

Explorer la distribution des intervalles d'estimations quantiles conditionnels : une application à l'estimation de coûts spécifiques de production du lait de vache dans l'Union européenne¹.

Dominique Desbois, UMR Economie publique, Inra-AgroParisTech, Université Paris-Saclay²

Résumé : Cet article utilise les outils d'analyse des données symboliques pour visualiser les intervalles d'estimations quantiles conditionnelles, en l'appliquant au problème de l'allocation des coûts en agriculture. Après avoir rappelé le cadre conceptuel de l'estimation des coûts de production agricole, la première partie présente le modèle empirique, l'approche de régression quantile et les techniques de traitement des données en intervalles utilisées comme outils d'analyse de données symboliques. La seconde partie présente l'analyse comparative des résultats économétriques du lait entre douze États membres européens, en utilisant l'analyse en composantes principales et le regroupement hiérarchique des intervalles d'estimation, en discutant de la pertinence des graphiques exploratoires obtenus pour les comparaisons internationales.

Mots-clés : production agricole, lait de vache, allocation des coûts, micro-économie, régression quantile, analyse des données symboliques, intervalles de confiance.

Exploring the distribution of conditional quantiles estimation ranges: an application to the estimation of specific costs of production of cow's milk in the European Union.

Abstract: This article uses symbolic data analysis tools to visualize conditional quantile estimation intervals, applying it to the problem of cost allocation in agriculture. After recalling the conceptual framework of the estimation of agricultural production costs, the first part presents the empirical model, the quantile regression approach and the interval data processing techniques used as symbolic data analysis tools. The second part presents the comparative analysis of the econometric results of milk between twelve European Member States, using the principal components analysis and the hierarchical grouping of the estimation intervals, by discussing the relevance of the exploratory graphs obtained for the international comparisons.

Keywords: input-output model, agricultural production cost, micro-economics, quantile regression, symbolic data analysis, interval estimates.

¹ Les estimations utilisées dans ce travail ont été effectuées lors de la préparation de la thèse de l'auteur (Desbois, 2015a), co-dirigée par Y. Surry et J.C. Bureau, bénéficiant du support du projet *Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture* du 7^e Programme-cadre de la Communauté européenne (FP7/2007-13, agrément n°212292). Cependant, ce texte n'engage que l'auteur, seul responsable d'éventuelles erreurs ou omissions.

² Pour toute correspondance : dominique.desbois@inra.fr et/ou desbois@agroparistech.fr.

“Applied economists increasingly want to know what is happening to an entire distribution, to the relative winners and losers, as well as to averages.”
(Angrist et Pischke, 2009)

I Introduction à la problématique d'estimation des coûts de production agricoles

Les réformes successives de la Politique agricole commune (Pac), l'intégration des agricultures des Etats membres résultant du processus d'élargissement de l'Union européenne (UE) suscitent tant dans le contexte de marchés concurrentiels que de marchés soumis à régulation, des besoins récurrents d'estimation des coûts de production des principaux produits agricoles. L'analyse des coûts de production agricoles, qu'elle soit rétrospective ou prospective, est également un outil d'analyse des marges des agriculteurs. Elle permet d'évaluer la compétitivité-prix des agriculteurs, un des éléments majeurs du développement ou du maintien de filières agroalimentaires dans certaines régions européennes. Ainsi, l'estimation des coûts de production apporte des éclairages certes partiels mais indispensables sur les questions que pose l'adaptation des agricultures européennes au contexte des marchés agricoles qu'ils soient nationaux, européens ou internationaux tant du point de vue de la régulation des échanges internationaux de produits agricoles (cf. les propositions de mesure pour lutter contre les déséquilibres de marché dans le cadre de l'après quotas pour le secteur laitier³), que des réformes successives de la Pac (cf. le débat sur les orientations de la future Pac en 2020⁴) ou des nouveaux défis lancés aux agricultures européennes par les facteurs environnementaux (changement climatique, gestion de l'environnement et de la biodiversité⁵).

Confronté plus directement aux risques de prix depuis la suppression des quotas laitiers en 2015⁶, les producteurs européens de lait de vache ayant peu d'opportunités de différenciation optent pour des stratégies de réduction des coûts, en cherchant soit à diminuer les coûts de structure en jouant sur le volume de la production, soit à diminuer les coûts spécifiques en optimisant la gestion des intrants ou en optant pour des itinéraires techniques à bas niveau d'intrants. Cependant, l'ajustement par les structures n'est pas toujours possible en raison des contraintes (gestion du troupeau, droits à produire, disponibilité) qui peuvent restreindre l'accès aux trois principaux facteurs de production que sont la terre (par exemple, en zones de montagne), le capital d'exploitation (conditions de financement) ou le travail, qu'il soit salarié ou non salarié. Par contre, l'ajustement sur les intrants spécifiques offre plus de souplesse comme le montre l'adoption de pratiques raisonnées conduisant à des économies sur les principaux postes de dépense comme l'alimentation animale et les frais vétérinaires. L'évolution des coûts spécifiques, non seulement globalement mais aussi par produits, constituent ainsi un indicateur important pour les éleveurs laitiers en matière de conduite technique du troupeau et d'ajustement de leurs mix-produit aux demandes des marchés agricoles, compte-tenu des ressources et des facteurs de compétitivité dont ils disposent.

Face à l'hétérogénéité des structures de production agricole et des choix productifs en Europe, comment restituer le maximum d'informations utiles dans l'estimation des coûts de production agricoles ? Compte-tenu de ces différents enjeux, dans des contextes soit *ex ante* d'élaboration de scénarii soit *ex post* d'évaluation de mesures concernant d'éventuelles options de politique publique agricole, nous devons être en mesure de fournir des informations comme le suggère (Angrist et Pischke, 2009) sur l'ensemble de la distribution des coûts de production, permettant ainsi de répondre aux besoins de simulations ou d'analyse d'impact au sein des différentes organisations communes de marché. Dans cette perspective, à partir du constat d'asymétrie et d'hétérogénéité de leur distribution

³ Cf. <http://agriculture.gouv.fr/etude-sur-les-mesures-contre-les-desequilibres-de-marche-quelles-perspectives-pour-lapres-quotas>

⁴ Cf. <https://www.sfer.asso.fr/source/Coll-trajectoire-2018/Programme-Future-of-CAP-30-05.pdf>

⁵ Cf. <http://agriculture.gouv.fr/lagriculture-et-les-forets-au-coeur-de-la-cop23>

⁶ Cf. *EU Milk Margin Estimate up to 2016*, n°16 : “Gross margins: a lot of instability and a record low level in third quarter of 2016”, https://ec.europa.eu/agriculture/sites/agriculture/files/rural-area-economics/briefs/pdf/016_en.pdf

empirique, nous proposons une méthodologie adaptée au problème de l'estimation des coûts spécifiques de production relativement aux principaux produits agricoles de référence dans un contexte européen où les exploitations agricoles demeurent de façon majoritaire multiproduits, malgré une prépondérance d'exploitations spécialisées dans certains secteurs plus intégrés de la production agricole. Dans ce contexte multiproductif, il est stratégique de générer pour chacun des principaux produits agricoles les estimations centrales de la distribution des coûts mais également les quantiles inférieurs ou supérieurs dans une optique de sélectivité des instruments de régulation des marchés agricoles pour l'élaboration et l'évaluation des politiques publiques.

En réponse à cette préoccupation, nous proposons une méthodologie d'estimation susceptible de fournir des informations sur l'ensemble de la distribution des coûts spécifiques de production relative aux principaux produits agricoles de référence dans un contexte européen. Afin de s'affranchir de la contrainte d'estimateurs en moyenne sensibles à l'asymétrie ou aux valeurs extrêmes des distributions d'intérêt et susceptibles de masquer les différences inter-structurelles, il convient de générer pour chacun des principaux produits agricoles non seulement les estimations médianes de la distribution des coûts mais également leurs quantiles inférieurs ou supérieurs. À cette fin, nous proposons d'utiliser une méthodologie permettant d'obtenir des estimations de ces quantiles de coûts spécifiques qui soient conditionnées par le mix-produit des exploitants agricoles (Desbois, Butault et Surry, 2017). Afin de juger de la pertinence de cette approche, nous appliquerons par la suite cette méthodologie pour l'estimation des coûts spécifiques du lait de vache, compte-tenu de sa place dans la valeur agricole produite par l'UE27⁷, sur un ensemble d'Etats européens (UE12) où ces productions sont significatives en 2006, année de référence choisie pour la période antérieure à la suppression des quotas laitiers.

Nous présentons tout d'abord le modèle empirique d'estimation des coûts spécifiques de production, dérivé d'une approche économétrique d'allocation des coûts, initialement développée par (Aufrant, 1983) proposant d'utiliser les données micro-économiques pour construire une matrice d'input-output (Divay et Meunier, 1980). Puis, nous introduisons la méthodologie d'estimation selon les quantiles conditionnels proposée par (Koenker et Basset, 1978) puis formalisée dans Koenker (2005). Ensuite, nous présentons les procédures d'analyse des données symboliques utilisées pour explorer les estimations empiriques en intervalles des distributions de quantiles conditionnels en se basant sur les concepts et méthodes fournis par l'approche symbolique principalement conceptualisée dans Bock et Diday (2000) puis synthétisée dans Billard et Diday (2006). Puis, nous présentons les graphiques issus des outils d'analyse des données symboliques appliqués aux intervalles d'estimation des quantiles conditionnels. Enfin, nous concluons sur la pertinence de cette approche appliquée à la production de lait de vache, proposant une extension de ce type d'analyse au niveau régional.

II Cadre conceptuel et aspects méthodologiques de l'allocation des coûts

Des enquêtes spécifiques aux grandes commodités agricoles sont conduites par atelier pour fournir des données détaillées sur les coûts opérationnels de production, comme celle pratiquée par l'Institut de l'Élevage sur les élevages spécialisés en bovins lait pour la France⁸. Cependant, ces enquêtes technico-économiques sont relativement coûteuses rendant leur généralisation à l'ensemble des exploitations laitières européennes financièrement insupportable. Aussi, ce travail se situe dans une problématique d'imputation des coûts des facteurs à des productions multiples, initiée à l'échelle européenne par les travaux de l'Inra (Butault, Hassan et Reignier, 1988) financés par la Commission des Communautés européennes (CCE), permettant d'estimer des coûts spécifiques aux produits sur

⁷ En 2015, la production de lait de vache en représentait environ 14,1 % de la valeur produite par les différentes agricultures européennes (d'après https://ec.europa.eu/agriculture/milk_en).

⁸ http://idele.fr/fileadmin/medias/Documents/CR_1250028-couts_production_BL-manuel_de_reference.pdf

la base du Réseau d'information comptable (RICA), enquête comptable harmonisée du point de vue de la définition des exploitations professionnelles et des agrégats comptables, techniques et financiers.

II.1 Le modèle d'estimation des coûts spécifiques de production liés aux marges brutes

Dans les systèmes de comptabilités agricoles de l'UE, l'enregistrement des charges s'effectue de manière agrégée au niveau de l'exploitation et ne fournit pas d'estimation directe des coûts de production supportés par cette exploitation pour chacune des spéculations agricoles entreprises. En revanche, la « fiche d'exploitation »⁹ de l'enquête RICA fournit individuellement par exploitation agricole à partir des enregistrements comptables le montant des produits bruts générés par les différentes spéculations et celui de l'ensemble des frais spécifiques, somme des achats d'intrants enregistrés. De sorte qu'il devient possible d'estimer par régression des frais spécifiques sur les produits bruts des coefficients d'affectation des dépenses aux principaux produits agricoles, que nous appellerons « coefficients spécifiques de production ».

Conformément aux spécifications du modèle général d'estimation des coûts de production (Desbois, Butault et Surry, 2013), le total des charges spécifiques (x_i), pour une exploitation agricole $i \in I$, est lié aux p produits bruts Y_i^j pour $j = 1, \dots, p$ par le modèle de régression suivant (Butault *et al.*, 1988) :

$$x_i = \sum_{j=1}^p (\alpha_j \times Y_i^j) + \epsilon_i \quad \text{pour } i = 1, \dots, n \quad (1)$$

où ϵ_i est un terme stochastique résiduel d'espérance nulle et à variance constante. Les coefficients s'interprètent comme un coefficient unitaire d'allocation de charges spécifiques proportionnel au produit brut des productions de l'exploitation agricole multiproduits.

Considérant la marge brute M_i de l'exploitation i comme différentiel entre la somme des p produits bruts $Y_i = \sum_{j=1}^p (Y_i^j)$ et le total de ses charges spécifiques x_i , les marges brutes sont exprimées en fonction des produits bruts par reparamétrisation du modèle (1) :

$$M_i = \sum_{j=1}^p (Y_i^j) - x_i = \sum_{j=1}^p (Y_i^j) - (\sum_{j=1}^p \alpha_j Y_i^j + \epsilon_i) = \sum_{j=1}^p (1 - \alpha_j) Y_i^j - \epsilon_i \quad (2)$$

En prenant l'espérance mathématique $E[.]$ de chacun des termes extrêmes de l'équation (2), nous obtenons :

$$E[M_i] = E[\sum_{j=1}^p (1 - \alpha_j) Y_i^j] - E[\epsilon_i] = \sum_{j=1}^p (1 - \alpha_j) E[Y_i^j] \quad (3)$$

du fait de la linéarité de l'opérateur $E[.]$ et de la spécification centrée du vecteur des aléas $E[\epsilon_i] = 0$.

Les produits Y_i^j étant positifs ou nuls par définition, nous en dérivons par identification le calcul d'espérance mathématique des p composantes-produits M_i^j de la marge brute, soit :

$$E[M_i^j] = (1 - \alpha_j) E[Y_i^j] = \gamma_j E[Y_i^j] \quad (4)$$

Ainsi, les valeurs estimées des coefficients $\hat{\gamma}_j = 1 - \hat{\alpha}_j$, pour $j = 1$ à p par la régression des moindres carrés ordinaires (MCO), fournissent les espérances conditionnelles des taux unitaires de marge brute de l'ensemble J des produits.

La décomposition linéaire de la marge brute nous conduit donc à estimer les coefficients spécifiques de production de l'équation stochastique (1) aux fins de comparaison :

- d'une part, selon la méthodologie de la régression des moindres carrés ordinaires (MCO) dont la solution $\hat{\mu}_{MCO}(i) = \sum_{j=1}^p \hat{\alpha}_j^{MCO} Y_i^j$ s'exprime classiquement en termes d'espérance conditionnelle ;

⁹ Le questionnaire servant à établir cette « fiche d'exploitation » et la méthodologie de l'enquête RICA sont disponibles à l'adresse : <http://www.agreste.agriculture.gouv.fr/enquetes/reseau-d-information-comptable-610/reseau-d-information-comptable>.

- d'autre part, selon celle de la régression quantile dont la solution $\hat{\mu}_q(i) = \sum_{j=1}^p \hat{\alpha}_j^{(q)} Y_i^j$ s'exprime en termes de quantiles conditionnels d'ordre q , afin de prendre en compte l'hétérogénéité intrinsèque de la distribution des frais spécifiques comme le montre la section suivante sur la méthodologie d'estimation quantile.

II.2 L'intérêt des quantiles conditionnels dans l'estimation des coûts de production agricole

La spécification standard du modèle des moindres carrés soulève certains problèmes qu'il serait hasardeux de négliger dans la perspective de l'établissement de référentiels sur les coûts de production, compte tenu des enjeux de compétitivité pour les différentes filières. Tout d'abord, dans un contexte d'utilisation mobilisant le RICA européen comme base empirique d'estimation des coûts spécifiques de production, les hypothèses stochastiques du modèle de Gauss-Markov peuvent ne pas être respectées : ainsi, l'asymétrie des distributions de frais spécifiques (concentration pour les valeurs plus faibles et dispersion des valeurs plus élevées que la moyenne, ou vice et versa) conduit au rejet de l'hypothèse de normalité des erreurs. En outre, compte tenu du mode de sélection spécifique à chacun des RICA nationaux (par exemple, le RICA français est une enquête administrée selon la méthode des quotas), les données comptables ne sont pas toujours recueillies selon un plan de sondage aléatoire stratifié permettant de délivrer selon un raisonnement inférentiel une estimation par intervalle basée sur une distribution paramétrique, même en se situant dans le cas asymptotique. L'estimation conditionnelle des quantiles a été développée dans (Koenker et Bassett, 1978) sous la dénomination de « *quantile regression* » afin de prendre en compte l'hétérogénéité de l'ensemble des valeurs d'une variable endogène x dans le contexte d'un modèle linéaire¹⁰. Lorsqu'on s'intéresse aux exploitations agricoles, cette méthode économétrique permet d'obtenir une distribution estimée des coûts spécifiques pour les principaux produits agricoles et ainsi complète les estimations obtenues par MCO, qui ne fournissent qu'une valeur moyenne (en termes d'espérance) de ces mêmes coûts. Au lieu d'une estimation par intervalle bâtie sur une hypothèse de normalité, le processus quantile fournit une distribution empirique des estimations sans avoir à formuler d'hypothèses sur la nature de cette distribution ni à suivre un plan de sondage aléatoire stratifié. Pour une variable aléatoire (va) continue x , le $q^{ième}$ quantile de la population est la valeur μ_q telle que x soit inférieure ou égale à μ_q avec la probabilité q : $q = \Pr[x \leq \mu_q] = F_x(\mu_q)$ (5)

où F_x est la fonction de répartition (FDR) de x donnant la probabilité cumulée d'une valeur sous la loi de x . Le $q^{ième}$ quantile se définit alors comme l'image de la valeur q par la fonction réciproque de la FDR : $\mu_q(x) = F_x^{-1}(q)$ (6)

En régression quantile, le $q^{ième}$ quantile conditionnel du coût de production x conditionné par l'ensemble des variables exogènes Y déterminant la consommation d'intrants est la fonction indicée $\mu_q(x|Y)$ telle que la variable aléatoire « x sachant Y » ($x|Y$) soit inférieure ou égale à $\mu_q(x|Y)$ avec la probabilité q . Ainsi, on peut définir formellement le $q^{ième}$ quantile conditionnel par l'expression suivante : $\mu_q(x/Y) = F_{x/Y}^{-1}(q)$ (7)

où $F_{x|Y}$ est la FDR de la loi de x conditionnée par Y .

La méthodologie d'estimation des quantiles conditionnels est celle présentée par (Desbois, Butault et Surry, 2017). Par analogie avec la régression MCO, la régression quantile peut être formulée comme suit (Cameron et Trivedi, 2006) :

$$x = Y'\beta^{(q)} + u_q \quad \text{avec} \quad u_q = Y'\alpha \times \epsilon_q \quad \text{et} \quad Y'\alpha > 0 \quad (8)$$

où $\beta^{(q)} = (\beta_1^{(q)}, \dots, \beta_j^{(q)}, \dots, \beta_p^{(q)})$ est le vecteur des coefficients des p variables explicatives, et ϵ_q un vecteur d'aléas identiquement et indépendamment distribués d'espérance nulle et de variance

¹⁰ Pour une présentation détaillée et concise de la régression quantile, voir (Koenker, 2005).

constante σ^2 . Sous cette hypothèse, le $q^{\text{ième}}$ quantile conditionnel du coût de production x , conditionné par Y et par les paramètres β et α , s'en déduit avec la forme analytique suivante :

$$\mu_q(x|Y, \beta, \alpha) = Y'[\beta + \alpha \times F_\epsilon^{-1}(q)] \quad (9)$$

où $F_\epsilon(q)$ est la fonction de répartition de ϵ_q . Ainsi, le quantile conditionnel d'ordre q dépendant linéairement de Y , l'estimation des paramètres converge vers $\beta + \alpha \times F_\epsilon^{-1}(q)$ et varie de façon monotone en fonction de l'ordre q du quantile. Si la variance σ^2 est constante, les quantiles conditionnels présentent des coefficients identiques ; sinon, leurs coefficients diffèrent d'un écart qui croît au fur et à mesure que l'ordre q du quantile augmente.

Suivant le modèle d'estimation des quantiles conditionnels pondérés par Koenker et Zhao (1994), la pondération ω des observations, définie par $\{\omega_i; i = 1, \dots, n\}$, est introduite dans la fonction de perte du problème de minimisation de la régression quantile comme suit :

$$\sum_{i; x_i \geq \beta y_i} [\omega_i q |x_i - \beta y_i|] + \sum_{i; x_i < \beta y_i} [\omega_i (1 - q) |x_i - \beta y_i|] \quad (10),$$

conduisant à l'estimation des paramètres du modèle (8) comme solution optimale du problème de minimisation de la fonction de perte (7) :

$$\hat{\beta}_\omega(q) = \underset{\beta \in \mathbb{R}^p}{\text{arg min}} \{ \sum_{i; x_i \geq y_i' \beta} [\omega_i q |x_i - y_i' \beta|] + \sum_{i; x_i < y_i' \beta} [\omega_i (1 - q) |x_i - y_i' \beta|] \} \quad (11).$$

Dans le cadre général présenté par D'Haultfoeuille et Givord (2014), plusieurs modèles doivent être distingués :

- i) $x = Y'\beta + u$ avec $u = K\epsilon$ aux résidus homoscédastiques ($V(\epsilon|Y) = \sigma^2$) désigné comme le modèle linéaire de quantile conditionnel à pente homogène (« *location shift model* »). Le cas où $Y'\alpha = K$ est constant, correspond à des quantiles conditionnels ne différant que par une constante $\mu_q(x|Y, \beta, \alpha) = Y'\beta + KF_\epsilon^{-1}(q)$, présentant tous la même pente et croissant de façon uniforme au fur et à mesure que l'ordre q du quantile augmente ;
- ii) $X = Y'\beta + (Y'\alpha)\epsilon$ avec $Y'\alpha > 0$ aux résidus hétéroscédastiques, désigné comme le modèle linéaire de quantile conditionnel à pente hétérogène (« *location-scale shift model* »). Le cas où $Y'\alpha > 0$ correspond à des pentes hétérogènes et croissantes $\mu_q(X|Y, \beta, \alpha) = Y'(\beta + \alpha\mu_q(\epsilon))$, impliquant des effets linéaire fixes $\gamma_q = \beta + \alpha\mu_q(\epsilon)$ relativement plus faibles pour les premiers quantiles et relativement plus élevés pour les derniers quantiles
- iii) $X = Y'\gamma_U$ avec U variable aléatoire indépendant de Y suivant une distribution uniforme sur l'intervalle $[0,1]$ telle que la fonction $u \rightarrow y'\gamma_u$ soit strictement croissante quel que soit y , désigné comme le modèle de quantile conditionnel à effet aléatoire (« *random coefficients model* »). U correspond à une composante aléatoire inobservée déterminant le rang de l'individu au sein de la distribution de X . Sous l'hypothèse distributionnelle d'invariance des rangs, hypothèse considérée comme forte dans la littérature scientifique, le coefficient aléatoire γ_q représenterait l'effet d'une modification marginale de Y pour les exploitations agricoles situées au q^e quantile de la distribution U , fonction des caractéristiques inobservées. Par exemple, cette hypothèse distributionnelle d'invariance des rangs revient à supposer, que les exploitations agricoles médianes ($q = 0,5$) en termes de productivité des intrants conserveraient ce rang, indépendamment des différents niveaux de production y_i enregistrés pour l'exploitation i .

II.3 Les procédures d'estimation et de test des quantiles conditionnels

L'estimateur des moindres carrés ordinaires (MCO) peut s'écrire comme solution d'un problème d'optimisation minimisant la somme des écarts quadratiques (en norme L_2) :

$$\hat{\beta}_{MCO} = \underset{\beta \in \mathbb{R}^p}{\text{arg min}} \left\{ \sum_i (x_i - y_i' \beta)^2 \right\} = \underset{\beta \in \mathbb{R}^p}{\text{arg min}} \{ \mathbf{e}' \delta^2 (X - Y' \beta) \} \quad (12)$$

où \mathbf{e} , est le vecteur directeur de la droite des constantes dans \mathbb{R}^n , espace des observations, et δ^2 , le vecteur des écarts quadratiques. De façon similaire, la régression quantile est définie pour chaque

quantile d'ordre q comme la solution d'un problème de minimisation de la somme des écarts en valeur

$$\text{absolue (en norme } L1) : \hat{\beta}(q) = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i \in \{i / x_i \geq y'_i \beta\}} q |x_i - y'_i \beta| + \sum_{i \in \{i / x_i \leq y'_i \beta\}} (1-q) |x_i - y'_i \beta| \right\} \quad (13)$$

pouvant s'écrire sous forme matricielle

$$\hat{\beta}(q) = \arg \min_{\beta \in \mathbb{R}^p} \left\{ q \mathbf{e}'(X - Y' \beta \geq 0) \delta^1 [\mathbf{X} - \mathbf{Y}' \beta] + (1-q) \mathbf{e}'(Y' \beta - X \geq 0) \delta^1 [\mathbf{Y}' \beta - \mathbf{X}] \right\} \quad (14)$$

avec $\mathbf{e}'(X - Y' \beta \geq 0)$, vecteur des valeurs indicatrices des observations i telles que $x_i - y'_i \beta \geq 0$, et δ^1 , le vecteur des écarts absolus.

Soit $x_i = y'_i \beta_q + e_i$ avec $e_i = u_i - v_i$, $u_i = e_i \mathbb{I}(e_i > 0)$, $v_i = |e_i| \mathbb{I}(e_i < 0)$, alors à l'instar de la régression $L1$ (Barrodale et Roberts, 1973), la régression quantile peut être formulée comme un problème primal d'optimisation linéaire, qui s'exprime sous forme matricielle comme suit :

$$\hat{\beta}(q) = \arg \min_{\beta \in \mathbb{R}^p, (u, v) \in \mathbb{R}^{n \times n}} \left\{ q \mathbf{1}' \mathbf{u} + (1-q) \mathbf{1}' \mathbf{v} \right\} \quad \text{sous la contrainte} \quad \mathbf{X} = \mathbf{Y}' \beta + \mathbf{u} - \mathbf{v} \quad (15)$$

Ce programme peut être reformulé en un problème dual d'optimisation équivalent :

$$\underset{z}{\text{Max}} \left\{ \mathbf{x}' \mathbf{z} \right\} \quad \text{sous la contrainte} \quad \mathbf{Y} \mathbf{z} = (1-q) \mathbf{Y} \mathbf{1} \quad \text{pour} \quad \mathbf{z} \in [0, 1]^n \quad (16)$$

Ainsi, les méthodes de résolution de problèmes d'optimisation linéaire développées pour la régression $L1$ (selon la médiane) s'étendent aisément à la régression quantile (Koenker et d'Orey, 1994). La méthode du simplexe (Dantzig, 1949) ayant une complexité algorithmique en $O(n^6)$, la méthode du « point intérieur » (Karmarkar, 1984) de complexité algorithmique $O(n^{3,5})$ lui est préférable en pratique dès que la taille de l'échantillon est importante. Pour les échantillons de grande taille, Portnoy et Koenker (1997) ont montré qu'une combinaison de l'algorithme du « point intérieur »¹¹ et d'un algorithme de lissage pour la fonction objectif (Madsen et Nielsen, 1993) rend les calculs de la régression quantile compétitifs avec ceux de la régression des moindres carrés.

Les quantiles conditionnels pondérés ont été proposés comme L -estimations¹² dans des modèles linéaires hétéroscédastiques par (Koenker et Zhao, 1994) définie par $\{\omega_i; i = 1, \dots, n\}$, la pondération ω des observations conduit à un schéma de régression quantile résolvant le problème de minimisation

$$\text{suivant : } \hat{\beta}_\omega(q) = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i \in \{i / x_i \geq y'_i \beta\}} \omega_i q |x_i - y'_i \beta| + \sum_{i \in \{i / x_i \leq y'_i \beta\}} \omega_i (1-q) |x_i - y'_i \beta| \right\} \quad (17).$$

La procédure d'estimation pondérée utilise l'implantation « prédicteur-correcteur » de l'algorithme primal-dual, proposée par Lustig *et al.* (1992).

Soit les conditions de régularité suivantes :

- i) Les distributions $F_i(x)$ des dépenses d'intrants pour un mix-produit donné y_i sont absolument continues avec des densités $f_i(x)$ continues et uniformément bornées sur $]0, +\infty[$ en $\xi_i = \mu_q(x|y_i)$;
- ii) $\Sigma_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i y'_i$ existe et est définie positive ;
- iii) $\Sigma_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_i(\xi_i) y_i y'_i$ existe et est définie positive ;
- iv) $\text{Sup}_{i=1, \dots, n} \|y_i\| \sim O(\sqrt{n})$, facteur de normalisation du mix-produit ;

¹¹ La pondération ω est introduite par l'instruction standard *weight* dans la procédure *QUANTREG* du logiciel *SAS 9.2*.

¹² Estimation robuste, telle que la médiane ou les moyennes tronquées, définie par une combinaison linéaire de statistiques ordinales.

Pollard (1991) montre que, sous les conditions i) et ii), $\hat{\beta}_q \xrightarrow{p} \beta_q$, l'estimateur converge en probabilité ; en outre, sous l'ensemble des conditions i), ii), iii) et iv), on obtient la normalité asymptotique, soit :

$$\sqrt{n}(\hat{\beta}_q - \beta_q) \xrightarrow{loi} N(0, q(1-q)\Sigma_1^{-1}\Sigma_0\Sigma_1) \tag{18}$$

Enfin, sous l'hypothèse d'égalité et d'indépendance des distributions $f_i(\xi_i) = f_\varepsilon(0)$, ce résultat se simplifie en :

$$\sqrt{n}(\hat{\beta}_q - \beta_q) \xrightarrow{loi} N(0, \sigma^2(q)\Sigma_0^{-1}) \text{ avec } \sigma(q) = \frac{\sqrt{q(1-q)}}{f_\varepsilon(0)} \tag{19}$$

En outre, sous les conditions iii) et iv), si les aléas attachés à la i^e observation $\varepsilon_i = x_i - y_i'\beta$ de distributions F_i identiques et indépendantes admettant une densité $f = F'$ telle que $f(F^{-1}(q)) > 0$ au voisinage de q , alors (Koenker et Bassett, 1982) :

$$\sqrt{n}(\hat{\beta}_q - \beta_q) \xrightarrow{loi} N(0, \omega^2(q, F)\Omega^{-1}) \text{ avec } \omega(q, F) = \frac{\sqrt{q(1-q)}}{f(F^{-1}(q))} \text{ et } \Omega = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i y_i' \tag{20}$$

Ces résultats peuvent être utilisés pour construire des intervalles de confiance pour les estimations selon trois procédures : fonction inverse de la densité, méthode des rangs ou algorithme de ré-échantillonnage. L'estimation selon la fonction inverse de la densité est la plus directe et la plus rapide des méthodes mais elle s'avère sensible à l'hypothèse de données identiquement et indépendamment distribuées (iid). Pour des données qui ne seraient pas iid, la méthode des rangs, qui calcule les intervalles de confiance en inversant le test des scores de rang, est préférable. Cependant, basée sur la méthode du simplexe, la méthode des rangs génère des temps de calcul importants pour les grands ensembles de données. La méthode du ré-échantillonnage, basée sur la technique du *bootstrap*, permet de s'affranchir de l'ensemble des hypothèses mais se révèle instable pour de petits échantillons. Compte tenu de la taille de l'échantillon RICA, de sa sélection non-aléatoire et de l'existence de trois sous-populations a priori distinctes (OTEXEs spécialisées ou non des exploitations laitières), nous avons opté pour la méthode du ré-échantillonnage, basée dans la procédure utilisée sur une technique du bootstrap marginal sur chaînes de Markov (*MCMB - Markov Chain Marginal Bootstrap*) car, sans hypothèse sur les distributions d'aléas, cette méthode donne des intervalles de confiance empiriques robustes en un temps de calcul raisonnable (He et Hu, 2002).

II.4 L'analyse symbolique des distributions de quantiles conditionnels de coûts spécifiques

II.4.1 ACP des distributions d'estimations quantiles conditionnelles

Les distributions quantiles conditionnelles comme objet symboliques

Soit un ensemble Δ de distributions quantiles conditionnelles $\{\delta_k, k = 1, \dots, K\}$, caractérisées par un ensemble de p quantiles conditionnels $\beta = \{\beta^1, \dots, \beta^j, \dots, \beta^p\}$. Estimée en ces p paramètres et localisée par un intervalle d'estimation ($[\underline{\beta}^j; \overline{\beta}^j]$, $\underline{\beta}^j$ étant le *minimum* et $\overline{\beta}^j$ le *maximum* de l'estimation du j^e quantile conditionnel), chacune de ces distributions d'intervalle d'estimation peut être considérée comme un objet symbolique.

Pour $p = 2$ paramètres, l'ensemble Δ de distributions δ_i est décrit par la matrice β des intervalles d'estimations quantiles conditionnelles, dont la k^e ligne correspondante au territoire k est égale à :

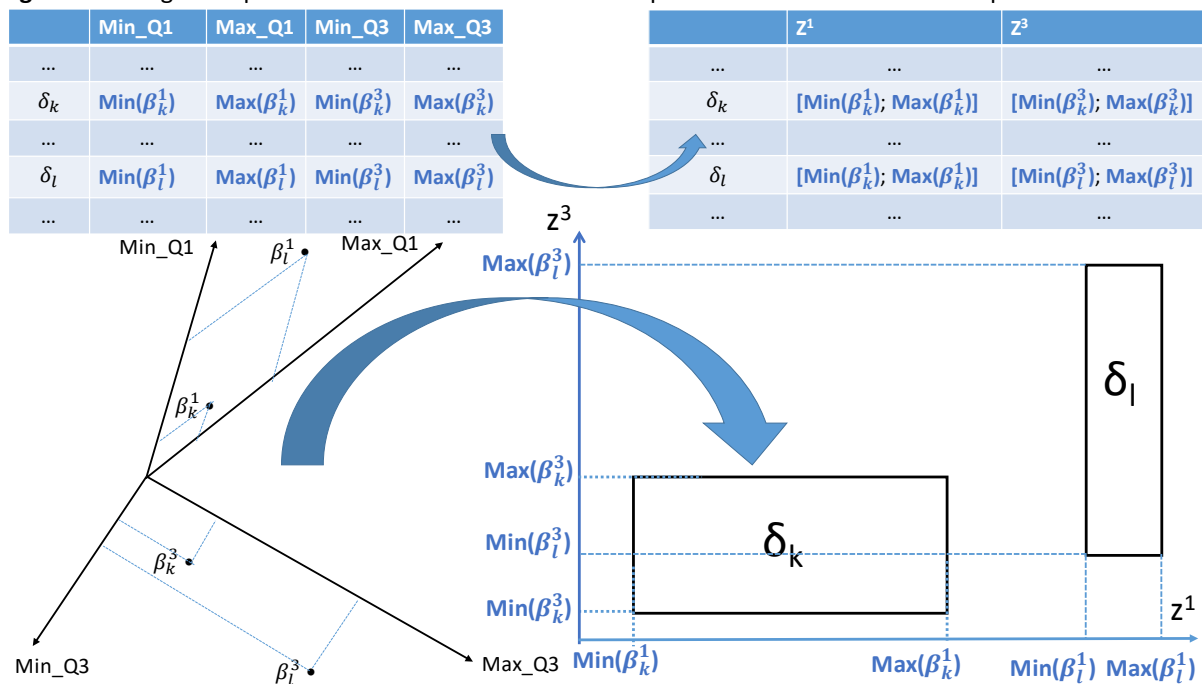
$$\beta_k = ([\underline{\beta}_k^1 ; \overline{\beta}_k^1] \quad [\underline{\beta}_k^2 ; \overline{\beta}_k^2]),$$

correspondant au vecteur Z_k à $p = 2$ composantes : $Z_k = \begin{pmatrix} z^1 & z^2 \\ \underline{\beta}_k^1 & \overline{\beta}_k^2 \\ \overline{\beta}_k^1 & \underline{\beta}_k^2 \\ \underline{\beta}_k^1 & \underline{\beta}_k^2 \\ \overline{\beta}_k^1 & \overline{\beta}_k^2 \end{pmatrix}$ qui peut être représenté

géométriquement par un hyper-pavé à $N = 2^p$ sommets dans l'espace des p composantes (cf.

figure 1, $p = 2; N = 4$). La matrice Z décrivant Δ est constituée en empilant les K vecteurs Z_k , représentant les distributions δ_k de quantiles conditionnels estimés.

Figure 1 : codage et représentation de distributions à deux quantiles conditionnels estimés par intervalle.



Source : d'après Lauro & Palumbo (2000) et Diday (2006).

L'ACP des extrema d'intervalles d'estimations quantiles conditionnelles

Dans l'extension de l'analyse en composantes principales (ACP) aux données d'intervalle proposée par (Cazes, Chouakria, Diday et Schektman, 1997), dénommée ACP des sommets (ACP-S), on effectue une ACP classique du tableau Z centré-réduit (ACP normée). De cette façon, les sommets des hyper-rectangles sont des vecteurs de \mathbb{R}^p , tandis que les estimations des quantiles conditionnels sont des éléments de \mathbb{R}^K . Ainsi, l'ACP-S fournit une représentation duale des distributions empiriques de coût spécifiques représentées par leurs intervalles d'estimations, qui sont les objets symboliques, et des quantiles conditionnels qui sont les descripteurs.

Comme en ACP classique, le sous-espace propre (optimal pour la représentation duale) est structuré par des axes orthonormés v_m ($1 \leq m \leq p$, maximisant la somme des carrés des coordonnées des sommets $\psi_m = Zv_m$ et satisfaisant dans \mathbb{R}^K à l'équation caractéristique aux vecteurs propres v_m et valeurs propres λ_m de la matrice $\frac{1}{N}Z'Z$:

$$\frac{1}{N}Z'Zv_m = \lambda_m v_m \tag{21}$$

L'analyse duale dans \mathbb{R}^p conduit à une équation similaire

$$\frac{1}{N}ZZ'w_m = \lambda_m w_m \tag{22}$$

ayant les mêmes valeurs propres non-nulles mais aux vecteurs propres w_m

tels que :

$$v_m = \lambda_m^{-1/2} Z'w_m \tag{23}$$

L'interprétation des axes de l'ACP-S s'effectue en fonction des quantiles conditionnels (variables de l'ACP-S) présentant les contributions les plus fortes. En ACP normée, les contributions à l'inertie des variables j à l'axe m se calculent comme le carré de la corrélation entre l'axe factoriel et la variable

(coordonnées factorielles) :

$$CTA_j^m = \frac{(\lambda_m^{1/2} v_j^m)^2}{\lambda_m} = (v_j^m)^2 \tag{24}$$

Les coordonnées des projections des extrémités des intervalles d'estimation des quantiles conditionnels (sommets $s(k)$) de la distribution empirique δ_k des coûts spécifiques (objet symbolique) sur les axes factoriels principaux sont fournies par la relation :

$$\psi_k^m = Z_k v_m \tag{25}$$

La représentation de la distribution empirique ω_i sur l'axe principal m est fournie par les projections des extrémités des intervalles quantiles estimés (hyper-rectangle d'imprécision maximale, HRIM). La

projection de l'HRIM sur un plan factoriel fournit un rectangle d'imprécision maximale (RIM) pour la distribution empirique représentée par l'objet symbolique δ_k .

Afin d'éviter les biais de projection sur-dimensionnant les rectangles projetés, Chouakria, Diday et Cazes (1998) ont proposé de retenir parmi les représentations celles dont les sommets sont les mieux représentés en utilisant comme critère la contribution relative (CTR) définie en termes de cosinus carré, soit pour une extrémité d'intervalle d'estimation $s(k)$ de la distribution quantile k :

$$CRT_{s(k)}^m = \sum_{j=1}^p (z_{s(k)}^m v_m)^2 / \sum_{j=1}^p z_{s(k)}^j{}^2 \quad (26)$$

Les problèmes de représentation des objets symboliques sont étudiés par Verde et De Angelis (1997) en termes de meilleurs ajustement d'enveloppes convexes. Afin de traiter de façon cohérente les extrema d'estimation (sommets) appartenant à la même distribution empirique de quantiles estimés (objet symbolique) et de représenter de façon optimale la taille et la forme de ces distributions, nous utiliserons l'ACP des objets symboliques, proposée par Lauro et Palumbo (2000).

L'ACP symbolique des intervalles d'estimation

Dans l'ACP-S basée sur les sommets, ceux-ci sont considérés comme des unités statistiques indépendantes. Afin de ne pas perdre d'information sur la taille et la forme des hyper-rectangles, Lauro et Palumbo (2000) introduisent une contrainte de cohésion entre les sommets. La méthode est basée sur la maximisation de la variance entre les objets symboliques. Il s'agit ici d'une application particulière de l'ACP par rapport à un espace de référence (D'Ambra et Lauro, 1982).

Soit A , la matrice booléenne indicatrice de l'appartenance des N extrema d'intervalle d'estimation aux K distributions empiriques. L'expression de la variance entre les objets symboliques est donnée par :

$$\frac{1}{N} Z' A (A' A)^{-1} A' Z \quad (27)$$

Soit, le projecteur orthogonal P_A associé à la matrice A sur le sous-espace de référence engendré par les vecteurs booléens codant l'appartenance des extrema aux différentes distributions :

$$P_A = A (A' A)^{-1} A' \quad (28)$$

Dans l'espace \mathbb{R}^N , les axes principaux d'inertie sont obtenus comme solution de l'équation suivante aux valeurs propres :

$$\frac{1}{N} Z' P_A Z \tilde{v}_m = \tilde{\lambda}_m \tilde{v}_m \quad (29)$$

où $\tilde{\lambda}_m$ et \tilde{v}_m sont respectivement les valeurs propres et les vecteurs propres associés à l'opérateur d'inertie $\frac{1}{N} Z' P_A Z$.

Les coordonnées de l'hyper-rectangle associé à la distribution empirique k sont alors calculées comme suit :

$$\tilde{\psi}_k^m = Z_k \tilde{v}_m \quad (30)$$

L'analyse dans \mathbb{R}^p revient à résoudre l'équation suivante aux valeurs propres :

$$(A' A)^{-1/2} (A' Z Z' A) (A' A)^{-1/2} \tilde{w}_m = \tilde{\lambda}_m \tilde{w}_m \quad (31)$$

où $\tilde{w}_m = (A' A)^{-1/2} A' Z \tilde{v}_m$ à la constante $\frac{1}{N}$ près.

Les contributions relatives (CRT) des paramètres j aux facteurs m sont définies de la même façon que pour l'ACP-S. Ces CRT sont également utilisées pour sélectionner les distributions empiriques à représenter sur les graphiques-plans factoriels :

$$CRT_k^m = \sum_{s \in \delta_k} \sum_{j=1}^p (z_s^j \tilde{v}_j^m)^2 / \sum_{s \in \delta_k} \sum_{j=1}^p (z_s^j)^2 \quad (32)$$

Comme dans l'ACP-S, les représentations des distributions empiriques en tant qu'objet symbolique sont construites par la méthode de visualisation des rectangles de surface couvrante maximale (*MCAR – Maximum Covering Area Rectangular*). Cependant, l'imposition de contraintes d'appartenance des extrema aux distributions permet de rendre leur représentation factorielle plus compacte.

Dans le cas de notre étude où le même nombre de quantiles conditionnels ont été estimés pour chaque distribution empirique, toutes les distributions empiriques ont le même nombre p de paramètres estimés alors elles ont le même nombre de sommets 2^p et $A' A = 2^p I_K$.

Ainsi, plutôt que maximiser la variance totale entre les sommets comme dans l'ACP-S, l'ACP des objets symboliques (ACP-OS) maximise la variance entre les objets symboliques que sont les distributions

d'intervalles d'estimation des quantiles conditionnels, représentée par la matrice $\frac{1}{N}Z'A(A'A)^{-1}A'Z$, où A est matrice booléenne de dimension $(N \times K)$ décrivant l'appartenance des N sommets aux K hyper-pavés représentant les distributions de quantiles estimés pour chacun des K territoires. Les solutions de cette maximisation sont données par l'équation caractéristique suivante

$$\frac{1}{N}Z'P_AZ\tilde{v}_m = \tilde{\lambda}_m\tilde{v}_m \quad (33)$$

où $P_A = A(A'A)^{-1}A'$ est la matrice du projecteur orthogonal sur le sous-espace engendré par les vecteurs de A , et \tilde{v}_m le vecteur propre associé à la valeur propre $\tilde{\lambda}_m$.

L'ACP partielle peut être également utilisée pour mieux souligner les différences entre objets symboliques. La section qui suit présente une ACP partielle où les sommets des hyper-rectangles sont tous rapportés à l'origine en soustrayant leur valeur minimum Min , pour permettre de mieux distinguer les caractéristiques de taille et de forme des différentes distributions

L'ACP partielle des étendues d'estimation

La transformation des données en étendues implique une translation affine où les minima $\{x_k^1, \dots, x_k^j, \dots, x_k^p\}$ sont tous rapportés à l'origine. Afin de mieux distinguer les tailles et les formes des hyper-rectangles associés aux descriptions, (Lauro et Palumbo, 2000) ont proposé une ACP partielle basée sur la translation à l'origine des données d'intervalle $e_k^j = \overline{x_k^j} - \underline{x_k^j}$ qui retient l'étendue comme information utile pour décrire la taille et la forme des objets symboliques. Le potentiel de description (*Description Potential – DP*, (Carvalho, 1997) est l'hyper-volume associé à la description de la k^e distribution empirique, domaine défini par le produit cartésien des étendues $E_k = e_k^1 \times \dots \times e_k^j \dots e_k^p$ des p paramètres associés à la description de l'objet symbolique δ_k . Sa mesure est définie par :

$$\pi(\delta_k) = \prod_{j=1}^p \tilde{e}_k^j \quad (34)$$

où $\tilde{e}_k^j = e_k^j/s_j$ est l'étendue normalisée par rapport au domaine $D^j = \{e_k^j; k \in K\}$ du descripteur Z^j . Cependant si la mesure de l'un des descripteurs tend vers zéro, alors le potentiel de description tend vers zéro. Pour pallier cet inconvénient, on utilise la mesure linéaire du potentiel de description (Carvalho, 1997), soit pour la distribution δ_k des intervalles estimés des quantiles conditionnels :

$$\sigma(\delta_k) = \sum_{j=1}^p \tilde{e}_k^j \quad (35)$$

Soit l'ensemble des instances des distributions empiriques de coûts spécifiques $\{\delta_1, \dots, \delta_k, \dots, \delta_K\}$ et X , la matrice de dimension $K \times p$ et de terme général $x_k^j = \sqrt{\tilde{e}_k^j}$, alors l'ACP en étendue des intervalles d'estimation est définie par la décomposition en facteurs du potentiel total de description linéaire $LDP = \sum_{k=1}^K \sigma(\delta_k)$, permettant une représentation géométrique différente de l'ACP-S des sommets.

Ainsi, la recherche d'un sous-espace de représentation optimale pour la taille et la forme de chaque objet symbolique est réalisée à partir d'une ACP non-centrée des maxima $\{\overline{x_k^1}, \dots, \overline{x_k^j}, \dots, \overline{x_k^p}\}$. Cette ACP des étendues (ACP-E) d'intervalle d'estimation décompose le critère

$$LDP = tr(X'X) = tr(XX') = \sum_{k=1}^K \sigma(\delta_k) \quad (36)$$

selon les équations caractéristiques suivantes : $X'Xt_m = \mu_m t_m$ (37)

et $XX'u_m = \mu_m u_m$ (38)

Ainsi, la somme des valeurs propres μ_m associées aux vecteurs propres t_m dans \mathbb{R}^K et u_m dans \mathbb{R}^p correspond à la décomposition factorielle du potentiel de description linéaire : $\sum_{m=1}^p \lambda_m = \sum_{k=1}^K \sigma(\delta_k)$. Les coordonnées factorielles de la représentation des distributions de coûts spécifiques dans le sous-espace optimal sont données par : $\phi_m = Xt_m$ (39)

La contribution absolue (CTA), rapport entre la coordonnée factorielle et la valeur propre, mesure la contribution de la distribution empirique des coûts spécifiques au potentiel de description de l'axe factoriel m ; elle est définie par : $CTA_k^m = \frac{(\phi_k^m)^2}{\mu_m}$ (40)

La contribution relative (CTR) mesure la qualité de représentation de la distribution empirique dans le sous-espace factoriel de représentation choisi :

$$CTR_k^m = \frac{\sum_{m=1}^{M^*} (\phi_k^m)^2}{\sum_{j=1}^p (x_k^j)^2} \quad (41)$$

L'interprétation des axes factoriels s'effectue selon les contributions (coordonnées factorielles) des quantiles conditionnels estimés pour les distributions empiriques de coûts spécifiques, en tant que descripteurs des objets symboliques : $CTA_j^m = (t_j^m)^2$ (42)

L'ACP des étendues peut être représentée par la projection des coordonnées factorielles des maxima. Les distributions décrites par les estimations de quantiles conditionnels, partagent des représentations en hyper-pavés similaires au plan de la taille et de la forme si elles sont projetées dans un même voisinage.

Si tous les termes de la matrice X sont positifs alors le premier vecteur propre u_1 et le facteur associé $\phi_1 = Xt_1$ sont positifs (Lauro et Palumbo, 2000). La première composante principale peut donc être interprétée comme un facteur de taille, tandis que les facteurs d'ordre supérieur ordonnent les distributions empiriques selon leurs caractéristiques de forme.

L'ACP des intervalles d'estimation en stratégie mixte

La stratégie mixte de l'analyse en composantes principales des objets symboliques (ACP-OS) combine l'ACP des sommets (ACP-S) et l'ACP des étendues (ACP-E) dans une approche en trois étapes pour prendre en compte les différences d'échelle et de forme entre les distributions empiriques de coûts spécifiques :

- i) ACP des étendues afin d'extraire les axes principaux qui restituent le mieux les échelles et les formes des distributions empiriques de quantiles conditionnels ;
- ii) Projection de Z en $\hat{Z} = P_A Z$ afin de prendre en compte les relations entre les différents *extrema*, compte-tenu des relations d'ordre entre les différents quantiles conditionnels de la distribution des coûts spécifiques ;
- iii) ACP des projections des lignes \hat{Z}_i sur le sous-espace de représentation optimale

$\Phi = \{\phi_1, \dots, \phi_m, \dots, \phi_{M^*}\}$ pour la représentation des échelles et des formes des distributions, avec l'opérateur de projection $P_\Phi = \Phi(\Phi'\Phi)^{-1}\Phi'$ (43)

La stratégie mixte d'analyse est donc basée sur la solution de l'équation caractéristique suivante :

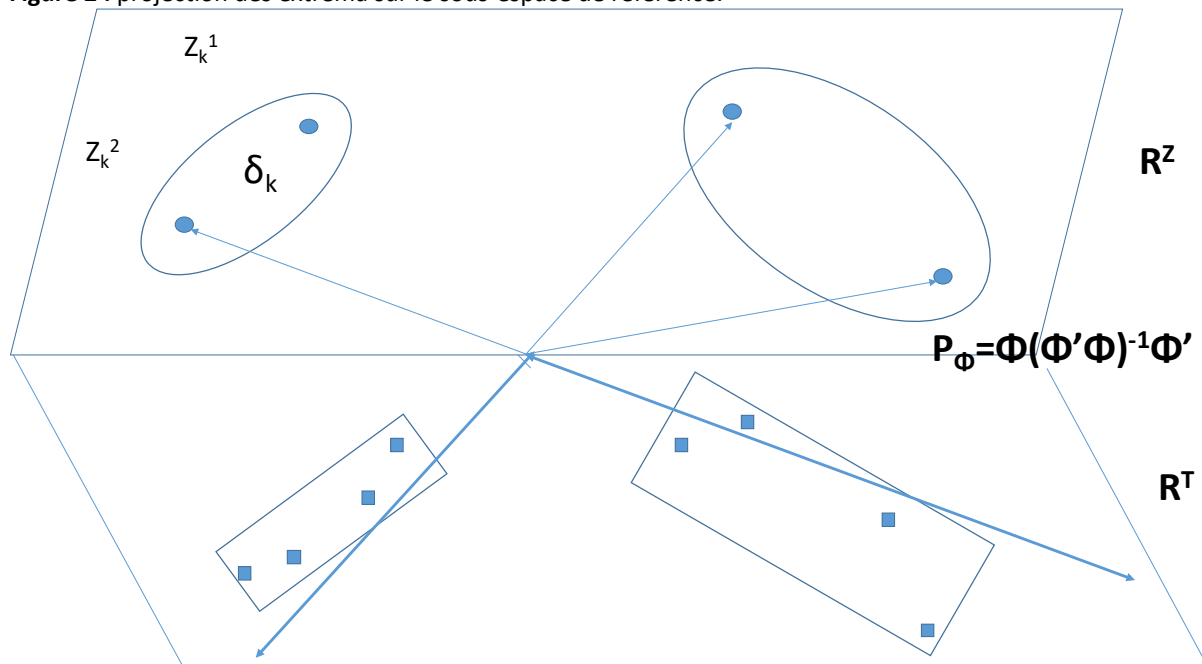
$$\hat{Z}' P_\Phi \hat{Z} = Z' A (A'A)^{-1/2} P_\Phi (A'A)^{-1/2} A' Z S_m = \rho_m S_m \quad (44)$$

où la matrice diagonale $(A'A)^{-1}$ est décomposée en un produit $(A'A)^{-1/2} (A'A)^{-1/2}$ pour des raisons de symétrie, avec respectivement ρ_m et S_m les valeurs propres et vecteurs propres associés répondant aux conditions d'orthonormalité usuelles.

L'interprétation des résultats de l'analyse dépend du choix de l'opérateur de projection P_Φ , dont le terme diagonal, interprétable comme un poids normalisé, est égal à :

$$\phi_k(\phi'_k\phi_k)^{-1}\phi'_k = \sum_{m=1}^{M^*} (\phi_k^m)^2 / \mu_m \tag{45}$$

Figure 2 : projection des extrema sur le sous-espace de référence.



Source : d'après Lauro & Palumbo (2000).

II.4.2 La classification automatique des distributions empirique de coûts spécifiques

Soit pour l'ensemble des distributions empiriques de coûts spécifiques $\Delta = \{\delta_1, \dots, \delta_k, \dots, \delta_n\}$ des pays k décrits en tant qu'objets symboliques par un ensemble de $p=6$ descripteurs¹³ qui sont les quantiles conditionnels $X = \{\tilde{Q}_{0,10}, \tilde{Q}_{0,25}, \tilde{Q}_{0,50}, Q_{0,75}, \tilde{Q}_{0,90}\} = \{x_1, \dots, x_j, \dots, x_p\}$. Les dissimilarités associées aux estimations par intervalles des quantiles conditionnels $\delta_k^j = [Min = \underline{x}_k^j; Max = \overline{x}_k^j]$ peuvent être calculées entre deux intervalles fermés du j^e quantile conditionnel $\delta_k^j = [\underline{x}_k^j; \overline{x}_k^j]$ et $\delta_l^j = [\underline{x}_l^j; \overline{x}_l^j]$ respectivement associés aux distributions caractérisant le pays k et le pays l , suivant les trois distances suivantes :

distance L_1 (somme des écarts absolus)¹⁴ $d_1(\delta_k^j, \delta_l^j) = |\underline{x}_k^j - \underline{x}_l^j| + |\overline{x}_k^j - \overline{x}_l^j|$ (46)

distance L_2 (somme des écarts quadratiques)¹⁵ $d_2(\delta_k^j, \delta_l^j) = \sqrt{(\underline{x}_k^j - \underline{x}_l^j)^2 + (\overline{x}_k^j - \overline{x}_l^j)^2}$ (47)

distance L_∞ (supremum des écarts)¹⁶ $d_\infty(\delta_k^j, \delta_l^j) = Sup\{(\underline{x}_k^j - \underline{x}_l^j); (\overline{x}_k^j - \overline{x}_l^j)\}$ (48)

¹³ Le choix d'un nombre réduit de descripteurs a été effectué pour des commodités de comparaison avec certaines approches graphiques plus classiques de (Desbois, Butault et Surry, 2013) et (Desbois, Butault et Surry, 2015) ; cependant, à l'instar de ces travaux antérieurs, il pourrait être étendu sans inconvénient à des ensembles de descripteurs de cardinalité $p=9$ (déciles), voire $p=99$ (centiles) si les objectifs d'analyse l'exigeaient.

¹⁴ Libellée « Type L1 » dans SCLUST.

¹⁵ Libellée « Euclidean » dans SCLUST.

¹⁶ Libellée « Hausdorff » dans SCLUST.

Pour chacune de ces distances métriques M sur \mathbb{R} , une dissimilarité entre distributions empiriques basées sur les écarts entre intervalles d'estimation des quantiles conditionnels peut être calculée selon

$$\text{un critère quadratique : } d(\delta_k, \delta_l) = \left(\sum_{j=1}^p d_M^2(\delta_k^j, \delta_l^j) \right)^{1/2} \quad (49)$$

La donnée d'une matrice de dissimilarités entre distributions empiriques nationales de coûts spécifiques permet d'appliquer directement les méthodes classiques de classification automatique basées sur des dissimilarités : méthodes de l'ultramétrique minimum (*single linkage*), de l'ultramétrique maximum (*complete linkage*), méthode du centroïde, et méthode de Ward.

Plusieurs procédures de classification automatique ont été développées pour les données à intervalle de valeurs. Parmi les premières procédures, Chavent (2000) propose une procédure de classification hiérarchique divisive (procédure DIV) sur données d'intervalle.

III Collecte des données et analyse distributionnelles des coûts spécifiques agricoles dans l'UE

III.1 Le RICA européen, le modèle, les agrégats comptables et les pays étudiés

Depuis son institution en 1965¹⁷, le RICA européen est défini par des règlements européens qui spécifient les modalités de mise en œuvre et leurs révisions dont la plus récente est le Règlement CE n°1217/2009 publié au JOE L328 du 15/12 2009 pour une entrée en vigueur le 01/04/2010. Avec le Recensement de l'Agriculture et les Enquêtes structures, il complète le tripode de l'acquis communautaire en matière de statistique agricole permettant de définir la population des exploitations agricoles, de suivre l'évolution de leurs structures productives, et enfin d'évaluer les variations de leur revenu. Orienté dès l'origine vers le suivi du revenu des agriculteurs dits « professionnels » et l'analyse du fonctionnement économique de leurs exploitations, il s'est progressivement imposé comme une base de données indispensable pour analyser ex ante et a posteriori l'impact des mesures de politique agricole, en particulier celles ayant trait aux réformes de la Politique agricole commune (PAC). Comme le souligne (Chantry, 1998 et 2003), le RICA européen est le fruit d'un processus d'adaptation et d'harmonisation de dispositifs nationaux préexistants au sein des Etats-membres¹⁸. Le RICA européen en tant que base de données s'alimente des RICA nationaux qui malgré l'harmonisation au niveau des concepts comptables et technico-économiques réalisée¹⁹ sous l'égide de la Direction Générale Agriculture (DG Agri), présentent un certain nombre de spécificités portant principalement sur la sélection de l'échantillon (plan d'échantillonnage, méthode de sélection, seuils de dimension économique) et la conduite de l'enquête. Pour chaque Etat-membre, les données de chaque exploitation (« fiche ») colligées au niveau européen (« fiche communautaire ») sont issues des données collectées au niveau national (« fiche nationale »). Pour certains Etats-membres comme la Belgique et les Pays-Bas, le questionnaire d'enquête du RICA national collecte plus de renseignements que prévus par le questionnaire virtuel du RICA européen. Inversement, pour d'autres Etats-membres, la « fiche européenne » incorpore comme données manquantes les possibilités d'exemption prévues par le règlement en raison de limitations ou de contraintes liés aux RICA nationaux. Cependant, en ce qui concerne les agrégats comptables mobilisés dans notre travail (produits bruts et charges spécifiques), les définitions sont harmonisées aussi bien en production végétales qu'animales ; les éléments de différenciation pouvant influencer sur les estimations via la pondération se situent surtout dans la population de référence des exploitations agricoles définies comme professionnelles (seuils de dimension économique) et dans la méthodologie d'échantillonnage (sélection aléatoire versus sélection sur quota).

¹⁷ Règlement n°79/65/CEE du 15 juin 1965.

¹⁸ Citons notamment les suivis des comptabilités agricoles organisés en Suède et en Finlande dès le début du XXIème siècle ainsi que le lancement du *Farm Business Survey* dès 1936 au Royaume-Uni.

¹⁹ Ce processus d'harmonisation est coordonné par l'unité L3 en charge des relations avec les opérateurs nationaux du RICA.

L'équation (8) constitue le modèle de base pour l'estimation des quantiles conditionnels des coûts directs spécifiques aux produits étudiés, dont le lait de vache notre produit d'intérêt. Variable endogène de la modélisation empirique notée X , les frais spécifiques²⁰ sont définis comme la somme :

- des intrants spécifiques aux cultures, soit les postes de dépense portant sur les semences et plants, les engrais et amendements, les produits phytosanitaires, et les autres coûts spécifiques aux cultures ;
- des intrants spécifiques au bétail, soit les postes de dépense portant sur les aliments pour herbivores, ceux pour granivores, et les autres coûts spécifiques aux animaux ;
- des frais spécifiques aux activités de foresterie.

Variables exogènes de la modélisation empirique, pour chacune des spéculations mises en œuvre par l'exploitation agricoles multi-produits, les produits bruts²¹ (notés Y) portent sur l'ensemble des productions tant végétales, qu'animales et produits animaux, voire forestières le cas échéant, avec la décomposition suivante en quinze agrégats : blé, autres céréales, cultures industrielles, protéagineux, oléagineux, productions horticoles, fruits, vins, autres productions végétales ou forestières, bovins, porcins, volailles, lait de vache, autres productions animales, autres produits bruts.

Les sous-populations d'exploitations agricoles choisies comme base d'estimation sont celles correspondant aux échantillons suivants du RICA européen : pour 2006, ont été sélectionnés les douze Etats-membres suivants : Allemagne, Autriche, Belgique, Danemark, Espagne, France, Hongrie, Italie, Pays-Bas, Pologne, Royaume-Uni et Suède, ensemble noté UE 12.

L'estimation quantile conditionnelle pondérée est réalisée au moyen du logiciel SAS, par la procédure QUANTREG associée à l'instruction WEIGHT, pour chacun des pays mais également pour chacune des classes de dimension.

III.2 Les caractéristiques distributionnelles des frais spécifiques agricoles

Selon (Angrist et Pischke, 2009), « Pour le meilleur ou pour le pire, 95 % des estimations en économétrie sont fournies par des moyennes » cependant « les économistes appliqués veulent de plus en plus savoir ce qui se passe, pas seulement en moyenne, mais pour l'ensemble de la distribution, les perdants comme les gagnants ». Ainsi, dans bien des études évaluatives comme prospectives, il est souvent utile de pouvoir comparer les résultats selon un nombre important de sous-populations, pour refléter l'hétérogénéité des populations étudiées et être en mesure de proposer des ajustements plus réalistes.

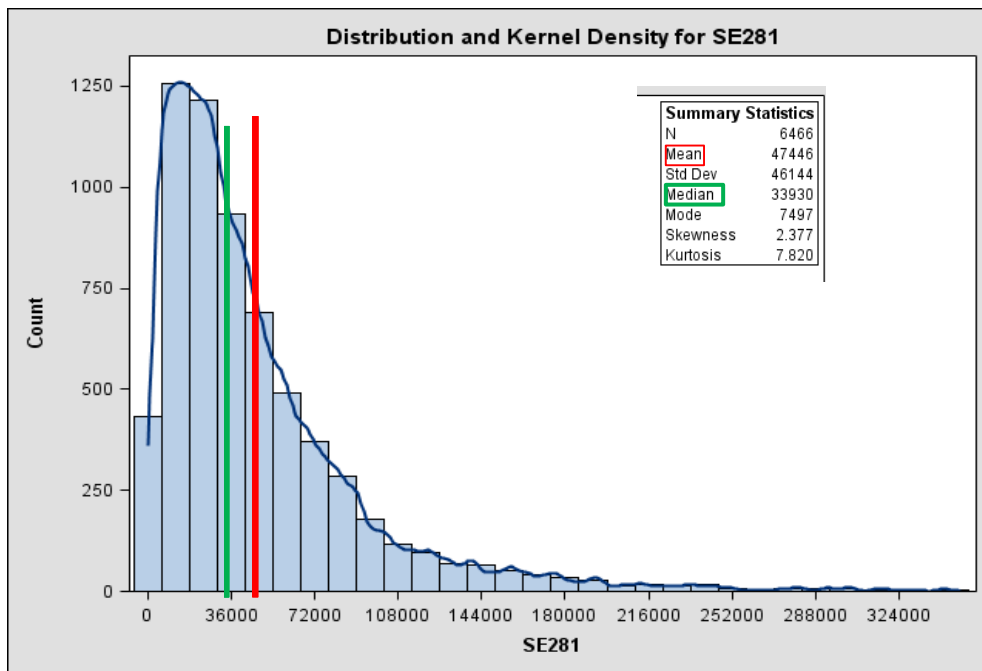
L'estimation non-paramétrique de la densité des coûts spécifiques par la méthode du noyau (Silverman, 1981) met en évidence l'asymétrie (2,377), signalée par le décalage entre la médiane (33 930 €) et la moyenne (47 446 €) excentrée par le poids des *extrema* par valeurs supérieures (cf. figure 3), représentant la distribution empirique du RICA français pour l'année civile 2006). Cette asymétrie, outre la dispersion des surfaces cultivées, révèle l'hétérogénéité sous-jacente dans la mobilisation des facteurs spécifiques de production.

Pour de telles distributions asymétriques, il est notoire que la médiane est un meilleur estimateur de la tendance centrale que la moyenne arithmétique. Cependant, bien souvent les distributions de coûts spécifiques ne sont pas unimodales, comme c'est le cas par exemple pour la distribution des frais spécifiques par unité de produit brut (cf. figure 4) que nous définissons comme les **coûts spécifiques**. D'autres valeurs peuvent alors être nécessaires pour mieux caractériser la forme de la distribution empirique des coûts spécifiques, comme celles des quartiles inférieur ($Q1=0,189$) et supérieur ($Q3=0,413$) apportant une information plus précise sur la distribution empirique des coûts spécifiques que celle donnée par l'unique estimation fournie par la moyenne.

²⁰ Les frais spécifiques sont enregistrés par le RICA européen sous le libellé de variable *SE281*.

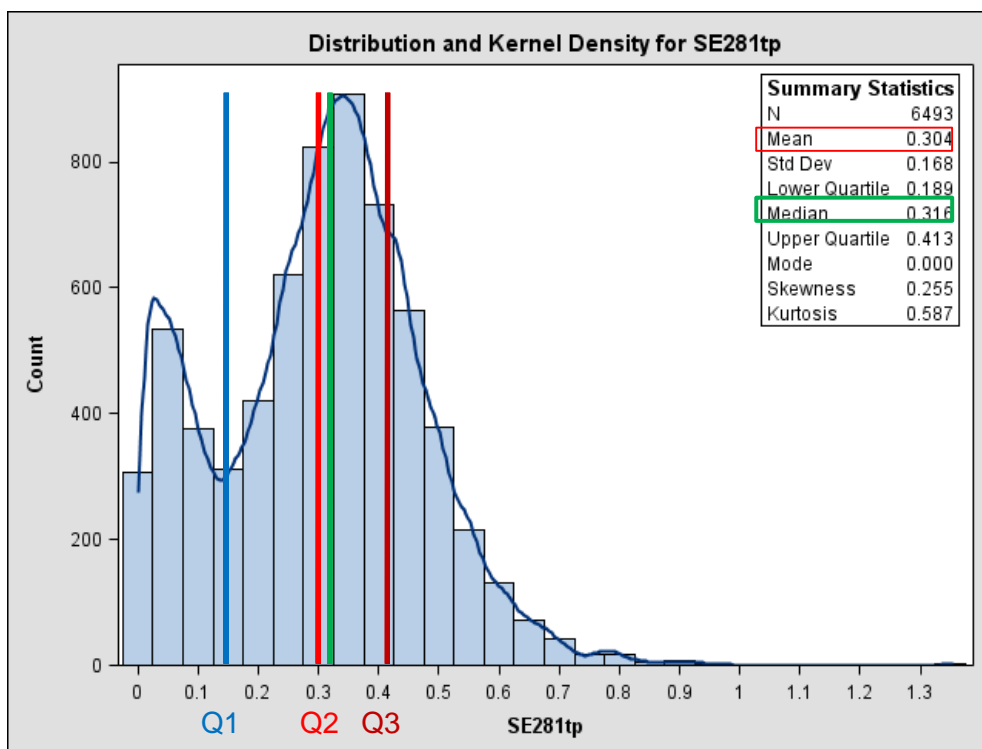
²¹ Le produit brut se définit, aux variations de stock près, comme la production brute totale de laquelle on soustrait le total des intra-consommations.

Figure 3 : frais spécifiques, distribution empirique du RICA français, 2006.



Source : traitement de l'auteur, d'après EU-RICA 2006.

Figure 4 : frais spécifiques par unité de produit brut, distribution empirique du RICA français, 2006.



Source : traitement de l'auteur, d'après EU-RICA 2006.

Tableau 1 : distributions nationales des frais spécifiques par exploitations, UE 12.

Pays	Echantillon	Moyenne	CV	Asymétrie	Aplatissement	D1	Q1	Médiane	Q3	D9	CQD*
Allemagne	6 750	63 420	261%	6,6	67,1	11 080	19 590	38 170	75 730	137 550	147%
Autriche	1 790	16 870	139%	6,5	86,7	3 500	5 840	10 430	19 700	37 350	133%
Belgique	1 040	74 150	134%	4,5	37,1	11 270	21 980	43 660	90 370	166 150	157%
Danemark	1 690	112 200	241%	3,4	23,1	4 670	10 810	34 620	155 180	314 640	417%
Espagne	7 740	9 520	585%	12,4	310,7	680	1 280	2 840	8 020	20 100	237%
France	6 510	39 310	160%	5,9	63,1	5 620	12 290	24 910	47 500	83 000	141%
Hongrie	1 690	14 850	1023%	7,5	85,7	1 070	1 880	4 350	10 240	25 460	192%
Italie	13 200	12 180	939%	14,4	314,5	700	1 320	2 670	7 200	20 860	220%
Pays-Bas	1 340	124 330	218%	3,4	17,1	9 870	25 350	56 100	138 300	294 040	201%
Pologne	11 000	7 010	383%	10,6	209,5	1 470	2 220	3 660	7 180	14 300	136%
Royaume-Uni	2 590	82 620	210%	7,9	97,8	14 300	23 090	44 220	93 150	177 050	158%
Suède	850	53 970	187%	8,5	111,6	7 300	15 840	28 850	67 030	122 760	177%
Ensemble	56 180	22 250	570%	9,0	135,9	1 010	2 010	5 050	18 070	51 490	318%

Source : traitement de l'auteur, d'après EU-RICA 2006.

Tout d'abord, si l'on compare les pays de l'ensemble UE12 en termes de tendance centrale, la médiane varie de 2 800 € (Espagne) à 10 500 € (Autriche) pour un premier groupe comprenant également la Pologne et l'Italie ; puis, on distingue un groupe intermédiaire avec la France et la Suède entre 25 000 € et 29 000 €, respectivement ; enfin le troisième groupe comprenant les autres pays avec des coûts médians par exploitation qui varient de 34 500 € pour le Danemark à 56 000 € pour les Pays-Bas, et comprenant également l'Allemagne, la Belgique et le Royaume-Uni.

En second lieu, si l'on compare ces pays de l'UE12 en termes de dispersion, le coefficient de variation (CV)²² se situe entre 134 % pour la Belgique, pays aux structures de production parmi les plus homogènes (avec l'Autriche, et la France et la Suède) et 1023 % pour la Hongrie qui apparaît avec l'Italie comme le pays le plus hétérogène.

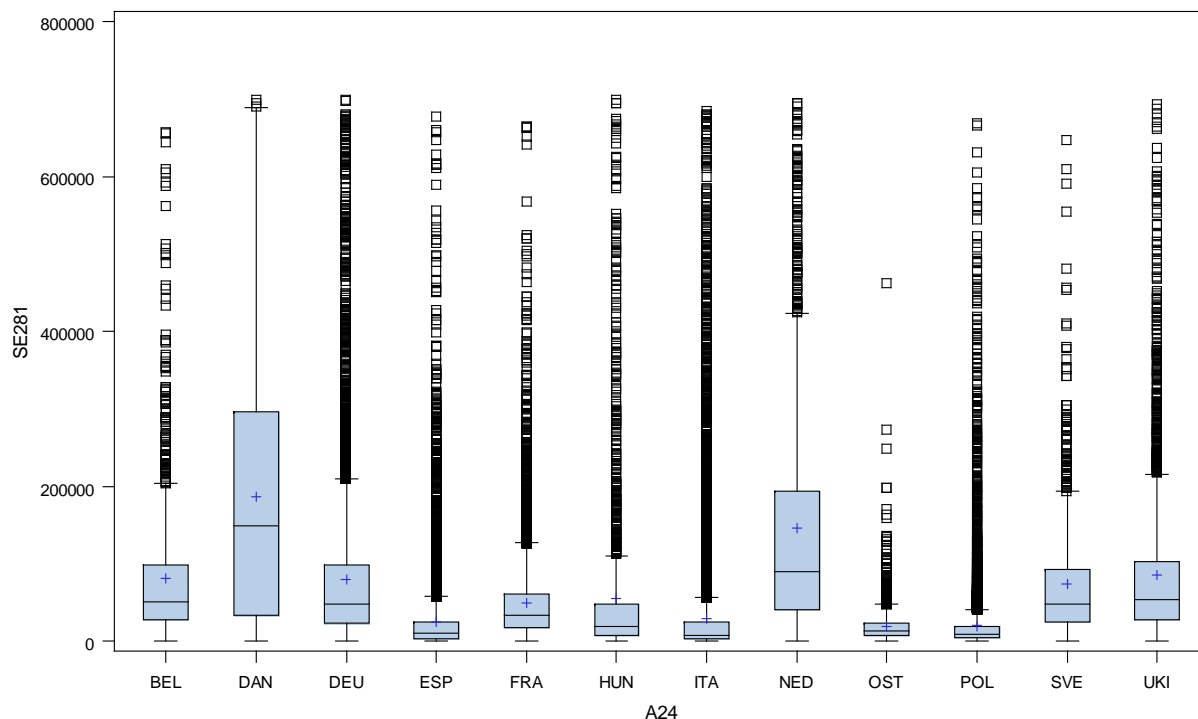
Bien que ces distributions soient toutes asymétriques par valeurs positives avec un nombre important de valeurs extrêmes, elle diffèrent cependant par leur forme : plus asymétriques par valeur positives (plus grande dispersion des valeurs supérieures à la médiane) comme en Italie, Espagne, Pologne ou Hongrie, d'un facteur d'asymétrie entre 5 et 6 ; ou moins asymétriques comme au Danemark, en Autriche, aux Pays-Bas, en Belgique ou en France (d'un facteur de 1,5 à 2,5), le Royaume-Uni présentant une asymétrie intermédiaire d'un facteur 4. En outre, le facteur d'aplatissement s'étend des minima, entre 17 pour les Pays-Bas, 23 pour le Danemark et 37 pour la Belgique, aux maxima entre 210 et 315 pour la Pologne, l'Espagne, ou l'Italie.

Pour des distributions très asymétriques et comportant des valeurs extrêmes, le coefficient quartile de dispersion (CQD)²³ est préférable au CV pour mesurer la dispersion relative : l'Autriche, la France et l'Allemagne (entre 130 et 150%) présentent la dispersion relative la plus faible, tandis que le Danemark possède la dispersion relative des coûts spécifiques la plus élevée (420%). En ignorant les valeurs extrêmes issues des exploitations héritières des structures collectives de plus en plus marginales dans sa production, le CQD ramène ainsi la dispersion relative de la Hongrie au niveau de celle de l'Italie.

Ainsi, les niveaux et les dispersions les plus faibles par exploitation se rencontrent dans des pays du Sud et de l'Est de l'Europe, avec des distributions plus asymétriques et aplaties, tandis que les niveaux et les dispersions les plus fortes sont observés dans des pays situés au Nord et à l'Ouest de l'Europe, avec des distributions moins asymétriques et plus concentrées que les précédentes. Les distributions de coûts spécifiques présentant de nombreuses valeurs extrêmes, il est plus pertinent d'utiliser les quartiles pour situer l'échelle des frais spécifiques, ainsi que l'écart interquartile ou le coefficient quartile de dispersion pour en mesurer la dispersion, plutôt que la moyenne, l'écart-type et le coefficient de variation sensibles aux poids de ces valeurs extrêmes.

²² Exprimé en pourcentage, le coefficient de variation rapporte la valeur de l'écart-type à la moyenne : $CV = \hat{\sigma}/\hat{\mu}$.

²³ Rapport de la dispersion interquartile au niveau médian, le coefficient quartile de dispersion $CQD = (Q_3 - Q_1)/Q_2$ fournit une mesure non-paramétrique de la dispersion relative.

Figure 5 : distribution des frais spécifiques de l'exploitation (SE 281 < 750 000 €) par pays, UE12.

Source : traitement de l'auteur, d'après EU-RICA 2006.

Le rapport entre les frais spécifiques et les produits bruts permet d'analyser la productivité des intrants et de la comparer à celle des autres facteurs de production. Dès lors, il est intéressant de pouvoir décrire par type de produit les différences structurelles du point de vue des coûts spécifiques entre pays où se situe la production : cet angle d'analyse est donc développé dans la suite de l'exposé.

IV Résultats économétriques : estimations nationales des coûts spécifiques

Ainsi que nous l'avons montré dans la section méthodologique, l'estimation selon les quantiles conditionnels permet de réaliser une imputation conditionnelle des coûts spécifiques par produits, autorisant la comparaison des différents ateliers dans le cadre d'une exploitation multi-produits sur la base de la marge brute, son complément au produit brut. Nous utilisons cette allocation conditionnelle pour fournir des estimations de coûts spécifiques permettant de répondre aux questions de mesure de compétitivité des exploitations agricoles, que posent la conception et l'évaluation ex ante ou ex post des différentes options de politique agricole. Dans le cadre du projet de recherche FACEPA, le choix retenu par les responsables en charge du programme *Knowledge Based Bio-Economy (KBBE)* du 7^{ème} PCRD s'est porté pour des raisons de faisabilité sur les commodités agricoles produites sur une échelle suffisamment large au niveau européen pour permettre des comparaisons entre pays, aux premiers rangs desquelles figure le lait de vache. Les estimations quantiles sont donc effectuées pour chacun des Etats-membres de l'ensemble UE12 afin de tester la différenciation nationale du cadre productif au plan européen. Nous avons choisi d'analyser les estimations obtenues sur l'année 2006²⁴, afin de pouvoir comparer les résultats de l'approche des quantiles conditionnels ultérieurement avec ceux de l'approche *Seemingly Unrelated Regressions Equations (SURE)*. Initialement proposée par (Zellner, 1962), cette dernière approche constitue la procédure standard d'estimation du modèle GECOM du projet FACEPA²⁵.

²⁴ L'analyse sur l'ensemble de la période fait l'objet de travaux en cours permettant d'adapter l'approche des estimations quantiles à une structuration en données de panel.

²⁵ Les résultats obtenus simultanément sur les autres productions font l'objet d'analyses en cours, menées parallèlement.

Ainsi, nous analysons par la suite les résultats obtenus en particulier pour le lait de vache, l'une des commodités retenues dans le cadre du projet FACEPA.

IV.1 Le lait de vache, analyse comparative des produits bruts entre douze pays de l'UE

En 2010, la part de l'UE27 dans la production mondiale de lait s'élève à 23%, devant les USA (15%) et l'Inde (8%)²⁶. Selon Eurostat, la production laitière européenne s'élevait en 2008 à 135 milliards de litres pour un cheptel de 24 millions de vaches laitières²⁷. Les pays étudiés figurent parmi les principaux pays producteurs en termes de quantités collectées, soit par ordre décroissant : l'Allemagne (20,4%), la France (17,7%), le Royaume-Uni (9,9%), les Pays-Bas (8,4%), l'Italie (7,8%), la Pologne (6,6%), l'Espagne (4,3%), le Danemark (3,4%), la Suède (2,2%), la Belgique (2,1%), et l'Autriche (2,0%), et la Hongrie (1,1%), soit 86% de la production européenne.

La hiérarchie des produits bruts observés au sein du RICA européen (tableau 2) apparaît conforme à la hiérarchie des statistiques nationales sur la production laitière²⁸, même si l'on constate certaines inversions de classement, entre l'Italie, les Pays-Bas et le Royaume-Uni, puis l'Autriche et la Suède.

Tableau 2 : lait de vache, distribution du produit brut par pays, UE12.

Pays	Population	D1	Q1	Médiane	Q3	D9	Moyenne	CQD*
Produits bruts (€)								
Allemagne	71 180	21 860	35 970	60 740	114 300	178 830	92 380	129%
Autriche	31 640	5 400	10 560	19 960	35 310	53 120	26 170	124%
Belgique	10 520	22 750	34 980	57 640	95 720	136 700	70 180	105%
Danemark	4 840	68 160	126 610	247 380	365 650	495 450	266 270	97%
Espagne	49 930	150	1 940	7 630	43 600	87 620	32 790	546%
France	80 760	31 190	47 180	70 490	102 500	140 510	80 110	78%
Hongrie	8 100	1 170	2 030	3 640	6 820	36 160	25 440	132%
Italie	55 480	2 210	6 830	18 050	61 920	178 610	69 530	305%
Pays-Bas	18 530	57 880	93 080	146 600	226 500	278 250	163 270	91%
Pologne	416 540	380	700	1 530	4 930	14 550	5 390	276%
Royaume-Uni	17 050	51 250	88 310	158 340	278 760	424 330	212 240	120%
Suède	6 640	21 270	43 170	86 520	137 810	230 680	108 930	109%
Ensemble	771 230	510	1 200	7 180	46 770	113 320	40 480	635%

Source : traitement de l'auteur, d'après EU-RICA 2006.

IV.2 l'analyse factorielle des distributions d'intervalles d'estimations

Le tableau 3 présente les principales estimations des quantiles conditionnels de coûts spécifiques et de marges brutes (D1 décile inférieur, Q1 quartile inférieur, Q2 médiane, Q3 quartile supérieur, D9 décile supérieur) pour le lait de vache, issues de la régression quantile et de la régression des moindres carrés ordinaires (MCO) pour les coûts spécifiques de production agricole (agrégat comptable SE281 du RICA européen) à partir d'une décomposition du produit brut en quinze agrégats (cf. supra III.1), pour le sous-ensemble de 12 pays européens retenus en 2006.

Nous pouvons vérifier immédiatement la complémentarité des estimations de coûts spécifiques et de marges brutes (cf. équation 4). Ainsi, pour l'Allemagne, le minimum de l'estimation de D1 (270,8 €) pour les coûts spécifiques est complémentaire à 1 000 € avec le maximum de l'estimation de D9 (729,2 €) pour les marges brutes ; de même, le maximum de l'estimation de D1 (283,2 €) pour les coûts spécifiques est complémentaire à 1 000 € avec le minimum de l'estimation de D9 (716,8 €) pour les marges brutes.

La visualisation des estimations des coûts spécifiques s'effectue selon le graphique de la figure 6, présentant les estimations de quantiles conditionnels par ordre croissant pour chacun des pays. Ce graphe permet d'identifier *de visu* trois types d'échelle distributionnelles : le premier, aux échelles de

²⁶ D'après *Focus on the Common Agricultural Policy*, Eurostat 2012.

²⁷ *EU Dairy Market 2008-2013*, ZMP 2009, d'après *European Dairy Association* (www.euromilk.org).

²⁸ Coefficient de corrélation $r = 0,98$.

coût les plus faibles et les moins dispersées (entre 200 € et 400 €), comprendrait l'Autriche (AUT), les Pays-Bas (NED), la France (FRA), l'Allemagne (DEU) et la Pologne (POL) ; le second, aux échelles de coût et de dispersion intermédiaires (entre 350 € et 600 €), serait composé de la Hongrie (HUN), du Royaume-Uni (UKI), le Danemark (DAN) et la Suède (SVE) ; enfin, le troisième aux échelles distributionnelles de coût les plus dispersées (entre 260 € et 780 €), comporterait l'Italie (ITA) et l'Espagne (ESP).

Tableau 3 : lait de vache, coûts spécifiques et marges brutes pour 1 000 € de produit brut, UE12.

Coûts spécifiques (€)						
Pays	D1 [min ; max]	Q1 [min ; max]	Q2 [min ; max]	Q3 [min ; max]	D9 [min ; max]	MCO [min ; max]
Allemagne	[270,8 ; 283,2]	[296,5 ; 306,7]	[331,7 ; 340,3]	[366,1 ; 375,5]	[396,1 ; 411,5]	[376,7 ; 384,1]
Autriche	[181,3 ; 208,1]	[226,7 ; 244,3]	[257,4 ; 277,0]	[298,3 ; 322,1]	[371,6 ; 406,6]	[300,8 ; 314,4]
Belgique	[214,2 ; 235,4]	[240,8 ; 278,6]	[288,4 ; 310,2]	[302,5 ; 333,9]	[326,4 ; 398,4]	[284,1 ; 311,6]
Danemark	[429,1 ; 445,9]	[457,9 ; 475,9]	[482,2 ; 497,2]	[503,7 ; 533,5]	[505,9 ; 557,7]	[489,7 ; 506,8]
Espagne	[254,5 ; 287,3]	[314,6 ; 346,4]	[420,7 ; 453,9]	[569,8 ; 598,0]	[643,4 ; 674,4]	[466,0 ; 484,6]
France	[210,3 ; 223,1]	[245,5 ; 256,7]	[282,9 ; 293,9]	[322,7 ; 336,5]	[371,8 ; 392,0]	[299,8 ; 309,1]
Hongrie	[353,6 ; 429,4]	[354,0 ; 433,0]	[466,0 ; 580,6]	[578,7 ; 699,1]	[712,7 ; 798,7]	[719,1 ; 747,3]
Italie	[244,4 ; 288,2]	[330,7 ; 361,3]	[443,4 ; 475,2]	[549,8 ; 575,2]	[589,1 ; 628,9]	[408,2 ; 418,6]
Pays-Bas	[207,8 ; 233,2]	[242,4 ; 261,4]	[277,8 ; 294,0]	[320,7 ; 333,9]	[350,1 ; 370,3]	[265,5 ; 309,9]
Pologne	[218,6 ; 230,8]	[261,3 ; 272,1]	[307,4 ; 317,4]	[357,6 ; 372,0]	[409,3 ; 428,3]	[327,7 ; 338,3]
Royaume-Uni	[363,0 ; 387,0]	[422,9 ; 443,5]	[470,7 ; 491,5]	[515,2 ; 531,4]	[571,9 ; 595,1]	[483,3 ; 495,9]
Suède	[368,6 ; 399,4]	[376,7 ; 398,5]	[399,3 ; 432,3]	[436,9 ; 461,5]	[447,5 ; 493,7]	[416,5 ; 438,2]
Marges brutes (€)						
Pays	D1 [min ; max]	Q1 [min ; max]	Q2 [min ; max]	Q3 [min ; max]	D9 [min ; max]	OLS [min ; max]
Allemagne	[588,5 ; 603,9]	[624,5 ; 633,9]	[659,7 ; 668,3]	[693,3 ; 703,5]	[716,8 ; 729,2]	[615,9 ; 623,3]
Autriche	[593,4 ; 628,4]	[677,9 ; 701,7]	[723,0 ; 742,6]	[755,7 ; 773,3]	[791,9 ; 818,7]	[685,6 ; 699,2]
Belgique	[601,6 ; 673,6]	[666,1 ; 697,5]	[689,8 ; 711,6]	[721,4 ; 759,2]	[764,6 ; 785,8]	[688,4 ; 715,9]
Danemark	[442,3 ; 494,1]	[466,5 ; 496,3]	[502,8 ; 517,8]	[524,1 ; 542,1]	[554,1 ; 570,9]	[493,2 ; 510,3]
Espagne	[325,6 ; 356,6]	[402,0 ; 430,2]	[546,1 ; 579,3]	[653,6 ; 685,4]	[712,7 ; 745,5]	[515,4 ; 534,0]
France	[608,0 ; 628,2]	[663,5 ; 677,3]	[706,1 ; 717,1]	[743,3 ; 754,5]	[776,9 ; 789,7]	[690,9 ; 700,2]
Hongrie	[201,3 ; 287,3]	[300,9 ; 421,3]	[419,4 ; 534,0]	[567,0 ; 646,0]	[570,6 ; 646,4]	[252,8 ; 280,9]
Italie	[371,1 ; 410,9]	[424,8 ; 450,2]	[524,8 ; 556,6]	[638,7 ; 669,3]	[711,8 ; 755,6]	[581,4 ; 591,8]
Pays-Bas	[629,7 ; 649,9]	[666,1 ; 679,3]	[709,0 ; 722,2]	[738,6 ; 757,6]	[766,8 ; 792,2]	[690,1 ; 734,5]
Pologne	[571,7 ; 590,7]	[628,0 ; 642,4]	[682,6 ; 692,6]	[727,9 ; 738,7]	[769,2 ; 781,4]	[661,7 ; 672,3]
Royaume-Uni	[404,9 ; 428,1]	[468,6 ; 484,8]	[508,5 ; 529,3]	[556,5 ; 577,1]	[613,0 ; 637,0]	[504,1 ; 516,7]
Suède	[506,3 ; 552,5]	[538,5 ; 563,1]	[567,7 ; 600,7]	[601,5 ; 623,3]	[600,6 ; 631,4]	[561,8 ; 538,5]

Source : traitement de l'auteur, d'après EU-RICA 2006.

Parmi les résultats auxquels peuvent être confrontés ces estimations, les parts du produit brut en coûts spécifiques estimées pour le lait de vache, cf. tableau A3.2 dans (Kleinhanss, Offerman, Butault et Surry, 2011), à partir du modèle standard GECOM de FACEPA (estimateur SURE), diffèrent sensiblement : sur onze Etats-membres de l'UE²⁹, tous les estimateurs quantiles conditionnels sont significativement corrélés à l'estimateur SURE du modèle GECOM de FACEPA (les niveaux de corrélation sont croissants du premier (D1) au neuvième (D9) décile conditionnel, ce dernier étant proche du niveau de corrélation atteint par l'estimateur MCO³⁰). En outre, l'ensemble des estimateurs quantiles conditionnels du premier décile conditionnel (D1) au neuvième décile conditionnel (D9) sont significativement corrélés³¹ avec la part de produit brut que représentent les coûts spécifiques (SHARE) des exploitations spécialisées en lait de vache (50% du produit brut, cf. (EU-DGAgri, 2013)), pour les douze pays de l'UE sélectionnés. Ces corrélations significatives démontrent la pertinence de l'approche d'estimation par les quantiles conditionnels pour estimer la part des coûts spécifiques dans le produit

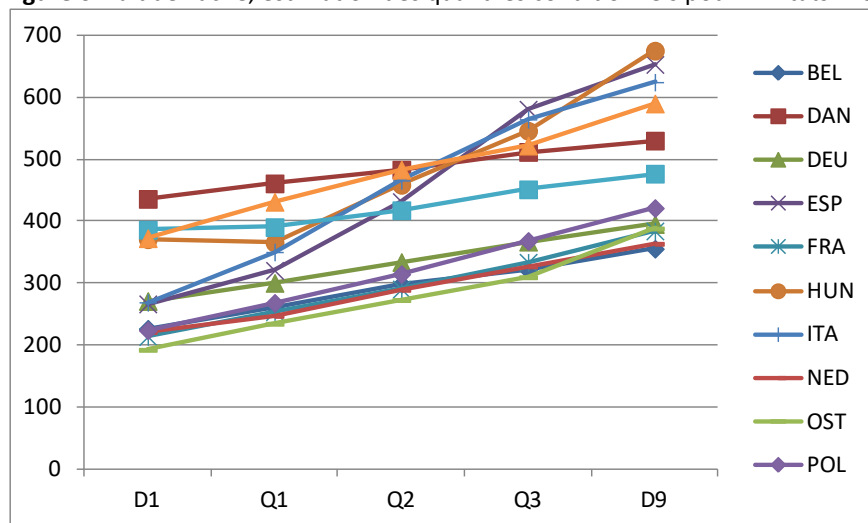
²⁹ L'Autriche n'ayant pas été incluse dans le rapport FACEPA.

³⁰ Les niveaux de corrélation des rangs de Spearman sont croissants de $corr(SURE, D1) = 0,62$ à $corr(SURE, D9) = 0,69$, comparables à $corr(SURE, MCO) = 0,72$.

³¹ Les niveaux de corrélation des rangs de Spearman vont croissants de $corr(SHARE, D1) = 0,74$ à $corr(SHARE, Q1) = corr(SHARE, Q2) = 0,87$ puis décroissants en $corr(SHARE, Q3) = corr(SHARE, D9) = 0,83$.

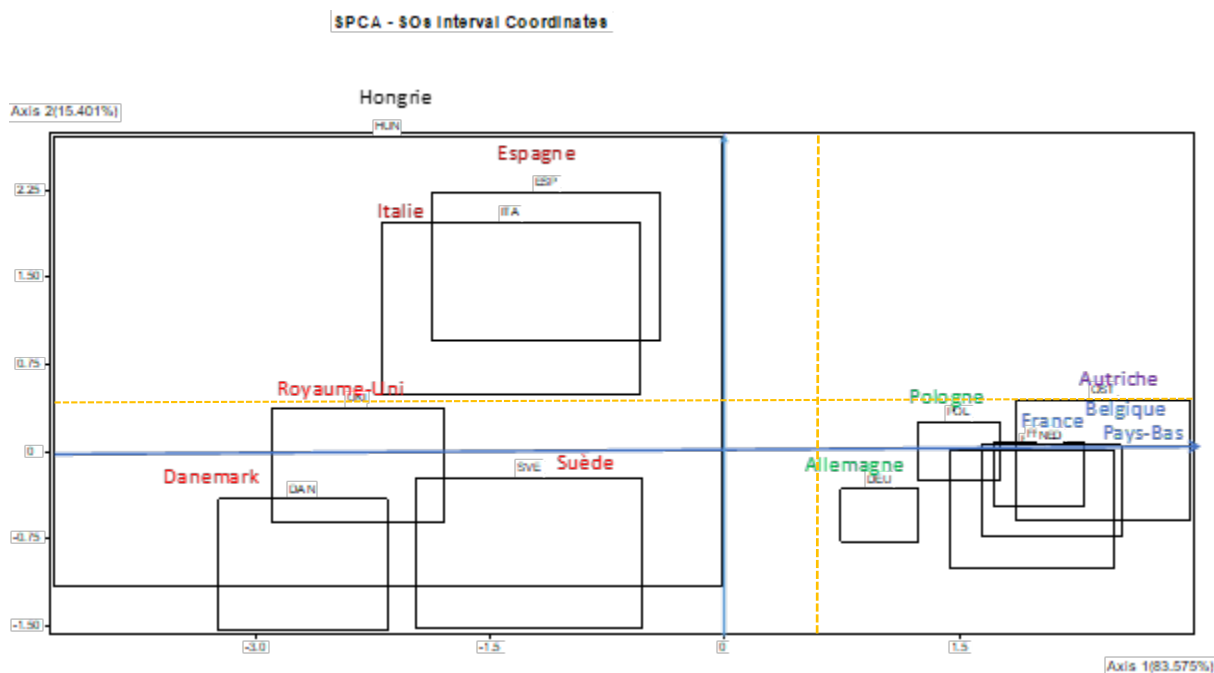
brut pour les exploitations spécialisées : les deux premiers quartiles conditionnels (Q1 et Q2) constituent les paramètres de coûts spécifiques susceptibles de caractériser au mieux ces exploitations spécialisées en lait de vache.

Figure 6 : lait de vache, estimation des quantiles conditionnels pour 12 Etats-Membres de l'UE (2006).



Source : traitement de l'auteur, d'après EU-RICA 2006.

Figure 7 : lait de vache, ACP des intervalles d'estimation quantile, plan factoriel F1xF2 des pays UE12 (2006).



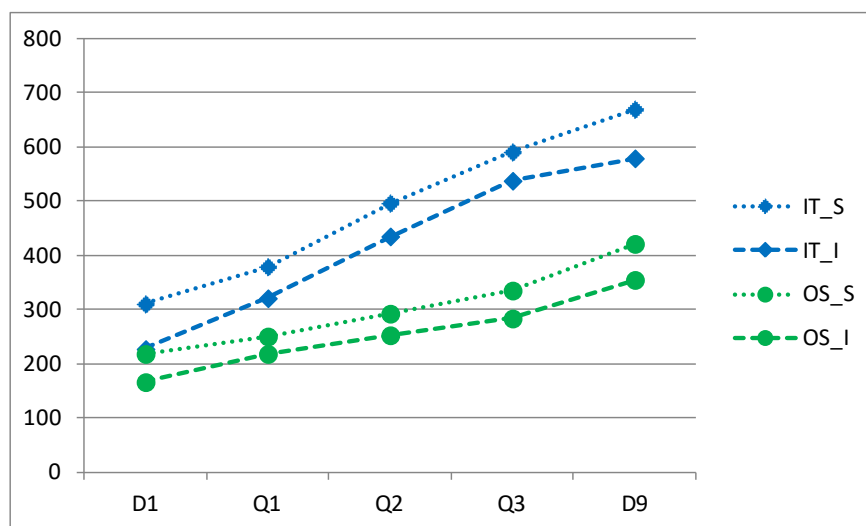
Source : traitement de l'auteur, d'après EU-RICA 2006.

L'ACP symbolique des estimations par intervalles des quantiles conditionnels (ACPSIQ) permet de préciser cette structure distributionnelle. La première composante principale F1 (Axis 1) représentant 84% de l'inertie, est corrélée à l'ensemble des quantiles conditionnels, la médiane conditionnelle (Q2) présentant le niveau de corrélation le plus élevé. La seconde composante principale F2 (Axis 2), représentant 15% de l'inertie, est surtout corrélée au troisième quartile (Q3) et au décile supérieur (D9). Le premier plan factoriel F1xF2 de l'ACPSIQ, représentant 99% de la variabilité des pays, permet d'identifier deux groupes de pays distincts se différenciant selon le niveau de l'estimation médiane conditionnelle (Q2) et les dimensions de leur échelle distributionnelle : d'une part, en F1>0, l'ensemble assez homogène {Allemagne, Autriche, Belgique, France, Pays-Bas, Pologne} à l'estimation médiane

comprise entre 273 € et 334 € et dont l'intervalle inter-décile [D1 ; D9] va de 193 € à 421 € ; d'autre part, l'ensemble {Hongrie, Espagne, Italie, Royaume-Uni, Suède, Danemark} plus hétérogène si l'on en juge par les dimensions des rectangles projetés selon les deux premières composantes (l'estimation médiane se situe entre 418 € et 484 €, mais l'intervalle inter-décile [D1 ; D9] va de 267 € à 676 €). Si l'on prend en compte la dispersion selon la seconde composante principale F2, alors cet ensemble peut être scindé en trois groupes : d'une part, l'Espagne et l'Italie, projetées positivement ($F2 > 0$) en raison de leur niveaux majeurs en Q3 et Q9 ; d'autre part, le Danemark, la Suède et le Royaume-Uni, projetés négativement ($F2 < 0.5$) ; enfin, la Hongrie dont la dispersion selon F2 recoupe les projections précédentes.

Parmi les différences qui peuvent être identifiées d'après le premier plan factoriel de l'ACPEIQ, signalons la séparation selon les deux premières composantes principales des rectangles projetés de l'Autriche (OST, quadrant $F1 > 0$) et de l'Italie (ITA, quadrant $F1 < 0$ et $F2 > 0$), séparables dans ces deux dimensions (séparations médianes en pointillés). La figure 8 confirme cette séparation des échelles distributionnelles pour l'ensemble des intervalles d'estimation des quantiles conditionnels. L'Autriche présente donc des coûts spécifiques de production significativement plus faibles que ceux de l'Italie pour l'ensemble des quantiles étudiés. Cependant, les coûts spécifiques des producteurs italiens sont plus hétérogènes que ceux des producteurs autrichiens : d'une part, l'écart inter-décile est de 196 € pour l'Autriche contre 356 € pour l'Italie ; d'autre part, les intervalles d'estimation des quartiles italiens sont disjoints tandis que, pour l'Autriche, les intervalles d'estimation de la médiane et du quartile supérieur se chevauchent ainsi que ceux du premier décile et du quartile inférieur.

Figure 8 : lait de vache, échelles distributionnelles de quantiles conditionnels de l'Autriche (OS) et de l'Italie (IT).



Source : traitement de l'auteur, d'après EU-RICA 2006.

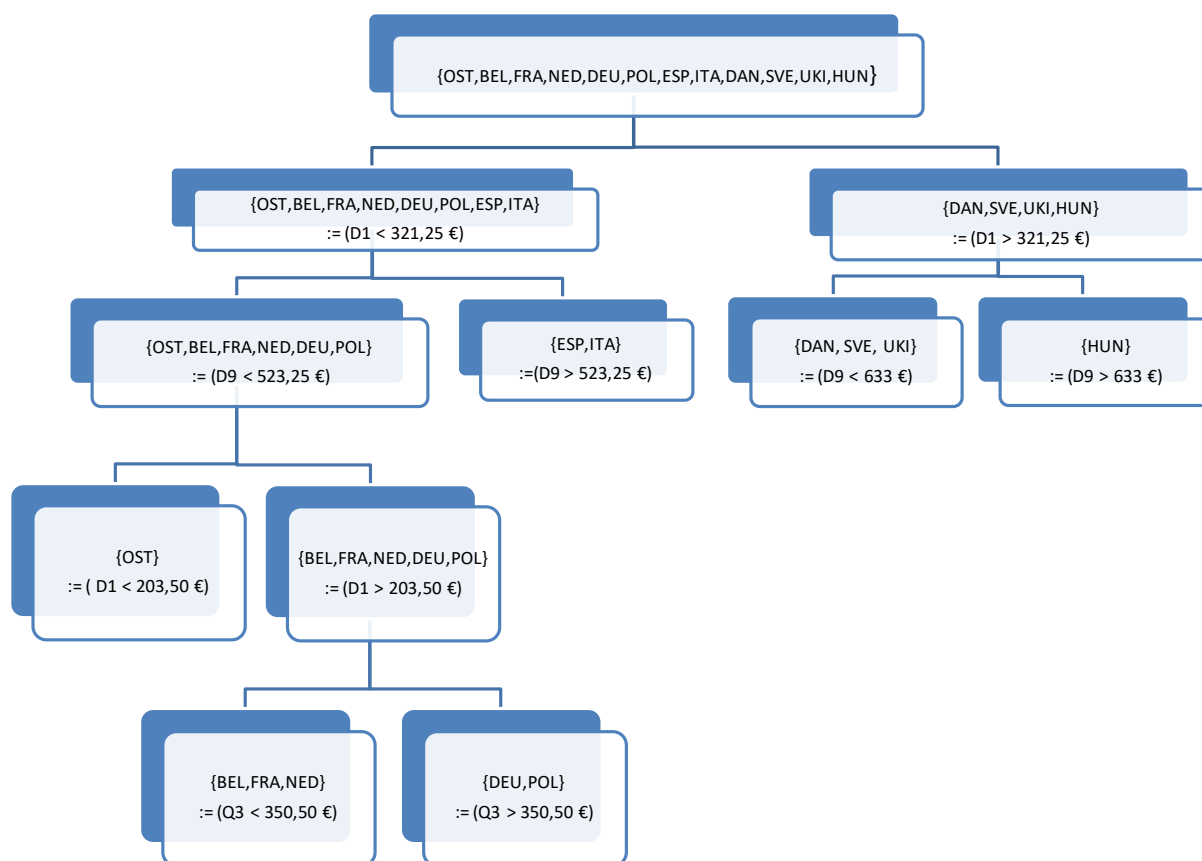
La classification descendante hiérarchique (DIV)³² permet de préciser la structure des coûts selon les classes de pays (figure 9). Tout d'abord, intervient une distinction majeure de localisation des échelles distributionnelles : d'une part, entre l'ensemble composé de la {Hongrie}=C1 associé à la classe {Danemark, Suède, Royaume-Uni}=C2 dont les estimations quartile inférieur (D1) de coûts spécifiques sont supérieures à 321,25 €, et les autres pays pour lesquels D1 est inférieur à cette valeur ; d'autre part, parmi ces autres pays, une distinction complémentaire doit être faite entre ceux dont l'estimation décile supérieure (D9) est supérieure à 523,25 €, soit la classe {Espagne, Italie}=C3, et ceux pour lesquels Q9 est inférieure à cette valeur ; ce dernier sous-ensemble se divise à son tour selon le niveau de l'estimation décile inférieure (D1) : d'une part, {Autriche}=C4 pour lesquels D1 est inférieur à

³² Algorithme de classification non supervisée sur le tableau des intervalles de confiance MBMC à 95% des estimations quantiles (logiciel SODAS 2.5).

203,5 € ; d'autre part, les pays restants {Belgique, France, Pays-Bas, Allemagne, Pologne} pour lesquels D1 est supérieur à cette valeur. Enfin, ces pays restants se subdivisent en deux classes selon le niveau de l'estimation quartile supérieure (Q3) : d'une part, {Belgique, France, Pays-Bas}=C5 pour lesquels Q3 est inférieur à 350,50 € ; d'autre part, {Allemagne, Pologne}=C6 pour lesquels Q3 est supérieur à ce seuil. Ainsi, pour le lait, nous constatons que ce sont les déciles et les quartiles qui constituent les niveaux de ségrégation entre ces six classes de pays : la classe C3={Espagne, Italie} avec l'hétérogénéité la plus importante entre le décile inférieur et supérieur s'opposant en termes distributionnels à la classe C2={Danemark, Suède, Royaume-Uni}, dont l'hétérogénéité est la moins importante au regard de l'intervalle qui sépare les déciles inférieurs des déciles supérieurs.

La définition de cette hiérarchie descendante peut ultérieurement être mobilisée aux fins de classement d'une distribution empirique issue d'un groupe d'exploitations en appliquant des règles décisionnelles : par exemple, si le premier décile conditionnel (D1) de coûts spécifique est inférieur à 203,50 € et si le dernier décile conditionnel (D9) de coûts spécifique est inférieur à 523,25 €, alors ce groupe d'exploitations présente une échelle distributionnelle semblable ou comparable à celle de l'Autriche (OST).

Figure 9 : lait de vache, coûts spécifiques pour 1 000 € de produit brut, classification des pays, UE12.



Source : traitement de l'auteur, d'après EU-RICA 2006.

V Discussion des résultats

L'hétérogénéité des distributions nationales de frais spécifiques recouvre le jeu combiné de différents facteurs de dispersion, dont celui de la dimension économique des exploitations qu'il conviendrait d'analyser. En effet, les pays européens étudiés n'ont ni la même composition en termes de dimension économique des exploitations agricoles, ni les mêmes seuils pour définir une exploitation

professionnelle. Ainsi, la dispersion des estimations quantiles de coûts spécifiques en Hongrie, Espagne ou Italie, recouvre vraisemblablement celles de structures de production très différentes tant dans leur dimension économique que par la technologie de production utilisée.

On ne peut exclure que les valeurs élevées de certaines estimations puissent dans certains cas provenir d'artefacts liés à la méthodologie d'estimation pour des pays où la production laitière soit corrélée avec d'autres ateliers de production présents sur l'exploitation. En effet, la taille de l'atelier laitier relativement aux autres ateliers, selon la spécialisation productive plus ou moins affirmée des exploitations, peut produire des artefacts issus des corrélations productives au niveau des orientations technico-économiques mixtes dans la mesure où le poids des coûts spécifiques liées aux autres ateliers conduirait, en fonction de la hiérarchie des coûts spécifiques, soit à un biais de sous-estimation pour des ateliers laitiers minoritaires par rapport aux produits de grandes cultures au détour de production moins important) soit à un biais de surestimation (pour des productions animales au détour de production plus important telle la production porcine).

Cependant, l'existence de coûts spécifiques très élevés peut également signaler le maintien de producteurs techniquement moins efficaces dans des zones moins favorables en raison de l'existence de mesures globales de soutien au revenu (Barkaoui, Daniel et Butault, 2009), voire de mesures agro-environnementales spécifiques à certains contextes productifs, en particulier celles visant au maintien de la production agricole sur certains territoires. A contrario, les estimations plus faibles peuvent signaler soit la présence d'élevages intensifs plus performants techniquement comme pour les producteurs de lait dans l'Ouest de la France, soit la présence de systèmes productifs basés sur des itinéraires techniques moins exigeants en intrants comme dans les zones de piémont et de montagne. Enfin, l'a priori d'un lait de vache comme matériau standard apte à toute transformation peut être plus ou moins remis en cause selon le type de production : lait conventionnel versus lait bio, lait de montagne versus lait de plaine. Si l'on abandonne cet a priori, l'hétérogénéité du matériau brut ou transformé peut alors être mise en relation avec les dimensions qualitatives de la production animale telles que, pour le lait de vache : le respect de certains standards de production voire de labels de qualité des produits (labels « Bio » ou « de montagne »), ou de caractéristiques biophysiques comme les teneurs en matières grasses et protéiques, plus ou moins favorables à certaines transformations, en particulier fromagères influent alors de manière significative sur sa valorisation (Desbois, 2015b).

Afin de stabiliser les estimations quantiles pour les spéculations minoritaires par rapport au lait de vache, nous proposons dans une première approche d'effectuer des estimations centrées sur des périodes triennales. La seconde approche proposée serait de déflater les produits bruts et charges à la fois des intrants et du produit par les indices de prix correspondants, afin d'éliminer cet effet de la variation de prix. La troisième solution vers laquelle va notre préférence serait de mettre en œuvre une estimation quantile sur données de panel qui nous permette de distinguer l'effet des variations temporelles de celui des variations individuelles. La mise en œuvre de ces estimations de panel dans le cadre de l'estimation par quantiles conditionnels peut être testée à l'aide de l'estimateur « intra » en incorporant directement des indicatrices temporelles et individuelles dans l'équation de la fonction d'objectif.

VI Conclusions

Sur la base du RICA européen, nous avons testé la faisabilité de la méthodologie d'estimation micro-économétrique des coûts spécifiques de production selon les quantiles conditionnels, et nous avons illustré sa pertinence pour prendre en compte l'hétérogénéité intrinsèque de ces distributions pour l'une des commodités majeures du marché européen, le lait de vache.

L'analyse de ces estimations permet de dégager des typologies de distributions nationales de coûts spécifiques. Les principaux pays producteurs sont situés dans un référentiel bidimensionnel à partir d'un schéma d'analyse en composantes principales des estimations par intervalles des quantiles conditionnels qui fournit un outil d'exploration des différences constatées entre échelles de distribution nationales selon leurs quantiles conditionnels respectifs. À partir des distributions d'intervalles estimés des quantiles conditionnels, les différences et similarités entre pays sont

également exploitées pour produire des classes de pays aux coûts comparables, au moyen d'une classification descendante hiérarchique. Les différences entre ces groupes de pays sont délimitées par des seuils exprimés selon les quantiles conditionnels en termes en part du produit brut. Ces seuils peuvent être mobilisés pour segmenter des populations d'exploitations afin d'analyser les effets différentiels de mesure de politique agricole. Ces analyses permettent donc d'identifier différents modèles d'échelle distributionnelle.

Les enseignements tirés de ces deux types d'analyses multidimensionnelles symboliques sont relativement convergents pour le lait de vache : le premier quartile (Q1) et la médiane (Q2) conditionnels constituent les paramètres de coûts spécifiques susceptibles de caractériser au mieux les exploitations spécialisées.

Nous traçons également les perspectives de poursuite et de valorisation de la méthodologie d'estimation selon les quantiles conditionnels dans le cadre d'une analyse input-output des agricultures européennes. Les estimations unitaires données en termes de part des coûts spécifiques dans le produit brut que nous avons privilégiées dans cet exposé peuvent être utilisées dans le cadre du calcul des marges brutes standards que ce soit au plan normatif pour fournir une base statistique à l'estimation de sinistres agricoles, mais également alimenter les matrices d'input-output du secteur agro-alimentaire particulières à un ensemble de pays de l'UE, voire à certains groupes de régions européennes, pour mettre en œuvre des analyses de sensibilité à d'éventuelles options de politique agricole par l'intermédiaire de matrices de comptes sociaux et environnementaux (Léon et Surry, 2009). Dans le contexte actuel de la suppression des quotas laitiers, la typologie nationale proposée pour le lait de vache pourrait être mobilisée pour effectuer des simulations visant à explorer le maintien de la production laitière en zone de montagne ou dans certaines régions intermédiaires.

Références bibliographiques

- Angrist J., Pischke J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 392 p.
- Aufrant M. (1983). *Les coûts de production des grands produits agricoles : évolution de 1970 à 1978*, Institut national de la statistique et des études économiques, Archives et Documents n°64, 246 p.
- Barkaoui A., Daniel K., Butault J.-P. (2009). Impact territorial de la réforme de PAC de 2003. Enjeux du découplage in *Politiques agricoles et territoires*, Editions Quae, 2009, pp. 119-140.
- Barrodale I., Roberts F.D.K. (1973). An improved algorithm for discrete l_1 linear approximation. *SIAM Journal of Numerical Analysis*, vol. 10, p. 839-848.
- Billard L., Diday E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*, 321 p.
- Bock H, Diday E. (2000). *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data*. Berlin, Springer-Verlag, 425 p..
- Bureau J. C., Cyncynatus M. (1991). Estimation de coûts de production et de coefficients input-output à partir de données comptables : méthodes et applications aux produits agricoles sur la base du RICA. *Economie et Sociologie Rurales*, Notes et Documents n°38, INRA, Département Économie et Sociologie Rurales, 87 p.
- Butault J.P., Hassan C.R., Reignier E. (1988). *Les coûts de production des principaux produits agricoles dans la CEE*. Luxembourg: Office of Official Publications of the European Communities.
- Cameron A. C., Trivedi P. K. (2005). *Microeconometrics. Methods and Applications*. Cambridge, Cambridge University Press.
- Carvalho F.A.T. (1997). Clustering of constrained symbolic objects based on dissimilarity functions, *Indo-French Workshop on Symbolic Data Analysis and its Applications*, Université de Paris IX.
- Cazes, P., Chouakria, A., Diday, E., Schektman, Y. (1997). Extension de l'analyse en composantes principales à des données de type intervalle, *Revue de statistique appliquée*, vol.45, n°3, pp. 5-24.
- Chavent M. (2000). Criterion-Based Divisive Clustering for Symbolic Objects, *Analysis of Symbolic Data*, eds. H.H.Bock, E.Diday, Springer, 2000.
- Chouakria, A., Diday, E., Cazes, P. (1998). An improved factorial representation of symbolic objects. In *Studies and Research, Proceedings of the Conference on Knowledge Extraction and Symbolic Data Analysis (KESDA'98)*, Luxembourg: Office for Official Publications of the European Communities, pp. 276-289.
- D'Ambra, L. et Lauro, C. N. (1982). Analisi in componenti principali in rapporto a un sottospazio di riferimento, *Rivista di Statistica Applicata*, n°15(1), pp. 51-67.
- Dantzig G.B. (1949). Programming in a linear structure, *Econometrica*, n°17, pp. 73-74.

- Diday E. (2006). Thinking by Classes in Data Science: the Symbolic Data Analysis Paradigm, *WIREs Computational Statistics*, vol. 8, pp. 171-205.
- Desbois D. (2015a). *Estimation des coûts de production agricoles : approches économétriques*. Thèse de doctorat ABIES-AgroParisTech, dirigée par J.C. Bureau et Y. Surry, 249 p.
- Desbois D. (2015b). La qualité du lait a-t-elle un prix ? Une estimation hédonique de la valeur des attributs spécifiques du lait de vache, *Revue d'Études en Agriculture et Environnement*, n°96, pp. 499-552.
- Desbois D., Butault J.-P., Surry Y. (2013). Estimation des coûts de production en phytosanitaires pour les grandes cultures. Une approche par la régression quantile, *Economie Rurale*, n° 333. pp. 27-49.
- Desbois D., Butault J.-P., Surry Y. (2015). Distribution des coûts spécifiques de production dans l'agriculture de l'Union européenne : une approche reposant sur la régression quantile, *Journées de Recherches en Sciences Sociales*, Nancy. 34 p.
- Desbois D., Butault J.-P., et Surry, Y. (2017). Distribution des coûts spécifiques de production dans l'agriculture de l'Union européenne : une approche reposant sur la méthode de régression quantile, *Economie rurale*, n° 361, pp. 3-22.
- D'Haultfoeuille X., Givord P. (2014). La régression quantile en pratique, *Economie et statistique*, n°71, pp. 85-111.
- EU-DGAgri (2013). *EU Dairy Farms Report 2012*, European Commission, Directorate General for Agriculture and Rural Development, 184 p.
- He, X. and Hu, F. (2002). Markov Chain Marginal Bootstrap, *Journal of the American Statistical Association*, 97, pp. 783–795.
- Karmarkar R. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, n°4, pp. 373-395.
- Kleinhans W, Offerman F., Butault J.-P., Surry Y. (2011) *Cost of Production Estimates for Wheat, Milk and Pigs in Selected EU Member States*, Johan-Heinrich von Thünen Institute, n°7/11, 31 p.
- Koenker R. (2005). *Quantile regression*. Cambridge, Cambridge University Press.
- Koenker R., Bassett G. (1978). Regression quantiles. *Econometrica*, vol. 46, pp. 33-50.
- Koenker R., Bassett G. (1982). Robust Tests for Heteroscedasticity Based on Regression Quantiles. *Econometrica*, vol. 50, n°1, p. 43-61.
- Koenker R., d'Orey V. (1994). Remark AS R92: A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores. *Applied Statistics*, vol. 43, p. 410-414.
- Koenker R., Zhao Q. (1994). L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics*, vol. 3, pp. 223-235.
- Lauro C.N., Palumbo F. (2000). Principal component analysis of interval data: a symbolic data analysis approach, *Computational Statistics*, 15 (1), pp. 73-87.
- Léon Y., Surry Y. (2009). Les effets d'entraînement du complexe agroalimentaire au niveau local, in *Politiques agricoles et territoires*, Editions Quae, pp. 21-48.
- Lustig I. J., Marsden R. E., Shanno D. F. (1992). On implementing Mehrotra's predictor-corrector interior-point method for Linear Programming. *SIAM Journal on Optimization*, n°2, pp. 435-449.
- Madsen K., Nielsen H. B. (1993). A Finite Smoothing Algorithm for Linear Estimation, *SIAM Journal on Optimization*, n°3, pp. 223–235.
- Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometrics Theory*, n°7, pp. 186-199.
- Portnoy S., Koenker R. (1997). The gaussian hare and the laplacian tortoise: Computation of squared-errors vs. absolute-errors estimators. *Statistical Science*, vol. 1, pp. 279-300.
- Silverman B.W. (1981). Using Kernel Density Estimates to Investigate Multimodality. *Journal of Royal Statistical Society B*, vol. 43, 1981, pp. 97-99.
- Verde, R., De Angelis, P. (1997). *Symbolic objects recognition on a factorial plan*, NGUS'97, Bilbao Spain.
- Verde R., Carvalho F.D., Lechevallier Y (2000) A Dynamical Clustering Algorithm for Multi-Nominal Data, *Data Analysis, Classification, and Related methods*, Springer Verlag, pp. 387-394.
- Zellner A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association* n°57, pp. 348–368.