GENERALIZED CROSS-ENTROPY ESTIMATION OF A VARYING-COEFFICIENTS MODEL OF COST ALLOCATION IN MULTI-PRODUCT FARMING

LUDO PEETERS AND YVES SURRY*

Abstract

Standard farm-accounting information is typically restricted to aggregate input expenditures, without revealing the allocation of these expenditures to the various farm enterprises. Given the diversified enterprise base of most farm businesses, a proper assessment of each enterprise's contribution to overall farm performance is severely obscured. This paper presents a new approach for estimating the unobserved input-output or (preferably) cost-allocation coefficients for multi-product farms. Specifically, a model with (non-randomly) varying coefficients is developed, which also allows for a more realistic appraisal of the "real-world" diversity of farm operations. The coefficients of the model are estimated using the Generalized Cross-Entropy (GCE) method. The GCE method proves to be an effective way to obtain unique estimates of farm-specific cost-allocation coefficients and to overcome the many practical problems usually encountered in earlier empirical works. The proposed GCE estimator is applied to a set of 2000-2001 accounting data for a small sample of beef-dairy farms located in Brittany, France.

Keywords: cost allocation, farm heterogeneity, varying coefficients, generalized cross entropy, aggregate data, small samples

Affiliations

Ludo Peeters, Limburg University, Hasselt, Belgium
+32-11-26 86 35
ludo.peeters@luc.ac.be

Yves Surry, Swedish University of Agricultural Sciences, Uppsala, Sweden
+46-18-67 17 95
yves.surry@ekon.slu.se

* Earlier versions of this paper were presented at the Organized Symposium on "Maximum Entropy Approaches to Modeling Agricultural Diversity in European Agriculture", held during the 10th EAAE Congress, Zaragoza, August 27-28, 2002, and at the Information and Entropy Econometrics Conference in Honor of Arnold Zellner, Washington, DC, September 19-21, 2003. Part of the research was conducted during the Summer of 2002, when the second author was visiting the Economics Department at Laval University, Quebec City, Canada. Partial funding of this research was provided by the "Breizh" Research Program, financed by the "Conseil Régional de Bretagne". The authors would like to thank Mr. Desbois, the editor, and three anonymous referees for their useful comments. The usual disclaimer applies.

1. Introduction

In spite of the trend towards increased specialization which has characterized farming in many European countries, most farms still have more than one enterprise. Yet, standard farm-accounting information is typically restricted to aggregate or whole-farm input expenditures, without revealing production costs per unit of each enterprise's output. Obtaining information on the per-unit cost of production for the individual enterprises, measured by so-called input-output or (preferably) cost-allocation coefficients, is particularly important, both from a business-management and agricultural-policy perspective. Specifically, individual farmers may need this kind of information for optimizing their production, for deciding whether to expand or contract particular enterprises or to establish a new (discontinue an existing) enterprise. Moreover, information on enterprise-level costs of production may be helpful in preparing activity budgets, planning yearly operations, applying for operational loans, and analyzing alternative marketing strategies (Hornbaker *et al.*, 1989). Also, policy-makers are showing a renewed interest in acquiring information on the costs of production, as it may considerably improve their capability of properly assessing the impacts of different Common Agricultural Policy scenarios (for example, the decoupling and the Single Farm Payment scheme brought in by the June 2003 CAP reform) on the economic performances and the competitiveness of different types of farms in the Member States and regions of the recently enlarged European Union (e.g., Pingault and Desbois, 2003).[1] Considering that the direct collection of enterprise-level information is difficult and requires costly farm surveys, econometric analysis may offer an attractive alternative for obtaining reliable, yearly updated estimates of farmers' production costs at an appreciably lower cost.

Although the cost-allocation problem is fairly simple, at least conceptually, recovering reliable enterprise-level information from whole-farm accounting data is not straightforward. Despite the fact that the issue continues to puzzle (at least part of) the agricultural-economics profession, hardly any new development in this area have been observed ever since a number of studies in the 1980s and early 1990s reported on several fruitless attempts to produce truly satisfactory or theoretically-consistent results. Most of these earlier studies involved multiple regressions, using a wide range of estimation techniques (e.g., Ordinary Least Squares, Generalized Least Squares, Inequality-Restricted GLS, Bayesian methods, etc.) and assumed fixed coefficients (e.g., Errington, 1989; Midmore, 1990; Moxey and Tiffin, 1994; Hallam *et al.*, 1999). However, the assumption of fixed coefficients implies that the cost allocations are for a notional "average" farm, expressing some central tendency of farm operations. As a result, variations in the cost (technology) structures across farms are ruled out *a priori*. Clearly, such assumptions are overly restrictive and preclude any realistic appraisal of the "real-world" diversity of farm operations. In addition, it is a well-known result (e.g., Judge *et al*., 1985) that using fixed coefficients may, under certain conditions, yield biased estimators, due to the unobserved effects of omitted variables (e.g., different managerial capacities of the farmers), aggregation (e.g., different compositions of the input and output categories used), and "measurement errors" (e.g., quality differences). Only a handful of papers have explicitly addressed the random nature of the cost-allocation coefficients (Dixon *et al*., 1984; Butault *et al*., 1988; Hornbaker *et al*., 1989). But, regardless of whether the coefficients are assumed fixed or random, a common observation is that the existing models usually show poor prediction accuracy and/or produce unacceptable and/or implausible results (e.g., zero or negative values of the coefficients, violation of the cost-accounting restrictions, negative or unrealistically small variances of the random coefficients, etc.).

The objective of this paper is to present a new methodology for estimating cost-allocation coefficients, which deviates from "traditional" approaches in two major respects. Firstly, we develop a model with (non-randomly) *varying coefficients*. This model should allow us to capture the unobserved farm heterogeneity and to provide likely ranges for the values of the cost-allocation coefficients. Secondly, we show how these farm-varying coefficients, while satisfying all relevant equality and inequality restrictions prescribed by basic cost-accounting rules, can be estimated by using the *Generalized Cross-Entropy* (GCE) method described in Golan *et al*. (1996), among others.

The proposed GCE estimator is particularly useful for our purposes. At least four distinguishing features of the GCE methodology will be exploited in this paper: (a) GCE is "immune" to degrees-of-freedom shortages, which makes it particularly suitable for handling ill-posed (underdetermined) estimation problems, where the number of unknown coefficients is larger than the number of observations; (b) GCE allows to impose equality and inequality restrictions on the coefficients in a flexible and straightforward manner; (c) GCE employs a formal mechanism for incorporating prior (pre-sample) information, thereby alleviating the deleterious effects of possibly ill-conditioned (co-linear) data; and (d) GCE computations are not very demanding and are relatively easy to implement. Hence, by applying GCE, most of the practical and methodological problems encountered in previous empirical studies can be overcome.

Yet, despite the many advantages of GCE, only a few researchers so far have adopted an entropy-based methodology for estimating models of cost allocation (e.g., Léon *et al*., 1999) or multi-output technologies (e.g., Lence and Miller, 1998a, 1998b; Oude Lansink, 1999) using aggregate input data. However, in all these instances the varying nature of the parameters is ignored, while parameter restrictions prescribed by either cost-accounting rules or economic theory are imposed only at a single (base) point. [2]

The remainder of this paper is divided into four sections. Section 2 presents the varying-coefficients model of cost-allocation, including the relevant (in)equality restrictions that need to be taken into account. Section 3 outlines the GCE formulation of the cost-allocation problem. Section 4 summarizes the empirical results of the GCE estimation, and provides a brief account of the validity of the model based on "post-estimation" analysis. Section 5 presents the conclusions and raises some issues for further research.

2. Varying-coefficients model of cost allocation

Assuming N inputs ($i = 1,\ldots,N$), K outputs ($k = 1,\ldots,K$), and T farms ($t = 1,\ldots,T$), we consider the following varying-coefficients model of cost-allocation, embedded in a standard stochastic framework:

$$x_{it} = \sum_{k=1}^{K} \beta_{ikt} y_{kt} + u_{it} \quad (1)$$

where x_{it} is the aggregate expenditure on input i by farm t, y_{kt} is the value of output k of farm t, β_{ikt} are farm-specific (varying) coefficients representing the unobserved expenditure by farm t on input i, to produce €1.00 worth of output k, and u_{it} is the usual disturbance term, which has zero mean and is identically and independently distributed across farms.

In estimating the coefficients of model (1), several methodological problems arise. The first difficulty is the problem of identification. Obviously, because the number of coefficients to be estimated (NKT) is greater than the number of observations (NT), the problem is ill-posed or underdetermined. A second complication relates to the fact that several equality (linear) and inequality restrictions must be imposed on the coefficients. Since all the variables are measured in monetary terms, the accounting restrictions dictate that the whole-farm input expenditures are identically equal to the whole-farm total value of output; that is, $\Sigma_i x_{it} = \Sigma_i \Sigma_k \beta_{ikt} y_{kt} = \Sigma_k (\Sigma_i \beta_{ikt}) y_{kt} = \Sigma_k y_{kt}$, and hence, $\Sigma_i u_{it} = 0$. By implication, the (column-wise) sums of all the enterprise-level coefficients, for each farm, are equal to one; that is,

$\Sigma_i \beta_{ikt} = 1 \ \forall k,t$. Also, if the dependent variable in (1) is non-negative, the coefficients should be non-negative as well; that is, $\beta_{ikt} \geq 0 \ \ \forall i,k,t$ if $x_{it} \geq 0$. Finally, the accounting restrictions imply that the "input-demand" equations are interdependent, which requires a system-of-equations approach.

In practice, however, most researchers would typically "solve" the identification problem by using a linear-in-parameters model with *fixed* coefficients – which can be viewed as a special case of (1), where $\beta_{ikt} = \overline{\beta}_{ik}$ (thereby reducing the number of unknown coefficients from *NKT* to *NK* < *T*), and would ignore the various restrictions on the coefficients and the accounting coherency of the input-demand equations altogether.

2.1 Randomly varying coefficients

An alternative way to overcome the identification problem, while still accounting for variations in the coefficients, can be found in using a random-coefficients modelling framework, whereby it is assumed that the farm-specific coefficients are drawn randomly from a particular (joint) distribution. Adopting a standard random-coefficients specification (Hildreth-Houck, 1968), the model in (1) can be reformulated as

$$x_{it} = \sum_{k=1}^{K} (\overline{\beta}_{ik} + \upsilon_{ikt}) y_{kt} + u_{it} = \sum_{k=1}^{K} \overline{\beta}_{ik} y_{kt} + \sum_{k=1}^{K} \upsilon_{ikt} y_{kt} + u_{it} = \sum_{k=1}^{K} \overline{\beta}_{ik} y_{kt} + \varepsilon_{it} \qquad (2)$$

where $\beta_{ikt} = \overline{\beta}_{ik} + \upsilon_{ikt}$, and $\varepsilon_{it} = \Sigma_k \upsilon_{ikt} y_{kt} + u_{it}$ is a "composite error" term, which is heteroskedastic. In other words, each individual coefficient, β_{ikt}, is the sum of a mean component, $\overline{\beta}_{ik}$, common to all farms, and a random component, υ_{ikt}. Basically, this is the approach that was also used by Dixon *et al*. (1984) and Hornbaker *et al*. (1989).[3] Such a random-coefficients model can be estimated under various (mild or stringent) conditions regarding the means and the variance-covariance structures of the random components.[4]

Given the specification in (2), the following accounting restrictions apply to the mean and random components of the model:

$$0 \leq \overline{\beta}_{ik} \leq 1 \ \ \forall i,k \quad \text{if } \ \overline{x}_i \geq 0 \qquad (3)$$

$$\bar{\beta}_{ik} + \upsilon_{ikt} \geq 0 \ \ \forall i,k,t \quad \text{if} \ \ x_{it} \geq 0 \tag{4}$$

$$\Sigma_i \bar{\beta}_{ik} = 1 \ \ \forall k \tag{5}$$

$$\Sigma_i (\bar{\beta}_{ik} + \upsilon_{ikt}) = 1 \ \ \forall k,t \tag{6}$$

$$\Sigma_i \upsilon_{ikt} = 0 \ \ \forall k,t \tag{7}$$

$$\Sigma_i u_{it} = 0 \ \ \forall t \tag{8}$$

Note that the restrictions on the random components in (7) are automatically implied by the accounting restrictions in (5) and (6).

Before proceeding, we should point to a particular problem that typically occurs when dealing with multi-output or mixed farms. Specifically, if we assume that mean coefficients do exist, it must be ensured that the average values of the farm-specific coefficients, over the farms in the sample, be equal to the corresponding mean coefficient. In other words, it must be ensured that $\Sigma_t \upsilon_{ikt} = 0$, given that $\bar{\beta}_{ik} = (1/T)\,\Sigma_t \bar{\beta}_{ikt} = (1/T)\,\Sigma_t\,(\bar{\beta}_{ik} + \upsilon_{ikt}) = \bar{\beta}_{ik} + (1/T)\,\Sigma_t \upsilon_{ikt}$. A problem arises, however, when all farms do *not* produce all the outputs considered. If there are zero observations on one or more outputs for some farms, $E_t(\beta_{ikt}) = \bar{\beta}_{ik}$ is not necessarily true because of the selectivity bias that may occur (e.g., Ray, 1985).[5] Therefore, we use $E_t[\,\beta_{ikt} \mid y_{kt} > 0\,] = E_t[\,\bar{\beta}_{ik} + \upsilon_{ikt} \mid y_{kt} > 0\,]$ as a measure of $\bar{\beta}_{ik}$, representing only the subset of farms effectively producing output *k*, which is unbiased only if $E_t[\,\upsilon_{ikt} \mid y_{kt} > 0\,] = 0$. Consequently, the random components υ_{ikt} should appear in the restrictions (4), (6) and (7) only for those farms that have a positive output *k*.

2.2 Non-randomly varying coefficients

Despite the advantages of the random-coefficients specification in (2), we prefer not to adopt this "classical" approach, but rather treat the farm-specific coefficients as *non-random* (non-stochastic) parameters to be estimated. This approach is motivated by the following considerations.

Firstly, a typical (but widely neglected) problem with the random-coefficients approach is that the farm-level coefficients are not *uniquely* defined. In a random-coefficients setting, the individual coefficients are usually "predicted", based on the second-order moments of the expectations vector, rather than estimated (Judge *et al.*, 1985, p. 807). However, the values of these predicted coefficients – occasionally referred to as Best Linear Unbiased Predictors, or BLUPs, for short (e.g., Hornbaker *et al*., 1989) – are in no way unique; at best they are only "suitable" values (Griffiths, 1972; Swamy and Mehta, 1975).[6]

Secondly, while random-coefficients models can in principle be estimated by means of traditional methods, such as Generalized Least Squares (e.g., Swamy, 1970) or Maximum Likelihood (e.g., Schwallie, 1982; Zaman, 2002), the practical implementation of the estimation procedure is not an easy task, often requiring *ad hoc* corrections and/or giving cause to computational problems – difficulties that were also reported by Dixon *et al*. (1984) and Hornbaker *et al*. (1989). Needless to say, that the estimation will become even more compounded, if not impossible, when several restrictions have to be imposed on the coefficients.[7]

Given the many problems with using a traditional random-coefficients specification, it will be shown in the next section that the application of GCE does allow to capture farm-level heterogeneity by means of a set of uniquely defined individual coefficients.

3. Generalized Cross-Entropy formulation

In order to avoid the practical problems described in the previous section, and to obtain estimates that satisfy the various restrictions at every point in the sample, the varying-coefficients cost-allocation model will be formulated as a GCE problem of the type proposed by Golan *et al*. (1996). While there is a fast growing body of literature with respect to information theory and entropy econometrics, the GCE method advocated here will only be discussed in the context of the present application.[8]

The implementation of GCE requires that the parameters of the model are specified as linear combinations of some predetermined and discrete support values and unknown probabilities or weights.[9] Furthermore, the estimation problem is converted into a constrained minimization problem, where the objective function, specified in equation (9) below, consists of the joint cross entropy.

Specifically, we define a set of unknown probability vectors $\mathbf{p}'_{ik} = [\ p_{ik,1},\dots, p_{ik,M}\ ]$ ($M \geq 2$), $\mathbf{w}'_{ikt} = [\ w_{ikt,1},\dots, w_{ikt,G}\ ]$ ($G \geq 2$), and $\boldsymbol{\mu}'_{it} = [\ \mu_{it,1},\dots,\mu_{it,G}\ ]$ ($G \geq 2$), and choose the corresponding support vectors $\mathbf{z}' = [\ z_1,\dots, z_M\ ]$, $\mathbf{r}' = [\ r_1,\dots, r_G\ ]$, and $\mathbf{e}' = [\ e_1,\dots,e_G\ ]$, for the mean coefficients $\bar{\beta}_{ik}$, the farm-varying components υ_{ikt}, and the residual terms u_{it}, respectively, where $\bar{\beta}_{ik} = \mathbf{z}'\mathbf{p}_{ik}$, $\upsilon_{ikt} = \mathbf{r}'\mathbf{w}_{ikt}$, and $u_{it} = \mathbf{e}'\boldsymbol{\mu}_{ikt}$. In addition, prior information is included through specifying the prior probability vectors $\mathbf{p}_{ik}^{o}$, $\mathbf{w}_{ikt}^{o}$, and $\boldsymbol{\mu}_{it}^{o}$, reflecting subjective information, informed "guesses", or any other sample and pre-sample information.

After appropriate reparameterization, the complete GCE optimization problem for the cost-allocation model, described by Eqs. (2) to (8), can then be formulated as

$$\underset{\mathbf{p},\mathbf{w},\boldsymbol{\mu}}{\text{Min}}\, CE = \sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{p}_{ik}' \ln\left(\frac{\mathbf{p}_{ik}}{\mathbf{p}_{ik}^{o}}\right) + \sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbf{w}_{ikt}' \ln\left(\frac{\mathbf{w}_{ikt}}{\mathbf{w}_{ikt}^{o}}\right) + \sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\mu}_{it}' \ln\left(\frac{\boldsymbol{\mu}_{it}}{\boldsymbol{\mu}_{it}^{o}}\right) \tag{9}$$

subject to

$$x_{it} = \sum_{k=1}^{K}(\mathbf{z}'\mathbf{p}_{ik} + \mathbf{r}'\mathbf{w}_{ikt})\, y_{kt} + \mathbf{e}'\boldsymbol{\mu}_{it} \;\; \forall i,t \tag{10}$$

$$\mathbf{z}'\mathbf{p}_{ik} + \mathbf{r}'\mathbf{w}_{ikt} \geq 0 \;\; \forall i,k,t \in \{t_k\}\, ,\;\; x_{it} \geq 0 \tag{11}$$

$$\mathbf{z}'\mathbf{p}_{ik} + \mathbf{r}'\mathbf{w}_{ikt} = 0 \;\; \forall i,k,t \notin \{t_k\} \tag{12}$$

$$\frac{1}{T_k}\sum_{t\in\{t_k\}}(\mathbf{z}'\mathbf{p}_{ik} + \mathbf{r}'\mathbf{w}_{ikt}) = \mathbf{z}'\mathbf{p}_{ik} \;\;\; \forall i,k \tag{13}$$

$$\sum_{i=1}^{N}(\mathbf{z}'\mathbf{p}_{ik} + \mathbf{r}'\mathbf{w}_{ikt}) = 1 \;\forall k,t \in \{t_k\};\;\; \sum_{i=1}^{N}\mathbf{z}'\mathbf{p}_{ik} = 1 \;\forall k\,;\;\; \sum_{i=1}^{N}\mathbf{e}'\boldsymbol{\mu}_{it} = 0 \;\forall t \tag{14}$$

$$\sum_{m=1}^{M} p_{ik,m} = 1 \;\forall i,k\,;\;\; \sum_{g=1}^{G} w_{ikt,g} = 1 \;\forall i,k,t\,;\;\; \sum_{g=1}^{G}\mu_{it,g} = 1 \;\forall i,t \tag{15}$$

where $\{t_k\} = \{t : y_{kt} > 0\} \subseteq \{t\}_1^T$ is the subset of the sampled farms, of size T_k, producing output *k*. Equation (9) denotes the cross-entropy objective, which is subject to the data-consistency constraints in (10), the non-negativity constraints for the farm-level coefficients in (11), the zero-output constraints in (12), the "mean-preservation" constraints in (13), and the accounting-balance constraints in (14). The constraints in (15) ensure that all unknown probabilities add up to one. Including additional constraints on the *mean* coefficients, $\mathbf{z}'\mathbf{p}_{ik}$, is not required if one chooses proper or theoretical bounds for the corresponding support ranges $\mathbf{z}$ (see Eq. (3)).

The principle of minimum cross entropy means that we are choosing, given the data-consistency and other constraints, the estimates of the unknown $\mathbf{p}_{ik}$, $\mathbf{w}_{ikt}$ and $\boldsymbol{\mu}_{it}$ that can be discriminated from the priors $\mathbf{p}_{ik}^{o}$, $\mathbf{w}_{ikt}^{o}$, and $\boldsymbol{\mu}_{it}^{o}$ with a minimum of difference (Golan *et al.*, 1996, p. 11). In other words, we are looking for the “least informative” (i.e., closest to the uniform) probability distributions that are consistent with the data and other constraints in (10)-(15) and with the prior information reflected in the support ranges and the prior probabilities. From the solution of this optimization program, the unknown mean coefficients, the farm-varying components, and the remaining residual terms can be calculated as follows: $\hat{\bar{\beta}}_{ik} = \mathbf{z}'\hat{\mathbf{p}}_{ik}$, $\hat{\upsilon}_{ikt} = \mathbf{r}'\hat{\mathbf{w}}_{ikt}$, $\hat{u}_{it} = \mathbf{e}'\hat{\boldsymbol{\mu}}_{ikt}$, with $\hat{\beta}_{ikt} = \mathbf{z}'\hat{\mathbf{p}}_{ik} + \mathbf{r}'\hat{\mathbf{w}}_{ikt}$. Hence, a total of *NK* + *NKT* + *NT* parameters (i.e., the *NK* mean and *NKT* farm-specific components, along with *NT* residual terms) are estimated, with only *NT* observations.[10]

Given the above formulation, several important advantages of GCE over the traditional random-coefficients approach become irrefutably apparent. Firstly, the idiosyncratic υ_{ikt}'s, capturing the farm-specific deviations from the mean, are essentially treated as “fixed” (i.e., non-random) parameters to be estimated rather than as random drawings from a continuous multivariate distribution. Secondly, by considering this “fixed-coefficients” heterogeneity (not to be confused with a fixed-coefficients model!), the estimated mean coefficients do not suffer from possible bias due to the correlation between the varying components and the regressors (e.g., Biørn, 2003, p. 73). Thirdly, the estimated $\hat{\upsilon}_{ikt}$'s are directly available, which makes it is possible to study the empirical distribution of the $\hat{\upsilon}_{ikt}$'s across farms, and to pick a particular farm to see whether its estimated β_{ikt} is above or below the estimated mean $\bar{\beta}_{ik}$ for the sample.

4. Empirical application

4.1 Data

The empirical application uses 2000-2001 accounting data for a sample of 38 “beef-dairy” farms, located in Brittany, France ($T = 38$).[11] The data were extracted from the European *Farm Accountancy Data Network* (FADN), which collates detailed information on whole-farm input costs and enterprise-level output values.

On the input side, we distinguish six ($N = 6$) broad input categories: (a) *Crop products*[12]; (b) *Fertilizers* (including those utilized in the fertilization of pasture lands); (c) *Phytosanitary and veterinary products*; (d) *Compound feeds*; (e) *Other (miscellaneous) inputs*; and (f) *Gross* *value added* (GVA). The costs associated with the first four input categories are typically variable (specific crops/livestock) costs, whereas the operational (direct) costs associated with the fifth input category are "semi-variable" in nature. Factoring in all five input categories means that we seek to allocate both variable and semi-variable input costs to the various enterprises. Furthermore, each farm's GVA is derived residually as total output value *minus* the sum of the costs associated with the other five input categories, and is treated as the sixth "input".[13] Three farms in the sample showed a negative value for the whole-farm GVA.

On the output side, four ($K = 4$) different categories are considered: (a) *Crops*; (b) *Beef*; (c) *Dairy*; and (d) *Pork*. Output values were taken directly from the raw FADN data. The output value associated with each livestock enterprise is defined as the "value of total production", which is equal to the sum of net production (sales *minus* purchases)[14] and stock variation. For the crops enterprise, the output value also includes on-farm use of crop products. All farms in the sample produced crops, beef, and dairy (except one farm producing only crops and beef), while 25 farms were also involved in pork production.[15]

In estimating the empirical model, we use two-year averages to level out possible extreme yearly variations in input costs and output values.[16] Summary statistics on whole-farm input expenditures and enterprise-level values of output for the farms in the sample are reported in table 1. All monetary values are expressed in €1,000. From the figures in table 1, it should be clear that we have a particularly heterogeneous sample, in terms of overall farm size (measured as whole-farm value of output or number of livestock units) as well as in terms of enterprise mixes and input uses.

4.2 Supports and prior information

The application of GCE requires that we specify support vectors for each parameter of the model. The choice of support values represents a way of imposing prior restrictions on the estimates, apart from the equalities and inequality restrictions that are directly specified in the constraint set.

In close interaction with the data (see summary statistics in table 1), the support vectors are defined as follows: for the mean coefficients, we choose the "natural" support vector $\mathbf{z'}$ = [0, 1] ($M = 2$), given the fact that the actual means of the whole-farm input expenditures (dependent variables) from the sample are all non-negative; for the varying components, the support vector is set as $\mathbf{r'}$ = [-2.5, 0, 2.5] ($G = 3$), wide enough to include all "plausible" values for the farm-specific coefficients;[17] for the error terms, the support vector is defined, following the "3-sigma" rule, as $\mathbf{e}' = [-3\sigma_i, 0, 3\sigma_i]$ $(G = 3)$, where σ_i is the sample standard deviation of the dependent variables (see Panel A of table 1).

In addition, we use prior information about the unknown probability vectors **p**, **w**, and **μ**, in the form of informative priors (for the mean coefficients) and non-informative, uniform, priors (for the farm-specific components as well as for the residual terms). In particular, we set $\mathbf{p}_{ik}^{o}{}' = [\, p_{ik,1}^{o}, p_{ik,2}^{o}\,] = [1 - \alpha_i, \alpha_i]$, where α_i is the average share of the expenditure on input i in the total output value (reported in Panel B of table 1),[18] and $\mathbf{w}_{ikt}^{o}{}' = \boldsymbol{\mu}_{it}^{o}{}' = [1/3, 1/3, 1/3]$. The choice of the informative priors $\mathbf{p}_{ik}^{o}$ is motivated on the ground that the sample averages of the whole-farm input-expenditure shares, $\alpha_i = (1/T)\Sigma_t (x_{it}/\Sigma_i x_{it}) = (1/T)\Sigma_t (x_{it}/\Sigma_k y_{kt})$, can be viewed as reasonable or "best possible" initial hypotheses with respect to the magnitudes of the corresponding mean coefficients for each farm enterprise.

As the GCE formulation results is a form of shrinkage estimator (Golan *et al*., 1996, p. 31), a larger weight will be assigned to the terms in the cross-entropy objective associated with the smallest of the prior probabilities. As a result, the probabilities $p_{ik,m}$ associated with the mean coefficients will be "shrunk" faster toward the priors, implying a more stable estimation. This property of the GCE estimator is considered to be particularly important, given the small size and heterogeneous composition of our sample.[19]

4.3 Empirical results

4.3.1 GCE estimates of mean coefficients

The GCE estimates of the cost-allocation coefficients are presented in table 2.[20] Panel A of table 2 shows the estimates of the *mean* coefficients obtained from the varying-coefficients model (henceforth labelled V-GCE estimates). Two coefficients have been set equal to zero *a priori*, for obvious reasons: no compound feeds go into the crops enterprise, and no fertilizers are utilized in the pork enterprise.[21] Overall, the magnitudes of the V-GCE estimates do have considerable face validity, and are broadly in line with prior expectations (a brief discussion of the validity of model will follow in Section 4.4 below).

For the purpose of comparison, Panels B of table 2 shows the estimates of the *fixed*-coefficients version of the model (henceforth referred to as F-GCE estimates).[22] The F-GCE estimates seem to be less plausible, for several reasons. Firstly, some of the coefficients are virtually zero (e.g., the coefficients associated with crop products/beef, and fertilizers/beef), while some other coefficients seem to be overly large (e.g., the coefficients corresponding to compound feeds/beef, and GVA/pork). Such highly questionable findings are likely to be indicative of the fact that the F-GCE estimates may be more heavily affected by the occurrence of influential data points in the (small) sample than the V-GCE estimates.

4.3.2 Statistical properties of the GCE estimates

Because the sampling properties of the bounded GCE estimator are generally unknown, we use a simple (percentile) bootstrap approach to assess the "precision" of the GCE estimates.[23]

The bootstraps are calculated by using a re-sampling procedure (rather than the error bootstrap), in order to preserve the heteroskedasticity present in the data. Specifically, we randomly draw 100 samples from the empirical *joint* distribution of the data (i.e., from the original sample), $F(\mathbf{x}, \mathbf{y})$, and form 100 new data matrices $[\mathbf{x}_b, \mathbf{y}_b]$, where $b = 1,\ldots,100$. Then, we re-estimate the (mean) coefficients 100 times using the newly-formed data matrices, yielding 100 different bootstrap estimates. The observed distributions of these estimates are then taken as approximations of the "true" distributions of the coefficients, from which 95 percent confidence intervals can easily be derived by selecting the 0.025 and 0.975 percentiles (i.e., by leaving out the outer 5 percent of estimated mean values).

Obviously, the bootstrap intervals reported in table 2 are very small, which is indicative of the fact that the V-GCE estimator produces quite stable results for the mean coefficients. Conversely, the bootstrap intervals for the F-GCE estimates are consistently larger (and in some cases close to the zero boundary) than their V-GCE counterparts. This finding clearly suggests that adopting a varying-coefficients specification of the model dramatically reduces the uncertainty about the values of the unknown mean coefficients.

4.3.3 Estimated farm-level coefficients

Our empirical findings are also clearly indicative of the fact that cost structures vary considerably across farms.

Table 3 reports summary statistics for the V-GCE estimates of the farm-specific coefficients. The dispersion across farms is measured as the inter-quartile range divided (standardized) by the corresponding median, rather than as the usual standard deviation or coefficient of variation (which are more vulnerable to outliers). The variation in the cost structures for the crops enterprise is consistently smaller than for the livestock enterprises. For the livestock enterprises, the dispersion for dairy and pork is much larger than for beef. Also, the coefficients for the minor input categories (i.e., crop products, fertilizers, and phyto and vet products) also display a wider dispersion than the other input categories (i.e., compound feeds, other inputs and GVA).

Given the large number of estimated farm-specific coefficients, it is unwieldy to report them all here. Nevertheless, in order to provide a more informative picture of the variations across farms, figure 1 shows the kernel density plots for the distributions of the estimated individual coefficients.[24] From these plots it can be seen that the estimated coefficients are generally not normally distributed; in many cases the observed distributions exhibit a strong asymmetrical and/or irregular shape, and also reveal some noticeable outliers, particularly in the case of the major beef and dairy enterprises.[25] A limited number of extreme values can be detected for the beef enterprise (e.g., compound feeds), which strongly affect the corresponding mean coefficient (in this particular case, the mean value, 0.328, is larger than the third-quartile value, 0.314). In only two instances we found a zero-boundary value for the individual coefficients associated with the use of crop products in beef production.

4.4 Model validation

In order to assess the overall performances of the varying- and fixed-coefficients models, we present some "post-estimation" results. In fact, the validation of the models should take into account both the "quality" (i.e., consistency with prior expectations) of the estimates and the within-sample "predictive power" (i.e., accuracy) of the models.

4.4.1 Cross-enterprise relationships

We begin with investigating to what extent the enterprise-level cost structures for the "representative" farm are similar or dissimilar, by looking more closely at the estimated mean coefficients across the various livestock enterprises. Table 4 presents dissimilarity matrices, based on simple Euclidean distances.

The dissimilarity indexes indicate that for the mean V-GCE estimates the various livestock enterprises are fairly similar in terms of their cost structures, and that the dissimilarity between the beef and dairy enterprises is smaller than between beef/dairy and pork. This result contrasts sharply with the corresponding F-GCE estimates, where it can be seen that the dissimilarities are excessively large. Also, the finding that the dissimilarity between dairy and beef is larger than between dairy and pork is highly questionable and inconsistent with results from other sources (e.g., Errington, 1989; personal communication by Desbois, RICA/SCEES, Paris). These observations lead us, once again, to the conclusion that the V-GCE estimates are superior over the F-GCE estimates.

Furthermore, strong positive correlations should emerge between the farm-specific coefficients (or the farm-varying components) across enterprises, since there is no *a priori* reason to believe that a farm's variation around the mean coefficient for one particular enterprise should be independent of its variation for the other enterprises (e.g., Dixon *et al.*, 1984; Ray, 1985). This cross-enterprise interdependency is confirmed by the calculated pair-wise rank correlations (based on Spearman's *rho*) for the V-GCE estimates, reported in table 5. This result also clearly highlights the fact that the assumption of zero covariances (as in the standard Hildreth-Houck model) would not be a realistic assumption.

4.4.2 Input substitutability/complementarities

Assuming that all farmers in the sample face the same input and output prices, the farm-level coefficients obtained should also be able to reflect prior expectations with respect to the diversity of "technology" (cost) structures across farms. In order to test for the existence of substitutability or complementarities among the various inputs, we again calculated pair-wise rank correlations (based on Spearman's *rho*). The results are presented in table 6.

We found that the results are largely consistent with prior expectations. The relationship of "substitutability" between crop products/fertilizers (for grazing cattle), and between crop products/phyto and vet products (for pigs), on the one hand, and compound feeds, on the other hand, is clearly visible from the strong negative correlations. Furthermore, the correlations between the on-farm use of crop products for livestock feeding, on the one hand, and phyto and vet products (where the latter category also includes feed additives, such as vitamins, drugs, antibiotics, etc.), on the other hand, turn out to be strongly positive, indicating complementarities. Finally, no significant correlations exist between variable and semi-variable inputs, whereas strong negative correlations emerge – not surprisingly – between GVA and most other inputs, except for fertilizers and compound feeds (for which we have no clear-cut explanation).

4.4.3 Implications of farm size

To analyze the relationship between farm size and cost structure, the farms in the sample are divided into three groups. The farm grouping is based on the quartiles of the empirical distributions of the total number of livestock units (table 1), where “small” and “large” farms correspond to the first and the fourth quartile, respectively, while medium-sized farms are those that fall within the inter-quartile range. Based on this classification, ten farms are to be considered as small, 18 farms as medium-sized, and ten farms as large. For convenience, it is assumed that size effects apply to the whole farm, regardless of the enterprise mix.

Table 7 reports the averaged estimates of the coefficients, for the three size classes of farms. The results show that there are substantial differences between small farms and large farms, particularly for the beef and dairy enterprises (though not so much for the crops and pork enterprises).

A comparison of the averaged estimates for small and large farms reveals that the per-unit costs for all specific (variable) livestock inputs, except for compound feeds in the beef and dairy enterprises, are consistently smaller for large farms than for small farms. Using the non-parametric Mann-Whitney *U* test, it was found that these differences are also statistically significant in many cases (and in some other cases, the MW-*U* test statistic is close to the borderline). On the other hand, large farms involved in beef and/or dairy production would use relatively more compound feeds. This result could be expected *a priori* (e.g., large farms are probably more “intensive” farms), but from our analysis it is not possible to disentangle the effects of substitution for crop products, on the one hand, and (technical and/or cost) efficiency gains, on the other hand. For example, large farms may be able to purchase compound feeds at a reduced price because of volume discounts. Another surprising result is that the use of other (misc.) inputs is barely dependent on farm size. Apparently, any expected “economies of size” in using semi-variable inputs do not appear to materialize. Finally, the results also indicate that the enterprise-level GVA increases with the size of the farm, particularly in dairy production and pork production (in the case of the pork enterprise, though, this increase is not statistically significant).

4.4.4 Model accuracy

The validity of a model should not only be evaluated based on the "plausibility" of the estimated (mean) coefficients, but also by looking at the accuracy of the model in tracking the observed whole-farm input expenditures. Two useful measures of accuracy are given by the pseudo-R^2 and the mean absolute percentage error (*MAPE*), which can be calculated for each input *i*. The pseudo-R^2 is defined as the square of the simple correlations between x_{it} and $\hat{x}_{it}$, while $MAPE_i = (1/T)\Sigma_t\left(|x_{it} - \hat{x}_{it}|/x_{it}\right)$, where x_{it} and $\hat{x}_{it}$ are the observed and "predicted" values of the whole-farm expenditures on input *i*, respectively.

The pseudo-R^2 and *MAPE* results for the varying- and fixed-coefficients models are presented in table 8. Apparently, the results are somewhat mixed. Given the fact that the V-GCE and F-GCE point estimates of the mean coefficients are markedly different, it is rather surprising to see that, in terms of within-sample prediction performances, one model is not strongly dominating the other. Only in the case of compound feeds, the F-GCE model is outperforming the V-GCE model in terms of the pseudo-R^2. On the other hand, the corresponding *MAPE* statistic for this input category is in favour of the V-GCE model. Finally, by looking at the results based on the individual coefficients from the V-GCE model it can be seen that a considerable portion of the variations in the observed whole-farm input uses is "explained" by the farm-varying effects, $\hat{\upsilon}_{ikt}$, particularly for the minor input categories of crop products, fertilizers, and phyto and vet products (the pseudo-R^2 for the various input categories ranges from 0.84 to a value close to one). Nevertheless, a considerable amount of unexplained "noise" stills remains, particularly for the major input categories of compound feeds, other inputs, and GVA.

5. Concluding remarks

In this paper, a varying-coefficients Generalized Cross-Entropy (GCE) modeling framework was developed for the estimation of the (unobserved) production costs for multi-product farms, using standard whole-farm accounting data. In the empirical part of the paper, the GCE estimator was applied to cross-sectional (average 2000-2001) data from a sample of 38 "beef-dairy" farms in Brittany, France. The empirical results are plausible at face value, which lead us to believe that the GCE methodology can fruitfully be applied in other agricultural-economic settings.

The main contributions of this paper are as follows. Firstly, unlike most previous studies dealing with cost allocation (or multi-output technologies) in the absence of enterprise-level input data, the focus of the present study was on accounting for the heterogeneity of farm operations. Secondly, the GCE approach turned out to be an effective way to overcoming the many practical and methodological problems reported in the cost-allocation literature. Specifically, we were able to obtain plausible estimates, showing the right signs, satisfying all the accounting restrictions, and being in line with prior expectations. Thirdly, it was shown that applying GCE is far less cumbersome and requires fewer computational efforts than other (traditional) estimation methods, in particular when dealing with farm-varying parameters and/or constraints in a system-of-equations context. Fourthly, we applied GCE to real farm data, whereas all previous entropy-based studies referenced in this paper have used experimental data in a Monte-Carlo setting. Finally, the framework developed in the present study can also be of relevance to researchers involved in input-output modelling.

A comparison of our estimation results with existing estimates from other sources was not possible. Besides, such a comparison was not considered to be sensible in the first place, since the data we used (for only 38 farms located in Brittany) can hardly be considered as a "representative" sample of dairy-beef farms in the EU. Based on some experimentations, we found that marginal changes in the data set (e.g., omitting the three farms with a negative GVA) would already lead to markedly different estimates of the mean coefficients.[26] Ideally, our estimation results should be compared with *actual* enterprise-level data for the same sample of farms. Unfortunately, though, such data were not available.

One probable shortcoming of the present modelling framework adopted in this paper is that it does not allow for estimating price-induced input substitution; that is, the model we use is not a cost function, because it is not an explicit function of input prices. Evidently, changes in relative input prices would change input costs and eventually lead to changes in input use. This means that the estimated cost-allocation coefficients have only short-run validity, and that, therefore, the *ex-ante* forecasting ability of the model is rather limited. While acknowledging this obvious disadvantage, we believe though that this should not be a severe problem, as the cost-allocation estimates can easily be updated from time to time to take account of changing prices and/or changing production relationships.

A number of important research issues still remain, however. One interesting avenue for future research would be to "explain" the variation in the coefficients or to "control for" concomitant variables (interaction terms) in the model, such as measures of managerial expertise and other structural farm characteristics, thereby allowing for systematic variations in the coefficients. Also, more work is needed to obtain improved estimates by, for example, creating more homogeneous input and output categories, using larger and more representative samples, exploiting the "richness" of panel data and/or employing more realistic prior information (e.g., based on expert opinions) on the support values of the cost-allocation coefficients.

Table 1
Summary statistics on whole-farm input costs and enterprise-level output values, two-year averages 2000-2001

	No. of farms	Minimum	Mean	Maximum	Standard deviation
		A. Input costs/Output values (€ 1,000)			
Inputs					
Crop products	37	0.00	8.99	39.60	8.48
Fertilizers	38	1.54	4.65	16.56	2.67
Phyto. and vet. products	38	2.30	11.13	31.62	6.89
Compound feeds	38	4.42	68.47	413.64	73.70
Other (misc.) inputs	38	24.62	64.12	246.06	43.05
Gross value added	38	-138.31	69.33	242.35	57.23
Total		52.84	226.68	678.49	122.75
Outputs					
Crops	38	0.80	17.88	50.48	12.05
Beef	38	1.86	42.10	506.44	88.04
Dairy	37	0.00	82.89	200.00	44.69
Pork	25	0.00	83.82	454.59	94.69
Total		52.84	226.68	678.49	136.25
		B. Input-cost/Output-value shares (%)			
Inputs					
Crop products	37	0.0	4.3	10.6	2.6
Fertilizers	38	0.3	2.8	12.3	1.7
Phyto. and vet. products	38	2.9	5.3	13.9	2.2
Compound feeds	38	4.4	29.5	95.7	18.2
Other (misc.) inputs	38	8.6	32.9	110.6	10.2
Gross value added	38	-143.0	25.3	53.1	19.8
Total			100.0		
Outputs					
Crops	38	0.2	9.4	33.9	6.6
Beef	38	0.7	18.6	91.4	24.1
Dairy	37	0.0	40.6	74.9	17.5
Pork	25	0.0	31.4	72.9	26.6
Total			100.0		
		C. Other structural characteristics			
Arable land (hectares)	38	0.00	2.71	7.08	1.70
Fodder land (hectares)	38	1.58	4.26	8.77	1.98
Livestock units (1,000 UGB)	38	4.21	17.74	82.68	13.97

Table 2
GCE point estimates of the mean coefficients and 95 percent bootstrap confidence intervals[a]

Inputs	Outputs			
	Crops	Beef	Dairy	Pork
	A. Varying coefficients (V-GCE)			
Crop products	0.066 (0.050/0.072)	0.040 (0.032/0.045)	0.035 (0.028/0.041)	0.058 (0.044/0.072)
Fertilizers	0.048 (0.036/0.051)	0.030 (0.023/0.036)	0.037 (0.030/0.044)	
Phyto. and vet. products	0.080 (0.064/0.091)	0.049 (0.042/0.055)	0.041 (0.035/0.046)	0.065 (0.053/0.077)
Compound feeds		0.328 (0.234/0.396)	0.298 (0.234/0.360)	0.285 (0.246/0.301)
Other (misc.) inputs	0.448 (0.374/0.458)	0.306 (0.249/0.324)	0.315 (0.260/0.343)	0.277 (0.237/0.292)
Gross value added	0.359 (0.342/0.468)	0.247 (0.212/0.350)	0.374 (0.213/0.362)	0.315 (0.302/0.391)
	B. Fixed coefficients (F-GCE)			
Crop products	0.114 (0.036/0.185)	0.003 (0.000/0.015)	0.050 (0.035/0.069)	0.040 (0.016/0.056)
Fertilizers	0.110 (0.001/0.186)	0.003 (0.000/0.046)	0.028 (0.004/0.042)	
Phyto. and vet. products	0.160 (0.075/0.206)	0.045 (0.032/0.049)	0.046 (0.034/0.066)	0.032 (0.022/0.044)
Compound feeds		0.637 (0.241/0.707)	0.283 (0.120/0.551)	0.226 (0.093/0.294)
Other (misc.) inputs	0.393 (0.253/0.569)	0.123 (0.078/0.369)	0.419 (0.287/0.517)	0.203 (0.055/0.303)
Gross value added	0.223 (0.080/0.412)	0.190 (0.136/0.337)	0.175 (0.002/0.401)	0.498 (0.354/0.728)

[a] The 95 percent bootstrap confidence intervals are calculated by applying the simple 0.025 and 0.975 percentile method.

Note: Since inputs and outputs are measured in monetary terms, the estimated mean coefficients presented in this table represent the average expenditure for input *i*, expressed in €, per € 1.00 value of output *k*.

Table 3
Summary statistics for the estimated farm-specific coefficients from the varying-coefficients model

Inputs	Outputs			
	Crops	Beef	Dairy	Pork
Crop products				
Mean	0.066	0.040	0.035	0.058
Standard deviation	0.005	0.014	0.019	0.032
Coefficient of variation	0.080	0.361	0.556	0.550
Median	0.066	0.044	0.034	0.054
First quartile	0.062	0.038	0.020	0.036
Third quartile	0.070	0.046	0.051	0.082
Dispersion[a]	0.111	0.175	0.892	0.851
Fertilizers				
Mean	0.048	0.030	0.037	
Standard deviation	0.006	0.020	0.024	
Coefficient of variation	0.133	0.660	0.639	
Median	0.046	0.025	0.033	
First quartile	0.044	0.022	0.019	
Third quartile	0.050	0.030	0.048	
Dispersion[a]	0.131	0.354	0.848	
Phyto. and vet. products				
Mean	0.080	0.049	0.041	0.065
Standard deviation	0.009	0.015	0.026	0.025
Coefficient of variation	0.113	0.298	0.637	0.384
Median	0.079	0.050	0.034	0.062
First quartile	0.075	0.043	0.024	0.042
Third quartile	0.084	0.053	0.053	0.079
Dispersion[a]	0.119	0.201	0.839	0.594
Compound feeds				
Mean		0.328	0.298	0.285
Standard deviation		0.080	0.063	0.033
Coefficient of variation		0.243	0.211	0.115
Median		0.311	0.296	0.282
First quartile		0.305	0.263	0.259
Third quartile		0.314	0.308	0.312
Dispersion[a]		0.029	0.152	0.190
Other (misc.) inputs				
Mean	0.448	0.306	0.315	0.277
Standard deviation	0.016	0.050	0.058	0.066
Coefficient of variation	0.035	0.162	0.184	0.239
Median	0.445	0.315	0.305	0.284
First quartile	0.440	0.311	0.284	0.245
Third quartile	0.449	0.321	0.334	0.311
Dispersion[a]	0.021	0.033	0.164	0.232
Gross value added				
Mean	0.359	0.247	0.274	0.315
Standard deviation	0.025	0.038	0.100	0.079
Coefficient of variation	0.069	0.153	0.363	0.252
Median	0.362	0.254	0.280	0.300
First quartile	0.358	0.245	0.252	0.259
Third quartile	0.372	0.260	0.328	0.371
Dispersion[a]	0.040	0.060	0.268	0.371

[a] The dispersion measure is defined as the inter-quartile range divided by the median.

Table 4
Dissimilarity matrices (based on Euclidean distances)[a] of enterprise-level cost structures

	Varying-coefficients model			Fixed-coefficients model		
	Beef	Dairy	Pork	Beef	Dairy	Pork
Beef	0	0.043	0.094	0	0.354	0.411
Dairy		0	0.076		0	0.323
Pork			0			0

[a] The (pair-wise) Euclidean distances are calculated simply as $D_{kh} = \sqrt{\Sigma_i (\bar{\beta}_{ik} - \bar{\beta}_{ih})^2}$, for all $k \neq h$.

Table 5
Cross-enterprise interdependency of the farm-specific coefficients, for livestock enterprises – Correlation analysis based on Spearman's *rho*

	Beef-Dairy	Beef-Pork	Dairy-Pork
Crop products	0.562**	0.512**	0.747**
Fertilizers	0.694**	n.a.	n.a.
Phyto. & vet. prod's	0.408*	0.234	0.285
Compound feeds	0.257	0.775**	0.868**
Other (misc.) inputs	0.311	0.814**	0.875**
Gross value added	0.538**	0.512**	0.752**

*, **. Correlation is significant at the 0.05 and 0.01 levels, respectively (two-tailed).

Table 6
Input substitutability/complementarities in livestock production – Correlation analysis based on Spearman's *rho*

	Crop products	Fertilizers	Phyto. & vet. products	Compound feeds	Other (misc.) inputs	Gross value added
			Beef			
Crop products	1	0.291	0.486**	-0.673**	0.251	-0.354*
Fertilizers		1	0.019	-0.661**	0.059	-0.126
Phyto. & vet. prod's			1	-0.190	0.185	-0.503**
Compound feeds				1	-0.265	0.188
Other (misc.) inputs					1	-0.397*
Gross value added						1
			Dairy			
Crop products	1	0.238	0.312	-0.443**	0.044	-0.408*
Fertilizers		1	0.077	-0.606**	-0.165	-0.017
Phyto. & vet. prod's			1	-0.163	-0.042	-0.499**
Compound feeds				1	-0.149	0.056
Other (misc.) inputs					1	-0.688*
Gross value added						1
			Pork			
Crop products	1	n.a.	0.780**	-0.791**	0.405*	-0.568**
Fertilizers		1	n.a.	n.a.	n.a.	n.a.
Phyto. & vet. prod's			1	-0.595**	0.278	-0.549**
Compound feeds				1	-0.373	0.399*
Other (misc.) inputs					1	-0.901**
Gross value added						1

*, **. Correlation is significant at the 0.05 and 0.01 levels, respectively (two-tailed).

Table 7
Implications of farm size – Test of differences between mean coefficients, by farm size, based on Mann-Whitney *U* test

	Crops			Beef		
Farm size	Small	Medium	Large	Small	Medium	Large
No. of farms involved	10	18	10	10	18	10
Crop products	0.066	0.065	0.066	0.046*	0.038	0.037*
Fertilizers	0.050*	0.047	0.047*	0.036**	0.030	0.024**
Phyto. & vet. products	0.078	0.080	0.081	0.050	0.051	0.045
Compound feeds				0.303**	0.349	0.315**
Other (misc.) inputs	0.447	0.445	0.452	0.317	0.290	0.322
Gross value added	0.359	0.363	0.353	0.248**	0.241	0.256**
	Dairy			Pork		
Farm size	Small	Medium	Large	Small	Medium	Large
No. of farms involved	10	17	10	4	13	8
Crop products	0.045*	0.034	0.026*	0.067	0.055	0.057
Fertilizers	0.053**	0.033	0.027**			
Phyto. & vet. products	0.048	0.039	0.037	0.075	0.065	0.060
Compound feeds	0.273	0.299	0.320	0.285	0.290	0.276
Other (misc.) inputs	0.317	0.309	0.325	0.294	0.276	0.270
Gross value added	0.264*	0.285	0.266*	0.279	0.313	0.336

*, **. Difference of means is significant at the 0.10 and 0.05 levels, respectively (two-tailed).

Table 8
Pseudo-R^2 and *MAPE* statistics for the varying- and fixed-coefficients models

Inputs	Varying-coefficients model		Fixed-coefficients model	
	Pseudo-R^2	*MAPE*	Pseudo-R^2	*MAPE*
Crop products	0.433 (0.997)	85.9% (7.8%)	0.370	82.9%
Fertilizers	0.238 (0.999)	88.8% (0.9%)	0.262	83.8%
Phyto. and vet. products	0.178 (0.997)	84.2% (4.8%)	0.139	77.6%
Compound feeds	0.230 (0.891)	65.2% (48.7%)	0.556	74.8%
Other (misc.) inputs	0.243 (0.889)	30.6% (23.1%)	0.246	34.2%
Gross value added	0.766 (0.843)	42.2% (13.4%)	0.682	45.7%

[a] The Pseudo-R^2 values associated with the estimated *individual* coefficients (i.e., the proportions of the variations in whole-farm input uses that is "explained" by the estimates of the individual coefficients, β_{ikt} , and the output levels, y_{kt}) are given in parentheses.

Figure 1

Kernel density plots of the V-GCE estimates of farm-specific coefficients

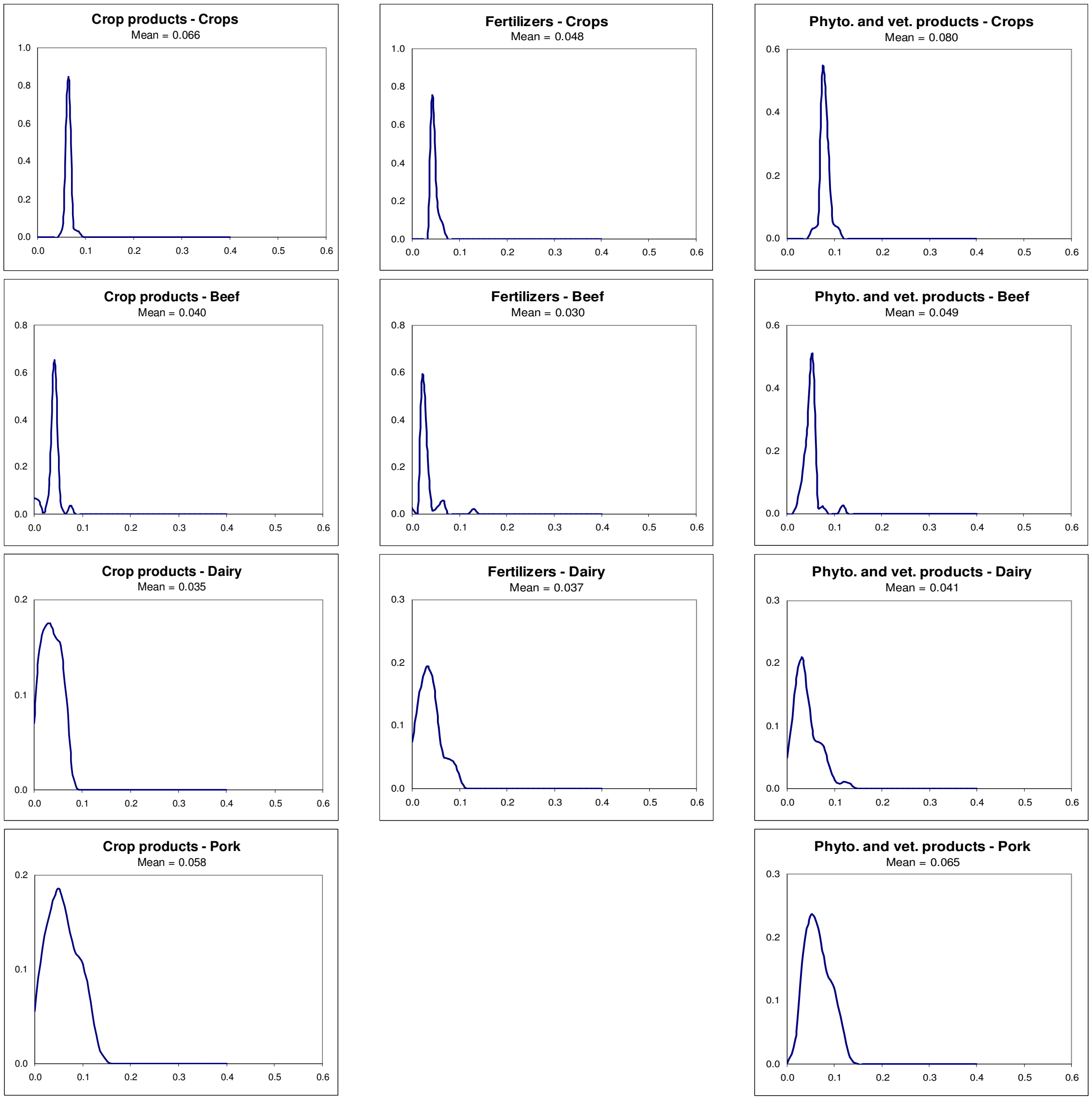
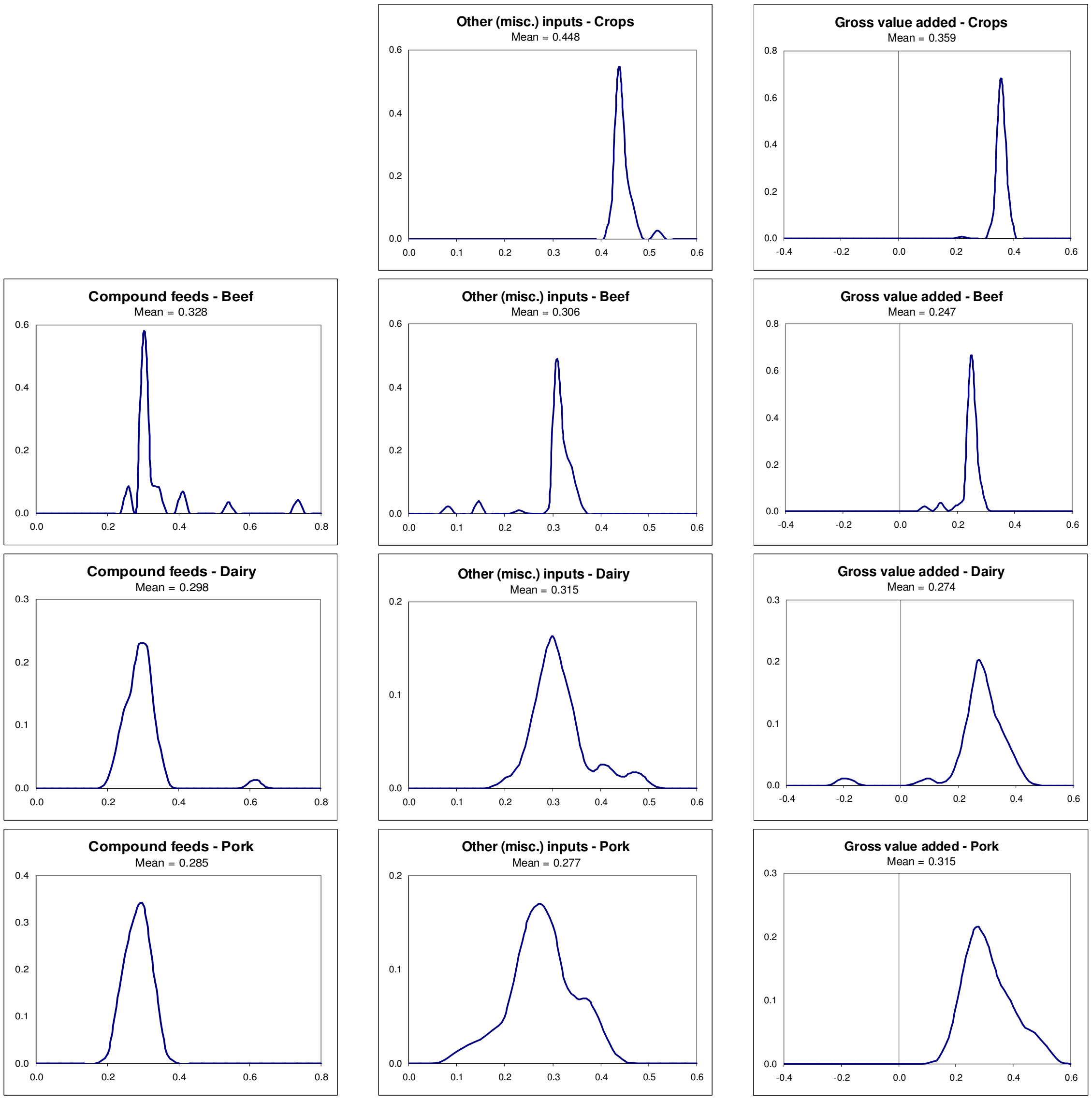


Figure 1
(continued)



References

Andrews, D.W.K. (1996). "Admissibility of the Likelihood Ratio test when the parameter space is restricted under the alternative." *Econometrica*, 64, 705-718.

Argilés, J.M. and Slof, E.J. (2001). "New opportunities for farm accounting." *The European Accounting Review*, 10, 361-383.

Biørn, E., Lindquist, K.G., and Skjerpen, T. (2003). "Random coefficients in unbalanced panels: an application on data from chemical plants." *Annales d' Économie et de Statistique*, 69, 55-85.

Bureau, J.C. and Cyncynatus, M. (1991). "Perspectives de l'amélioration de la méthode d'estimation des coûts de production à partir du RICA." *INRA Actes et Communications*, 5, 33-62.

Butault, J.P., Hassan, C.R., and Reignier, E. (1988). *Les coûts de production des principaux produits agricoles dans la CEE*. Luxembourg: Office of Official Publications of the European Communities.

Dixon, B.L., Batte, M.T., and Sonka, S.T. (1984). "Random coefficients estimation of average total product costs for multiproduct firms." *Journal of Business and Economic Statistics*, 2, 360-366.

Dixon, B. and Hornbaker, R.H. (1992). "Estimating the technology coefficients in linear programming models." *American Journal of Agricultural Economics*, 74, 1029-1038.

Errington, A. (1989). "Estimating enterprise input-output coefficients from regional farm data." *Journal of Agricultural Economics*, 40, 52-56.

Errington, A.J., Jones, P., Giles, A.K., Hallam, D., and Tranter, R.B. (1992). *Enterprise Gross Margin Determination and Forecasting: Final Report*. Unpublished Report to DGVI, Commission of the European Communities, Department of Agricultural Economics and Management, University of Reading.

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

Golan, A., Perloff, J.M., and Shen, Z. (2001). "Estimating a demand system with non-negativity constraints: Mexican meat demand." *The Review of Economics and Statistics*, 8, 541-550.

Greene, W. (2004). *Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models*. Department of Economics, Stern School of Business, New York University, mimeo (www.stern.nyu.edu/~wgreene).

Griffiths, W.E. (1972). "Estimation of actual response coefficients in the Hildreth-Houck random coefficient model." *Journal of the American Statistical Association*, 67, 633-635.

Hallam, D., Bailey, A., Jones, Ph., and Errington, A. (1999). "Estimating input use and production costs from farm survey panel data." *Journal of Agricultural Economics*, 50, 440-449.

Hildreth, C. and Houck, J.P. (1968). "Some estimators for a linear model with random coefficients." *Journal of the American Statistical Association*, 63, 583-594.

Hornbaker, R.H., Dixon, B.L., and Sonka, S.T. (1989). "Estimating production activity costs for multi-output firms with a random coefficient regression model." *American Journal of Agricultural Economics*, 71, 167-177.

Judge, G.G, Griffiths, W.E., Hill, R.C., Lütkepohl, H., and Lee, T.C. (1985). *The Theory and Practice of Econometrics*, 2nd edition. New York: Wiley.

Lence, S.H. and Miller, D.J. (1998a). "Recovering output-specific inputs from aggregate input data: a generalized cross-entropy approach." *American Journal of Agricultural Economics*, 80, 852-867.

Lence, S.H. and Miller, D.J. (1998b). "Estimation of multi-output production functions with incomplete data: a generalised maximum entropy approach." *European Review of Agricultural Economics*, 25, 188-209.

Léon, Y., Peeters, L., Quinqu, M., and Surry, Y. (1999). "The use of the maximum entropy method to estimate input-output coefficients from regional farm accounting data." *Journal of Agricultural Economics*, 50, 425-439.

Midmore, P. (1990). "Estimating input-output coefficients from regional farm data: a comment." *Journal of Agricultural Economics*, 41, 108-111.

Moxey, A. and Tiffin, R. (1994). "Estimating linear production coefficients from farm business survey data: a note." *Journal of Agricultural Economics*, 45, 381-385.

O'Donnell, C.K., Rambaldi, A.N., and Doran, H.E. (2001). "Estimating economic relationships subject to firm- and time-varying equality and inequality constraints." *Journal of Applied Econometrics*, 16, 709-726.

Oude Lansink, A. (1999). "Generalised maximum entropy estimation and heterogeneous technologies." *European Review of Agricultural Economics*, 26, 101-115.

Pingault, N. and Desbois, D. (2003). "Estimations des coûts de production des principaux produits agricoles à partir du RICA." *Notes et Études économiques*, No. 19, 9-51.

Preckel, P.V. (2001). "Least squares and entropy: a penalty function perspective." *American Journal of Agricultural Economics*, 83, 366-377.

Ray, S.C. (1985). "Methods of estimating the input coefficients for linear programming models." *American Journal of Agricultural Economics*, 67, 660-665.

Scandizzo, P.L. (1990). "The estimation of input-output coefficients: methods and problems." *Richerche Economiche*, 44, 455-474.

Schwallie, D.P. (1982). "Unconstrained maximum likelihood estimation of contemporaneous covariances." *Economics Letters*, 9, 359-364.

Shen, E.Z. and Perloff, J.M. (2001). *Maximum Entropy and Bayesian Approaches to the Ratio Problem*. Discussion Paper (mimeo).

Swamy, P.A.V.B. (1970). "Efficient inference in a random coefficient regression model." *Econometrica*, 38, 311-33

Swamy, P.A.V.B and Mehta, J.S. (1975). "Bayesian and non-Bayesian analysis of switching regressions and of random coefficient regression models." *Journal of the American Statistical Association*, 70, 593-602.

Swamy, P.A.V.B and Tavlas, G. (1995). "Random coefficients models: theory and applications." *Journal of Economic Surveys*, 9, 165-197.

Zaman, A. (2002). "Maximum likelihood estimates of the Hildreth-Houck random coefficients model." *Econometrics Journal*, 5, 237-262.

Endnotes

[1] In passing, it should be noted that the varying-coefficients modelling framework proposed in this paper may also be useful in other contexts, such as "stochastic" input-output modelling. Since most researchers use fixed input-output coefficients, they seem to be more involved in the results of the application of input-output models rather than in the intricacies of their statistical properties (Scandizzo, 1990). Conversely, consideration of the variation in the input-output coefficients allows for addressing the important question of the distributional properties of the Leontief inverse and the associated direct and indirect linkages between farm-enterprise operations and environmental and other key indicators.

[2] It should be noted that an earlier, entropy-based paper by Oude Lansink (1999) is only seemingly similar to the present one. Specifically, his approach amounts to estimating separate or *independent* farm-specific models, assuming *fixed* coefficients.

[3] The specifications used by Dixon *et al*. (1984) and Hornbaker *et al*. (1989) deviate from the familiar, simple specification in (2) only in the sense that they included one or more concomitants, or interaction variables, in the means of the random coefficients: $\beta_{ikt} = \beta_{1,ik} + \beta_{2,ik} z_t + \upsilon_{ikt}$, where z_{kt} is, for example, the overall size of farm *t*. While the model we propose can easily be extended to accommodate for such generalizations, this is beyond the scope of the present paper.

[4] Usually, the mean of the random components is set equal to zero, $E(\mathbf{\upsilon}_i) = 0$, while the variance-covariance structure is represented by the matrix $E(\mathbf{\upsilon}_i \mathbf{\upsilon}_i') = \mathbf{A}_{ii}$, where $\mathbf{A}_{ii}$ may be either a diagonal matrix, implying zero covariances (e.g., Hildreth-Houck, 1968), or a positive semi-definite matrix, implying non-zero covariances (e.g., Schwallie, 1982; Swamy and Mehta, 1975).

[5] Moreover, Monte-Carlo results in Dixon *et al*. (1984), among others, indicated that the reliability of the estimates declines in a random-coefficients model when the proportion of zero observations on the associated regressors rises.

[6] To our knowledge, the economic literature is strikingly silent about alternative methods for obtaining unique individual coefficients. One possible way, which has been proposed by O'Donnell *et al*. (2001), among others, is to apply a Singular-Value Decomposition (SVD) method. However, we believe that using SVD has some obvious shortcomings, in the sense that the individual coefficients are derived from a set of "pre-estimated" *mean* coefficients (e.g., by OLS or Bayesian methods). To the extent that the estimated mean coefficients are biased, these biases will evidently also "contaminate" the individual coefficients in one way or another.

[7] Recently, Greene (2004) proposed the use of the Maximum Simulated Likelihood (MSL) method to overcome the many computational difficulties with random-coefficients models. For our purposes, however, the MSL approach does not give much comfort, as the problem of finding unique estimates still remains.

[8] We use the Generalized Cross-Entropy (GCE) estimator rather than the Generalized Maximum Entropy (GME) estimator. This choice is primarily motivated by the observation that the GCE estimates are generally less sensitive to changes in prior information (i.e., the support intervals and/or the number of support values) supplied by the researcher. The sensitivity of the results to the prior information is a standard concern in GME/GCE as well as in Bayesian inference. However, a full treatment of the sensitivity issue is beyond the

scope of the present paper. In the context of cost-allocation models the sensitivity of GME was already clearly demonstrated in Léon *et al*. (1999).

[9] As rightly pointed out by Preckel (2001), it is not appropriate to view the weights attached to the support values as genuine "probability" distributions, since the results of the estimation problem are merely a set of values for the coefficients and the errors for each observation.

[10] Although the farm-level coefficients β_{ikt} could in principle be estimated directly, we found it more convenient to estimate the mean components, $\bar{\beta}_{ik}$, and the farm-level components, υ_{ikt} , as separate quantities (to ensure the "mean preservation"), and then calculate the individual β_{ikt}'s simply by summing the mean and farm-level components.

[11] More precisely, the type of farms in the sample is defined as "cattle-dairying, rearing and fattening combined" (FT43), which is a sub-category of "specialist grazing livestock" (FT4). Based on some preliminary analyses, we decided to select only those beef-dairy farms with an economic size greater than 48 European Standard Units (ESUs). These farms are to be considered as large-sized farms in a Brittany context.

[12] In the FADN, inputs are subdivided into those purchased and those produced and used on the farm. The latter are included as cost items to the extent that they are *saleable* farm products, valued at their *opportunity cost* (i.e., farm-gate price). For the crops enterprise, the input category of "Crop products" includes purchased and farm-produced seeds and seedlings, planting materials, etc. For the livestock enterprises, this input category includes purchased coarse fodder, purchased and farm-produced coarse grains and oilseeds, and expenditures on the use of common pastures, grazing and forage land. Farm-produced fodder (e.g., litter and straw, silage, grazed pasture, etc.) is excluded, however. On the other hand, the input category of "Compound feeds" includes purchased compounds, dried grass, dried sugar beet pulp, oilcakes, fish and meat meal, milk and dairy products, minerals, and products for the preservation and storage of feeding stuffs.

[13] Borrowing the FADN terminology (e.g., Argilés and Slof, 2001), GVA is defined here as the value of total production *minus* the sum of intermediate consumption (i.e., specific crop/livestock costs), overheads (i.e., machinery and building costs, energy, contract work, and other direct inputs) and external factors (i.e., operating rents and interests, paid labour), where the overheads and external factors are subsumed under the heading "other (miscellaneous) inputs". In other words, GVA is equal to the sum of personal drawings, depreciation (i.e., loan repayments and reinvestments), interests (on borrowings), and taxes. Accordingly, GVA can grossly be defined as the return to the farmer's labour and land *plus* asset-ownership costs and taxes, *net* of subsidies and capital grants.

[14] Sales also include on-farm uses of saleable milk products. However, milk suckled by calves and farm-produced fodder crops are excluded from the sales figures.

[15] Due to data limitations, it is implicitly assumed that a zero enterprise-level value of output automatically implies corresponding zero input use.

[16] Given the fact that we observed a zero value for the whole-farm input (i.e., crop products) expenditures for only one farm in the sample, it was not considered worthwhile using a GCE-Tobit specification of the model

[17] The occurrence of negative values for the whole-farm GVA (observed for merely three farms in the sample) means that one or more of the farm-specific coefficients can be outside the [-1, 1] range. In general, the relevant width of the support range for the values of the farm-specific coefficients is not known *a priori*. However, under the (reasonable) assumption that the negative GVA of a farm is allocated "approximately" in an equiproportional way to the various enterprises (according to the enterprise-level values of production), the magnitudes of the enterprise-level GVA coefficients should not be too far away from those observed for the whole-farm GVA (reported in Panel A of table 1), where the minimum value is -1.43. Choosing the [-2.5, 2.5] range, and taking the boundary values zero and one for the corresponding mean coefficients, this would imply that the minimum possible values for the farm-specific, enterprise-level GVA coefficients are -2.5 and -1.5, respectively (note that an equiproportional allocation would mean that all the enterprise-level GVA coefficients are exactly equal to -1.43). Moreover, the accounting restrictions would preclude large deviations from the observed whole-farm values anyway (dampening effect).

[18] The priors $\mathbf{p}_{ik}^{\text{o}}$ are determined as follows: $p_{ik,1}^{\text{o}}\,(0) + p_{ik,2}^{\text{o}}\,(1) = \alpha_i$, $p_{ik,2}^{\text{o}} = \alpha_i$, hence, $p_{ik,1}^{\text{o}} = 1 - \alpha_i$.

[19] It is a well-known result that GCE and other shrinkage estimators perform substantially better than do estimators that are not based on informative priors, such as Generalized Maximum Entropy (GME) (Shen and Perloff, 2001). Under GME, uniform prior weights are placed on the parameter support values and the focal point of the prior information is the midpoint of the bounded parameter space. However, based on Monte-Carlo analysis, Lence and Miller (1998b) were able to show that the average estimates are roughly midway between the true values and the centres of the associated support intervals. In addition, they found that the GME bias typically increases as the prior information departs from the "truth". In order to examine the sensitivity of the GCE estimator, we performed several model runs with different support ranges for the farm-varying components, while preserving the [0, 1] bounds on the mean coefficients. It turned out that moderately widening the support ranges had only minor to negligible effects on the GCE estimates (not reported here to conserve space).

[20] All the results have been generated by using the GAMS (Generalized Algebraic Modeling System) solver CONOPT3 for non-linear programming problems.

[21] Hence, the total number of parameters to be estimated is 3,286; that is, (6×4) – 2 = 22 mean coefficients ($\bar{\beta}_{ik}$) *plus* 22×(38+38+37+25) = 3,036 farm-specific terms (υ_{ikt}) *plus* 6×38 = 228 residual terms (u_{it}). Given the dimensions *M* = 2 and *G* = 3 of the support vectors, this amounts to a total of 9,836 probability estimates.

[22] The fixed-coefficients specification of the model would simply imply that all the terms involving the varying components, υ_{ikt} , would have to taken out from the entropy objective (9) and the constraint set (10)-(15).

[23] Given the bounded nature of the GCE estimator used here, it may not be asymptotically normal (i.e., the parameter space is restricted under the alternative). As a result, drawing statistical inferences based on asymptotic standard errors, such as in Golan *et al*. (2001), is not admissible (e.g., Andrews, 1996).

[24] The densities are calculated non-parametrically, by using a Gaussian kernel with an adaptive bandwidth parameter set as $h = 0.9\,\gamma T^{-0.2}$ and γ = min{standard deviation of $\hat{\beta}_{ikt}$, inter-quartile/1.34}.

[25] These kernel density results may raise some doubts on the various tests employed in the literature (e.g., Dixon *et al*., 1992) for testing the dependence between the coefficients and the regressors, which usually assume normally-distributed random components.

[26] To conserve space, this paper only reported the results for the full sample. The results for the 35 farms with positive GVA are available from the authors upon request.