

ESTIMATION D'UN MODÈLE À PARAMÈTRES VARIABLES PAR LA MÉTHODE D'ENTROPIE CROISÉE GÉNÉRALISÉE ET APPLICATION À LA RÉPARTITION DES COÛTS DE PRODUCTION EN AGRICULTURE¹

Ludo PEETERS (), Yves SURRY(**)*

()Limburg University, Hasselt, Belgique*

*(**) Swedish University of Agricultural Sciences, Uppsala, Suède*

Introduction

En économie de la production ou dans l'établissement des tableaux ou modèles entrées-sorties (E/S), on a souvent besoin de désagréger les secteurs multi-produits par catégories de produits ou de ventiler des données agrégées entre plusieurs secteurs ou régions. Ce problème se pose en agriculture. Il revient à estimer des coefficients techniques (dénommés aussi coefficients input/output ou entrées/sorties (E/S)) ou comme dans ce travail, des paramètres de répartition des coûts², à partir de données agrégées. L'information nécessaire pour entreprendre une telle tâche est rarement disponible et dans bien des cas, il faut faire appel à des procédures hybrides pour générer les données recherchées.

Malgré sa simplicité conceptuelle, l'estimation de la répartition des coûts à partir de données agrégées de comptabilité d'entreprises n'est pas aussi facile qu'on pourrait le croire. Ce dernier problème a surtout attiré l'attention des économistes agricoles³ depuis une vingtaine d'années et a donné lieu à la réalisation de plusieurs études. Toutefois, leurs résultats sont peu satisfaisants et/ou théoriquement incohérents (Errington, 1989; Midmore, 1990; Moxey and Tiffin, 1994; Hallam *et al.*, 1999). La plupart de ces études reposent sur des méthodes de

¹ Des versions préliminaires de cette communication ont été présentées à la 10^{ème} Conférence de l'Association Européenne des Économistes Agricoles, Saragosse, 27-28 août 2002 et à la Conférence "Entropy Econometrics" en l'honneur de Arnold Zellner, Washington, 19-21 septembre 2003. Ce travail a été partiellement financé par le projet de recherche "Breizh" (Région Bretagne). Nous aimerions remercier Dominique Desbois (RICA-SCEES) et Yves Léon pour leurs commentaires pertinents. Toutes erreurs restantes sont de la responsabilité des auteurs.

² L'approche de modélisation à paramètres variables que nous proposons dans ce travail est transposable à d'autres situations, telles que la spécification "stochastique" de modèles E/S (Scandizzo, 1990).

³ Signalons cependant que cette question a été abordée par des experts de l'INSEE il y a plus d'une vingtaine d'années (voir Divay et Meunier, 1980). De même, il faut également noter le travail récent de Pingault et Desbois (2003) sur la performance économique et la compétitivité de différents types d'exploitations agricoles dans les pays membres de l'Union européenne.

régressions multiples faisant appel à un large éventail de méthodes d'estimation⁴ et en supposant que les coefficients à estimer sont fixes. Toutefois, cette hypothèse implique que les répartitions des coûts caractérisent une firme idéale "moyenne", exprimant quelques tendances centrales des activités des firmes étudiées. Par conséquent, les variations dans les structures de coûts (ou technologies) ne sont pas permises *a priori*. De plus, il est bien connu (Judge *et al.*, 1985) que l'utilisation d'un modèle à coefficients fixes peut produire, sous certaines conditions, des estimations biaisées qui s'expliquent en partie par les phénomènes suivants : a) effets non observés liés à des variables omises influençant la conduite des entreprises concernées (par exemple caractéristiques des dirigeants de l'entreprise tels que l'âge, le niveau d'éducation, qualités managériales...), b) effets d'agrégation (compositions différentes des intrants et des catégories utilisées de produits) et c) erreurs de mesure (données erronées ou différences dans la qualité des données, etc...). Seuls quelques articles ont explicitement traité de la nature "aléatoire" des coefficients de répartition des coûts (Dixon *et al.*, 1984; Butault *et al.*, 1988; Hornbaker *et al.*, 1989). Cependant, quelle que soit la nature fixe ou aléatoire, ces méthodologies ont en partage une mauvaise qualité de prédiction et/ou des valeurs estimées peu vraisemblables, voire irréalistes⁵.

Cette communication a deux objectifs. Tout d'abord, nous présentons une procédure utilisant la notion d'entropie croisée généralisée (ECG) pour obtenir des coefficients de répartition des coûts à partir de données comptables d'entreprises. La procédure proposée comporte néanmoins un élément novateur : elle permet d'estimer des coefficients de répartition des coûts *spécifiques* à chaque firme. Pour arriver à un tel résultat, nous sommes amenés à estimer un modèle à paramètres variables qui permet de prendre en compte l'hétérogénéité non observable des firmes étudiées. De plus, nous montrons qu'il est possible d'estimer des coefficients de répartition des coûts à l'aide d'une méthode d'estimation de type ECG - similaire à celle proposée, entre autres, par Golan *et al.* (1996) - qui respecte les différentes contraintes résultant des règles comptables sur les coûts. Le deuxième objectif de cette étude est d'appliquer

⁴ Les méthodes d'estimation le plus souvent utilisées dans ces études furent les moindres carrés ordinaires et généralisés avec ou sans contraintes d'inégalités et les procédures d'estimation bayésienne.

⁵ En fait, beaucoup de ces coefficients prennent des valeurs négatives ou nulles et bien souvent, on constate une violation des contraintes liées aux règles comptables sur les coûts.

ce modèle de répartition des coûts à paramètres variables à un échantillon d'exploitations agricoles bretonnes.

Quatre caractéristiques spécifiques de la méthode ECG seront exploitées dans cette étude : (a) cette approche n'est pas affectée par le manque de degrés de liberté, ce qui la rend particulièrement adaptée pour traiter de tous les problèmes d'indétermination quand le nombre des paramètres inconnus est plus grand que le nombre d'observations ; (b) elle permet d'imposer des contraintes d'égalité et d'inégalité sur les paramètres d'une manière flexible et simple; (c) elle emploie un mécanisme formalisé pour incorporer toute information préalable, permettant ainsi de réduire les effets nuisibles de données qui seraient co-linéaires; et (d) sa mise en oeuvre est relativement facile. Par conséquent, en appliquant la méthode ECG, la plupart des problèmes pratiques et méthodologiques rencontrés dans les études antérieures sont éliminés.

Malgré ces nombreux avantages, seuls quelques chercheurs ont jusqu'à maintenant adopté cette méthode reposant sur la notion d'entropie pour estimer des modèles de répartition des coûts (Léon *et al.*, 1999) ou pour analyser des technologies multi-produits (Lence et Miller, 1998a et 1998b ; Oude Lansink, 1999) à partir de données agrégées sur les intrants. Cependant, dans tous ces exemples, la nature variable des paramètres est ignorée, tandis que les contraintes liées aux paramètres, qui seraient issues soit de la théorie économique, soit de règles comptables sont en général imposées à un seul point (de référence) de l'échantillon.

La communication est constituée de cinq parties. Après une partie introductive, on présente le modèle de répartition des coûts à paramètres variables et les contraintes pertinentes qui lui sont liées. En troisième partie, on introduit la formulation du modèle de répartition des coûts en utilisant la notion ECG. Les résultats empiriques de l'estimation par la méthode ECG du modèle de répartition des coûts, ainsi que sa validité "*ex post*" sont développés dans la quatrième partie. On conclut cette communication en récapitulant les principaux résultats de ce travail et en donnant quelques pistes à suivre pour résoudre certaines questions de recherche.

2. Modèles de répartition des coûts à paramètres variables

Soit N intrants utilisés par T firmes produisant K biens. En adoptant un cadre probabiliste, le modèle de répartition des coûts à paramètres variables s'écrit formellement comme suit:

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

où x_{it} représente les dépenses par la firme t en intrant i , y_{kt} est la valeur totale du bien k produit par la firme t , β_{ikt} est un paramètre variable, spécifique à chaque firme, qui s'interprète comme la dépense non observable, engagée par la firme t et relative à l'intrant i , nécessaire pour produire une unité monétaire du bien k ⁶, et u_{it} est un terme résiduel stochastique, également spécifique à chaque firme avec une espérance nulle. On suppose aussi que le résidu u_{it} est distribué de manière identique, mais indépendante entre firmes. Les coefficients β_{ikt} doivent être positifs ou nuls pour tout niveau de production positif du bien k . Le fait d'avoir des coefficients β_{ikt} variables se justifie essentiellement par l'hétérogénéité des firmes.

Lorsqu'on veut estimer les coefficients du modèle représenté par l'expression (1), plusieurs problèmes d'ordre méthodologique surviennent. Tout d'abord, il faut résoudre un problème d'identification, car le nombre de paramètres à estimer (NKT) est supérieur au nombre d'observations (NT). La deuxième difficulté concerne l'estimation du modèle (1), car ventiler la dépense totale relative à un intrant entre plusieurs produits nécessite d'imposer des contraintes (d'égalité et d'inégalité) aux coefficients β_{ikt} . Comme toutes les variables sont mesurées en unités monétaires, l'identité comptable équilibrant recettes et dépenses totales doit être satisfaite pour chaque firme. Cette condition a les conséquences suivantes sur les paramètres du modèle (1):

1) $\sum_i x_{it} = \sum_i \sum_k \beta_{ikt} y_{kt} = \sum_k (\sum_i \beta_{ikt}) y_{kt} = \sum_k y_{kt}$ et 2) la somme (en colonne) des coefficients associés à la production d'un bien k pour chaque firme t doit être égale à l'unité. Par conséquent, $\sum_i \beta_{ikt} = 1 \forall k, t$.

De plus, si les variables dépendantes du modèle (1) sont positives ou nulles, les coefficients propres à chaque activité de production qui lui sont associés doivent, dans la plupart des cas, être également positifs ou nuls. Cette restriction s'écrit formellement comme suit : $\beta_{ikt} \geq 0 \forall i, k, t$ si $x_{it} \geq 0$. Enfin, les équations linéaires expliquant les dépenses relatives à chaque

⁶ L'autre interprétation de ce paramètre est de le considérer comme un coefficient input/output qui indiquerait le montant de l'intrant i nécessaire à la production d'une unité du bien k .

intransant sont interdépendantes entre elles, nécessitant d'utiliser une "approche-système" pour estimer le modèle (1).

En pratique, on pourrait "résoudre" ce problème classique d'identification en employant un modèle linéaire à paramètres *fixes* – qui peut être considéré comme un cas particulier du modèle (1) où $\beta_{ikt} = \bar{\beta}_{ik}$ (permettant ainsi de diminuer le nombre inconnu de paramètres de NKT à $NK < T$). Cela éviterait d'avoir à imposer des restrictions aux coefficients à estimer et à s'assurer de la cohérence comptable équilibrant recettes et dépenses dans les équations expliquant les dépenses en intrants.

2.1 paramètres variables aléatoires

Un autre moyen de résoudre le problème d'identification, tout en tenant compte du fait que les coefficients β_{ikt} sont variables, est d'utiliser une modélisation à paramètres aléatoires, avec la particularité suivante : les coefficients propres à chaque firme seraient tirés de manière aléatoire d'une probabilité de distribution (jointe) particulière⁷. La mise en œuvre d'une telle approche peut s'effectuer en adoptant un modèle standard à coefficients aléatoires similaire à celui développé par Hildreth et Houck (1968). Dans ce cas, on suppose que les coefficients β_{ikt} sont, d'une part, constitués de deux éléments – un élément fixe commun à toutes les firmes et dénoté $\bar{\beta}_{ik}$, et un élément variable et aléatoire propre à chaque firme, dénommé v_{ikt} . Les paramètres β_{ikt} sont liés aux deux éléments $\bar{\beta}_{ik}$ et v_{ikt} par la relation suivante :

$$\beta_{ikt} = \bar{\beta}_{ik} + v_{ikt} . \quad (2)$$

Il est important de noter que la partie fixe du coefficient β_{ikt} peut être assimilée à une valeur moyenne ou espérance des paramètres individuels β_{ikt} (voir la note de bas de page 8 pour une justification de cette propriété).

Une substitution des coefficients β_{ikt} par l'équation (2) dans l'expression (1) conduit à définir un nouveau modèle à paramètres variables et aléatoires, défini comme suit :

$$x_{it} = \sum_{k=1}^K (\bar{\beta}_{ik} + v_{ikt}) y_{kt} + u_{it} = \sum_{k=1}^K \bar{\beta}_{ik} y_{kt} + \sum_{k=1}^K v_{ikt} y_{kt} + u_{it} = \sum_{k=1}^K \bar{\beta}_{ik} y_{kt} + \varepsilon_{it} \quad (3)$$

⁷ Pour des raisons de commodité, nous dénommons cette interprétation du modèle (1) l'approche de modélisation à "paramètres aléatoires".

où $\varepsilon_{it} = \sum_k v_{ikt} y_{kt} + u_{it}$ est un résidu aléatoire “composite” hétéroscédastique. Ce type de modèle à coefficients aléatoires peut être estimé sous certaines conditions (plus ou moins strictes) concernant les moyennes et la matrice de variance-covariance des éléments aléatoires v_{ikt} ⁸. Des applications du modèle (3) à l’agriculture ont été effectuées par Dixon *et al.* (1984) pour déterminer des coefficients input/output et par Hornbaker *et al.* (1989) dans le cadre de fonctions de coûts multi-produits⁹.

Etant donné la spécification (3), les contraintes suivantes doivent être respectées par les paramètres individuels et moyens du modèle de répartition des coûts, ainsi que par ses termes d’erreur :

$$\bar{\beta}_{ik} \geq 0 \quad \forall i, k \quad (4)$$

$$\bar{\beta}_{ik} + v_{ikt} \geq 0 \quad \forall i, k, t \quad (5)$$

$$\sum_i \bar{\beta}_{ik} = 1 \quad \forall k \quad (6)$$

$$\sum_i (\bar{\beta}_{ik} + v_{ikt}) = 1 \quad \forall k, t \quad (7)$$

$$\sum_i v_{ikt} = 0 \quad \forall k, t \quad (8)$$

$$\sum_i u_{it} = 0 \quad \forall t \quad (9)$$

Il faut noter que les restrictions (8) liées aux éléments aléatoires v_{ikt} sont automatiquement satisfaites par les contraintes comptables (6) et (7). De même, en combinant les conditions de non-négativité (4) et (5), et comptables (5) et (6), on est assuré que les coefficients moyens $\bar{\beta}_{ik}$ et ceux spécifiques à chaque firme, β_{ikt} sont inférieurs ou égaux à l’unité.

Avant de procéder à l’estimation du modèle (3), on doit mettre l’accent sur le problème rencontré lorsqu’on a affaire à une entreprise possédant une technologie multi-produits. En effet,

⁸ En général, les espérances des éléments aléatoires v_{ikt} sont supposées être égales à zéro, $E(v_i) = 0$, tandis que la structure des variances-covariances est représentée par la matrice $E(v_i v_i') = A_{ii}$, où A_{ii} peut être, soit une matrice diagonale, ce qui implique des covariances égales à zéro (voir Hildreth-Houck, 1968), soit une matrice semi-définie positive, auquel cas les covariances ne sont pas nulles (voir Schwallie, 1982; Swamy and Mehta, 1975).

⁹ Les spécifications développées par Dixon *et al.* (1984), et Hornbaker *et al.* (1989) sont légèrement différentes du modèle (3), car elles incluent une ou plusieurs variables exogènes influençant les moyennes des coefficients aléatoires. Par suite, les paramètres individuels β_{ikt} sont alors expliqués par la formule suivante : $\beta_{ikt} = \beta_{1,ik} + \beta_{2,ik} z_t + v_{ikt}$, où z_t est, par exemple, la taille de la firme t . Le modèle que nous proposons peut prendre en compte de telles généralisations, mais elles ne seront pas cependant traitées dans cette communication.

si nous supposons que les paramètres moyens $\bar{\beta}_{ik}$ existent bien, il faut s'assurer que les moyennes des paramètres individuels (spécifiques à chaque firme), calculées pour l'ensemble de l'échantillon sont égales au coefficient moyen correspondant¹⁰. En d'autres termes, on doit être sûr que $\sum_t v_{ikt} = 0$, sachant que $\bar{\beta}_{ik} = (1/T) \sum_t \bar{\beta}_{ikt} = (1/T) \sum_t (\bar{\beta}_{ik} + v_{ikt}) = \bar{\beta}_{ik} + (1/T) \sum_t v_{ikt}$. Toutefois, une difficulté apparaît lorsque toutes les firmes ne produisent *pas* tous les biens considérés. En effet, si l'on observe des niveaux nuls de production pour un ou plusieurs biens et pour quelques firmes, la condition $E_t(\beta_{ikt}) = \bar{\beta}_{ik}$ n'est pas nécessairement remplie parce qu'un biais de sélection se produit (Ray, 1985)¹¹. C'est pour cette raison que nous utilisons le terme $E_t[\beta_{ikt} | y_{kt} > 0] = E_t[\bar{\beta}_{ik} + v_{ikt} | y_{kt} > 0]$ comme une mesure de $\bar{\beta}_{ik}$, correspondant seulement au sous-ensemble des entreprises produisant effectivement le bien k . Ce dernier terme est sans biais si et seulement si $E_t[v_{ikt} | y_{kt} > 0] = 0$. En conséquence, les restrictions (5), (7) et (8) ne s'appliquent qu'à une firme produisant un niveau positif du bien k .

2.2 Paramètres à coefficients variables non aléatoires

En dépit des avantages de la spécification du modèle à paramètres aléatoires représentée par l'expression (3), nous n'avons pas opté pour une approche "classique", mais plutôt adopté une stratégie d'estimation qui considérerait les paramètres individuels comme *non-aléatoires* (ou non stochastiques). Nous justifions une telle interprétation des paramètres β_{ikt} pour les raisons suivantes. Tout d'abord, un problème spécifique (mais souvent ignoré) à l'approche "paramètres aléatoires" réside dans le fait que les paramètres individuels ne sont pas *uniques*. En effet, avec une telle approche, les coefficients individuels ne sont pas estimés explicitement, mais habituellement *prédits* en utilisant les moments du deuxième ordre associés au vecteur de leurs moyennes (espérance), $\bar{\beta}_{ik}$ (voir, par exemple, Judge *et al.*, 1985, p. 807). Cependant, ces valeurs prédites – souvent appelées "meilleurs prédicteurs linéaires sans biais"¹² ou "MPLSB"

¹⁰ Nous appelons également cette condition la contrainte de "préservation de la moyenne".

¹¹ En effectuant des simulations de Monte Carlo, Dixon *et al.* (1984), entre autres, ont remarqué également que la qualité des valeurs estimées des paramètres se détériore dans un modèle à paramètres aléatoires lorsque la proportion des valeurs nulles dans les observations des régresseurs augmente.

¹² Ce terme est traduit de l'expression anglaise "*best linear unbiased predictors*" ou BLUP.

(Hornbaker *et al.*, 1989) – ne sont en aucun cas uniques. Au mieux, ce ne sont que des valeurs estimées "appropriées" (Griffiths, 1972; Swamy et Mehta, 1975)¹³.

En deuxième lieu, alors que les modèles à paramètres aléatoires peuvent en principe être estimés à l'aide de procédures traditionnelles telles que les moindres carrés généralisés (MCG) (Swamy, 1970) ou par des techniques de maximum de vraisemblance (Schwallie, 1982; Zaman, 2002), leur mise en oeuvre n'est pas une tâche aisée, nécessitant souvent entre autres choses des corrections *ad-hoc*, et/ou générant des problèmes numériques de résolution¹⁴. Il va de soi que l'estimation devient encore plus difficile (et même impossible), lorsque plusieurs restrictions sont imposées aux paramètres¹⁵.

En raison des nombreux problèmes liés à l'approche traditionnelle à "paramètres aléatoires", nous allons montrer dans la partie suivante que l'application de la méthode d'estimation reposant sur la notion d'entropie croisée généralisée (ECG) permet de prendre en compte l'hétérogénéité des firmes au moyen d'un ensemble de paramètres définis de manière unique.

3. Formulation du modèle de répartition des coûts et notion d'entropie généralisée croisée

Dans cette partie, nous allons re-formuler le modèle de répartition des coûts à paramètres variables en proposant une méthode d'estimation reposant sur la notion d'entropie croisée

¹³ A notre connaissance, la littérature économique est incroyablement silencieuse sur les méthodes d'estimation capables de générer des valeurs estimées uniques des paramètres individuels. Une stratégie possible, qui a été proposée, entre autres, par O'Donnell *et al.* (2001) consiste à appliquer la méthode de décomposition des valeurs singulières (DVS). Toutefois, nous pensons que l'emploi de cette méthode comporte des inconvénients évidents, car elle génère des coefficients individuels qui sont dérivés d'un ensemble de coefficients *moyens* "pré-estimés" (obtenus par les moindres carrés ordinaires ou par des méthodes bayésiennes) qui peuvent être biaisés. En fonction de leur importance, ces biais "contamineront" également d'une manière ou d'une autre les paramètres individuels.

¹⁴Toutes ces difficultés ont été notées par plusieurs économètres, et plus particulièrement par Dixon *et al.* (1984), et Hornbaker *et al.* (1989).

¹⁵ Récemment, Greene (2004) a proposé d'utiliser la méthode d'estimation par le maximum de vraisemblance simulé (MVS) pour surmonter toutes les difficultés rencontrées dans les modèles à coefficients aléatoires. Cependant, pour les besoins de cette étude, l'approche MVS n'est pas le palliatif souhaité, car on ne sait toujours pas trouver des valeurs estimées uniques pour les paramètres.

généralisée et similaire à celle proposée par Golan *et al.* (1996). Notre propos dans ce travail est d'expliquer la méthode ECG appliquée à l'estimation du modèle (3)¹⁶.

Pour mettre en œuvre la méthode ECG, il faut exprimer les paramètres du modèle (3) comme une combinaison linéaire de probabilités ou poids inconnus¹⁷ et de quelques valeurs discrètes, appelées "valeurs support", se situant dans un intervalle prédéterminé. De plus, la procédure d'estimation devient alors un problème de minimisation sous contraintes où la fonction objectif représentée par l'expression (10) ci-dessous consiste en une entropie croisée jointe.

Pour exposer cette approche ECG, nous utilisons une notation comprenant un ensemble de vecteurs de probabilités inconnues définies comme suit: $\mathbf{p}'_{ik} = [p_{ik,1}, \dots, p_{ik,M}]$ ($M \geq 2$), $\mathbf{w}'_{ikt} = [w_{ikt,1}, \dots, w_{ikt,G}]$ ($G \geq 2$), et $\boldsymbol{\mu}'_{it} = [\mu_{it,1}, \dots, \mu_{it,G}]$ ($G \geq 2$). Nous définissons également les vecteurs support suivants: $\mathbf{z}' = [z_1, \dots, z_M]$, $\mathbf{r}' = [r_1, \dots, r_G]$, et $\mathbf{e}' = [e_1, \dots, e_G]$, qui sont respectivement associés aux paramètres moyens $\bar{\beta}_{ik}$, composantes propres à chaque firme, v_{ikt} , et aux termes résiduels u_{it} . De plus, ces divers vecteurs de probabilités et de valeurs support sont liés aux paramètres du modèle de répartition des coûts et de ses termes résiduels par les relations suivantes :

$$\bar{\beta}_{ik} = \mathbf{z}' \mathbf{p}_{ik}, \quad v_{ikt} = \mathbf{r}' \mathbf{w}_{ikt}, \quad \text{et} \quad u_{it} = \mathbf{e}' \boldsymbol{\mu}_{it}.$$

Enfin, des informations préalables (ou "valeurs de référence préalables") sur les différentes probabilités que nous venons de définir sont incorporées dans des vecteurs de probabilités, dénotés \mathbf{p}^o_{ik} , \mathbf{w}^o_{ikt} , et $\boldsymbol{\mu}^o_{it}$, reflétant, soit des connaissances subjectives, soit des conjectures bien établies ou n'importe quelle information préalable sur l'échantillon de données.

¹⁶ Nous préférons plutôt utiliser la procédure d'estimation ECG que sa parente MEG (maximum d'entropie généralisée). Ce choix est principalement motivé par le fait que les estimateurs ECG sont en général moins sensibles aux changements affectant les informations préalables (c.à.d. les intervalles et/ou le nombre de valeurs support) fournies par l'économètre. La sensibilité des résultats économétriques par rapport à ces informations est un problème courant aussi bien dans les procédures MEG/EGC qu'en inférence bayésienne. Une étude exhaustive de cette question va bien au-delà du champ d'étude du présent travail. Pour une analyse complète de ce problème dans le contexte d'un modèle de répartition des coûts, nous suggérons au lecteur de se reporter à Léon *et al.* (1999).

¹⁷ Comme l'a montré à juste titre Preckel (2001), il n'est pas approprié de considérer les poids associés aux valeurs support comme de véritables éléments issus d'une distribution de "probabilités" parce que les résultats de l'estimation sont simplement un ensemble de valeurs estimées pour les paramètres et les termes résiduels (correspondant aux éléments aléatoires) associés à chaque observation. Ainsi, le rôle de la distribution des poids est tout simplement de servir d'intermédiaire pour exprimer les valeurs souhaitables du paramètre et du résidu (voir aussi Golan *et al.*, 1996, p.86 sur cette question). Par suite, l'entropie joue simplement le rôle d'une fonction de pénalité.

Après une reparamétrisation appropriée du modèle de répartition des coûts décrit par les expressions (3) à (9), le problème d'optimisation qui lui est associé peut être reformulé comme suit :

$$\text{Min}_{\mathbf{p}, \mathbf{w}, \boldsymbol{\mu}} CE = \sum_{i=1}^N \sum_{k=1}^K \mathbf{p}'_{ik} \ln \left(\frac{\mathbf{p}_{ik}}{\mathbf{p}_{ik}^o} \right) + \sum_{i=1}^N \sum_{k=1}^K \sum_{t=1}^T \mathbf{w}'_{ikt} \ln \left(\frac{\mathbf{w}_{ikt}}{\mathbf{w}_{ikt}^o} \right) + \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\mu}'_{it} \ln \left(\frac{\boldsymbol{\mu}_{it}}{\boldsymbol{\mu}_{it}^o} \right) \quad (10)$$

sous les contraintes suivantes :

$$x_{it} = \sum_{k=1}^K (\mathbf{z}' \mathbf{p}_{ik} + \mathbf{r}' \mathbf{w}_{ikt}) y_{kt} + \mathbf{e}' \boldsymbol{\mu}_{it} \quad \forall i, t \quad (11)$$

$$\mathbf{z}' \mathbf{p}_{ik} + \mathbf{r}' \mathbf{w}_{ikt} \geq 0 \quad \forall i, k, t \in \{t_k\} \quad (12)$$

$$\mathbf{z}' \mathbf{p}_{ik} + \mathbf{r}' \mathbf{w}_{ikt} = 0 \quad \forall i, k, t \notin \{t_k\} \quad (13)$$

$$\frac{1}{T_k} \sum_{t \in \{t_k\}} (\mathbf{z}' \mathbf{p}_{ik} + \mathbf{r}' \mathbf{w}_{ikt}) = \mathbf{z}' \mathbf{p}_{ik} \quad \forall i, k \quad (14)$$

$$\sum_{i=1}^N (\mathbf{z}' \mathbf{p}_{ik} + \mathbf{r}' \mathbf{w}_{ikt}) = 1 \quad \forall k, t \in \{t_k\}; \quad \sum_{i=1}^N \mathbf{z}' \mathbf{p}_{ik} = 1 \quad \forall k; \quad \sum_{i=1}^N \mathbf{e}' \boldsymbol{\mu}_{it} = 0 \quad \forall t \quad (15)$$

$$\sum_{m=1}^M p_{ik,m} = 1 \quad \forall i, k; \quad \sum_{g=1}^G w_{ikt,g} = 1 \quad \forall i, k, t; \quad \sum_{g=1}^G \mu_{it,g} = 1 \quad \forall i, t \quad (16)$$

où $\{t_k\} = \{t : y_{kt} > 0\} \subseteq \{t\}_1^T$ est le sous-ensemble, de dimension T_k , des firmes échantillonnées, produisant le bien k . L'équation (10) est la fonction objectif à minimiser, qui prend en compte le critère d'entropie croisée généralisée. Cette fonction objectif est optimisée tout en respectant un ensemble de six contraintes représentées par les équations (11) à (16). La première de ces contraintes traduit le fait que les coefficients du modèle sont estimés en tenant compte des égalités comptables permettant de répartir les coûts de chaque firme entre ses diverses activités de production. Les deux contraintes suivantes permettent de s'assurer que tous les coefficients individuels et propres à chaque firme sont non négatifs (expression (12)) ou égaux à zéro (expression (13)) pour un niveau nul de production d'une activité donnée. La contrainte de "préservation de la moyenne" est imposée aux coefficients par l'intermédiaire de la relation (14). L'identité comptable équilibrant recettes et dépenses pour chaque firme est satisfaite au travers

des restrictions représentées par l'expression (15). Enfin, les conditions (16) représentent les contraintes de normalisation stipulant que la somme des probabilités inconnues constituant un vecteur donné est égale à 1. Il n'est pas nécessaire d'inclure des contraintes supplémentaires sur les coefficients *moyens*, $\mathbf{z}'\mathbf{p}_{ik}$, aussi longtemps qu'on sélectionne des limites (théoriques) appropriées définissant les intervalles des valeurs support \mathbf{z} (voir ci-dessous).

Le principe de minimisation de l'entropie croisée implique que nous allons déterminer des estimations des vecteurs-paramètres (probabilités) inconnus \mathbf{p}_{ik} , \mathbf{w}_{ikt} et $\boldsymbol{\mu}_{it}$, qui s'écartent des vecteurs de référence préalables et connues, \mathbf{p}_{ik}^o , \mathbf{w}_{ikt}^o , et $\boldsymbol{\mu}_{it}^o$ tout en minimisant les différences correspondantes (Golan *et al.*, 1996, p. 11). En d'autres termes, nous recherchons la distribution de probabilité la "moins riche en information" (c.à.d. la plus proche de la distribution uniforme), qui est à la fois cohérente avec les données et les autres restrictions représentées par les expressions (11) à (16) et avec l'information préalable correspondant aux intervalles des valeurs support et des probabilités de référence connues au départ. La résolution de ce programme d'optimisation sous contraintes permet donc d'estimer les paramètres moyens inconnus ($\bar{\beta}_{ik}$), les éléments propres à chaque firme (v_{ikt}) et les termes résiduels (u_{it}) restants à l'aide des expressions suivantes : $\hat{\beta}_{ik} = \mathbf{z}'\hat{\mathbf{p}}_{ik}$, $\hat{v}_{ikt} = \mathbf{r}'\hat{\mathbf{w}}_{ikt}$, et $\hat{u}_{it} = \mathbf{e}'\hat{\boldsymbol{\mu}}_{it}$, avec $\hat{\beta}_{ikt} = \mathbf{z}'\hat{\mathbf{p}}_{ik} + \mathbf{r}'\hat{\mathbf{w}}_{ikt}$. En conséquence, un nombre de paramètres égal à $NK + NKT + NT$ (correspondant aux NK paramètres moyens, NKT composantes individuelles propres à chaque firme et les NT erreurs aléatoires) doit être estimé avec seulement NT observations¹⁸.

Étant donné la formulation que nous venons de développer pour estimer un modèle de répartition des coûts à coefficients variables, il est possible de faire apparaître plusieurs avantages importants de la procédure d'estimation ECG par rapport aux procédures traditionnelles d'estimation de modèles à coefficients aléatoires. Premièrement, les éléments particuliers v_{ikt} , représentant les écarts des paramètres individuels par rapport à leur moyenne, sont interprétés

¹⁸ Les coefficients individuels β_{ikt} pourraient en principe être estimés directement, mais il est plus commode d'obtenir les éléments fixes (moyens), $\bar{\beta}_{ik}$, et les éléments propres à chaque firme, v_{ikt} , comme des termes distincts (permettant ainsi de maintenir le principe de "conservation de la moyenne"), et de calculer ensuite les coefficients individuels β_{ikt} par une simple opération de sommation de l'élément fixe (moyen) et des éléments propres à chaque firme.

comme des coefficients "fixes" ou "non-aléatoires", qui doivent être estimés pour chaque firme (ce qui, d'une certaine manière, revient à utiliser la technique de régression à variables dichotomiques pour prendre en compte un phénomène d'hétérogénéité non-observable dans une modélisation de données de panel à effets fixes). En deuxième lieu, en considérant une hétérogénéité à coefficients fixes (qui ne doit pas être confondu à un modèle à paramètres fixes !), les paramètres moyens estimés ne souffrent pas des biais possibles, qui pourraient résulter d'une corrélation entre les éléments variables et les régresseurs (BiØrn, 2003, p.73). Troisièmement, les éléments estimés v_{ikt} sont directement disponibles; ce qui permet d'étudier la distribution empirique des v_{ikt} à travers toutes les firmes, et de retenir une firme particulière pour laquelle ses paramètres estimés $\hat{\beta}_{ikt}$ se situent au-dessus ou au-dessous de la valeur moyenne $\bar{\beta}_{ik}$ estimée pour l'échantillon.

4 Application empirique

4.1. Données utilisées

Le modèle de répartition des coûts de production défini précédemment est estimé en employant un échantillon réduit de 34 exploitations agricoles bretonnes ($T = 34$) à orientation laitière et viande bovine¹⁹. Les observations utilisées sont issues des bases de données du Réseau d'information comptable agricole (RICA) qui collecte, au niveau de chaque pays membre de l'Union européenne, des informations complètes et détaillées sur les dépenses totales en intrants et les valeurs des produits de chaque activité de production.

¹⁹ Pour être plus exact, le type d'exploitations agricoles retenues est défini comme étant une "exploitation bovine spécialisée-orientation lait, élevage et viande combinés" (TF43), qui est une sous-catégorie du groupe "exploitations spécialisées herbivores" (Commission Européenne, 2000). Nous avons pris en compte les exploitations dont la dimension économique était supérieure à 48 UDE (unités de dimension économique) de marge brute standard (MBS). Ces exploitations sont considérées comme étant de grande dimension dans le contexte breton. Par suite, nous avons retenu un échantillon de 38 exploitations. Toutefois, une analyse exploratoire des données à l'aide des boîtes à moustaches (*box-plots*) (voir Chambers *et al.*, 1983) permet de déceler des "valeurs extrêmes" pour quatre fermes, ces dernières étant définies comme étant des observations associées aux dépenses totales en intrants prenant des valeurs supérieures à trois interquartiles et donc situés en dehors de la boîte. Nous avons décidé d'éliminer ces quatre observations, nous retrouvant au final avec un échantillon exploitable de 34 fermes.

Nous répartissons les intrants²⁰ en six grandes catégories définies comme suit : a) produits végétaux consommés sur l'exploitation (intra-consommation) ; b) engrais (y compris ceux destinés à la fertilisation des prairies) ; c) produits phytosanitaires et vétérinaires; d) aliments composés, e) autres intrants divers; et f) valeur ajoutée brute (VAB)²¹. Les dépenses liées aux quatre premiers groupes d'intrants sont définitivement des coûts variables (propres à chaque activité végétale ou animale), tandis que les achats correspondant aux autres intrants divers (groupe (e)) sont considérés comme étant "semi-variables" par nature. Classifier les dépenses en intrants en cinq catégories revient explicitement à répartir les coûts variables et semi-variables entre les différentes activités de production. De plus, la valeur ajoutée de chaque exploitation agricole est déterminée de manière résiduelle en déduisant de la valeur totale de la production finale l'ensemble des coûts variables et semi-variables. En résumé, nous pouvons traiter la VAB comme étant le sixième "facteur de production".

Les recettes de production sont réparties en quatre activités de production, soit $K = 4$), à savoir : a) produits végétaux ; b) viande de bœuf; c) lait et d) viande de porc. La production en valeur de chaque activité est reprise directement des bases de données du RICA. Pour les activités

²⁰ Dans les comptes du RICA, les intrants sont répartis entre ceux qui sont achetés et ceux qui sont *intraconsommés*, c'est à dire produits et utilisés sur la ferme en tant que consommations intermédiaires. Les intrants intraconsommés sont évalués à leur coût d'opportunité, c'est à dire au prix-départ de l'exploitation dans la mesure où ce sont des produits agricoles vendables. Pour les activités de production végétale, la catégorie d'intrants intitulé "produits végétaux" comprend :

- les semences et plants achetés ou produits sur l'exploitation ;
- les achats pour production végétale .

Pour les activités de production animale, ce type d'intrants inclut les éléments suivants :

- les aliments du bétail grossiers, les céréales auto-produites et les dépenses associées à l'utilisation de pâtures et de prairies fourragères ou à la mise en pension, les fourrages produits sur la ferme (foin, ensilage, paille et autres) en sont cependant exclus puisque cette production en général intra-consommée n'est pas estimée;
- les aliments du bétail concentrés comprend les aliments achetés qu'ils soient simples (céréales, aliments déshydratés, farine, poudre de lait, tourteaux, minéraux, urée)ou composés (i.e. constituant toute la ration animale), et les aliments complémentaires (mélanges d'aliments concentrés simples et de fourrages grossiers ou de céréales produits sur la ferme) ;
- les achats pour production animale, en particulier les dépenses pour la préservation et le stockage des aliments fourragers.

²¹ Si nous nous référons à la terminologie du RICA (voir Argilès et Slof, 2001) la VAB est définie comme la valeur de la production totale *moins* la totalité des consommations intermédiaires (coûts spécifiques aux activités végétales et animales), les frais généraux (dépenses d'équipement et de construction, dépenses d'énergie, travail à façon, et autres coûts directs), et les dépenses externes (constituées des charges de location, des intérêts et des charges salariales). Les deux postes de coûts – frais généraux et externes – sont réunis dans notre travail sous le vocable "autres intrants divers". En d'autres termes, la valeur ajoutée brute est égale à la somme des retraits personnels, du poste "amortissements" (estimation des amortissements incluant les ré-investissements et les remboursements de prêts), des intérêts (sur emprunts) et des impôts. En conséquence, la VAB peut être grossièrement assimilée à la rémunération des facteurs de production de l'exploitant, à savoir le travail familial, la terre utilisée et les coûts du capital appartenant à l'exploitant, *nets* des subventions et aides en capital.

animales, la valeur de la production est égale à la somme de la production nette et de la variation des stocks²². Les activités végétales sont caractérisées par une valeur de la production, qui inclut à la fois les ventes et les intra-consommations évaluées aux prix-départ de l'exploitation. Toutes les exploitations agricoles appartenant à l'échantillon produisent des produits végétaux, de la viande de bœuf et du lait, et 25 d'entre elles ont une activité porcine²³

Toutes les données utilisées, exprimées en unités monétaires et mesurées en milliers d'euros, correspondent à la moyenne des deux années comptables 2000 et 2001. Cette option permet de réduire l'influence possible de fluctuations annuelles extrêmes des dépenses en intrants et des productions des activités de production de chaque exploitation. Le tableau 1 fournit des statistiques d'ensemble sur les dépenses en intrants et les productions de chaque activité pour toutes les exploitations agricoles appartenant à l'échantillon. Une lecture de ces indicateurs montre bien l'hétérogénéité importante de l'échantillon, que ce soit en termes de dimension économique (mesurée par la valeur totale de la production ou par la taille du troupeau), de la combinaison des activités de production ou de la consommation des intrants.

4.2. Valeurs support et informations préalables

La mise en oeuvre de la méthode ECG nécessite de spécifier des valeurs support pour tous les vecteurs des paramètres du modèle de répartition des coûts. En procédant de cette manière et en choisissant ces valeurs support, on impose un ensemble de restrictions préalables sur les paramètres à estimer, qui s'ajoutent aux conditions d'égalité ou d'inégalité s'appliquant à la spécification empirique du modèle de répartition des coûts.

Les valeurs support sont sélectionnées en fonction des données utilisées (voir partie B du tableau 1 sur les indicateurs statistiques) et définies comme suit :

i) pour chaque coefficient moyen, $\bar{\beta}_{ik}$, nous choisissons un vecteur support, $\mathbf{z}' = [0, 1]$ avec $M = 2$, les moyennes observées des parts des dépenses totales associées à chaque intrant étant positives ou nulles, et inférieures à l'unité. On peut donc raisonnablement supposer que les

²² Les ventes comprennent également les utilisations de produits laitiers vendables. Toutefois, la consommation de lait d'allaitement pour les veaux et les productions à la ferme de fourrages sont exclues des chiffres de ventes.

²³ En raison du manque de données disponibles, nous supposons implicitement qu'une activité caractérisée par un niveau nul de production n'est associée aucune dépense en intrants.

Tableau 1 : Dépenses en intrants et valeur de la production
(moyennes 2000-2001)

	Nombre d'exploitations agricoles	Minimum	Moyenne	Maximum	Ecart type
A. Dépenses en intrants et valeur de la production(€ 1 000)					
Intrants					
Produits végétaux	33	0,00	9,38	39,60	8,82
Engrais	34	2,22	4,74	16,56	2,74
Produits phytosanitaires et vétérinaires	34	2,30	10,65	31,62	6,61
Aliments composés	34	4,42	54,26	173,70	40,90
Autres intrants divers	34	24,62	63,02	246,06	44,33
Valeur ajoutée brute	34	9,20	84,06	242,35	55,14
Total		65,61	226,10	678,49	127,00
Produits					
Produits végétaux	34	0,80	17,69	49,16	11,29
Viande bovine	34	1,86	29,99	506,44	44,24
Lait	34	28,80	84,75	200,00	42,43
Viande de porc	25	0,00	93,68	454,59	95,65
Total	34	65,61	226,10	678,49	127,00
B. Parts des dépenses en intrants et parts de la valeur de la production (%)					
Intrants					
Produits végétaux	33	0,0	4,2	8,7	2,2
Engrais	34	0,8	2,6	8,0	1,8
Produits phytosanitaires et vétérinaires	34	2,9	4,8	11,3	1,8
Aliments composés	34	4,4	22,5	59,8	10,6
Autres intrants	34	12,4	29,6	62,1	11,7
Valeur ajoutée brute	34	10,0	36,2	53,1	11,8
Total			100,0		
Produits					
Produits végétaux	34	0,2	9,1	33,9	6,3
Viande bovine	34	0,7	14,9	75,0	18,2
Lait	34	17,3	40,9	64,8	14,0
Viande de porc	25	0,0	35,1	72,9	25,7
Total			100,0		
C. Autres caractéristiques structurelles					
Surface arable (hectares)	34	0,00	2,70	6,60	1,56
Surface fourragère (hectares)	34	1,77	4,28	8,77	1,98
Cheptel (1000 UB)	34	4,21	15,90	50,94	9,36

UB . Unités bétail

parts définis au niveau de chaque activité de production seront en *moyenne* positives et inférieures à l'unité.

ii) le vecteur support pour les éléments spécifiques à chaque exploitation est $\mathbf{r}' = [-1, 0, 1]$ avec $G = 3$. Cette sélection de valeurs support semble "naturelle", car elle implique, en liaison avec la condition (12), que tous les paramètres individuels (propres à chaque ferme) sont positifs et inférieurs à l'unité. Par contre, si nous éliminons la contrainte (12) de non-négativité pour l'intrant VAB (dont les dépenses correspondantes sont déterminées de manière résiduelle), les valeurs support, que nous avons choisies pour le vecteur \mathbf{r} n'empêchent pas que quelques

paramètres individuels (y compris la VAB) puissent prendre des valeurs situées à l'extérieur de l'intervalle $[0, 1]$. Par conséquent, l'intervalle de définition du vecteur support \mathbf{r} est assez large pour couvrir toutes les situations qui peuvent être rencontrées en réalité pour les coefficients individuels²⁴.

iii) Le vecteur des valeurs support u_{it} est spécifié en adoptant la règle des "trois sigmas"²⁵. En conséquence, $\mathbf{e}' = [-3\sigma_i, 0, 3\sigma_i]$ avec $G = 3$ où σ_i est l'écart type de la variable dépendante.

De plus, on utilise des "valeurs de référence" pour les vecteurs de probabilités inconnues \mathbf{p} , \mathbf{w} , et $\boldsymbol{\mu}$. Celles-ci se présentent sous forme d'informations préalables pour les probabilités associées aux paramètres moyens et de valeurs uniformes ou sans contenu en information pour celles liées aux composantes variables et spécifiques à chaque ferme, et pour les résidus. Plus précisément, nous imposons à \mathbf{p}_{ik}^0 les valeurs support suivantes : $\mathbf{p}_{ik}^0 = [p_{ik,1}^0, p_{ik,2}^0] = [1 - \alpha_i, \alpha_i]$, où α_i est le coefficient budgétaire total moyen de l'intrant i (apparaissant dans la partie B du tableau 1). Pour les deux autres vecteurs de "référence" de probabilités, \mathbf{w}_{ikt}^0 et $\boldsymbol{\mu}_{it}^0$, nous adoptons le jeu de valeurs suivant : $[1/3, 1/3, 1/3]$. Le choix des valeurs support \mathbf{p}_{ik}^0 se justifie par le fait que la moyenne du coefficient budgétaire de l'intrant i , $\alpha_i = (1/T)\sum_t (x_{it}/\sum_i x_{it}) = (1/T)\sum_t (x_{it}/\sum_k y_{kt})$, peut être considérée comme une hypothèse initiale raisonnable ou la plus "plausible" par rapport à l'éventail des valeurs possibles que peuvent prendre les paramètres moyens correspondants associés à chaque activité de production. Comme la procédure d'estimation ECG s'apparente à une forme d'estimateur rétrécisseur (*shrinkage estimator*) (Golan *et al.*, 1996, p. 31), un plus grand poids sera attribué aux termes de la fonction d'entropie croisée correspondant aux valeurs les plus faibles des probabilités de référence. Par suite, les probabilités associées aux paramètres moyens se "rapprocheront" plus rapidement des probabilités de référence, ce qui implique une procédure d'estimation plus stable. Cette dernière

²⁴ La possibilité de rencontrer des valeurs négatives pour les observations (par exemple la VAB totale) complique sans aucun doute le choix des bornes inférieure et supérieure appropriées pour les valeurs support des paramètres individuels associés à chaque ferme. Plus précisément, cela implique qu'un ou plusieurs paramètres individuels peuvent prendre des valeurs situées à l'extérieur de l'intervalle $[-1, 1]$. Plus généralement, l'étendue pertinente de l'intervalle de définition des valeurs support des coefficients individuels n'est pas connue *a priori*. A partir de quelques expérimentations, on a trouvé (et comme cela était prévisible) qu'un élargissement systématique de l'intervalle des valeurs support associées aux éléments variables et spécifiques à chaque ferme avait pour effet de faire tendre les estimations des coefficients moyens vers les valeurs établies préalablement.

²⁵ Cette règle est d'une certaine manière une extension de l'inégalité de Bienaymé-Tchébitcheff, qui repose sur l'utilisation de deux écart-types pour déterminer l'étendue des valeurs prises par une variable aléatoire.

propriété de l'estimateur ECG est jugée particulièrement importante au regard de la taille réduite et de la structure hétérogène de notre échantillon²⁶

4.3. Résultats empiriques

Le modèle de répartition des coûts a été estimé en utilisant le solveur CONOPT3 du logiciel GAMS utilisé pour résoudre les problèmes et modèles de programmation non linéaires. Comme aucune activité végétale n'utilise des aliments composés et qu'aucune consommation d'engrais n'intervient dans la production de viande porcine, les coefficients de répartition des coûts (moyens et individuels) correspondants sont *a priori* fixés à zéro²⁷.

4.3.1 Valeurs estimées des paramètres moyens

Les estimations ECG des paramètres $\bar{\beta}_{ik}$ sont présentées au tableau 2. Les valeurs estimées des coefficients moyens correspondant au modèle à paramètres variables (que nous dénotons par la suite "modèle V-ECG") figurent dans la partie A de ce tableau. Dans l'ensemble, les estimations V-ECG sont plausibles et conformes à nos attentes. Une brève discussion de la validation du modèle est reportée à la section 4.4.

Nous avons fait figurer dans la partie B du tableau 2 les valeurs estimées des coefficients correspondant au modèle de répartition des coûts à coefficients fixes (que nous désignons par la suite "modèle F-ECG")²⁸. Les résultats économétriques relatifs aux estimations F-ECG semblent être moins plausibles, car les valeurs obtenues pour certains coefficients sont presque égales à

²⁶ Il est bien connu en économétrie que les estimateurs de type ECG et estimateurs rétrécisseurs ont une performance bien supérieure à celle des estimateurs qui ne font pas appel à des valeurs préalables à contenu informationnel, tels que la méthode d'estimation MEG (Shen et Perloff, 2001). Avec cette dernière procédure, des poids uniformes établis préalablement sont affectés à chaque valeur support des paramètres et le point central de cette information préalable est la médiane de l'espace borné des paramètres. Cependant, en utilisant des simulations de Monte Carlo, Lence et Miller (1998b) ont été en mesure de montrer que les estimations moyennes se situaient en général à mi-chemin entre les vraies valeurs des paramètres et les points médians des intervalles des valeurs support correspondants. De plus, ils ont trouvé que les biais obtenus avec la procédure MEG avaient tendance à augmenter lorsque les informations préalables s'écartaient des "vraies" valeurs.

²⁷ Le nombre total de paramètres à estimer est 3 020 se décomposant comme suit : $(6 \times 4) - 2 = 22$ paramètres moyens ($\bar{\beta}_{ik}$) plus $22 \times (34 + 34 + 34 + 25) = 2\,794$ éléments spécifiques à chaque exploitation agricole (v_{ikt}) plus $6 \times 34 = 204$ résidus (u_{it}). Etant donné les dimensions $M = 2$ and $G = 3$ des vecteurs des valeurs support, nous avons à déterminer un ensemble de 9 038 probabilités ou poids.

²⁸ L'estimation du modèle de répartition des coûts à coefficients fixes revient à éliminer tous les termes impliquant les éléments individuels propres à chaque ferme, v_{ikt} , de la fonction d'entropie (expression (10)), ainsi que les contraintes représentées par les relations (11) à (16).

zéro, tandis que d'autres tendent à prendre des valeurs irréalistes parce que trop grandes (voir en particulier les coefficients correspondant aux intrants "produits végétaux", "engrais" et "produits phytosanitaires et vétérinaires" dans l'activité de production végétale, ainsi que les coefficients associant d'un côté dépenses d'aliments composés et activité "viande bovine" et de l'autre VAB et viande porcine).

4.3.2. Propriétés des estimateurs EGC des coefficients moyens

Comme les propriétés des estimateurs contraints de type EGC ne sont généralement pas connues, on a eu recours à une technique de bootstrap pour analyser leur précision. Pour ce faire, on fait appel à la méthode simple de bootstrap par percentiles (Efron and Tibshirani, 1993)²⁹.

Les estimations bootstrap sont calculées à l'aide d'une procédure de ré-échantillonnage des données, plutôt que de prendre des estimations bootstrap effectuées sur les résidus, afin de préserver leur caractère hétéroscédastique. Plus précisément, nous procédons, dans un premier temps, à un tirage aléatoire de 250 échantillons obtenus à partir d'une distribution *jointe* empirique $F(\mathbf{x}, \mathbf{y})$ des données observées, correspondant aux observations initiales de l'échantillon, puis nous construisons 250 échantillons de matrices d'observations $[\mathbf{x}_b, \mathbf{y}_b]$, où $b = 1, \dots, 250$. Avec ces nouveaux jeux artificiels de données, nous ré-estimons les paramètres (moyens) 250 fois, permettant ainsi d'obtenir un nombre semblable de valeurs estimées des paramètres. Les distributions empiriques de ces estimations vont alors constituer des approximations des "vraies" distributions des paramètres, à partir desquelles des intervalles de confiance à 95% peuvent être facilement dérivés en choisissant les seuils 0,025 et 0,975 (c.à.d. en écartant les 5% des valeurs estimées moyennes situées à l'extérieur de l'intervalle de confiance).

Les intervalles de confiance obtenus par bootstrap et présentés au tableau 2 sont très resserrés, ce qui semble indiquer que les estimateurs V-EGC produisent des résultats assez stables pour les paramètres moyens. Par contre, les intervalles de confiance obtenus à partir des échantillons bootstrap pour les estimations F-ECG sont systématiquement plus grands (et dans quelques cas près de la valeur limite nulle) que leurs homologues V-ECG. Ce résultat suggère clairement que

²⁹ En raison de la nature contrainte de l'estimateur EGC que nous utilisons, celui-ci n'est pas asymptotiquement normal. Par conséquent, il est impossible, comme l'avait fait Golan *et al.* (2001), de faire des inférences statistiques en utilisant les écarts type asymptotiques. Sur cette question, voir Andrews (1996).

Tableau 2: Estimations ECG des coefficients moyens et intervalles de confiance à 95% obtenus par bootstrap^a

Intrants	Produits			
	Produits	Viande bovine	Lait	Viande de porc
A. paramètres variables (V-ECG)				
Produits végétaux	0,053 (0,042/0,061)	0,033 (0,024/0,044)	0,021 (0,015/0,029)	0,071 (0,051/0,092)
Engrais	0,043 (0,035/0,052)	0,029 (0,023/0,039)	0,042 (0,036/0,048)	
Produits phytosanitaires et vétérinaires	0,060 (0,048/0,072)	0,043 (0,036/0,051)	0,022 (0,016/0,033)	0,078 (0,059/0,096)
Aliments composés		0,339 (0,200/0,467)	0,199 (0,167/0,236)	0,267 (0,250/0,294)
Autres intrants divers	0,394 (0,359/0,432)	0,258 (0,167/0,236)	0,317 (0,269/0,374)	0,238 (0,191/0,270)
Valeur ajoutée brute	0,450 (0,398/0,491)	0,298 (0,269/0,374)	0,399 (0,336/0,449)	0,346 (0,309/0,404)
B. Paramètres fixes (F-ECG)				
Produits végétaux	0,100 (0,037/0,199)	0,006 (0,000/0,027)	0,056 (0,034/0,081)	0,038 (0,001/0,057)
Engrais	0,117 (0,006/0,192)	0,008 (0,002/0,050)	0,028 (0,011/0,047)	
Produits phytosanitaires et vétérinaires	0,140 (0,054/0,200)	0,042 (0,019/0,054)	0,046 (0,032/0,067)	0,034 (0,023/0,052)
Aliments composés		0,493 (0,199/0,620)	0,145 (0,081/0,219)	0,287 (0,261/0,339)
Autres intrants divers	0,305 (0,218/0,444)	0,195 (0,089/0,358)	0,370 (0,259/0,517)	0,227 (0,045/0,334)
Valeur ajoutée brute	0,337 (0,205/0,470)	0,257 (0,199/0,383)	0,356 (0,237/0,431)	0,414 (0,312/0,600)

^a Les intervalles de confiance à 95% sont calculés en utilisant la méthode bootstrap simple percentile.

Note: Comme les intrants et productions sont mesurés en unités monétaires (euros), les coefficients moyens estimés présentés dans ce tableau représentent les dépenses moyennes en intrant i nécessaires pour produire une unité de produit k exprimé en euros.

l'adoption d'une spécification à paramètres variables diminue considérablement les incertitudes liées aux valeurs des coefficients moyens inconnus.

4.3.3 Estimations des coefficients individuels

Les résultats empiriques montrent que la structure des coûts varie considérablement d'une exploitation à l'autre. Ainsi, en se référant aux informations statistiques du tableau 3, nous

observons que la variabilité de la structure des coûts, mesurée par l'écart interquartile rapporté à la médiane correspondante, est systématiquement plus faible pour l'activité de production végétale que pour les activités animales. Pour ces dernières, la dispersion des coûts est plus forte pour les activités "porcs" et "lait" que pour l'activité "viande bovine". De même, les paramètres associés aux intrants mineurs (à savoir les intra consommations végétales, les engrais et produits phytosanitaires et vétérinaires) sont caractérisés par une dispersion plus étendue que celle des coefficients associées aux autres catégories d'intrants (aliments composés, autres intrants et VAB).

Le nombre important d'estimations des coefficients individuels ne nous permet pas de les présenter tous. Néanmoins, afin de fournir une photographie plus riche et plus objective de la variation inter-exploitation de ces coefficients individuels, nous présentons en figure 1 les représentations graphiques de leurs distributions empiriques approximées par une méthode non paramétrique (méthode d'estimation par le noyau). Ces graphiques montrent clairement que les valeurs estimées des paramètres individuels n'ont en général pas une distribution normale³⁰. Un grand nombre de ces distributions empiriques sont caractérisées par une forme fortement irrégulière et/ou asymétrique³¹. Dans plusieurs cas, nous trouvons également des valeurs nulles pour les paramètres individuels, et plus particulièrement pour les petites catégories d'intrants telles que les intra consommations de produits végétaux et les produits vétérinaires et phytosanitaires utilisés pour la production de lait.

Enfin, les interdépendances entre les coefficients individuels mesurées par les coefficients de corrélation, que nous ne présentons pas par manque de place, s'avèrent être importantes et fortement positives. Cette tendance est également confirmée par le fait que les matrices de variances-covariances associées à chaque catégorie d'intrants (voir note de bas de page n° 8) sont semi définie positives. De toute évidence, ce résultat tend à conforter l'affirmation selon laquelle il n'y a aucune raison *a priori* de penser que la variabilité autour du paramètre moyen pour une

³⁰ Ce résultat resterait valide si nous disposions d'échantillons de plus grande taille.

³¹ Les résultats sur les distributions empiriques non paramétriques peuvent conduire à émettre quelques doutes sur les différents tests statistiques employés dans la littérature (Dixon *et al.*, 1992) visant à évaluer la dépendance entre les paramètres estimés et les régresseurs qui supposent habituellement une distribution normale des composantes aléatoires.

Tableau 3: Indicateurs statistiques pour les coefficients individuels dans le modèle V-ECG

Intrants	Produits			
	Produits végétaux	Viande bovine	Lait	Viande porcine
Produits végétaux				
Moyenne	0,053	0,033	0,021	0,071
Ecart type	0,005	0,010	0,022	0,035
Coefficient de variation	0,104	0,308	1,045	0,499
Médiane	0,053	0,034	0,015	0,065
Premier quartile	0,050	0,031	0,000	0,046
Troisième quartile	0,056	0,040	0,044	0,102
Dispersion ^a	0,115	0,246	3,022	0,855
Engrais				
Moyenne	0,043	0,029	0,042	
Ecart type	0,005	0,010	0,022	
Coefficient de variation	0,128	0,345	0,527	
Médiane	0,042	0,028	0,036	
Premier quartile	0,040	0,026	0,026	
Troisième quartile	0,044	0,031	0,051	
Dispersion ^a	0,106	0,182	0,709	
Produits phytosanitaires et vétérinaires				
Moyenne	0,060	0,043	0,022	0,078
Ecart type	0,007	0,009	0,027	0,025
Coefficient de variation	0,116	0,208	1,226	0,321
Médiane	0,059	0,042	0,014	0,075
Premier quartile	0,057	0,038	0,006	0,054
Troisième quartile	0,061	0,044	0,025	0,095
Dispersion ^a	0,065	0,126	1,348	0,560
Aliments composés				
Moyenne		0,339	0,199	0,267
Ecart type		0,057	0,039	0,039
Coefficient de variation		0,169	0,198	0,145
Médiane		0,328	0,206	0,275
Premier quartile		0,321	0,165	0,240
Troisième quartile		0,331	0,221	0,286
Dispersion ^a		0,031	0,270	0,169
Autres intrants divers				
Moyenne	0,394	0,258	0,317	0,238
Ecart type	0,011	0,032	0,053	0,063
Coefficient de variation	0,029	0,124	0,166	0,264
Médiane	0,392	0,260	0,304	0,234
Premier quartile	0,389	0,258	0,291	0,210
Troisième quartile	0,396	0,265	0,330	0,256
Dispersion ^a	0,018	0,030	0,127	0,195
Valeur ajoutée brute				
Moyenne	0,450	0,298	0,399	0,346
Ecart type	0,013	0,026	0,056	0,079
Coefficient de variation	0,028	0,086	0,139	0,228
Médiane	0,450	0,303	0,393	0,345
Premier quartile	0,446	0,290	0,370	0,292
Troisième quartile	0,457	0,308	0,443	0,412
Dispersion ^a	0,023	0,061	0,187	0,346

^a L'indice de dispersion est l'écart inter-quartile divisé par la médiane. Cette mesure a l'avantage d'être peu influencée par les données extrêmes.

Figure 1: Distribution des coefficients individuels obtenus pour le modèle V-GCE

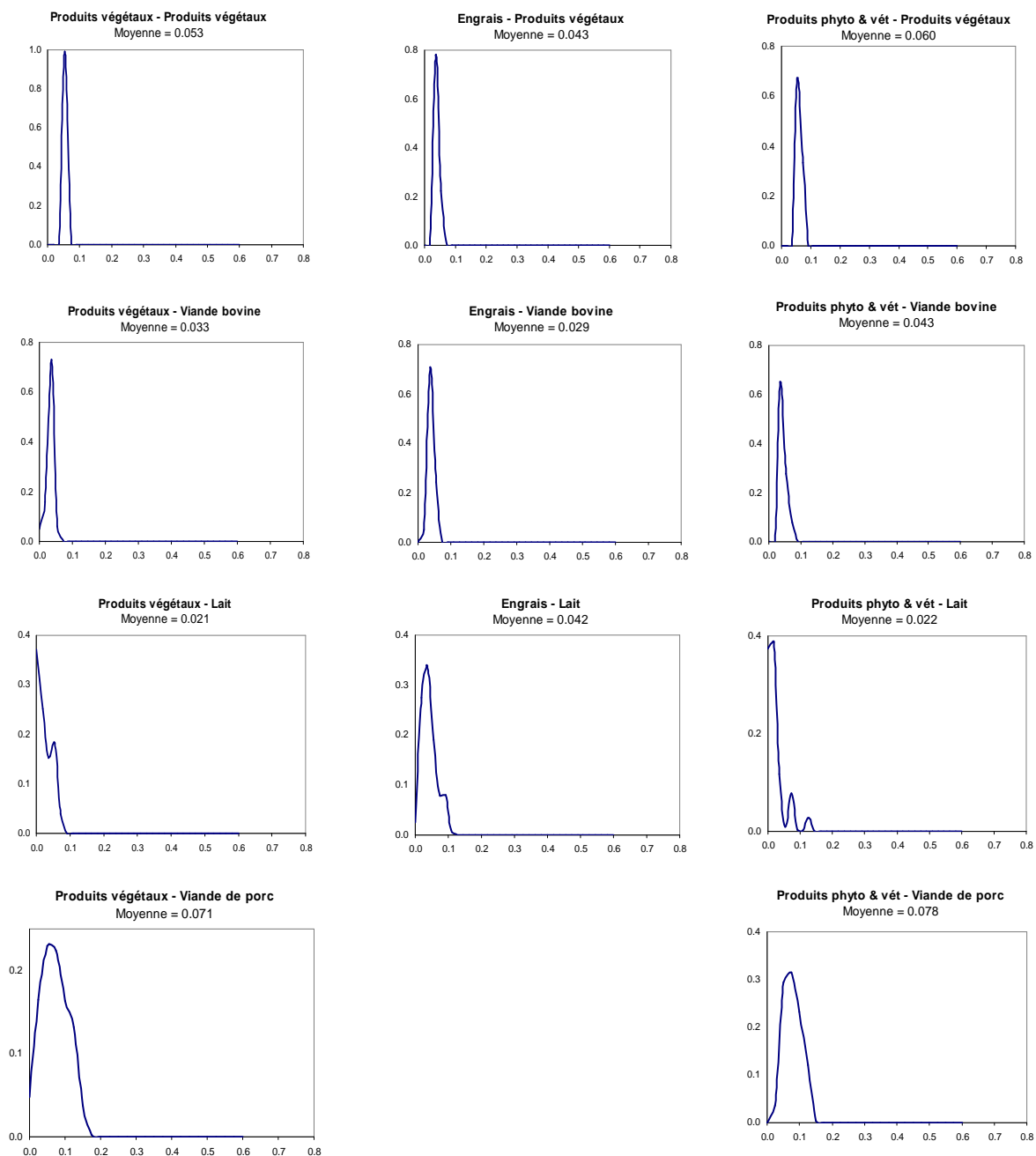
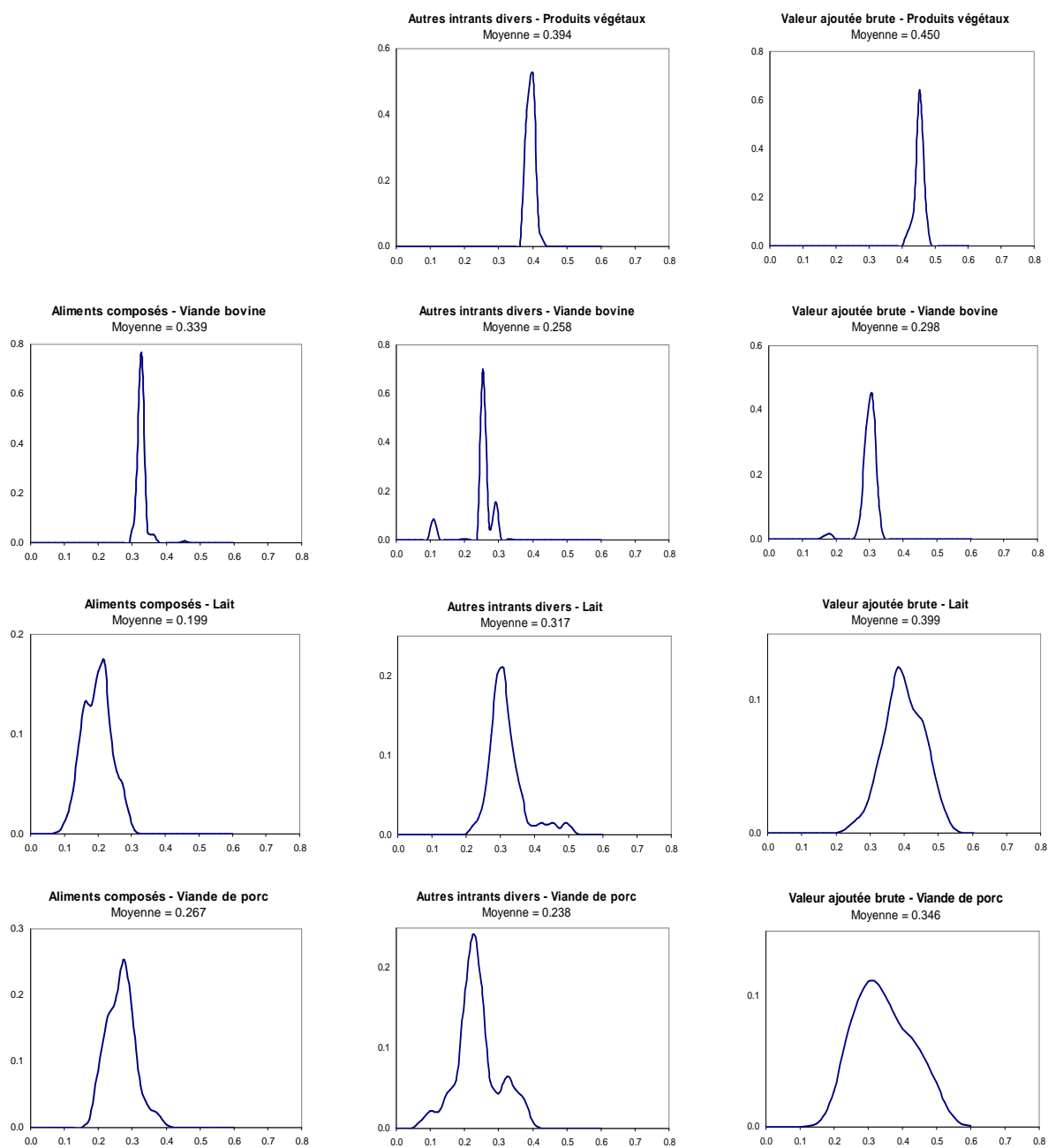


Figure 1 (suite)



activité de production donnée est indépendante de toutes les autres variations qui caractériseraient d'autres activités de production (Dixon *et al.*, 1984 ; Ray, 1985). De plus, le résultat montre clairement que l'hypothèse de covariances nulles sous-tendant le modèle standard développé par Hildreth et Houck ne semble pas être réaliste.

4.4. Validation du modèle

La validation du modèle que nous effectuons prend en compte à la fois la "qualité" (c.à.d. la conformité par rapport aux valeurs attendues *a priori*) des estimations et le "pouvoir prédictif" inter-échantillon (appelé également "précision") du modèle. Cela nous permet d'évaluer ainsi la performance d'ensemble des coefficients variables estimés par la méthode ECG.

4.4.1. Complémentarités/substitutions entre intrants

Si nous supposons que les prix des intrants et des produits sont les mêmes pour tous les exploitants agricoles de l'échantillon, les coefficients individuels obtenus devraient être conformes à ce que nous savons *a priori* sur la diversité des structures de coûts existant parmi les exploitations agricoles. Ainsi, pour tester l'existence de relations de substitution ou de complémentarité entre intrants, nous calculons les coefficients de corrélation des rangs reposant sur la statistique *rho* de Spearman. Les résultats de cet exercice figurent dans le tableau 4³².

Dans l'ensemble, les résultats empiriques sont conformes à ce que l'on attendait. Ainsi, quelle que soit l'activité de production animale, la corrélation entre, d'un côté, les aliments composés, et de l'autre, les utilisations d'engrais, de produits végétaux, et des produits vétérinaires et phytosanitaires est négative. Ces résultats confirment bien l'existence de relations techniques de substitution entre ces intrants. De plus, les complémentarités qui doivent exister

³² L'utilisation de coefficients de corrélation peut parfois conduire à émettre de fausses conclusions sur les dépendances entre variables, car cet indicateur ne peut contrôler tous les autres facteurs influençant lesdites relations. Pour remédier à ce dernier problème, on pourrait estimer les coefficients de corrélations *partiels*. Toutefois, la mise en œuvre d'une telle procédure est hors de question car elle engendre un très grand nombre de corrélations que l'on peut difficilement interpréter, même avec un nombre limité de variables. Cependant, en effectuant quelques expérimentations (en contrôlant par exemple la dimension des exploitations), nous trouvons que les relations statistiques estimées pour chaque ferme entre les coefficients de répartition des coûts restent très robustes.

Tableau 4 : Complémentarités/substitutions entre intrants utilisés dans les activités animales -Analyse des corrélations à partir du coefficient *rho* de Spearman

	Produits végétaux	Engrais	Produits phyto.et vét.	Aliments composés	Autres intrants divers	Valeur ajoutée brute
Viande bovine						
Produits végétaux	1	0,535**	0,414*	-0,845**	0,102	-0,133
Engrais		1	0,027	-0,706**	-0,104	0,087
Produits phyto. & vét.			1	-0,239	-0,214	-0,393*
Aliments composés				1	-0,159	0,033
Autres intrants					1	-0,220
Valeur ajoutée brute						1
Lait						
Produits végétaux	1	0,269	0,709**	-0,619**	0,135	-0,450**
Engrais		1	0,281	-0,580**	-0,108	-0,159
Produits phyto. & vét.			1	-0,434*	-0,061	-0,527**
Aliments composés				1	-0,268	0,171
Autres intrants					1	-0,547**
Valeur ajoutée brute						1
Viande porcine						
Produits végétaux	1	n,a,	0,742**	-0,762**	0,364	-0,509**
Engrais		1	n.d.	n. d.	n. d.	n. d.
Produits phyto. & vét.			1	-0,604**	0,175	-0,449*
Aliments composés				1	-0,279	0,182
Autres intrants					1	-0,820**
Valeur ajoutée brute						1

*, **, La corrélation est respectivement significative à un niveau de 0,05 and 0,01 % (tests bilatéraux).

n. d. : non disponible

entre les intraconsommations de produits végétaux et l'emploi de produits phytosanitaires et vétérinaires³³ sont bien vérifiées pour toutes les activités animales, comme le montrent les valeurs positives prises par la statistique *rho* de Spearman. Enfin, il est impossible de discerner des relations significatives entre intrants variables et semi-variables, mais l'on note l'existence de corrélations négatives entre la valeur ajoutée brute (VAB) et la plupart des autres intrants, à l'exception des engrais et des aliments composés.

4.4.2 Rôle joué par la dimension des exploitations agricoles

La relation entre la dimension des exploitations agricoles et la structure des coûts est analysée en regroupant les observations de l'échantillon en trois catégories. Le critère de

³³La catégorie des produits vétérinaires comprend des additifs alimentaires tels que les vitamines, les antibiotiques et les autres médicaments.

répartition des observations repose sur les quartiles des distributions empiriques du nombre total de têtes de cheptel (mesurée en unités bétail) apparaissant dans la partie C du tableau 1. Les groupes d'exploitations que l'on considère être de "petite" ou "grande" dimension correspondent respectivement au premier et quatrième quartiles, tandis que les exploitations de taille moyenne se situent à l'intérieur de l'intervalle interquartile. En utilisant cette typologie, les exploitations agricoles de notre échantillon se répartissent comme suit : neuf fermes sont de petite dimension, seize de taille moyenne et neuf sont considérées comme grandes.

Le tableau 5 présente les estimations moyennes des paramètres pour les trois classes d'exploitations. Les résultats indiquent qu'il y a des différences substantielles entre les exploitations de petite et grande dimension. Cette tendance est surtout prononcée pour les activités de production laitière et de viande bovine (et à un degré bien moindre pour les activités de production végétale et porcine).

Si l'on approfondit l'analyse, on se rend compte qu'hormis les aliments composés utilisés dans les productions laitières et de viande bovine, la plupart des coûts unitaires estimés pour tous les intrants spécifiques (variables) utilisés dans les activités de productions animales sont systématiquement inférieurs dans les exploitations de grande dimension par rapport à ceux des exploitations de petite dimension. Les tests effectués à l'aide de la statistique non paramétrique U de Mann-Whitney permettent de conclure que la différence est significative dans beaucoup de cas (et pour certains cas, la statistique U prend même des valeurs proches de la valeur critique). Par contre, les exploitations de grande dimension produisant du lait ou de la viande bovine utiliseraient une plus grande proportion d'aliments composés. Ce résultat est prévisible, car les exploitations agricoles de grande dimension sont probablement plus intensives, mais il n'est pas possible à partir de notre analyse d'isoler les effets de substitution propres aux productions végétales des gains d'efficacité (techniques et/ou liés aux coûts). Ainsi, les fermes de grande taille peuvent être en mesure d'acheter des aliments composés à des prix plus bas en raison des réductions offertes sur les volumes achetés. Un autre résultat quelque peu surprenant tient dans le fait que la consommation d'autres intrants semble être indépendante de la dimension des exploitations. Les "économies de taille" attendues pour l'utilisation des intrants semi-variables

Tableau 5 : Rôle joué par la dimension des exploitations agricoles. Test de différences entre paramètres moyens estimés, par dimension d’exploitation, en utilisant la statistique U de Mann-Whitney

Taille des exploitations	Produits végétaux			Viande bovine		
	Petite	Intermédiaire	Grande	Petite	Intermédiaire	Grande
Nombre d’exploitations concernées	9	16	9	9	16	9
Produits végétaux	0,056**	0,051	0,052**	0,043***	0,031	,028***
Engrais	0,043*	0,043	0,042*	0,038**	0,027	0,025**
Produits phyto. & vét.	0,064	0,059	0,059	0,048*	0,042	0,039*
Aliments composés				0,316***	0,354	,355***
Autres intrants	0,394	0,393	0,397	0,261	0,248	0,271
Valeur ajoutée brute	0,442	0,454	0,450	0,293	0,297	0,302
Taille des exploitations	Lait			Viande porcine		
	Petite	Intermédiaire	Grande	Petite	Intermédiaire	Grande
Nombre d’exploitations concernées	9	16	9	3	14	8
Produits végétaux	0,042***	0,015	0,011***	0,082	0,069	0,070
Engrais	0,058**	0,039	0,031**			
Produits phyto. & vét.	0,045*	0,014	0,013*	0,085	0,078	0,075
Aliments composés	0,167***	0,209	0,212***	0,254	0,265	0,275
Autres intrants	0,317	0,313	0,327	0,246	0,235	0,242
Valeur ajoutée brute	0,371	0,411	0,407	0,332	0,353	0,338

*, **, *** : différence des moyennes significative pour des niveaux respectifs de 10%, 5% et 1%. (tests bilatéraux).

ne se concrétiseraient donc pas toutes. Enfin, les résultats montrent également que les paramètres associés à VAB s’accroissent avec la taille des exploitations. Cette relation qui n’est pas statistiquement significative est plus prononcée pour les activités laitière et de viande bovine.

4.4.3 Précision du modèle

La validation du modèle peut se faire aussi en analysant la précision avec laquelle il reproduit les valeurs observées des dépenses en intrants. Le pseudo- R^2 et l’erreur absolue moyenne relative (ou *MAPE* pour *Mean Absolute Percentage Error*) sont les deux indicateurs de précision que nous avons calculés pour chaque variable endogène (c.à.d. pour chaque dépense en intrants). Le pseudo- R^2 est défini comme le carré du coefficient de corrélation entre x_{it} et \hat{x}_{it} , tandis que $MAPE_i = (1/T)\sum_t (|x_{it} - \hat{x}_{it}|/x_{it})$, où x_{it} et \hat{x}_{it} sont respectivement les valeurs observées et “prédites” des dépenses totales consacrées aux achats de l’intrant i .

Les pseudo- R^2 et les valeurs de l'indicateur *MAPE* pour les modèles V-ECG and F-ECG sont présentés au tableau 6. Dans le cas du modèle V-ECG, les valeurs obtenues pour le pseudo- R^2 et calculées pour des paramètres *moyens* estimés sont comprises entre 0,30 et 0,77, alors que les pseudo- R^2 obtenus pour le modèle F-ECG prennent des valeurs variant entre 0,43 et 0,87. De même, les valeurs prises par l'indicateur *MAPE* sont comprises entre 32,6% et 77,9% pour la spécification V-ECG, et entre 26,7% et 54,9% pour la spécification F-ECG.

En s'appuyant sur ces derniers résultats, on peut être tenté de conclure que le modèle F-ECG est plus précis que le modèle V-ECG. Toutefois, cette conclusion a besoin d'être remise en cause pour au moins quatre raisons. En premier lieu, la supériorité apparente du modèle F-ECG en termes de "qualité d'ajustement" n'est pas vraiment surprenante. En effet, en ignorant l'hétérogénéité des exploitations, on oblige les coefficients fixes (moyens) à saisir la variation des données avec le risque de prendre des valeurs estimées qui tendent à s'éloigner des "vraies" valeurs des paramètres. Ensuite, les augmentations du pseudo- R^2 ou les baisses de l'indicateur *MAPE* observées pour le modèle F-ECG par rapport au modèle V-ECG sont modestes, et plus particulièrement pour les intrants les plus importants (aliments composés, autres intrants et VAB), même si les structures sous-jacentes des coûts moyens résultant des coefficients (moyens) des deux modèles V-ECG et F-ECG sont très différentes. Troisièmement, si nous examinons les deux indicateurs de précision calculés à partir des coefficients *individuels* du modèle V-ECG, il ne fait aucun doute que la précision de ce dernier modèle s'améliore substantiellement : les valeurs prises par le pseudo- R^2 sont comprises entre 0,94 et 1, tandis que l'indicateur *MAPE* prend des valeurs variant entre un minuscule 0,7% et 8,1% pour les catégories d'intrants les moins importantes (consommation de produits végétaux, engrais, produits phytosanitaires et vétérinaires), et entre 17,2% et 22,3% pour les principaux intrants (aliments composés, autres intrants divers et VAB). Alors qu'un tel résultat était quelque peu prévisible *a priori*, il montre tout de même que l'hétérogénéité non observée entre les exploitations, qui est mesurée par les éléments individuels et variables \hat{v}_{ikt} ,³⁴ joue un rôle prééminent pour "expliquer" les dépenses totales en intrants. Enfin, ici, l'utilité du modèle ne tient pas tant à son pouvoir prédictif qu'à sa capacité à produire des *estimations* plausibles des coefficients moyens de répartition des coûts,

³⁴ On peut aussi soutenir que ces éléments mesurent également les "interactions" qui existent entre les valeurs observées des productions et les caractéristiques spécifiques et non observables des fermes.

Tableau 6: Pseudo- R^2 et statistique $MAPE$ pour les modèles V-ECG et F-ECG

Intrants	Modèle à paramètres variables (V-ECG)				Modèle à paramètres fixes (F-ECG)	
	Paramètres <i>individuels</i> estimés		Paramètres <i>moyens</i> estimés			
	Pseudo- R^2	$MAPE$	Pseudo- R^2	$MAPE$	Pseudo- R^2	$MAPE$
Produits végétaux	0,997	8,1%	0,544	77,9%	0,693	54,9%
Engrais	0,999	0,7%	0,296	43,0%	0,427	34,9%
Produits phyto. et vét.	0,998	3,2%	0,597	35,9%	0,783	26,7%
Aliments composés	0,976	22,3%	0,769	38,7%	0,873	37,1%
Autres intrants divers	0,943	17,2%	0,620	32,6%	0,640	31,2%
Valeur ajoutée brute	0,962	22,1%	0,722	38,8%	0,750	33,7%

permettant de prendre en compte les "effets purs" des volumes de production et/ou des combinaisons productives.

5. Conclusion

Dans cette communication, nous développons une approche de modélisation à paramètres variables reposant sur la notion d'entropie croisée généralisée (ECG) en vue d'estimer des coûts de production (non observables) pour des firmes caractérisées par des activités multi-produits. Dans la partie empirique de ce travail, nous utilisons un échantillon de 34 exploitations bretonnes à orientation laitière et viande bovine, pour lesquelles nous disposons de données comptables moyennes se rapportant à la période 2000-2001. Les résultats empiriques obtenus sont, de prime abord, plausibles, ce qui nous conduit à penser que la procédure d'estimation de type ECG pourrait être appliquée avec un certain succès à d'autres situations économiques et agricoles.

Les apports de cette communication peuvent se résumer comme suit. Tout d'abord, et à l'inverse des études antérieures qui ont analysé la répartition des coûts (ou les technologies multi-produits) en l'absence de données individuelles pour chaque activité de production, le point central de ce travail a été de prendre en compte l'hétérogénéité des firmes (des exploitations agricoles dans notre cas). Deuxièmement, la procédure d'estimation se fondant sur la notion d'entropie croisée généralisée (ECG) semble être une façon efficace de surmonter les problèmes méthodologiques et pratiques rencontrés dans les travaux relatifs à la répartition des coûts. En

particulier, nous sommes en mesure d'obtenir des valeurs estimées des paramètres qui sont plausibles, ayant les bons signes, satisfaisant toutes les contraintes liées à l'identité comptable équilibrant recettes et dépenses, tout en restant généralement cohérents avec nos intuitions de départ. Troisièmement, ce travail montre clairement qu'une méthode d'estimation utilisant la notion d'ECG est bien plus facile à mettre en œuvre et bien plus performante en moyens de calcul que les autres procédures d'estimation conventionnelles. Cela est d'autant plus vrai que nous avons affaire à l'estimation d'un système d'équations comportant des paramètres variables et individuels, et devant en plus satisfaire plusieurs contraintes. Le quatrième apport de ce travail a été de procéder à une application réelle de la méthode EGC à des exploitations agricoles, alors que toutes les études précédentes citées dans cette communication utilisant la notion d'entropie ont eu recours à des données expérimentales générées par des simulations de Monte Carlo. Enfin, l'approche de modélisation et d'estimation, que nous venons d'exposer, peut être aussi pertinente pour tout chercheur travaillant sur les modèles input-output.

L'approche de modélisation adoptée dans cette communication ne permet pas d'estimer des effets de substitution induits par les prix et c'est une limite de notre travail. En fait, le modèle utilisé n'est pas une fonction de coût, car il ne dépend pas explicitement des prix des intrants. Evidemment, des changements dans les prix relatifs des intrants induiraient des variations de coûts et éventuellement conduiraient à des modifications des consommations d'intrants. Cela signifie que les coefficients de répartition des coûts que nous avons estimés n'ont qu'une validité de court terme et que, par conséquent la capacité de prédiction *ex-ante* du modèle est plutôt limitée. Bien que nous reconnaissons cet inconvénient évident, nous pensons que cela ne constitue pas un problème grave, car les estimations de répartition des coûts peuvent être facilement remises à jour pour prendre en compte les variations de prix et/ou les changements dans les relations de production.

Cependant, des questions importantes de recherche subsistent encore. Une voie intéressante pour des recherches futures serait "d'expliquer" la variation des coefficients en intégrant ou contrôlant dans le modèle quelques variables comme le facteur "gestion de l'exploitation" ou des caractéristiques structurelles des fermes (telles que la dimension). En procédant de cette manière, on permettrait aux paramètres de varier systématiquement. De même,

on pourrait obtenir des améliorations dans les estimations des paramètres en adoptant les actions suivantes : en créant, par exemple, des catégories d'intrants et de produits plus homogènes, en utilisant de plus grands échantillons représentatifs, en exploitant la "richesse" des données de panel et en ayant recours à des informations préalables plus réalistes (telles que des opinions d'expert) sur les valeurs support des coefficients de répartition des coûts.

Références

- [1] Andrews, D.W.K. (1996). "Admissibility of the Likelihood Ratio test when the parameter space is restricted under the alternative." *Econometrica*, 64, 705-718.
- [2] Argilés, J.M. et Slof, E.J. (2001). "New opportunities for farm accounting." *The European Accounting Review*, 10, 361-383.
- [3] BiØrn, E., Lindquist, K.G., et Skjerpen, T. (2003). "Random coefficients in unbalanced panels: an application on data from chemical plants." *Annales d' Économie et de Statistique*, 69, 55-85.
- [4] Butault, J.P., Hassan, C.R., et Reignier, E. (1988). *Les coûts de production des principaux produits agricoles dans la CEE*. Luxembourg: Office of Official Publications of the European Communities.
- [5] Chambers, J., Cleveland, W., Kleiner, B., et Tukey, P. (1983). *Graphical Methods for Data Analysis*. Wadsworth.
- [6] Commission Européenne (2000) Réseau d'information comptable agricole. Guide méthodologique. http://europa.eu.int/comm/agriculture/rica/pdf/site_fr.pdf.
- [7] Divay, J-F et F., Meunier (1980). Deux méthodes de confection du tableau entrées-sorties. *Annales de l'INSEE*, 37, 59-111.
- [8] Dixon, B.L., Batte, M.T. et Sonka, S.T. (1984). "Random coefficients estimation of average total product costs for multiproduct firms." *Journal of Business and Economic Statistics*, 2, 360-366.
- [9] Dixon, B. et Hornbaker, R.H. (1992). "Estimating the technology coefficients in linear programming models." *American Journal of Agricultural Economics*, 74, 1029-1038.
- [10] Efron, B. et Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability, 57. New York: Chapman & Hall.
- [11] Errington, A. (1989). "Estimating enterprise input-output coefficients from regional farm data." *Journal of Agricultural Economics*, 40, 52-56.
- [12] Golan, A., Judge, G., et Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: John Wiley & Sons.
- [13] Golan, A., Perloff, J.M., et Shen, Z. (2001). "Estimating a demand system with non-negativity constraints: Mexican meat demand." *The Review of Economics and Statistics*, 8, 541-550.

- [14] Greene, W. (2004). *Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models*. Department of Economics, Stern School of Business, New York University (www.stern.nyu.edu/~wgreene).
- [15] 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.
- [16] Hallam, D., Bailey, A., Jones, Ph., et Errington, A. (1999). "Estimating input use and production costs from farm survey panel data." *Journal of Agricultural Economics*, 50, 440-449.
- [17] Hildreth, C. et Houck, J.P. (1968). "Some estimators for a linear model with random coefficients." *Journal of the American Statistical Association*, 63, 583-594.
- [18] Hornbaker, R.H., Dixon, B.L., et 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.
- [19] Judge, G.G, Griffiths, W.E., Hill, R.C., Lütkepohl, H., et Lee, T.C. (1985). *The Theory and Practice of Econometrics*, 2nd edition. New York: Wiley.
- [20] Lence, S.H. et 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.
- [21] Lence, S.H. et 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.
- [22] Léon, Y., Peeters, L., Quinqu, M., et 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.
- [23] Midmore, P. (1990). "Estimating input-output coefficients from regional farm data: a comment." *Journal of Agricultural Economics*, 41, 108-111.
- [24] Moxey, A. et Tiffin, R. (1994). "Estimating linear production coefficients from farm business survey data: a note." *Journal of Agricultural Economics*, 45, 381-385.
- [25] O'Donnell, C.K., Rambaldi, A.N., et Doran, H.E. (2001). "Estimating economic relationships subject to firm- and time-varying equality and inequality constraints." *Journal of Applied Econometrics*, 16, 709-726.
- [26] Oude Lansink, A. (1999). "Generalised maximum entropy estimation and heterogeneous technologies." *European Review of Agricultural Economics*, 26, 101-115.
- [27] Pingault, N. et Desbois, D. (2003). "Estimations des coûts de production des principaux produits agricoles à partir du RICA." *Notes et Études économiques*, No. 19, 9-51.

- [28] Preckel, P.V. (2001). "Least squares and entropy: a penalty function perspective." *American Journal of Agricultural Economics*, 83, 366-377.
- [29] Ray, S.C. (1985). "Methods of estimating the input coefficients for linear programming models." *American Journal of Agricultural Economics*, 67, 660-665.
- [30] Scandizzo, P.L. (1990). "The estimation of input-output coefficients: methods and problems." *Richerche Economiche*, 44, 455-474.
- [31] Schwallie, D.P. (1982). "Unconstrained maximum likelihood estimation of contemporaneous covariances." *Economics Letters*, 9, 359-364.
- [32] Shen, E.Z. et Perloff, J.M. (2001). "Maximum entropy and Bayesian approaches to the ratio problem." *Journal of Econometrics*, 104, 289-313.
- [33] Swamy, P.A.V.B. (1970). "Efficient inference in a random coefficient regression model." *Econometrica*, 38, 311-33
- [34] Swamy, P.A.V.B et 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.
- [35] Zaman, A. (2002). "Maximum likelihood estimates of the Hildreth-Houck random coefficients model." *Econometrics Journal*, 5, 237-262.