Estimating input allocation from heterogeneous data sources: A comparison of alternative estimation approaches

Kamel Louhichi[*]**, Florence Jacquet and Jean Pierre Butault**

Abstract

This paper proposes the use of the Generalized Maximum Entropy (GME) method to estimate input allocation in multi-crop systems using heterogeneous data sources (farm accountancy data and cropping practices survey data). The aim is to explore the role of well-defined a priori information in improving the accuracy of GME estimation. The performance of the GME method is compared afterward to a Bayesian approach—Highest Posterior Density (HPD)—to assess their accuracy when reliable non-sample (prior) information is used and investigate their usefulness for reconciling heterogeneous data sources. Both approaches are applied to a given set of farm accounting data which reports information on input allocation between alternative input uses. The estimation results show that the use of well-defined prior information from external data source improves GME estimates even though this performance is not always significant. It also appears that the Bayesian (HPD) approach could be a good alternative to the GME estimator. HPD provides results that are close to the GME method with the advantage of a straightforward and transparent implementation of the a priori information.

Keywords: *Input allocation; prior information; Generalized Maximum Entropy; Highest Posterior Density*

Introduction

Over the last decade, there has been a substantial increase in the demand for tools to assess the impact of EU (European Union) policies and technological innovations on agricultural sustainability. In fact, knowing how farmers' decision making would impact crop-level input use and which policy instrument could be used to influence this decision are important issues from a policy-maker perspective. However, the information on input output coefficients (or cost-allocation coefficients) needed to capture policy impacts and to represent technologies in an explicit way is not available from the Farm Accountancy Data Network (FADN). FADN data provides only total costs and total input use per input category, without indicating the input use (and unit costs) of each (crop and animal) output. To overcome this lack of information, most studies in the EU have used either linear programming, based on the minimisation of the sum of the absolute residuals (Koenker and Basset, 1978), or regressions approaches (multiple-regression, OLS technique or Generalized Least Squares)[1] (Ray, 1985; Errington, 1989; Bureau and Cyncynatus, 1991). The difficulty of linear programming is that it leads to unacceptable and corner solutions or zero values. The limit of regression approaches, as pointed out by Midmore (1990), is their incapacity to ensure the non-negativity of the estimated input coefficients. To deal with this, Moxey and

[*] INRA-UMR Economie publique, klouhichi@grignon.inra.fr

[1] For an overview see Léon et al. (1999).

Tiffin (1994) suggested applying the Bayesian estimation while Léon et al. (1999) argued the use of the Generalised Maximum Entropy (GME) estimation. Both approaches are based on the application of a set of restrictions to ensure the non-negativity and the adding up of input coefficients (Gocht, 2008). Léon et al. (1999) compared these two approaches and found that is difficult to discriminate between them. They also pointed out that the GME estimated coefficients are very sensitive to the design of either parameter and/or error support set and further encouraged researchers to be very careful about selecting suitable support values. Along these same lines, Paris and Caputo (2001) indicated that if the parameter estimates are sensitive to the *a priori* information (i.e., support points), then it is probable that policy implications will be affected too.

In this paper, we further investigate this issue by using additional information from an external data source to set *informative*[2] *(more reliable) a priori* expectations about the estimated parameters. The aim is to explore the role of well-defined *a priori* information in improving the accuracy of GME estimation. More specifically, we test if the use of non-sample information from existing cropping practice surveys to set support points sufficiently improves the reliability of the input allocation estimation from FADN data, or, on the contrary, is the improvement negligible, in which case it might be better to define vague (ad hoc) support points to save time and resources. In fact, if the use of additional information from a data source external to FADN plays a role in improving the accuracy of estimation, this procedure could be duplicated throughout the EU, relying on already existing cropping practice surveys at the national level in most EU member states. The second aim of this paper is to compare the performance of the GME method to the Highest Posterior Density (HPD) estimator which makes it possible to incorporate *a priori* information into the estimation process. We seek to evaluate the accuracy of these two approaches when informative *a priori* information is used and to investigate their usefulness for reconciling heterogeneous data sources in a theoretically sound way.

In section 2, the two alternative estimation approaches are described. In section 3, the GME approach is applied to a given set of farm accounting data in a French region (a sample of 533 farms located in the Department of Meuse, France) to examine the sensitivity of model outcomes to the design of support values and illustrate the robustness of the GME approach when *informative a priori* information is available. Two sets of supports values were tested: the first one involves *uninformative (i.e., ad hoc) a priori* information and the second one is based on *informative a priori* information from the Cropping Practices Survey Data (Agreste, 2006). The incorporation of the prior was done only through the support points (i.e., not as new constraints in the system) in order to increase the impact of sample data on the estimation process. In section 4, the outcomes of the GME approach are compared with those derived from HPD estimator. In the final section, we discuss the interest of *informative a priori* information in making estimation potentially more efficient and the usefulness of the GME approach to reconcile heterogeneous data sources.

[2] An *informative prior* expresses specific, definite information about a variable. In the opposite, an uninformative *prior* expresses vague or general information about a variable (Gelman et al, 2003).

Estimation approaches: specification and formulation

A literature review has revealed that numerous modelling approaches have been developed to predict crop-specific input information from data involving farm-level input use and crop-level land use (Chambers and Just, 1989; Just et al., 1990; Shumway et al., 1984; Just et al., 1983; Ray, 1985; Errington, 1989; Midmore 1990; Moxey and Tiffin, 1994; Lence and Miller, 1998a; Lence and Miller, 1998b; Léon et al., 1999; Hansen and Surry, 2006; Gocht, 2008). The common used form for representing input-output relationships in these approaches is the linear form with noise represented as follows:

$$x = \mathrm{Ay} + \mathrm{u} \tag{1}$$

where **x** is a (m×1) vector of total input use in monetary terms, **y** is a (n×1) vector of monetary output, **A** is a (m×n) matrix of unknown input output coefficients (defined by a_{ij} which represent the amount of input ***i*** required per unit of output ***j***), and **u** is an (n×1) vector of noise or error term distributed randomly (Errington, 1989). This linear function imposes a common technology on the whole sample.

In some cases, expenditures on some input uses are equal to zero (or missing) for certain farms. In this case Golan et al. (1996) advise the use of Tobit variant of the linear static model in which the observations are ordered as follows:

$$x = \mathrm{Ay} + \mathrm{u} \Rightarrow \begin{cases} \mathrm{x} = \mathrm{Ay} + \mathrm{u} & if\ x > 0 \\ \mathrm{x} \geq \mathrm{Ay} + \mathrm{u} & if\ \mathrm{x} = 0 \end{cases} \tag{2}$$

To solve this linear inverse problem with noise, several modelling approaches have been developed. In this investigation, we have focused on two alternative approaches: Generalized Maximum Entropy (GME) and Highest Posterior Density (HPD) estimators. The common specification of both approaches is the incorporation of *a priori* information into the estimation process. The problem is that this *a priori* information is often limited or unavailable and in such cases the analyst has to decide to use uninformative (ad hoc) prior or apply subjective *a priori* expectations defined as a weighted average of support values[3]. Our purpose is, therefore, to assess the consequences of these different ways of setting priors on model outcomes, and, in turn, to show how *well defined a priori* information can improve the accuracy and the reliability of forecasting.

Generalized Maximum Entropy (GME)

The Maximum Entropy (ME) principle is used in a wide variety of fields to estimate and make inferences when information is incomplete, highly scattered, and/or inconsistent (Kapur and Kesavan, 1992). The philosophy underlying this approach is to uses *all*, and *only* the information available for the estimation problem at hand (Jaynes, 1957). It provides a more flexible framework that can handle the use of all available information, regardless of how scarce and incomplete it is, along with empirical knowledge to predict the most reliable outcome (De Fraiture, 2003). Golan et al. (1996) have proposed the Generalized Maximum Entropy (GME) approach based on the ME principle to overcome two empirical problems that hamper traditional econometrics for

[3] Generally, prior expectations are defined as a weighted average of support values. In the GME case, the weights are probabilities following a uniform distribution (Heckelei and Britz, 2000).

parameter estimation: multi-collinearity and ill-posed problems (i.e., when the number of parameters to estimate is greater than the number of observations). This approach allows empirical specification and estimation of underdetermined models as well as inclusion of prior knowledge in a technically straightforward way, making estimates potentially more efficient (Jansson, 2007). Apart from solving traditional estimation problems, the GME approach has been used to deal with three well-known issues. The first one is to allocate input to production activities from data involving total input use (Lence and Miller, 1998a; Miller and Plantinga, 1999; Zhang and Fan, 2001). The second is to disaggregate technological and economic data (e.g., Howitt and Reynaud, 2003) and the last one is to fill gaps and reconcile conflicting data sources (Robillard and Robinson, 2003).

The application of the GME to the linear estimation problem is based on the re-parameterization of the unknown vectors as:

$$\begin{aligned} A &= pZ \\ u &= wV \end{aligned} \tag{3}$$

where **Z** and **V** are the matrix of parameter and error support points provided by the user based on previous research, economic theory, researcher intuition or other knowledge sources, and **p** and **w** are the vector of unknown probabilities which are determined by solving the following maximum entropy measure:

$$H[p, w] = -p' \ln p - w' \ln w \tag{4}$$

Subject to

$$x = y(pZ) + (wV) \tag{4.1}$$

$$1' p_k = 1 \ \forall k; \qquad 1' w_{k'} = 1 \ \ \forall k' \tag{4.2}$$

$$p_k \geq 0 \ \ \forall k; \qquad w_{k'} \geq 0 \ \ \forall k' \tag{4.3}$$

where **K** and **K’** are the number of support points associated to unknown parameters and error term, (4.1) is the data-consistency constraint, (4.2) is the adding-up or normalization constraint which ensures that probabilities appropriately sum to one and (4.3) is the non-negativity condition. The objective function (4) attains an unconstrained maximum when all elements of **p** and **w** have, respectively, the value 1/K and 1/K’, that is to say when the probabilities are uniform.

The general formulation of the linear Tobit model using the GME formalism can now be stated as:

$$\max H(p, w_1, w_2) = -\sum_k \sum_i \sum_j p_{kij} \ln p_{kij} - \sum_f \sum_{k'} \sum_i w_{1fk'j} \ln w_{1fk'i} - \sum_f \sum_{k'} \sum_i w_{2fk'j} \ln w_{2fk'i} \tag{5}$$

Subject to:

Data-consistency constraints

$$\begin{cases} x_{fi} = \sum_j a_{ij} y_{fj} + u_{1fi} & \forall i \ and f = 1, \dots, F_1 \\ x_{fi} \geq \sum_j a_{ij} y_{fj} + u_{2fi} & \forall i \ and f = 1, \dots, F_2 \end{cases} \tag{5.1}$$

$$a_{ij} = \sum_k p_{kij} z_{kij} \qquad \forall i, j \tag{5.2}$$

$$u_{1fi} = \sum_{k'} w_{1fk'i} v_{1fk'i} \qquad \forall i \ and f = 1, \dots, F_1 \tag{5.3}$$

$$u_{2fi} = \sum_{k'} w_{2fki} v_{2fk'i} \quad \forall i \; and f = 1,\dots,F_2 \tag{5.4}$$

Adding-up or normalization constraints

$$\sum_{k} p_{kij} = 1 \quad \forall i,j \tag{5.5}$$

$$\sum_{k'} w_{1fk'i} = 1 \quad \forall i \; and f = 1,\dots,F_1 \tag{5.6}$$

$$\sum_{k'} w_{2fk'i} = 1 \quad \forall i \; and f = 1,\dots,F_2 \tag{5.7}$$

Accounting restriction

$$\sum_{i} a_{ij} = 1 \tag{5.8}$$

Non-negativity conditions

$$\mathrm{a}_{ij} \geq 0; \quad \mathrm{p}_{kij} \geq 0; \; \mathrm{w}_{1fk'j} \geq 0\,; \mathrm{w}_{2fk'j} \geq 0 \tag{5.9}$$

where $\mathbf{x}_{\mathrm{fi}}$ is the total cost of input *i* paid by farm *f*, $\mathbf{y}_{\mathrm{fj}}$ is the total value of output *j* produced by farm *f*, $\mathbf{a}_{\mathrm{ij}}$ is the unknown input output coefficients which represent the amount of input *i* required per unit of output *j*, $\mathbf{u}_{\mathrm{fi}}$ (i.e., $\mathbf{u_1}$ and $\mathbf{u_2}$) is the error term which is specific to each input *i* and to each farm *f*. $\mathbf{F_1}$ are the farms with positive observations for input *i*, $\mathbf{F_2}$ are the farms with zero observations and $F_1 + F_2 = F$. **K** and ***K'*** are the numbers of discrete support points, $\mathbf{z}_{\mathrm{fi,k}}$ and $\mathbf{v}_{\mathrm{fi,k'}}$ (i.e., $\mathbf{v_1}$ and $\mathbf{v_2}$) are the matrices of the support points and $\mathbf{p}_{\mathrm{ij}}$ and $\mathbf{w}_{\mathrm{fi}}$ (i.e., $\mathbf{w_1}$ and $\mathbf{w_2}$) are their unknown probabilities, respectively. The two extreme support values for each parameter and error term constitute the support bounds. This model runs under the assumption of common technology for whole sample farms (i.e., $\mathbf{a}_{\mathrm{ij}}$ are common across all farms even for those with zero input expenditures).

The principle of Maximum Generalized Entropy consists of selecting values of **A** and **u** whose distributions **p** and **w** maximize the function **H** in (5), subject to the data-consistency constraints (5.1-4), the normalization constraints (5.5-7), the accounting constraint (5.8) and the non-negativity condition (5.9). The additional accounting restriction (5.8) is imposed for each type of output *j* in order to ensure that total cost and total revenue at farm level are equal. Doing so means that all the inputs are taken into account *simultaneously*[4]. This is achieved by introducing a residual input category 'value added' as suggested by Léon et al. (1999) with corresponding monetary input coefficients equal to the difference between the total revenue and the sum of all other monetary input coefficients across input categories. Similar to other input categories, value added is restricted to be positive, assuming that, for each type of output *j*, averaged (across all farms) total cost cannot exceed total revenue. The solution of this optimization problem yields values for **p** and **w**, which are used to compute the unknown parameters **A** and the error term **u**.

The main advantages of using GME estimation method are its desirable properties such as: it does not require distributional error assumptions; it may be used with small samples and with many highly correlated covariates; it allows imposing nonlinear and inequality constraints. Despite these advantages, the performance of GME approach

[4] Considering the linear models in (1) simultaneously, for each farm *f*, implies that we assume that the errors *u* are "contemporaneously" correlated (i.e., for each individual farm), but uncorrelated across the farms (Peeters and Surry, 2000).

remains extremely sensitive to the design of support points (i.e., the number and the value (or "spacing") of supports) as it can strongly impact estimation results. In this study, we attempt to tackle this problem by using transparent and informative *a priori* information to define support points.

Highest Posterior Density (HPD)

The Highest Posterior Density estimation was proposed by Heckelei et al. (2005) as an alternative to entropy methods for deriving solutions to underdetermined system of equations. They argued that the main advantage of this approach is that it allows a more direct and straightforwardly interpretable formulation of available *a priori* information and a clearly defined estimation objective. HPD estimation is a Bayesian approach, in which the model parameters are treated as stochastic outcomes. In this context, the method distinguishes between the prior density p(A), which summarizes *a priori* information on parameters and the Likelihood function L(A|x), which represent information obtained from the data in conjunction with the assumed model. The combination of the prior density and the Likelihood function results in posterior density which can be expressed as (e.g. Zellner 1971, p.14).

$$h(A|x) \propto (p(A)L(A|x))$$

where **h** denotes posterior density, $\propto$ is the proportionality, **A** is the coefficient matrix to estimate and **x** is the vector of total input use.

The value of **A** that maximizes **h(A|x)** taking into account the data-consistency constraints (5.1) is the *Highest Posterior Density* (HPD) estimate of A. Thus, the main difference between entropy approaches and the Bayesian approach (HPD) is that the entropy techniques do not need to pre-specify and regularize a Likelihood function.

Empirical data: sample and non-sample (prior) information

The given two approaches are applied to a sample of 533 farms located in the Department of Meuse (France) for the year 2006. The advantage of this dataset is that the input costs per production activity are available (Table 1). These input costs are used to validate the results of estimation methods and are not included in the estimation process. As can be seen in Table 1, the data is distinguished according to five input categories, including value-added (I = 5) and eight outputs (J = 8). All input costs and outputs are expressed in monetary terms (in Euros).

Table 1. Observed input costs (as an average across all sample farms)

	Wheat	Winter barley	Spring barley	Maize grain	Peas	Rape	Sunflower	Rape for biodiesel
Input	144(37)	136(36)	111(41)	120(54)	35(39)	165(49)	59(34)	165(51)
Seed	52(21)	55(23)	56(18)	150(49)	52(40)	29(16)	96(32)	29(15)
Pesticide	135(41)	133(39)	74(30)	78(36)	90(43)	188(52)	87(36)	185(50)
Insurance	7(5)	8(5)	8(5)	9(8)	11(8)	23(12)	16(13)	23(12)

Note: Standard deviations of variables are given in parenthesis *Source: Meuse database, 2006*

To make GME estimation of the input coefficients operational, we need to define support points for the unknown input output coefficients as well as for the error vector.

As pointed out by several studies (Golan et al, 1996; Paris and Howitt, 1998; Heckelei and Britz, 2000), the determination of support points in the context of GME is an important issue, as it can strongly affect model outcomes. To define the number of support points, their bounds, spacing, and the implied *prior* expectation, we have made the following assumption:

- For the error term **u** (i.e., $\mathbf{u_1}$ and $\mathbf{u_2}$) we use the common assumption where three support points (i.e., K' = 3) are symmetrically defined around zero and bounded by the so-called "three-sigma rule" (Pukelsheim, 1994).

$$v_i = \left[-3\hat{\sigma}_i, 0, +3\hat{\sigma}_i\right] \ \forall i$$
$$\textit{Where}$$
$$\hat{\sigma}: \text{sample standard deviation of each input category } i \qquad (6)$$

- For the residual "added value" input category (i.e., i=added value), we follow the Léon et al' (1999) proposal in which 11 support points (i.e., K=11) are chosen, bounded between zero and one and equally spaced with a distance of 0.1 (i.e., we assume *a priori* expectation of 0.5 because this category incorporates the remuneration of all fixed factors which can easily account for up fifty percent of the total revenue for each products).

$$z_{"add\_value",j} = s_{"add\_value",j}$$
$$\textit{where} \qquad (7)$$
$$s_{"add\_value",j} = [0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1] \quad \forall j, i = add\_value$$

- For all other input categories, 11 discrete support points with uniform distribution are selected. The corresponding values for these support points are defined in two different ways in order to compare GME outcomes with and without informative priors.
  - Firstly, termed **GME_NP**, the support values are arbitrary defined with wide bounds following the Golan et al. assumption when no prior knowledge is available. They argue that "wide bounds may be used without extreme risk consequences if our knowledge is minimal and we want to ensure that **Z** contains **A**. Intuitively, increasing the bounds increases the impact of the data and decreases the impact of the support" (Golan et al., 1996; p.138). Hence, the selected support values are bounded between zero and one and equally spaced with a distance of 0.1, since the unknown parameter **A** falls within this interval.
  - Secondly, termed **GME_WP**, we use information on input allocation from data source external to sample data to define the support values. This data source is the Cropping Practices Survey Data *"Enquêtes pratiques Culturales"* (CPSD). It is a French survey and database that contains information on input use for major field crops taking into account the heterogeneity in terms of soil type. It has been carried out by the SSP *(Service de la statistique et de la prospective; Ministère de l'Agriculture et de la Pêche)* every four years since 1986. In 2006, this survey covered around 11 arable crops grown in 18,000 fields located in 21 administrative regions (Agreste, 2006). As no input prices are included in this database, a calculation procedure was developed to compute input costs by crop using average input prices drawn from Teyssier (2005). The calculated averaged input costs, reported in Table 2, are then used to compute the averaged input

output matrix A° (i.e., $a°_{ij}$) of the CPSD sample. This matrix is, in turn, used as *informative prior* to set: the center of the support points at $a°_{ij}$ (i.e., expected values of the estimated parameters; also called *a priori* expectations); the bounds at $\lfloor a°_{ij} \pm \sigma \rfloor$; and the spacing at $\lfloor (a°_{ij} \pm \sigma)/5 \rfloor$. The standard deviation (σ) was set to 0.5 for all inputs and products as the CPSD does not provide straightforward means to calculate it. To assess the impact of this value on model outcomes, a sensitivity analysis was carried out and presented in the results section.

$$
\begin{aligned}
& z_{ij} = s_{i,j} a°_{ij} \\
& \textit{where} \\
& s_{i,j}: [0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5] \ \forall j, i \neq add\_value \\
& a°_{ij}: \text{ exp}\textit{ected input output coefficients derived from CPSD}
\end{aligned}
\tag{8}
$$

Table 2. Input costs derived from Cropping Practices Survey

	Wheat	Winter barley	Spring barley	Maize grain	Peas	Rape	Sunflower	Rape for biodiesel
Input (€/ha)	145	138	124	144	44	167	58	167
Seed	50	65	65	125	80	37	86	37
Pesticide	99	138	178	81	121	158	93	158
Insurance	7	8	8	9	11	23	16	23

Source: Cropping Practices Survey Data

To apply the HPD estimation to our case we followed the Heckelei et al., (2005) assumption. Furthermore, in order to make methods comparable we used the same prior information as in GME estimator. Let us assume that the matrix A° (i.e., $a°_{ij}$) drawn from the CPSD is the prior mean and that the prior density function has the following form: vec(**A**) ~ N(vec(**A°**),Σ). The covariance matrix Σ is set equal to a diagonal matrix with elements $(\text{vec}(\mathbf{A°})\sigma)^2$, the square taken element-wise. σ is the standard deviation and it was set to 0.5 for all inputs and products to ensure consistency with GME approach. Taking natural logs and restricting the objective function to terms that are relevant for the optimization leads to the following estimation problem:

$$
\begin{aligned}
& \min \ [\text{vec}(\mathrm{A} - \mathrm{A°})]' \, \Sigma^{-1} [\text{vec}(\mathrm{A} - \mathrm{A°})] \\
& \text{subject to:} \\
& \begin{cases} \mathrm{x} = \mathrm{Ay} + \mathrm{u} & if \ \ \mathrm{x} > 0 \\ \mathrm{x} \geq \mathrm{Ay} + \mathrm{u} & if \ \mathrm{x} = 0 \end{cases} \\
& \mathrm{l'A} = 1
\end{aligned}
\tag{9}
$$

where $\sum^{-1}$ is the covariance metrics, A is the coefficient matrix to estimate, A° is the prior and u is the error term. The solution of this optimization problem yields values for the unknown parameters **A** and the error term **u**.

Estimation results and discussion

This section presents the results of the estimation experiments in relation to the different hypotheses formed in the previous section. These experiments were programmed in GAMS (*General Algebraic Modeling System*) language and were solved numerically using the solver CONOPT[5]. Three estimation experiments were performed and compared: (i) GME estimation - with *informative a priori* information (GME_WP); (ii) GME estimation – with *uninformative a priori* information (GME_NP); and (iii) HPD estimation - with *informative a priori* information.

First, we compare the results of GME models with and without *informative a priori* information to show the role of well-defined prior in improving the reliability of estimation. Then, the sensitivity of the GME_WP outcomes to support bounds is presented. Finally, the prediction accuracy of the two alternative estimators (GME and HPD) with *informative a priori* information are assessed and discussed.

Accuracy criterions

To examine the prediction accuracy of the alternative estimation approaches, we use three familiar criterions: the Weighted Absolute Percentage Error (WAPE), the Pearson's correlation coefficient (R) and the Normalized entropy criterion. Their general specifications and formulas are presented below.

1. Weighted Absolute Percentage Error (WAPE) measures the accuracy of fitted values. It has been widely used as a performance measure in forecasting since it is easy to interpret and understand. Since, this is a weighted measure; it does not have the same problems as Mean Absolute Percentage Error (MAPE) such as over-skewing due to very small or zero observed values. It usually expresses accuracy as a percentage, and is defined by the following formula.

$$WAPE = \frac{\sum_{i=1}^{N} \left| \frac{x_i - \hat{x}_i}{\hat{x}_i} \right| \hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i} * 100 \qquad (10)$$

where

x_i : *predicted value*

$\hat{x}_i$: *observed value*

N : *number of observations*

Forecasting is best when WAPE is close to 0.

2. Pearson's correlation coefficient (R) is a measure of the strength and the direction of the linear relationship between two measurable variables. It is always between -1 and +1, where -1 means a perfect negative, +1 a perfect positive relationship and 0 means the perfect absence of a relationship. Pearson's correlation is calculated according to various formulas. In this application, we apply the commonly used formula that follows:

[5] Full details are available from the authors.

$$R = \frac{\sum x_i \hat{x}_i - \frac{\sum x_i \sum \hat{x}_i}{N}}{\sqrt{\left[(\sum (x_i)^2 - \frac{(\sum x_i)^2}{N})(\sum (\hat{x}_i)^2 - \frac{(\sum \hat{x}_i)^2}{N}) \right]}} \quad (11)$$

where

x_i : *predicted value*

$\hat{x}_i$: *observed value*

N : *number of observations*

3. Normalized entropy (information) criterion (S) is a measure of the relative information content of the estimated parameters. The normalized entropy measure for I×J unknown parameters (I inputs and J outputs) and K the number of support points, is defined by Golan et al. (1996b, p. 93) as follows:

$$S(p) = \frac{-\sum_{ijk} p_{ijk} Lnp_{ijk}}{IJLn(K)} \quad (12)$$

where

$s(p) \in (0,1)$

$IJLn(K)$ represent maximum uncertainty

where p is the probabilities of supports for parameters a_{ij} and K the corresponding number of support points. S(p) ranges between zero and one; S(p) = 0 reflects perfect knowledge in the parameter distribution and S(p) = 1 corresponds to an uninformative uniform distribution (i.e., it reflects complete ignorance about the parameter distribution). Similar normalized measures reflecting the information in each one of the *i, j* distributions can also be defined. According to Léon et al. (1999), the greater the normalized entropy measure S(p), the better the estimator. This means that the "superior" model would yield a solution for the recovered cost-allocation coefficients that is more "uniform" or closer to the *a priori* expectations defined by the support values.

Along similar lines, the informational content of the noise or error component can be assessed through the normalized entropy measure, for *I×F* errors (*I* inputs and F farms) and K' the number of support points for the error term, defined by Golan et al. (1996b, p.93) as follows:

$$S(w) = \frac{-\sum_{fik'} w_{fik'} Lnw_{fik'}}{IFLn(K')} \quad (13)$$

Where

$s(w) \in (0,1)$

$IFLn(K')$ represent maximum uncertainty

The normalized entropy criterion (S) cannot be computed for the HPD approach, as it does not employ the principle of support points.

These indicators are complementary, as they measure at the same time the deviation (WAPE), the strength and the direction of the relationship (R) as well as the degrees of uniformity of distribution (S) between the observed and the predicted values.

GME results with and without informative prior

The results of GME estimations with and without *informative* prior are presented in Table 3. In line with our expectations, the GME with *informative a priori* information (GME_WP) performs better, measured by the sum of the Weighted Absolute Percentage Error (WAPE) over all input categories and products (i.e., outputs). The generated WAPE from GME_WP model is smaller by around 17%. The performance of GME_WP model prevails as well when attention is paid to individual input categories. It has lowest WAPE for all input categories, except for pesticide.

Table 3 also depicts the Pearson's correlation coefficient between estimated and observed input allocation. It appears that the two models provide good fits, as they outcome strong positive correlations for all input categories exceeding 0.85. It illustrates as well that the generated correlations by input category are very close across models; it is thus difficult to judge their relative performance using this criterion at this aggregated level.

Table 3. Performance of the two GME models in predicting input allocation

	Weighted Absolute Percentage Error (WAPE)		Pearson's correlation coefficient (R)	
	GME_WP	GME_NP	GME_WP	GME_NP
Fertilizer	25.9606	30.3481	0.9425	0.9443
Seed	37.5651	51.8074	0.8728	0.8773
Pesticide	31.3911	28.6614	0.9337	0.9379
Insurance	33.9981	45.3990	0.8720	0.8762
Added value	16.5267	19.8749	0.9765	0.9772
Sum	145.4417	176.0907	4.5975	4.6128

Source: model results

Despite mixed results at the aggregated levels, we can infer a consistently superior performance of the GME_WP model regarding the single input category and product. This is shown in Table 4 which reports the deviation of estimated coefficients by input category and product averaged across all farms from their observed counterparts. From this Table, it appears that the WAPE are smaller for the GME_WP model in 32 out of the 40 cases (i.e., 80% of the cases).

Table 4. Weighted Absolute Percentage Error between observed and estimated input coefficients by input category and product

		Observed Input Coefficients	Weighted Absolute Percentage Error (WAPE)	
			GME_WP	GME_NP
Fertilizer	Wheat	0.172	27.11	30.61
	W. barley*	0.194	39.79	54.03
	S. barley	0.166	31.57	27.40
	Maize grain*	0.166	35.52	46.39
	Peas	0.066	16.85	18.76
	Rape	0.206	25.82	29.23
	Sunflower*	0.110	37.59	53.83
	Rape for biodiesel	0.226	29.68	26.92
Seed	Wheat	0.064	34.10	42.68
	W. barley	0.079	16.45	18.28
	S. barley	0.084	25.67	28.93
	Maize grain	0.258	37.78	53.46
	Peas	0.090	31.69	27.43
	Rape	0.036	33.36	42.11
	Sunflower	0.184	16.51	17.90
	Rape for biodiesel	0.040	29.14	30.29
Pesticide	Wheat	0.161	38.29	47.60
	W. barley	0.189	32.25	31.44
	S. barley	0.111	36.25	48.29
	Maize grain	0.108	20.22	21.45
	Peas	0.151	25.50	34.94
	Rape	0.235	47.51	56.33
	Sunflower	0.160	30.92	29.65
	Rape for biodiesel	0.256	32.72	61.09
Insurance	Wheat	0.008	14.62	22.29
	W. barley	0.012	25.81	28.79
	S. barley	0.012	34.60	50.94
	Maize grain	0.011	30.66	26.41
	Peas	0.019	31.77	37.26
	Rape	0.029	16.36	19.12
	Sunflower	0.028	23.73	32.28
	Rape for biodiesel	0.032	31.54	44.49
Add value	Wheat	0.595	34.90	34.50
	W. barley	0.527	35.57	48.19
	S. barley	0.627	15.35	21.99
	Maize grain	0.456	24.90	27.71
	Peas	0.674	33.41	53.79
	Rape	0.494	29.45	25.54
	Sunflower	0.518	32.70	37.17
	Rape for biodiesel	0.445	15.86	19.19
Sum			145.4417	176.0907

For these crops very few observations were available. *Source: model results*

This Table shows also that in most cases the WAPE between observed and estimated input coefficients is low, below 30%, for both models. Exceptions are winter barley, maize grain, peas and sunflower products. This finding is, however, not surprising since only a few observation are available for these three products.

The superiority of GME *with informative prior* is confirmed as well when we examine input allocation by single farm, input category, and product. The GME_WP model performs better in 6,700 out of the 10,863 cases (around 60% of the cases).

The inspection of the normalised entropy criterion reported in Table 5 reveals that the informational content of the error components is invariant across models. This is not unexpected, as the support values for ***w*** have not been changed. However, the normalized entropy of the estimated parameters *S(p)* varies significantly between the two models, and it is greater in the GME with *informative prior* (i.e., GME_WP) showing the relative superiority of this model. Indeed, the normalized entropy measure for the GME_WP is about 0.96 which is closer to one (the upper bound of entropy) while in the case of the GME_NP it is around 0.65. This finding is in line with our intuition that the estimated coefficients would more likely be closer to the *a priori* expected values from the Cropping Practices Survey Data than to unspecified values between zero and one. To conclude, the use of ad hoc (vague) support points would make the estimation task much easier as it saves time in terms of collecting external (i.e., non-sample) information from other database or studies; however it generates less satisfactory results. It remains to be known whether or not the variation of the support ranges (i.e., interval) has an impact on model estimates when reliable prior are used. This is examined in section 4.3.

Table 5. Normalized entropy criterion

	GME_WP	GME_NP
S(p)	0.9593	0.6539
S(w)	0.9853	0.9860

Source: model results

Sensitivity of GME results to support bounds

The aim of this analysis is to assess the sensitivity of the GME model outcomes to support bounds (i.e., end points) when *informative a priori* expectations (i.e., GME_WP) are used. More specifically, we seek to show whether or not moving away from the *a priori* expectation weakens the GME estimates. To do so, we shift the support bounds from a minimum of 25% (a lower percent does not lead the model to converge) to a 100% maximum of the *a priori* expectation while maintaining the support values equally and symmetrically spaced in each case. Five different designs of the support sets, including the base case design set (design n°3), are selected and described in Table 6. Because no *a priori* information is included for the error term and the residual “added value” category, we shall focus on the input allocation matrix. The results of the sensitivity analysis are compared with the results for the initial support set (base case, design set n°3).

Table 6. Designs of the support set for the input allocation matrix

	Input categories	Number of support points	Type of spacing	Selected support set
Design n°1	Fertilizer Seed Pesticide Insurance	11	Symmetries	±25% a^0_{ij}
Design n°2				±40% a^0_{ij}
Design n°3 (base case)				±50% a^0_{ij}
Design n°4				±75% a^0_{ij}
Design n°5				±100% a^0_{ij}

Table 7 presents the WAPE accuracy indicator for the various support sets as well as its percentage deviation from the base case (design n°3). This it is possible to see that when the prior expectation is well defined, the GME estimator seems insensitive to the change of support ranges. In fact, doubling the support bounds (in comparison to the base case) induces less than a 3.5% change in the WAPE summed over input categories and products. This small change reveals, however, that decreasing the support bounds impacts negatively estimation results (designs n°1 and n°2). Inversely, increasing the bounds slightly improves estimation (designs n°4 and n°5). This is not surprising as the estimated input coefficients are oriented by the *a priori* expectations but not all of them are close to this point. This is consistent with Golan et al (1996) who concluded that "increasing the bounds increases the impact of the data and decreases the impact of the support".

Table 7. Weighted Absolute Percentage Error (WAPE) for different support point designs

	Design n°1	Design n°2	Design n°3 (base case)	Design n°4	Design n°5
Fertilizer	26.36 (1.5)	26.17(0.8)	25.9606	25.82(-0.6)	26.03(0.3)
Seed	39.12(4.1)	39.28(4.4)	37.5651	38.75(3.1)	39.60(5.1)
Pesticide	33.80(7.7)	32.86(4.5)	31.3911	28.52(-10)	26.97(-16.4)
Insurance	33.87(-0.4)	33.88(-0.3)	33.9981	34.42(1.2)	34.99(2.8)
Added value	17.26(4.4)	16.99(2.7)	16.5267	15.90(-4)	15.84(-4.3)
Sum	150.39(3.4)	149.19(2.5)	145.4417	143.41(-1.4)	143.43(-1.4)

Note: numbers in parenthesis are percentages deviations to the base case (design 3) *Source: model results*

This insensitiveness is confirmed by detailing WAPE according to input category and product as shown in Table 8, as well as examining two other indicators (i.e., Normalized entropy indicators (Table 9) and Pearson's correlation coefficient (Table 10). Doubling the support bounds provokes a slight change in the levels of these indicators.

However, from Table 8, it appears that the change in support bounds matters for the products with few observations such as winter barley and peas as well as for those that have a large variability of input costs across farms; i.e., a high standard deviation of input costs (cf. Table 1), such as rape for biodiesel and spring barley.

Table 8. Weighted Absolute Percentage Error (WAPE) for different support point designs

		Design n°1	Design n°2	Design n°3	Design n°4	Design n°5
Fertilizer	Wheat	27.4082	27.2693	27.1097	27.0677	27.3906
	W. barley	41.3832	41.6499	39.7937	40.8840	41.4410
	S. barley	34.1700	33.1559	31.5746	28.4549	26.7032
	Maize grain	35.2570	35.3406	35.5194	36.1421	36.8946
	Peas	17.5497	17.2761	16.8509	16.2989	16.3330
	Rape	26.0048	25.9054	25.8178	26.0079	26.5353
	Sunflower	39.3077	39.6032	37.5938	38.7410	39.6341
	Rape for biodiesel	32.0036	31.0778	29.6754	27.0724	25.8046
Seed	Wheat	33.9240	33.9680	34.0996	34.5506	35.1322
	W. barley	17.2350	16.9252	16.4473	15.7799	15.7362
	S. barley	25.8813	25.7769	25.6740	25.7979	26.2319
	Maize grain	40.0772	40.2681	37.7837	39.8196	41.1127
	Peas	34.5883	33.4314	31.6897	28.3897	26.6383
	Rape	33.1893	33.2245	33.3558	33.7855	34.3385
	Sunflower	17.2892	16.9590	16.5076	15.7101	15.5392
	Rape for biodiesel	30.3974	29.8744	29.1429	27.8740	27.4346
Pesticide	Wheat	38.2961	38.4685	38.2939	39.6159	40.5457
	W. barley	34.2440	33.4667	32.2534	29.8198	28.4009
	S. barley	35.8032	35.9221	36.2464	37.3763	38.7243
	Maize grain	21.2013	20.8870	20.2198	19.3401	19.0385
	Peas	25.4073	25.3980	25.4993	26.0071	26.5557
	Rape	48.9425	48.9448	47.5093	47.2836	46.8341
	Sunflower	33.0321	32.2572	30.9206	28.4345	27.1430
	Rape for biodiesel	32.6979	32.6561	32.7249	32.9411	33.3359
Insurance	Wheat	15.2919	15.0733	14.6189	14.0435	13.9757
	W. barley	26.0674	25.9616	25.8126	25.8364	26.2091
	S. barley	37.3002	37.3155	34.5962	35.9554	37.0673
	Maize grain	33.0396	32.1092	30.6645	27.7899	26.1917
	Peas	31.8069	31.7712	31.7729	31.8486	32.0717
	Rape	16.9985	16.7375	16.3627	15.9292	16.0411
	Sunflower	24.5492	24.1553	23.7272	22.9559	22.4507
	Rape for biodiesel	31.3610	31.6164	31.5391	32.9336	34.2540
Sum		150.39	149.19	145.4417	143.41	143.43

Source: model results

Table 9. Pearson's correlation coefficient (R) for sensitivity designs

	Design n°1	Design n°2	Design n°3 (base case)	Design n°4	Design n°5
Fertilizer	0.9415	0.9419	0.9425	0.9439	0.9448
Seed	0.8742	0.8744	0.8728	0.8734	0.8741
Pesticide	0.9317	0.9325	0.9337	0.9364	0.9384
Insurance	0.8706	0.8711	0.8720	0.8742	0.8761
Added value	0.9760	0.9762	0.9765	0.9771	0.9775
Sum	4.5940	4.5960	4.5975	4.6051	4.6110

Source: model results

Table 10. Normalized entropy indicators for sensitivity designs

	Design n°1	Design n°2	Design n°3 (base)	Design n°4	Design n°5
S(p)	0.8892	0.9086	0.9593	0.9738	0.9783
S(w)	0.9852	0.9853	0.9853	0.9857	0.9860

Source: model results

Comparison of alternative estimation models

This section presents the empirical results of the two alternative models: GME and HPD. According to the WAPE accuracy indicator, reported in Table 11, the two approaches come up with very similar results. The GME_WP slightly outperforms the HPD model while looking to the WAPE summed over all farms, input categories, and products. However, its performance does not prevail as, on the one hand, the differences between WAPEs are not significant (i.e., less than 5%) and, on the other hand, the lowest WAPE varies between models when attention is paid to individual input categories.

Regarding the Pearson's correlation coefficient (R), the obtained results show that the values are higher than 0.850 for all inputs and with the two model specifications, showing the likeness and the good predictive power of the used methods.

Table 11. Performance of the two alternative models in predicting input allocation

	Weighted Absolute Percentage Error (WAPE)		Pearson's correlation coefficient (R)	
	GME_WP	HPD	GME_WP	HPD
Fertilizer	25.9606	27.1524	0.9425	0.9402
Seed	37.5651	36.9203	0.8728	0.8714
Pesticide	31.3911	37.9461	0.9337	0.9292
Insurance	33.9981	33.6610	0.8720	0.8699
Added value	16.5267	18.7598	0.9765	0.9755
Sum	145.4417	154.4397	4.5975	4.5863

Source: model results

The examination of model outcomes by input category and product confirms the quite similar performance of the two alternative approaches. In fact, according to the WAPE indicator reported in Table 12, the GME_WP outperforms the HPD approach in 60% (i.e., 24 out of 40) of the cases, but in most of these cases the differences are very slight in absolute terms.

The likeness of both approach is also confirmed when we compare the bias by single farm, input category, and product. The GME approach outperforms in only 6,191 out of the 10,843 cases (about 55%) compared to HDP.

From this comparison, it appears that the Baysian (HPD) approach could be a good alternative to GME estimators as it gives results close to GME with a straightforward and transparent implementation of *a priori* information. As explained in Heckelei et al. (2008), contrary to GME approach which requires a specification of support points and their reference distributions, as well as their final weighting implied by the maximum entropy criterion, the implementation of HPD approach necessitates the definition of only the *a priori* expectations which is much easier and significantly reduces the computational demand.

Table 12. Weighted Absolute Percentage Error between observed and estimated input coefficients for alternative approaches

		Observed Input Coefficients	Weighted Absolute Percentage Error (WAPE)	
			GME_WP	HPD
Fertilizer	Wheat	0.172	27.11	28.12
	W. barley*	0.194	39.79	38.81
	S. barley	0.166	31.57	38.55
	Maize grain*	0.166	35.52	35.11
	Peas	0.066	16.85	18.98
	Rape	0.206	25.82	26.52
	Sunflower*	0.110	37.59	36.41
	Rape for biodiesel	0.226	29.68	36.08
Seed	Wheat	0.064	34.10	33.77
	W. barley	0.079	16.45	18.90
	S. barley	0.084	25.67	26.42
	Maize grain	0.258	37.78	35.63
	Peas	0.090	31.69	39.43
	Rape	0.036	33.36	33.05
	Sunflower	0.184	16.51	19.05
	Rape for biodiesel	0.040	29.14	32.21
Pesticide	Wheat	0.161	38.29	38.14
	W. barley	0.189	32.25	37.97
	S. barley	0.111	36.25	35.23
	Maize grain	0.108	20.22	22.43
	Peas	0.151	25.50	25.49
	Rape	0.235	47.51	48.46
	Sunflower	0.160	30.92	36.80
	Rape for biodiesel	0.256	32.72	32.50
Insurance	Wheat	0.008	14.62	16.94
	W. barley	0.012	25.81	26.59
	S. barley	0.012	34.60	33.94
	Maize grain	0.011	30.66	36.79
	Peas	0.019	31.77	31.55
	Rape	0.029	16.36	18.39
	Sunflower	0.028	23.73	26.16
	Rape for biodiesel	0.032	31.54	31.25
Add value	Wheat	0.595	34.90	42.80
	W. barley	0.527	35.57	35.56
	S. barley	0.627	15.35	17.62
	Maize grain	0.456	24.90	25.70
	Peas	0.674	33.41	32.71
	Rape	0.494	29.45	35.17
	Sunflower	0.518	32.70	32.51
	Rape for biodiesel	0.445	15.86	17.76
Sum			145.4417	154.4397

Source: model results

Conclusion

In this paper, in a first step, we assessed the outcomes of the GME approach, with and without informative prior to shed light on the role of well-defined *a priori* information in improving the reliability of input allocation estimations. The GME approach showed better results when consistent *a priori* information (GME_WP) was used, even though this performance is not always significant. The superiority of

GME_WP confirmed mainly at disaggregated level (i.e., when we examined input allocation by single farm, input category and product) and for the products with very few observations. The GME_WP model outperforms in 60% out of over 10,853 cases when compared by farm, input and product. Moreover, the GME estimator seems insensitive to the change in support bounds when the prior expectation is well defined. However, since the accuracy improvement related to the use of reliable prior is not very apparent, it might be better in certain cases, mainly when these prior are not readily available, to define ad hoc (vague) support points (i.e., GME without prior) to save time and resources. That is, the trade-off between accuracy of estimations, cost of data collection, and the computational burden related to each approach has to be considered by the model developer according to context. This is in line with the proposal of Howitt and Reynauld (2003), who suggested the use of uninformative uniform distributions as prior when no a priori information on estimated parameters is available.

In a second step, we compared the outcomes of GME to a Bayesian approach (HPD) which makes it possible to incorporate *a priori* information into the estimation process. The main finding is that the Bayesian approach could be a good alternative to the GME estimator. In fact, the GME approach outperforms HPD in only 55% of the cases. This means that, not only is the difference among approaches not big enough but there are also a large number of cases (45%) where the HPD gives better estimations. This is in line with Heckelei et al. (2005; 2008) who suggested the use of this approach instead of GME to solve undetermined system of equations. HPD provides results that are close to the GME method with the advantage of a straightforward and transparent implementation of the *a priori* information. In addition, from this investigation it appears that both the GME and the HPD approaches are very useful for reconciling heterogeneous data sources (i.e., farm accountancy data and cropping practices surveys) in a theoretically sound way.

Based on our finding that well-defined *a priori* information plays a role in improving the accuracy of GME estimation, we shall implement this procedure at a large scale to estimate input allocation in France using FADN and CPRD data sources. The estimated coefficients will be used in a supply model for policy analysis based on a mathematical programming approach.

Acknowledgements

The authors would like to thank the two anonymous referees and the editor of the review for their helpful comments. They also thank the financial support from the "Systèmes de Production en Grandes Cultures, Environnement Politiques Publiques (POPSY)" (ANR-08-STRA 12) research project, funded by the "Agence Nationale de la Recherche" (ANR).

References

Agreste, 2006. *Enquêtes pratiques Culturales 2006*. Service de la Statistique et de la Prospective (SSP) du MAAP.

Bureau J.C, Cyncynatus, M., 1991. Perspectives d'amélioration de la méthode d'estimation des coûts de production à partir du RICA. In: *les coûts de production des principaux produits agricoles de la Communauté européenne en 1984-1985-1986*, (RICA). INRA/INSEE, Paris.

Chambers, R.G., Just R.E., 1989. Estimating multi-output technologies. *American Journal of Agricultural Economics*, 71, 980-995.

De Fraiture, C., 2003. *The use of entropy optimization principles in parameter estimation: applications to global water and food modeling.* Ph.D. Dissertation submitted to the Faculty of the Graduate School of the University of Colorado, Department of Civil Environmental Architectural Engineering, Colorado, 165p.

Errington, A., 1989. Estimating Enterprise Input Output Coefficients from Regional Farm Data. *Journal of Agricultural Economics*, 40, 52-56.

Gelman, A., Carlin, J. B., Stern, H.S., Rubin, D.B. 2003. Bayesian Data Analysis, 2nd edition, CRC Press.

Gocht A., 2008. Estimating input allocation for farm supply models. In *Modelling of Agricultural and Rural Development Policies.* 107th EAAE Seminar, Sevilla, Spain, January 29th -February 1st, 2008.

Golan, A., Judge, G., Miller, D., 1996. *Maximum entropy econometrics: Robust estimation with limited data*. New York: John Wiley & Sons

Hansen, H., and Surry, Y. 2006. Estimating the cost allocation for German agriculture: an application of the maximum entropy methodology. 46th Annual Conference of German Association of Agricultural Economists, October 4-6.

Heckelei, T., Britz, W., 2000. Positive Mathematical Programming with Multiple Data Points: A Cross-Sectional Estimation Procedure. *Cahiers d'Economie et Sociologie Rurales*, 57, 28-50.

Heckelei, T., Wolff, H., 2003. Estimation of constrained optimization models for agricultural supply analysis based on Generalised Maximim Entropy. *European Review of Agricultural Economics,* 30, 27–50.

Heckelei, T., Mittelhammer, R., Britz, W., 2005. *A Bayesian Alternative to Generalized Cross Entropy*. Parma, Monte Universita Parma Editore, Italy, Parma.

Heckelei, T., Mittelhammer R., Jansson T., 2008. A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models. *Agricultural and Resource Economics, Discussion Paper*, Institute for Food and Resource Economics, Bonn. http://www.ilr1.uni-bonn.de/agpo/publ/dispap/download/dispap08_02.pdf

Howitt, R.E., Reynaud, A., 2003. Spatial disaggregation of agricultural production data using maximum entropy. *European Review of Agricultural Economics*, 30, 359-387.

Jansson, T., 2007. *Econometric specification of constrained optimization models*. Dissertationen der landwirtschaftlichen Fakultät, Universität Bonn. Published online: Download thesis.

Jaynes, E. T., 1957. Information Theory and Statistical Mechanics," *Physics Review,* 106, 620-630.

Just, R., Zilberman, D., Hochman, E., 1983. Estimation of multi-crop production functions. *American Journal of Agricultural Economics* 65, 770-780.

Just, R., Zilberman, D., Hochman, E., Bar-Shira, Z., 1990. Input allocations in multicrop system. *American Journal of Agricultural Economics* 72, 200-209.

Kapur, J. N., Kesavan, H.K., 1992. *Entropy Optimization principles with applications*. Academic Press, Inc., San Diego, CA, 405p.

Koenker, R., Basset, G., 1978. Regression Quantiles. *Econometrica*, 46, 33-50.

Lence, S. H., Miller, D. J., 1998a. Recovering Output-Specific Input from Aggregate Input Data: A Generalized Cross-Entropy Approach. *American Journal of Agricultural Economics* 80, 852-867.

Lence, S. H., Miller D. J., 1998b. Estimation of Multi-Output Production Functions with Incomplete Data: A Generalized Maximum Entropy Approach. *European Review of Agricultural Economics*, 25, 188-209.

Léon Y., L. Peeters, Quinque M., and Surry Y., 1999. The use of maximum entropy 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.

Miller, D. J., Plantinga A. J., 1999. Modeling land use decisions with aggregated data. *American Journal of Agricultural Economics*, 81, 180-194.

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

Moxey, A., and Tiffin, R. 1994. Estimating linear production coefficients from Farm Business Survey Data: A note. *Journal of Agricultural Economics*, 45, 381-385.

Paris Q., and Caputo M. R., 2001. *Sensitivity of the GME Estimates to Support Bounds*. Working Paper No. 01-008. Department of Agricultural and Resource Economics, University of California, 18p.

Paris, Q., Howitt, R. E., 1998. Analysis of Ill-Posed Production Problems Using Maximum Entropy. *American Journal of Agricultural Economics*, 80, 124-138.

Peeters, L. and Surry, Y., 2000. Incorporating price-induced innovation in a symmetric generalised McFadden cost function with several outputs. *Journal of Productivity Analysis,* 14, 53–70.

Perloff, J. M., Karp, L. S. and Golan, A. 2007. Estimating Market Power and Strategies. Cambridge: Cambridge University Press.

Pukelsheim, F., 1994. The three sigma rule. *American Statistician*, 48, 88–91.

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

Robillard, A. S., Robinson, S., 2003. Reconciling Household Surveys and National Accounts Data using a cross Entropy Estimation Method. *Review of Income and Wealth*, 49, 395-406.

Shumway, C.R., Pope, R.D., Nash, E.K., 1984. Allocatable Fixed Inputs and Jointness in Agricultural Production: Implications for Economic Modeling. *American Journal of Agricultural Economics,* 66, 72-78.

Teyssier D., 2005. Index des prix et des normes agricoles 2005-2006. 21 éditions. Lavoisier, 200p.

Zellner, A., 1971. *Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Zhang, X., Fan S., 2001. Estimating Crop-Specific Production Technologies in Chinese Agriculture: A Generalized Maximum Entropy Approach. *American Journal of Agricultural Economics,* 83, 378-388.