calibration method in low information conditions. The purpose is to estimate a cost function that is capable of reproducing exactly the behavior of a single farm given information on output prices and quantities, and on yields in terms of land and possibly other inputs. PMP has three phases: calibration, estimation and simulation. The objective of the calibration phase is to estimate the marginal cost of each output, but as pointed out by Heckelei and Wolff (2003), this phase can be integrated in the estimation phase by means of Lagrangean multipliers. Also if there is no constraint on input quantity, the marginal cost is equal to the output price; that is what is assumed in this paper. We will return on this point later. The estimation phase is concerned with the method of estimation of the coefficients of a special type of cost function. The cost function is special because its only arguments are output quantities while inputs are not modeled explicitly. Howitt (1995) argues that such a function can in fact capture the decreasing marginal productivity of input provided the matrix of second derivatives (with respect to output quantities) of the function is positive semi-definite. The simulation phase uses the estimated cost function in a profit maximization program to simulate the farm's behavior when some condition changes, such as output price or yield. PMP calibrates automatically, that is, in the simulation phase, given the original output prices, the original output quantities are recovered exactly. Several papers have demonstrated the usefulness of the approach to model agricultural supply, e.g. Röhm and Dabber (2003). One of the key assumptions of PMP is that farm behavior is optimum at the observed data point. Another key assumption in PMP, implemented through the GME estimator (Paris and Howitt, 1998) is that much information on the farm behavior is already contained in a single data point (a single farm), provided the researcher can "guide" the estimation by means of prior information.

Paris (2001) suggested SPEP as a multiple farms extension of PMP that treats inputs explicitly. SPEP has the same three phases as PMP but resorts to a quadratic cost function that is explicit in input prices. As demonstrated by Britz et al., (2003), SPEP has several drawbacks: the simulation phase has no straightforward economic interpretation and the quadratic cost function does not have the regularity properties that it should (see e.g. Chambers, 1988). Furthermore SPEP is impractical: it is devised as a complementary slackness (or equilibrium) problem instead of a profit maximizing problem. That means that each constraint in the problem must have a dual counterpart; that makes introduction of additional, in particular non-economic, constraints difficult.

In a more econometric register, agricultural supply has often been analyzed using a production or profit function approach. Notable and recent examples are Moro and Sckokai (1999) and Chambers and Just (1989). In this paper, we are interested in a cost function approach essentially because it is better suited to capture the FADN data: prices can be seen as marginal cost and input demand (derived through Sheppard's lemma) can be decomposed per output. The latter is a special property of the cost function presented in this paper. An additional advantage of the cost function, comparatively to the production function or the profit function, is the possibility to include constraints on quantities directly in the estimation process as shown in Heckelei and Wolff (2003).

Several functional forms are available in the literature, but none is very much suited to the FADN data. The well-known Translog cost function can be extended straightforwardly to the multi-output case. However in multi-farm agricultural data most farms do not produce all the products. The Translog cannot accommodate zero output quantities because it is expressed in logarithm of output quantities. The Symmetric Generalized McFadden cost function introduced by Kumbhakar (1994) is similar to a quadratic cost function. Without modifications, quadratic cost functions do not satisfy the standard regularity conditions (Chambers, 1988) because (among other things) they are not directly linearly homogenous in input prices. That problem has been addressed in the Symmetric Generalized McFadden cost function by dividing the quadratic form in input prices by the sum of input prices. Although apparently a simple and attractive solution, the derivatives become messy and guaranteeing the concavity in input prices is more complicated. In the modified quadratic-Leontieff cost function that is presented in the next section, that problem is addressed by using the square root of the input prices in the quadratic term as suggested by Paris and Howitt (2001). That cost function has however useful additional properties; they will become apparent in the next section. Finally the Hildreth-Houck functional form as used in Peeters and Surry (2003) does not make use of all the information available in a FADN sample.

3. Specification of the cost function

The modified quadratic-Leontieff cost function can be stated as follows:

$$C_n\left(X_n,R_n\right)=\left(\sum_j f_{jn}X_{jn}+\sum_{j'}X_{j'n}\sum_j q_{j'j}X_{jn}/2\right)\sum_i R_{in}$$

$$+\left(\sum_i g_{in}R_{in}+\sum_{i'}\sqrt{R_{i'n}}\sum_i s_{i'i}\sqrt{R_{in}}\right)\sum_j X_{jn} \qquad \textbf{(1)}$$

$$+\sum_j X_{jn}\sum_i d_{ji}R_{in}+\left(\sum_j \varepsilon_{jn}X_{jn}\right)\left(\sum_i \mu_{in}R_{in}\right)$$

The indexes are *j* for the J outputs, *i* for the I inputs, and *n* for the N farms in the sample. The lower-case Latin letters *f, q, g, s* and *d* represent unknown coefficients. The *q* and *s* coefficients must be symmetric: $q_{jj'}=q_{j'j}$ and $s_{ii'}=s_{i'i}$, otherwise the same variable would have two distinct coefficients; for example the variable $X_{jn}X_{j'n}$ would have two coefficients $q_{jj'}$ and $q_{j'j}$. The upper-case Latin letters represents data: *X* is for outputs, *R* for prices. The terms in ε and μ are stochastic and determine the stochastic structure of the functions to estimate (see below); they may represent unobservable variables or measurement errors. Similarly to PMP, two sets of equations are derived from the cost function; both will be used in the estimation of the coefficients of the cost function.

The first is the set of marginal cost equations:

$$mc_{jn}=\sum_{i=1}^{I}\left(f_{jn}+\sum_{j'=1}^{J}q_{jj'}X_{j'n}+g_{in}+\sum_{i'=1}^{I}s_{ii'}\sqrt{R_{i'n}}\Big/\sqrt{R_{in}}+d_{ji}+\varepsilon_{jn}\mu_{in}\right)R_{in}\geq P_{jn}. \qquad \textbf{(2)}$$

The term *mc* stands for marginal cost, while *P* represents the output prices. One of the “positive” ideas of PMP is that a farmer has produced up to the point where marginal cost equals the marginal return. Provided there are no constraints on input availability and perfect expectation on prices, the marginal return is the output price. For each output, each farm fetches a different unit price than the other farms. This is not surprising as under the name of one output, each farm produces in fact a somewhat different product: the quality or the variety may be different, or some farms may have specific selling contracts. Farms sometimes produce rather heterogeneous crops that are small in terms of the farm revenue, but their nature varies widely across farms. Some farms may produce a small amount of one such crop while it is fetching a price much higher than its neighbor who produces a large quantity of a product with an identical name. In the Belgian agriculture, some bean or potato crops are typical examples. As a result, we prefer excluding those crops from the sample. For the same reason, in the present context, it is advisable to remove small crops from the sample rather than to lump them together.

The second set of equations comes from the total input use *B* over the farm, derived through Shephard’s lemma:

$$B_{in}=\sum_{j=1}^{J}\left(f_{jn}+\tfrac{1}{2}\sum_{j'=1}^{J}q_{jj'}X_{j'n}+g_{in}+\sum_{i'=1}^{I}s_{ii'}\sqrt{R_{i'n}}\Big/\sqrt{R_{in}}+d_{ji}+\varepsilon_{jn}\mu_{in}\right)X_{jn}. \qquad \textbf{(3)}$$

Shephard’s lemma also has a positive interpretation: in the same way as the farmer increases production until marginal cost equals output price (or does not produce), the farmer buys inputs until his marginal willingness to pay equals the input price (or does not buy at all). Examination of the input use function shows that it can be decomposed into a set of *I* equations corresponding to the technical coefficients A_{ijn} (input use per unit of output, i. e. inverses of yield):

$$B_{in} = \sum_{j=1}^{J} A_{ijn} X_{jn} \text{, thus}$$

$$A_{ijn} = f_{jn} + \tfrac{1}{2} \sum_{j'=1}^{J} q_{jj'} X_{j'n} + g_{in} + \sum_{i'=1}^{I} s_{ii'} \sqrt{R_{i'n}} \Big/ \sqrt{R_{in}} + d_{ji} + \varepsilon_{jn} \mu_{in} . \qquad \textbf{(4)}$$

This is the "decomposition" property of $C(X,R)$. It is very attractive because an analytical expression for A_{ijn} makes it possible to incorporate a lot of technological information into the cost function since A_{ijn} is a data available in the FADN for some inputs such as land and fertilizers. This should help to model better diversified farms. For some inputs, the allocation per output is not known; this is the case most notably for labor and capital. For those inputs, we can still resort to equation **(3).** As summarized in the next table, the function defined by **(1)** complies with all the standard properties of a cost function.

Table 1. Properties of the modified quadratic-Leontieff cost function

Property	Completion of the property
Continuous in *R*	$C(X,R)$ is a sum of terms in *R* and $\sqrt{R}$
Non decreasing in *R*: $\partial C/\partial R \geq 0$	$\partial C/\partial R = B$: observed input demands are positive by construction
Concave in *R*	The matrix of second derivatives of $C(X,R)$ with respect to input prices is negative semi-definite as long as $s_{ii'} \geq 0$ for $i \neq i'$
Homogeneity of degree 1 in *R*	$C(X,\lambda R) = \lambda C(X,R) \qquad \forall \lambda > 0$
Non decreasing in *X*: $\partial C/\partial X \geq 0$	$\partial C/\partial X \geq P$: observed marginal costs are no smaller than observed prices
No fixed costs	$C(0,R) = 0$
Non negativity: $C(X,R) > 0$ for $X > 0$ and $R > 0$	$C(0,0) = 0$, *C* is non-decreasing both in *X* and in *R*

In contrast to PMP, $C(X,R)$ is not required to be convex in output *X* (the matrix of second derivative with respect to *X* is not required to be positive semi-definite) because concavity in input prices implies convexity in output quantities by duality. From an estimation point of view, imposing convexity in output requires resorting to some form of Cholesky decomposition (see e.g. Paris and Howitt, 1998), while imposing concavity in input prices only requires imposing sign constraints as indicated in the above table.

There are at most (J + JI) equations per farm, but when a farm does not supply an output, all the equations corresponding to that output (one marginal cost equations and *I* technical coefficient equations) are omitted from the estimation process. If simulation is the objective, it is however necessary to estimate the coefficients that are specific to these equations. Some hypotheses have to be taken regarding the level of price P and technical coefficients A_{ijn} that a farm would have obtained if it had supplied that output. Paris (2001) suggests taking the sample mean.

Over a sample of N farms, there are NJ+J(J+1)/2+NI+I(I+1)/2+IJ coefficients to estimate including the restrictions of symmetry. Again, some coefficients may be skipped if some output is not supplied. In the sample of the present paper, a typical farm supplies 6 outputs and uses 4 inputs. In that case, the model is identified with as little as three farms, provided they all produce all the outputs with all the inputs. Because there are as many coefficients to estimate as first and second-order derivatives, $C\left(X,R\right)$ is a flexible functional form.

The model is similar to a panel data model in the sense that for each output and input there is a coefficient f_{jn} and g_{in} respectively that is farm-specific. Those coefficients are intended to capture the effect of relevant explanatory variables for which we do not have data, e.g., management capacity or some forms of capital.

4. Ordinary Least Squares and Generalized Maximum Entropy Estimators

Both PMP and the Generalized Maximum Entropy (GME) estimator were developed for cases of shortage of data. The typical PMP problem is under-identified in the sense that there are more coefficient to estimate than there are equations, therefore only an entropy estimator may be applied (the robustness of such estimator is however little explored). Extending PMP to make use of information on input prices as in SPEP and in the present paper requires more data than in the original PMP, but only the input prices, which are included in the FADN data set. Access to the FADN implies access to a sample of several hundreds or thousands of farm accounting data in each EU member state. Therefore, it seems that the sample size is large enough to use more conventional techniques, such as least squares, and there is little need for the GME estimator anymore.

However, sample size is a concern in a different sense: the larger the sample size, the more heterogeneous the sample, the less likely the farms share a similar technology. In other words, with a small homogenous sample, we may expect that there indeed exists common parameters (the *q*, *s* and *d* coefficients) summarizing a technology that is similar across the farms in the sample. But as the geographical extent of the sample increases, that hypothesis becomes more and more restrictive. The presence of error terms ensures that estimation is always feasible, thus the investigator must be relatively confident about the homogeneity of the sample because estimation does not reveal a lack of homogeneity. Otherwise, one must resort to estimating one set of coefficients per farm. In the above sets of equations **(2)** and **(4)**, that amounts to appending an index *n* to the *q*, *s* and *d* coefficients.

When one is willing to impose a common technology restriction, that is, imposing that the *q*, *s* and *d* coefficients are the same over the sample, the system of marginal cost and technical coefficients equations **(2)** and **(4)** can be estimated using Ordinary Least Squares (OLS). When one is not willing to make that hypothesis, a GME technique can be used as described by Golan et al., (1996). In this section we present an application of the GME and OLS estimators for a farm-specific estimation and in section 6 we compare their results.

Both the OLS and the GME estimators can be described as an optimization process conditional on the following three sets of constraints corresponding to equation sets **(2)**, **(4)** and to the negative semi-definiteness of the matrix of second derivatives with respect to prices of the cost function:

- If $Xr_{jn} > 0$: $\quad P_{jn} = \sum_{i=1}^{I}\left(\hat{f}_{jn} + \sum_{j'=1}^{J} \hat{q}_{jj'} X_{j'n} + \hat{g}_{in} + \sum_{i'=1}^{I} \hat{s}_{ii'} \sqrt{R_{i'n}} \Big/ \sqrt{R_{in}} + \hat{d}_{ji} \right) R_{in} + \hat{\varepsilon}_{jn}$
- If input i is used by farm n for output j:

$$A_{ijn} = \hat{f}_{jn} + \tfrac{1}{2} \sum_{j'=1}^{J} \hat{q}_{jj'} X_{j'n} + \hat{g}_{in} + \sum_{i'=1}^{I} \hat{s}_{ii'} \sqrt{R_{i'n}} \Big/ \sqrt{R_{in}} + \hat{d}_{ji} + \hat{\mu}_{ijn}$$

- $\hat{s}_{ii'} \geq 0,\ i \neq i'$.

The residuals $\hat{\varepsilon}_{jn}$ and $\hat{\mu}_{ijn}$ correspond to a reparametrization of the error terms in the sets of equations **(2)** and **(4)** respectively. The set of OLS coefficient estimates is defined as the set $\left\{ \hat{f}_{LS}, \hat{q}_{LS}, \hat{g}_{LS}, \hat{s}_{LS}, \hat{d}_{LS}, \hat{\varepsilon}_{LS}, \hat{\mu}_{LS} \right\}$ solution to $\min_{\hat{f},\hat{q},\hat{g},\hat{s},\hat{d},\hat{\varepsilon},\hat{\mu}} \sum_{n}\sum_{j}\left(\hat{\varepsilon}_{jn}^{2} + \sum_{i} \hat{\mu}_{jin}^{2} \right)$. It is stated here to underline the similarity with the GME estimator.

The set of GME coefficient estimates is defined as the set $\left\{ \hat{f}_{ME}, \hat{q}_{ME}, \hat{g}_{ME}, \hat{s}_{ME}, \hat{d}_{ME}, \hat{\varepsilon}_{ME}, \hat{\mu}_{ME}, \hat{p} \right\}$ solution to $\min_{\hat{p}} \sum_{k}\sum_{n}\sum_{j}\sum_{i} \hat{p}_{jink} \log\left(\hat{p}_{jink}\right)$ such that $\hat{\gamma}_{jin} = \sum_{k} \hat{p}_{jink} z_{jink}$ where γ represents any of the *f*, *q*, g, *s*, or *d* coefficients or of the ε or μ errors (adjusting the indexes where necessary). The *k* estimated probabilities $\hat{p}_{jink}$ must sum to one (over *k*) and must be non-negative. The parameter z_{jink} is the k[th] support point for coefficient γ_{jin}. The restriction $\hat{\gamma}_{jin} = \sum_{k} \hat{p}_{jink} z_{jink}$ states that all the coefficients γ_{jin} of the model are defined as a convex combination of their support points z_{jink}. The estimated probabilities $\hat{p}_{jink}$ are the weights of this convex combination. The support points are therefore a way to incorporate prior information into the GME estimator. The support points have to be specified whether the investigator possesses such prior information or not; some support points can therefore be quite arbitrary. The support points are generally ordered from smallest to greatest; of particular interest are the first and the last support points since they constitute the lower and upper limits of the parameter estimate. To take an extreme example, if only one support point is specified, it is equivalent to imposing that the coefficient estimate is equal to the support point.

The interesting feature of entropy maximization from an econometric point of view is that it makes possible estimation of under-identified systems of equations. We say that a system is identified when it has more equations than coefficients to estimate. Otherwise, it is under-identified and can only be estimated by GME. That is the case when each farm has its own sets of coefficients, not only the f_{jn} and g_{in} coefficients, but also the *q*, *s* and *d* coefficients. Only the GME estimator can be applied in under-identified systems. When the common technology restriction is not imposed, the GME estimator described above is equivalent to farm-by-farm estimation over the sample (i.e., there are *N* independent estimation processes). In identified systems, the only reason to apply a GME estimator is to incorporate prior information on the possible values of the coefficients.

In the definitions of the OLS and GME estimators, no hypothesis has been made on the distribution of the error terms ε and μ or on the structure of their covariance matrix. Regarding the latter, it seems reasonable to assume that there is no correlation between farms but that all the error terms within one farm are correlated between them (E denotes the expectation operator):

$$E\varepsilon_{jn}\varepsilon_{j'n'} = \begin{cases} 0 \; if \; n \neq n' \\ \sigma_{jn}^2 \; if \; n = n' \end{cases}, \; E\mu_{ijn}\mu_{i'j'n'} = \begin{cases} 0 \; if \; n \neq n' \\ \tau_{ijn}^2 \; if \; n = n' \end{cases}, \text{ and } E\varepsilon_{jn}\mu_{ij'n'} = \begin{cases} 0 \; if \; n \neq n' \\ \upsilon_{ijn} \; if \; n = n' \end{cases}.$$

It seems also reasonable to assume an extended version of heteroskedasticity, that is, the whole covariance matrix of the errors is farm-specific, i.e. all the above variances and covariances have an index *n*. These assumptions are rather weak and preclude the use of White or Newey-West robust estimator of the covariance matrix or feasible Generalized Least Squares techniques (see e.g. Mittelhammer et al., 2000). If we wanted to use a Generalized Least Squares estimator we would have to make stronger hypotheses on the covariance matrix of the errors that could lead to inconsistency if they were not true. Regarding the GME estimator, it is known that it is consistent (provided the support points are not misspecified) when the covariance matrix of the errors is unspecified. It is not clear whether GME would retain that property if we would specify an incorrect covariance matrix of the errors.

The drawback of such weak assumptions is that the conventional methods for inference is available. However, bootstrap methods can still be used. Furthermore, even if we had made stronger hypotheses on the covariance matrix of the errors, we would still have had to resort to bootstrap technique for calculating the confidence interval of the elasticity measures derived from this model. A final advantage of bootstrap methods is that no hypothesis needs to be made on the distribution of the errors, neither for the OLS nor for the GME estimator.

Finally, with both estimators, it is easy to incorporate additional restrictions to account for specific policy aspects. One can add equations describing the subsidy mechanism of Agenda 2000 such as in Moro and Sckokai (1999), or to impose that land use is fixed for each farm such as in Heckelei and Wolff (2003).

5. The data

The data that the model needs are the yields for each input per output (inverse of the A_{ij} coefficients), the output prices *P*, the output quantities *X*, and the input prices *R*. That data are available from the FADN sample at farm level, but the available information on input prices has to be supplemented for the needs of the model. The FADN sample does not have input prices directly, but has expenses per variable input; however, there is not always a measure of quantity. The land quantity per output is simply the acreage and is well documented, but depending on national regulations the FADN sample may report expenses based only on an official lease price for land which may differ widely from the actual price paid. For the present paper, each farm is assigned an estimated lease price for land based on (i) the share of leased land in total land for that farm, (ii) the official lease price for land (both data from the FADN sample), and (ii) the sale prices per district (from the Belgian National Institute for Statistics).

For fertilizers prices, the available information is the total expenses and the quantities of each N, P and K fertilizer. If one assumes that the prices of these fertilizers are the same across outputs inside a single farm, then it is possible to estimate their prices on the basis of a least square or entropy estimator. The FADN sample does not provide information on quantities phytosanitary, seeds, and hired services inputs; for each of these inputs, only the total expenses per output are included in the FADN sample. That sample also includes some data on labor, capital and equipment inputs. We have momentarily excluded those inputs from the cost function and we assume that their impacts are properly captured through the outputs and through the farm-specific coefficients f_{jn} and g_{in}. In other words, it is assumed that they always remain in fixed proportions to the included inputs. Future versions of this paper will include a more proper treatment those inputs, in particular through equation **(3)** for the inputs that are not allocated per output.

In this paper, we use the 37 crops farms of the year 2000 FADN sample of the Belgian Walloon Loam region. The sample information is summarized in the following two tables.

Table 2. Output descriptive statistics

Crop	Output X (T)	Average P (€)	Land (ha)	N (u)	P (u)	K (u)	Nbr farms
Winter Wheat	9 030	130.4	11 031	203 441	8 169	13 550	36
Winter Barley	1 088	124.1	1 507	24 900	5 412	8 043	17
Beet Sugar	31 557	53.0	4 647	67 390	35 640	85 520	36
Potato	9 725	63.8	1 940	32 665	14 571	44 612	13
Green Pea	228	260.0	287	81	2 200	2 475	3
Chicoree	8 884	48.4	1 942	10 907	8 339	23 731	20

Table 3. Input descriptive statistics

Input	Total B	Average R (€)	Nbr of farms that use input i in crop j					
			W. Wheat	W. Barley	Beet Sugar	Potato	Green Pea	Chicoree
Land	21 354 (ha)	0.292	36	17	36	13	3	20
Nytrogen	339 384 (u)	0.008	36	17	35	13	1	17
Potassium	74 331 (u)	0.028	9	8	27	12	1	14
Phosphor	177 931 (u)	0.042	9	8	31	12	1	15

6. Results

We consider that the sample described in the previous section is homogenous, that is, the 37 crops farms share the same technology. That hypothesis is motivated by agronomic and climatic considerations, the small size of the region (about 3000 km²) and its agricultural history (a long stretch of cereals and beet sugar cultivation). In that section, we compare an OLS estimator that imposes equality of the *q*, *s* and *d* coefficients over the sample with a GME estimator that does not impose such equality. The test is that since the sample is homogenous, the farm-specific GME estimates should not significantly differ from the OLS estimates. More explicitly, if the assumption of homogenous technology is imposed, then both OLS and GME should produce similar coefficient estimates. When the homogenous technology assumption is not imposed in GME, if the sample is indeed homogenous, the same estimates should still obtain. If they do not, then we cannot know whether the assumption is wrong or GME is unable of extracting the common elements from each farm technology.

However, contrarily to OLS, GME requires the specification of support points, and it is well-known that GME is consistent only if the support points are well-specified. For that reason, we present results with different choices of support points.

The results of the OLS estimation of $C\left(X,R\right)$ are presented in Table 4 under the form of elasticities. We have chosen this presentation because it appeared more intuitive than presenting the coefficient estimates directly. Table 4 presents the elasticity for a quantity (output supply or input use) on the left hand side column with respect to a price on the top row. The corresponding 95% confidence intervals are represented with the 2.5 and 97.5 percentiles (P 2.5 and P 97.5 respectively). Italics indicate significance at the 95% level. Percentiles have been computed by bootstrap (1000 replications); no distributional assumption has been made at any point of estimation or inference. The fourth row of each output indicates the number of times the (quantity – price) pair has been observed; for example, out of the 36 farms producing wheat, 16 also produce barley.

Table 4. OLS Elasticities

Quantity	Price	WW	WB	SB	Po	GP	Chi	Land	NF	PF	KF
W. Wheat	P 2.5	0.49	-1.99	-4.72	-2.06	-0.19	-0.50	-6.09	0.00	0.00	0.00
	Estim	***1.55***	**-0.10**	**-0.05**	**-0.28**	**0.00**	**0.51**	***-3.05***	**0.00**	**0.01**	**0.01**
	P 97.5	6.23	0.58	1.14	1.54	0.10	4.16	-2.39	0.00	0.02	0.02
	N	36	16	35	13	3	20	36	36	9	9
W. Barley	P 2.5	-0.57	0.96	-0.34	-0.24	-0.06	-0.14	-3.70	0.00	0.00	0.00
	Estim	**-0.03**	***1.40***	**-0.03**	**-0.09**	**-0.02**	**0.05**	***-3.10***	**0.00**	**0.00**	**0.00**
	P 97.5	0.15	2.03	0.53	0.15	0.03	0.19	-1.46	0.00	0.00	0.01
	N	16	17	17	4	2	6	17	17	8	8
Beet Sugar	P 2.5	-5.50	-1.10	0.59	-3.60	-0.13	-5.47	-5.66	0.00	0.00	-0.01
	Estim	**-0.05**	**-0.09**	***2.52***	**-0.75**	**0.02**	**-0.45**	***-3.54***	**0.00**	**0.01**	**0.01**
	P 97.5	1.18	1.63	9.31	0.81	0.27	0.55	-1.54	0.01	0.01	0.01
	N	35	17	36	12	3	19	36	35	27	31
Potato	P 2.5	-1.90	-1.67	-2.57	1.15	-0.06	-1.37	-11.23	0.00	0.00	0.00
	Estim	**-0.26**	**-0.54**	**-0.55**	***4.57***	**0.00**	**-0.48**	***-4.90***	**0.00**	**0.01**	**0.01**
	P 97.5	1.38	0.85	0.67	9.80	0.03	1.37	-1.19	0.01	0.03	0.03
	N	13	4	12	13	1	8	13	13	12	12
Green Pea	P 2.5	-0.09	-0.21	-0.04	-0.05	0.50	-0.10	-2.93	0.00	0.00	0.00
	Estim	**0.00**	**-0.08**	**0.01**	**0.00**	***0.58***	**-0.03**	***-2.29***	**0.00**	**0.00**	**0.00**
	P 97.5	0.05	0.11	0.09	0.02	0.71	0.01	-1.52	0.00	0.00	0.00
	N	3	2	3	1	3	2	3	1	1	1
Chicoree	P 2.5	-0.28	-0.56	-1.99	-0.76	-0.18	0.64	-8.32	0.00	0.00	0.00
	Estim	**0.27**	**0.17**	**-0.18**	**-0.23**	**-0.04**	***1.34***	***-3.37***	**0.00**	**0.00**	**0.01**
	P 97.5	2.13	0.71	0.21	0.74	0.02	3.90	-2.15	0.00	0.01	0.02
	N	20	6	19	8	2	20	20	17	14	15
Land	P 2.5	1.21	0.41	1.41	1.10	0.34	0.53	-14.98	0.001	0.01	0.01
	Estim	***1.82***	***0.61***	***1.87***	***1.92***	***0.38***	***0.68***	***-12.47***	***0.002***	***0.01***	***0.02***
	P 97.5	4.35	0.88	3.29	2.93	0.44	1.00	-10.87	0.01	0.03	0.03
	N	36	17	36	13	3	20	37	37	37	37
N Fertilizer	P 2.5	0.88	0.30	0.93	0.81	0.17	0.34	-11.64	-0.61	0.01	0.01
	Estim	***1.52***	***0.47***	***1.27***	***1.61***	***0.25***	***0.43***	***-9.24***	***-0.29***	***0.01***	***0.01***
	P 97.5	4.28	0.69	2.44	2.58	0.30	0.58	-7.94	-0.09	0.05	0.06
	N	36	17	36	13	3	20	37	37	37	37
P Fertilizer	P 2.5	2.42	0.73	3.55	2.29	0.30	0.99	-66.89	0.003	-8.04	0.02
	Estim	***3.66***	***1.49***	***5.63***	***4.55***	***3.44***	***1.83***	***-39.18***	***0.01***	***-3.39***	***0.04***
	P 97.5	5.25	2.51	9.23	7.06	6.57	3.38	-17.01	0.07	-1.60	0.15
	N	27	13	27	12	2	14	28	28	28	28
K Fertilizer	P 2.5	1.39	0.56	2.49	1.03	0.30	0.70	-111.37	0.002	0.01	-8.64
	Estim	***4.56***	***0.96***	***5.90***	***10.97***	***3.11***	***2.22***	***-50.75***	***0.01***	***0.05***	***-4.19***
	P 97.5	10.30	1.45	11.72	32.08	7.71	4.36	-12.74	0.08	0.16	-0.62
	N	31	16	31	12	3	17	32	32	32	32

From Table 4, it is apparent that output elasticities with respect to output prices are significantly different from zero only for own prices. All the own price elasticities of output are positive as expected. Elasticities of output with respect to input prices are significantly different from zero only for land prices, but not for fertilizers prices. Input demand elasticities on the other hand are significantly different from zero with respect to all the output or input prices. Elasticity of input with respect to output prices is positive for all inputs as expected. The elasticities of input with respect to own prices are negative and significant as expected. The cross elasticities of fertilizer quantities with respect to fertilizer prices are positive, significant and small. Elasticity of input with respect to land price is negative for all inputs. To understand the latter, interpretation in terms of cross-section are to be favored over in dynamic ones. In a dynamic setting, a positive elasticity would indicate that when price of land increases, at constant land productivity, it might be reasonable for the farmer to increase fertilizer intensity so as to substitute some land for fertilizer. In a cross-section sample, it simply means that the crops on the most expensive lands tend to be the ones with the least fertilizer intensity. The reason might be that those lands are not very productive while they are expensive because they are located near cities. Such a "cross-section interpretation" does not preclude the dynamic interpretation that could be inferred with time series data.

To compare the elasticities derived from the GME estimates with those of Table 4 derived from the OLS estimates, it is important to consider two aspects of GME estimation. First, the choice of support points and second, whether the system is identified or not. We suggest three choices of support points in each identification condition. Therefore, the elasticities derived from the OLS estimates can be compared with six sets of elasticities derived from the GME estimates. To specify a set of support points, we first specify its center and then its range.

Three choices of support points appear relevant. First, OLS-centered support points can be used when the researcher believes that although each farm has a specific technology (and thus specific coefficients), there are enough similarities across farms that OLS estimates can be seen as a kind of average. The range of the support points can be calculated using the "three sigma" rule recommended by Golan et al., (1996) where sigma is the estimated standard deviation of the coefficients estimates. The standard deviation has been estimated using bootstrap, as for the 95% confidence interval of the elasticities. In identified systems (imposing homogeneity), such a choice of support points should lead to GME estimates identical to OLS; in unidentified systems, the GME estimates may differ widely from the OLS ones.

A second choice for the support points is zero, corresponding to a prior belief that no regressor explains the variations of the dependent variable. For the range of those zero-centered support points, we have chosen an arbitrary [-1000; +1000], but the estimated elasticities are robust to a change in the range of a factor 100. It is important to realize that not only the range and the center of the support points intervals are important, but also the relative size of the range; therefore, multiplying the range by a factor of 100 may not change the estimates because it does not change the relative size of each coefficient support points range. Nevertheless, it corresponds to a case in which the investigator has no prior belief on the effect of the explanatory variables on the dependent variables.

A third possible choice of support points follows an idea indicated by Howitt (1995) that a priori each output is independent from the others and all the equations pass through the origin, that is, all the intercepts f_{jn} and g_{in}, the cross-product terms d_{ji} and the off-diagonal terms $q_{jj'}$ and $s_{ii'}$ for $j \neq j'$ and $i \neq i'$ are assumed to be zero. The support points ranges for these coefficients can be an arbitrary [-1000; +1000]. For the remaining coefficients, q_{jj} and s_{ii}, we substitute the zero values in the marginal cost and input demand equations **(2)** and **(3)** and we solve for q_{jj} and s_{ii} in terms of the observed outputs X and input prices R. We write these solutions $\tilde{q}_{jj}$ and $\tilde{s}_{ii}$; they can be taken as the centre of the support points interval, and the range can be assigned by adding or subtracting a multiple of their observed sample standard deviation to the support point center (e.g. the three sigma rule). In a farm-by-farm estimation process, each farm has then its own support points $\tilde{q}_{jjn}$ and $\tilde{s}_{iin}$; therefore, this choice of support points only applies to heterogeneous samples. To impose the assumption of homogenous sample, and therefore identical q, s and d coefficients across the sample, the support point centre can be taken equal to the sample average of the $\tilde{q}_{jjn}$ and $\tilde{s}_{iin}$.

To compare the elasticities derived from the GME estimates with those of Table 4 derived from the OLS estimates, we report the percentage of the former that fall within the 95% confidence interval of the latter. These percentages are reported in Table 5: each choice of support points is reported on the rows of Table 5. The columns indicate whether the homogeneity assumption has been imposed, that is, whether the system of the equations **(2)** and **(4)** is identified (left hand side column) or not (right hand side column). The results with the OLS-centered support points indicate that the sample is indeed homogeneous since elasticities derived from farm-specific *q*, *s* and *d* coefficients are fully (100%) compatible with elasticities derived imposing the same *q*, *s* and *d* coefficients for all the farms. Those results also show, as expected, that in these conditions the GME estimates coincide perfectly with the OLS ones. The homogeneity of the sample can also be seen with the results derived from data-based support points since the numbers of elasticities compatible with the OLS ones are quite similar between the identified and under-identified systems. These numbers are quite high (above 80%), indicating that the hypothesis of zero effect across outputs is not in fact very wrong. On the other hand, elasticities based on zero-centered support points deviates substantially from the OLS ones in the farm-by-farm estimation case, but not in the identified system case. That indicates that when the investigator has little prior information about the coefficients of the cost function (all support points set to zero), there is not enough information in the data in the under-identified case, and the GME estimator performs poorly.

Table 5. GME estimates: percent of elasticities within OLS 95% confidence interval

	Identified systems: Same *q*, *s* and *d* coefficients for all the farms	Under-identified systems: Farm-specific *q*, *s* and *d* coefficients
OLS-centered support points, support range +/- 3 sigma	100	100
Zero-centered support points, support range from +/- 10 to +/- 100 0000	98	42
Data-based support points, support range +/- 3 sigma	87	81

7. Partial Conclusions and Discussions

One interesting aspect of PMP is that inference may be drawn from limited information. Using a multi-output multi-input cost function, we extend the PMP approach to cases where some information on input prices is available. That cost function has a decomposition property that makes it particularly suitable for FADN data sets.

We also show how ordinary least squares can be used when the investigator has reasons to believe that a sample is sufficiently homogenous so that a common technology exists across several farms. We empirically test GME. In identified systems, results are identical or very close to the OLS ones. In under-identified systems (farm-specific coefficients), the choice of support points determines how close the estimated elasticities are from the OLS ones. With OLS-centered support points the elasticities are identical, as should be expected from a sample that is truly homogenous. With data-based support points, assuming that cross-terms are zero, the elasticities are still remarkably close to the OLS ones. With zero-centered support points, the elasticities deviate substantially from the OLS ones. Those results are an indicator of the sensitivity of GME to the support points: in identified systems, the information in the sample is sufficient to “correct” for misspecified support points, to some extent. That is not the case for under-identified systems, so that even if GME is the only available estimator, its results should be taken with caution. In other words, in heterogeneous samples where OLS estimates are meaningless, even if GME can indeed be used to estimate the coefficients of the cost function, the estimates may not approach their true value.

There remains a series of important problems with our approach. The first is that the information contained in one year may not be revealing a supply curve, but merely differences between farms reflecting some form of heterogeneity (output, input, management…). That problem can be solved by using several years of data instead of a single year, thus resorting to panel data methods. The FADN sample is a panel data, therefore the extension consists in applying the appropriate techniques to estimate the system of equations presented in section 3 of the present paper. Second, there is an issue about the scale of measurement of the variables used in estimation: when changing the units of any of the variables in the system of first order conditions described in section 3, the residuals of the corresponding equations also change scale. Since the least squares estimator consists in minimizing the sum of squares of these residuals, the estimated parameters will change in some non-linear way. The general solution is to resort to the generalized method of moments estimator because such an estimator is basically a weighting of variable devised to minimize the variance of the estimates. Whether this problem also affects the entropy estimator has to be explored. Third, the system of first order conditions suffers from endogeneity: prices and outputs are jointly determined, as well as yields and outputs. This problem can be solved by resorting to instrumental variable techniques. The Generalized Method of Moments is also an instrumental variables method. Fourth, it might be necessary to take uncertainty into account: when taking production decisions, the farmer does not know what prices or yields he will obtain. Resorting to panel data would let us link current output with last year prices, possibly solving at the same time the endogeneity problem. Finally, there are two issues on inputs. Primarily, inputs for which there are no data on the allocation per output, especially labor and some forms of capital, could be incorporated easily in the estimation process. Secondly, the possibility to integrate some constraint on input quantity, especially for family labor or some form of credit, should be examined. Integrating such constraints in the estimation process can be done as indicated in section 2.

References

Britz, W., Heckelei, T. and Wolff, H. (2003). Symmetric Positive Equilibrium Problem: A Framework For Rationalizing Economic Behavior With Limited Information: Comment. American Journal of Agricultural Economics 85: 1078–81.

Chambers, R. G. (1988). Applied Production Analysis: A Dual Approach. New-York: Cambridge University Press.

Chambers, R. G. and Just, R. E. (1989). Estimating multioutput technologies. American Journal of Agricultural Economics 71: 980-95.

Golan, A., Judge, G. and Miller, D. (1996). Maximum Entropy Econometrics. Chichester UK: Wiley.

Heckelei, T. and Wolff, H. (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy. European Review of Agricultural Economics 30: 27-50.

Howitt, R. E. (1995). Positive Mathematical Programming. American Journal of Agricultural Economics 77: 329-42.

Kumbhakar, S. C. (1994). A Multiproduct Symmetric Generalized McFadden Cost Function. Journal of Productivity Analysis 5: 349-57.

Mittelhammer, R. C., Judge, G. G. and Miller, D. J. (2000). Econometric Foundations. Cambridge: Cambridge University Press.

Moro, D. and Sckokai, P. (1999). Modelling the CAP Arable Crop Regime in Italy : Degree of Decoupling and Impact of Agenda 2000. Cahiers d'économie et sociologie rurales 53: 49-73.

Paris, Q. (2001). Symmetric Positive Equilibrium Problem: A Framework for Rationalizing Economic Behavior with Limited Information. American Journal of Agricultural Economics 83: 1049-61.

Paris, Q. and Howitt, R. E. (1998). An Analysis of Ill-posed Production Problems Using Maximum Entropy. American Journal of Agricultural Economics 80: 124-38.

Paris, Q. and Howitt, R. E. (2001). Symmetric Positive Equilibrium Problem. Agricultural Sector Modelling and Policy Information Systems, Proceedings of the 65th EAAE Seminar, March 29-31, Bonn. Kiel: Vauk Verlag.

Peeters, L. and Surry, Y. (2003). Maximum entropy estimation of a restricted Hildreth-Houck random-coefficients model. A Conference in Honor of Arnold Zellner: Recent Developments in the Theory, Method, and Application of Information and Entropy Econometrics: American University, Washington, D.C. Paper available at http://www.american.edu/academic.depts/cas/econ/.

Röhm, O. and Dabbert, S. (2003). Integrating Agri-Environmental Programs into Regional Production Models: An extension of positive mathematical programming. American Journal of Agricultural Economics 85: 254-65.