Exploring the distribution of conditional quantiles estimation ranges: an application to the estimation of specific costs of production of pig in the European Union

Dominique Desbois

Exploring the distribution of conditional quantiles estimation ranges: an application to the estimation of specific costs of production of pig in the European Union.

Dominique Desbois[1], UMR Economie publique, Inra-AgroParisTech (dominique.desbois@inra.fr)

Abstract: This communication 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 pig 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, pig, micro-economics, quantile regression, factor analysis and hierarchic clustering of interval estimates.

[1] INRA-AgroParisTech, 16 rue Claude Bernard, F-75231 PARIS CEDEX 05. Mél : dominique.desbois@inra.fr.
This communication is an adaptation of some of the author's work done during the preparation of a thesis (Desbois, 2015), co-directed by Y. Surry and JC. Bureau, supported by the Farm Accountancy Cost Estimation and Policy Analysis project. of European 7th Framework Program of the European Community (FP7 / 2007 2013, Approval No. 212292). This mention does not imply the approval of the persons and organizations mentioned, the author assuming full responsibility for the text.

18th ASMDA Conference Proceedings, 11-14 June 2019, Florence, Italy

"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 to the problem of estimating agricultural production costs.

The successive reforms of the Common Agricultural Policy (CAP), the integration of the agricultural systems of the Member States resulting from the enlargement process of the European Union (EU), both in the context of competitive markets and markets subject to regulation recurrent needs to estimate the production costs of major agricultural products. The analysis of agricultural production costs, whether retrospective or prospective, is also a tool for analyzing farmers' margins. It makes it possible to evaluate the price competitiveness of farmers, one of the major elements of the development or maintenance of agro-food chains in certain European regions. Thus, the estimation of production costs provides some partial but essential insights into the questions posed by the adaptation of European agriculture to the context of agricultural markets, whether national, European or international, both from the point of view of the regulation of agricultural products. international trade in agricultural products (see the proposals for measures to combat market imbalances in the post-quota dairy sector[2]), and the successive reforms of the CAP (see the debate on future CAP in 2020[3]) or new challenges for European agriculture caused by environmental factors (climate change, environmental and biodiversity management[4]).

Confronted more directly with price risks since the abolition of production quotas in 2015[5], European producers with few opportunities for differentiation opt for cost reduction strategies, seeking either to reduce structural costs by playing on the volume of production, either to reduce specific costs by optimizing the management of inputs or opting for low-input technical routes. However, structural adjustment is not always possible due to constraints (herd management, rights to produce, availability) that can restrict access to the three main production factors of land (e.g. mountain areas), working capital (financing conditions) or work, whether salaried or self-employed. On the other hand, the adjustment on specific inputs offers more flexibility as shown by the adoption of reasoned practices leading to savings on the main items of expenditure such as animal feed and veterinary fees. The evolution of specific costs, not only globally but also by product, thus constitutes an important indicator for pig farmers in terms of technical management of the herd and adjustment of their product mix to the demands of the agricultural markets, taking into account the resources and competitiveness factors available to them.

Given these different issues, in contexts either ex ante scenario development or ex post evaluation of measures concerning possible agricultural public policy options, we must be able to provide information as suggested (Angrist and Pischke, 2009) across the entire distribution of production costs, thus making it possible to meet the needs of simulations or impact analysis within the various common organizations of the market. In this perspective, from the observation of asymmetry and heterogeneity of their empirical distribution, we propose a methodology adapted to the problem of the estimation of the specific costs of production relative to the main agricultural reference products in a European context where farm holdings remain predominantly multi-commodity, despite a preponderance of

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

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

4 http://agriculture.gouv.fr/lagriculture-et-les-forets-au-coeur-de-la-cop23

5 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

specialized farms in some more integrated sectors of agricultural production. In this multi-product context, it is strategic to generate for each of the main agricultural products the central estimates of the cost distribution, but also the lower or higher quantiles with a view to selectivity of the instruments for regulating agricultural markets for production, and evaluation of public policies.

Given the heterogeneity of agricultural production structures and productive choices in Europe, how can the maximum amount of information be used to estimate agricultural production costs? In response to this concern, we propose an estimation methodology that can provide information on the overall distribution of specific production costs for the main agricultural reference products in a European context. In order to overcome the constraint of average estimators, sensitive to the asymmetry or the extreme values of the distributions of interest and likely to mask the inter-structural differences, it is necessary to generate for each of the main agricultural products not only the median estimates of cost distribution but also lower or higher quantiles. To this end, we propose using a methodology to obtain estimates of these quantiles of specific costs that are conditioned by the product mix of farmers (Desbois, Butault and Surry, 2017). In order to demonstrate the relevance of this approach, we will then apply this methodology to estimate the specific costs of pig, given its place in the world production by the EU28[6], on a set of twelve European states (EU12) where these productions are significant in 2006, the base year chosen for the period.

We first present the empirical model for estimating the specific costs of production, derived from an econometric cost allocation approach, initially developed by (Aufrant, 1983) proposing to use microeconomic data to build an input-output matrix [Divay and Meunier, 1980). Then, we introduce the estimation methodology according to the conditional quantiles proposed by (Koenker and Basset, 1978). Next, we present the symbolic data analysis procedures used to explore the empirical estimates of conditional quantile distribution intervals based on the concepts and methods provided by the symbolic approach (Bock and Diday, 2000). Then, we present the graphs from the tools of analysis of the symbolic data applied to the estimation intervals of the conditional quantiles. Finally, we conclude on the relevance of this approach applied to the production of pig, proposing an extension of this type of analysis at the regional level.

II Conceptual framework and methodological aspects of cost allocation

Surveys specific to large agricultural commodities are conducted by workshop to provide detailed data on operational production costs, such as that used by the French pig Institute (IFIP) on specialized pig producers for France[7]. However, these technical and economic surveys are relatively expensive, making their generalization to all European pig farms financially unbearable. Also, this work is situated in a problem of attribution of the costs of the factors to multiple productions, initiated on a European scale by Inra works (Butault, Hassan and Reignier, 1988) financed by the Commission of the European Communities (CEC) , allowing to estimate on the basis of the Accounting Information Network (FADN),

[6] In 2017, pig production produced by the 28 European countries accounted for 20% in weight of pigmeat produced at the world level (according to https://ec.europa.eu/agriculture).

[7] https://www.ifip.asso.fr/fr/resultats-economiques-elevages-de-porc.html

harmonized accounting survey from the point of view of the definition of the professional holdings and the accounting, technical and financial aggregates.

II.1 The empirical model for estimating the specific costs of production

In EU agricultural accounting systems, the recording of charges is done at the farm level and does not provide a direct estimate of the production costs incurred by that farm for each of the agricultural crops undertaken. On the other hand, the farm holding sheet[8] of the FADN survey provides individually by farm from the accounting records the amount of the gross products generated by the various speculations and that of the set of specific costs, sum of the purchases of recorded inputs. So that it becomes possible to estimate by regression the specific costs on the gross products of the allocation coefficients of expenditure to the main agricultural products, which we will call "specific coefficients of production".

The definition of the gross margin M_i of the farm holding *i* as a difference between the sum of the gross products Y_i and the sum x_i of the specific fees: $M_i = Y_i - x_i$,

by linearly decomposing the sum of the specific costs according to each production *j* :

$$x_i = \sum_{j=1}^{p} \gamma_j Y_i^j + \varepsilon_i \quad with \quad \varepsilon_i \quad i.i.d. \qquad (1)$$

implies:

$$M_i = \sum_{j=1}^{p} Y_i^j - \sum_{j=1}^{p} \gamma_j Y_i^j + \varepsilon_i = \sum_{j=1}^{p} (1-\gamma_j) Y_i^j + \varepsilon_i$$

Thus, the allocation of the specific costs of the farm holding *i* to the set ***J*** of the productions carried out by this conceptual model makes it possible, because of the complementation to the unit, to deduct the unit rates of gross margin $\hat{\alpha}_j$ from the estimate of the specific production coefficients $\hat{\gamma}_j$ for each of the *J* productions envisaged: $\hat{\alpha}_j = (1-\hat{\gamma}_j) \; j = 1, \cdots, p$.

The linear decomposition of the gross margin leads us to estimate the specific production coefficients of the stochastic equation (1) for comparison:

- on the one hand, according to the Ordinary Least Squares (OLS) regression methodology, the solution of which $\hat{\mu}_{MCO}(i) = \sum_{j=1}^{p} \hat{\gamma}_j^{MCO} Y_i^j$ is conventionally expressed in terms of conditional expectation;
- on the other hand, according to that of the quantile regression whose solution $\hat{\mu}_q(i) = \sum_{j=1}^{p} \hat{\beta}_j^{(q)} Y_i^j$ is expressed in terms of conditional quantiles of order *q*, in order to take into account the intrinsic heterogeneity of the distribution of specific costs as shown in the following section on the estimation methodology.

II.2 The interest of conditional quantiles in the estimation of agricultural production costs

---

[8] The questionnaire used to establish this farm holding sheet and the methodology of the FADN survey are available at: : http://wwww.agreste.agriculture.gouv.fr/enquetes/reseau-d-information-comptable-610/reseau-d-information-comptable.

The standard specification of the least squares model raises certain problems that it would be risky to neglect in the perspective of the establishment of benchmarks on production costs, taking into account the challenges of competitiveness for the various sectors. Firstly, in a context of use mobilizing the European FADN as an empirical basis for estimating the specific production costs, the stochastic assumptions of the Gauss-Markov model may not be respected: thus, the asymmetry of the distributions of specific costs (concentration for lower values and dispersion of values higher than average, or vice versa) leads to rejection of the assumption of normality of errors. In addition, given the selection method specific to each national RICA (for example, the French FADN is a survey administered according to the quota method), the accounting data are not always collected according to a stratified random sampling plan allowing to deliver inferential reasoning an interval estimation based on a parametric distribution, even in the asymptotic case.

The conditional estimation of quantiles was developed in (Koenker and Bassett, 1978) under the name of "quantile regression" in order to take into account the heterogeneity of the set of values of an endogenous variable *x* in the context of a linear model. When looking at farms, this econometric method yields an estimated distribution of specific costs for major agricultural products and thus complements the estimates obtained by OLS, which provide only an average value (in terms of expectation) of these same costs. Instead of an interval estimate built on a normality assumption, the quantile process provides an empirical distribution of the estimates without having to make assumptions about the nature of this distribution or to follow a stratified random sampling design. For a continuous random variable *x*, the q^{th} quantile of the population is the value μ_q such as *x* is less than or equal to μ_q with the probability *q* : $q = \Pr[x \leq \mu_q] = F_x(\mu_q)$ **(2)**

where F_x is the cumulative distribution function (CDF) of *x* giving the cumulative probability of a value under the law of *x*. The $q^{ième}$ quantile is then defined as the image of the value q by the reciprocal function of the CDF : $\mu_q(x) = F_x^{-1}(q)$ (**3**)

In quantile regression, the q^{th} conditional quantile of the cost of production ***x*** conditioned by all the exogenous variables ***Y*** determining input consumption is the indexed function $\mu_q(x|\boldsymbol{Y})$ such as the random variable « *x* knowing ***Y*** » $(x|\boldsymbol{Y})$ is less than or equal to $\mu_q(x|\boldsymbol{Y})$ with the probability *q*. Thus, we can formally define the q^{th} conditional quantile by the following expression :

$$\mu_q(x/Y) = F_{x/Y}^{-1}(q) \quad \textbf{(4)}$$

where $F_{x|\boldsymbol{Y}}$ is the CDF of *x*'s probability law conditioned by ***Y***.

Following (Cameron and Trivedi, 2005), suppose that the data generating process is a linear model with multiplicative heteroscedasticity:

$$\mathbf{x} = \mathbf{Y'\beta} + \mathbf{u} \quad \text{avec} \quad \mathbf{u} = \mathbf{Y'\alpha} \times \boldsymbol{\varepsilon} \quad \text{et} \quad \mathbf{Y'\alpha} \rangle 0 \quad \textbf{(5)},$$

where $\varepsilon \sim i.i.d.$ $(0, \sigma^2)$ is an identically and independently distributed random vector of zero mean and constant variance σ^2.

Under this hypothesis , $\boldsymbol{\mu_q(x/Y,\beta,\alpha)}$, the q^{th} conditional quantile of the production cost *x* conditioned by ***Y*** and the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, is analytically deduced as follows :

$$\mu_q(\mathbf{Y},\boldsymbol{\beta},\boldsymbol{\alpha}) = \mathbf{Y}'\left[\boldsymbol{\beta} + \alpha \times F_\varepsilon^{-1}(q)\right] \qquad \text{where } F_\varepsilon \text{ is the CDF of the random error } \boldsymbol{\varepsilon}.$$

Thus, for a data generating process following a linear model with multiplicative heteroscedasticity (i.e. ***u=Y'α×ε***), the q^{te} conditional quantile of the cost of production *x* conditioned by exogenous factors ***Y*** is linear in $\boldsymbol{Y}$. Moreover, the estimation of the parameters associated with the q^{th} quantile regression converges towards $\beta+\alpha \times F_\varepsilon^{-1}(q)$ and therefore behaves monotonically with respect to the evolution of the quantile *q*, depending on the quantile function of the residues , $F_\varepsilon^{-1}(q)$.

Following a typology of models presented in (Givord and D'Haultfoeuille, 2013), several models can be distinguished:

i) $X = Y'\beta + u$ with $u = K\epsilon$ with homoscedastic residues ($V(\epsilon|Y) = \sigma^2$) designated as the linear model of homogeneous slope conditional quantile ("location shift model"). The case where $Y'\alpha = K$ is constant, corresponds to conditional quantiles differing only by a constant ($\mu_q(X / Y, \beta, \alpha) = Y'\beta + KF_\varepsilon^{-1}(q)$, all showing the same slope and growing uniformly as the *q* order of the quantile increases;

ii) $X = Y'\beta + (Y'\alpha)\varepsilon$ with $Y'\alpha > 0$ with heteroscedastic residues, referred to as the heterogeneous-slope conditional quantile linear model ("location-scale shift model"). the case where $Y'\alpha > 0$ corresponds to heterogeneous and increasing slopes $\mu_q(X|Y, \beta, \alpha) = Y'\left(\beta + \alpha\mu_q(\epsilon)\right)$, involving fixed linear effects $\gamma_q = \beta + \alpha\mu_q(\varepsilon)$ relatively lower for the first quantiles and relatively higher for the last quantiles;

iii) $X = Y'\gamma_U$ with U random variable independent of Y according to a uniform distribution over the interval [0,1] such that the function $u \longrightarrow y'\gamma_u$ is strictly increasing whatever *y* , is designated as the random coefficient model. U corresponds to an unobserved random component determining the rank of the individual within the *X* distribution. Under the distribution invariance assumption of ranks, hypothesis considered strong in the scientific literature, the random coefficient γ_q would represent the effect of a marginal change in *Y* for farms at the q^{th} quantile of the *U* distribution, based on unobserved characteristic . For example, this distributional assumption of rank invariance is equivalent to assuming that median farms ($q = 0{,}5$) in terms of input productivity would maintain this rank, regardless of the different levels of production y_i registered for farm holding *i*.

II.3 Estimation and test procedures

The Ordinary Least Squares (OLS) estimator can be written as a solution to an optimization problem that minimizes the sum of the squared deviations (in L_2 norm) :

$$\hat{\beta}_{MCO} = \underset{\beta\in\Re^p}{\arg\min}\left\{\sum_i (x_i - y_i'\beta)^2\right\} = \underset{\beta\in\Re^p}{\arg\min}\left\{\mathbf{e}'\boldsymbol{\delta}^2(X - Y'\beta)\right\}$$ where **e**, is the director vector of the constant line in $\mathbb{R}^n$, space of observations , and δ^2, the vector of quadratic differences. Similarly, the quantile regression is defined for each quantile of order *q* as the solution of a problem of minimizing the sum of the deviations in absolute value (in norm L_1):

$$\hat{\beta}(q) = \underset{\beta\in\Re^p}{\arg\min}\left\{\sum_{i\in\{i / x_i \geq y_i'\beta\}} q|x_i - y_i'\beta| + \sum_{i\in\{i / x_i \leq y_i'\beta\}} (1-q)|x_i - y_i'\beta|\right\} \qquad \textbf{(7)}$$

which can be written in matrix form as follows:

$$\hat{\beta}(q) = \underset{\beta\in\Re^p}{\arg\min}\left\{q\mathbf{e}'(X - Y'\beta \geq 0)\boldsymbol{\delta}^1[\mathbf{X} - \mathbf{Y}'\boldsymbol{\beta}] + (1-q)\mathbf{e}'(Y'\beta - X \geq 0)\boldsymbol{\delta}^1[\mathbf{Y}'\boldsymbol{\beta} - \mathbf{X}]\right\},$$

with $\boldsymbol{e}'(\mathrm{X}-\mathrm{Y}'\beta \geq 0)$, vector of indicator values of observations i such as $x_i - y'_i\beta \geq 0$, and δ^1, the vector of absolute deviations.

Let be $x_i = y_i'\beta_q + e_i$ with $e_i = u_i - v_i, u_i = e_i\mathbb{I}(e_i > 0)$, $v_i = |e_i|\mathbb{I}(e_i < 0)$, then, like the L1 regression (Barrodale and Roberts, 1973), quantile regression can be formulated as a primal problem of linear optimization, which is expressed in matrix form as follows:

$$\hat{\beta}(q) = \underset{\beta\in\Re^p,(u,v)\in\Re^{n\times n}}{\arg\min} \{q\mathbf{1'u} + (1-q)\mathbf{1'v}\} \quad \text{under the constraint} \quad \mathbf{X} = \mathbf{Y'\beta} + \mathbf{u} - \mathbf{v} \quad (\mathbf{8})$$

This program can be reformulated into a dual problem of equivalent optimization:

$$\underset{z}{Max}\,\{\mathbf{x'z}\} \text{ under the constraint } \mathbf{Yz} = (1-q)\mathbf{Y1} \text{ for } z \in [0,1]^n \quad (\mathbf{9})$$

Thus, the linear optimization problem solving methods developed for the L1 (median) regression easily extend to quantile regression (Koenker and d'Orey, 1994). The simplex method (Danzig, 1949) has an algorithmic complexity in $\mathrm{O}(n^6)$), the "interior point" method (Karmarkar, 1984) of algorithmic complexity $\mathrm{O}(n^{3,5})$ is preferable in practice as soon as the size of the sample is important. For large samples, Portnoy and Koenker (1997) showed that a combination of the "interior point" algorithm[9] and a smoothing algorithm of (Madsen and Nielsen, 1993) for objective function makes quantile regression computationally competitive with least squares regression .

The weighted conditional quantiles have been proposed as L-estimates in linear heteroscedastic models by (Koenker and Zhao, 1994) defined by $\{\omega_i; i = 1,\cdots,n\}$, the weighting ω of observations leads to a quantile regression scheme solving the following minimization problem:

$$\hat{\beta}_\omega(q) = \underset{\beta\in\Re^p}{\arg\min}\left\{\sum_{i\in\{i/x_i\geq y_i'\beta\}} \omega_i q|x_i - y_i'\beta| + \sum_{i\in\{i/x_i\leq y_i'\beta\}} \omega_i(1-q)|x_i - y_i'\beta|\right\} \quad (10).$$

The weighted estimation procedure uses the "predictor-corrector" implementation of the primal-dual algorithm proposed by (Lustig et al., 1992).
Let assume the following regularity conditions:

i) The distributions $F_i(x)$ of input expenditures for a given product mix are absolutely continuous with densities $f_i(x)$ continuous and uniformly bounded on $]0,+\infty[$ at $\xi_i = \mu_q(x|y_i)$;

ii) $\Sigma_0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n y_i y_i'$ exists and is positively defined ;

iii) $\Sigma_1 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_i(\xi_i) y_i y_i'$ exists and is positively defined ;

iv) $Sup_{i=1,\ldots,n}\|y_i\| \sim O(\sqrt{n})$, as a product mix normalization factor ;

(Pollard, 1991) shows that under conditions i) and ii), $\hat{\beta}_q \xrightarrow{p} \beta_q$, the estimator converges in probability; in addition, under the set of conditions i), ii), iii) et iv), we obtain the asymptotic normality, that is: In addition, under conditions iii) and iv), if the hazards attached to the

[9] The weighting ω is introduced by the standard instruction *weight* into the *QUANTREG* procedure of the *SAS 9.2* software.

$$\sqrt{n}\left(\hat{\beta}_q - \beta_q\right) \underset{loi}{\rightarrow} N\left(0, q(1-q){\Sigma_1}^{-1}\Sigma_0\Sigma_1\right) \quad (11)$$

Finally, under the assumption of equality and independence of distributions $f_i(\xi_i) = f_\varepsilon(0)$, this result is simplified as follows:

(12) $\sqrt{n}\left(\hat{\beta}_q - \beta_q\right) \underset{loi}{\rightarrow} N\left(0, \sigma^2(q){\Sigma_0}^{-1}\right)$ avec $\sigma(q) = \frac{\sqrt{q(1-q)}}{f_\varepsilon(0)}$

In addition, under conditions iii) and iv), if the hazards attached to the i^{th} observation $\varepsilon_i = x_i - y_i'\beta$ of identical and independent distributions F_i , admitting a density $f = \dot{F}$ such as $f\left(F^{-1}(q)\right) > 0$ at the neighborhood of q, then (Koenker et Bassett, 1982) shows :

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

These results can be used to construct confidence intervals for estimates using three procedures: inverse density function, rank method, or resampling algorithm. The inverse density function estimation is the most direct and the fastest method, but it is sensitive to the hypothesis of identically and independently distributed data (iid). For data that is not iid, the rank method, which calculates confidence intervals by reversing the rank score test, is preferred. However, based on the simplex method, the rank method generates significant computation times for large datasets. The resampling method, based on the bootstrap technique, makes it possible to overcome all assumptions but is unstable for small samples. Given the size of the FADN sample, its non-random selection and the existence of three distinct a priori sub-populations (specialized or non-pig ToFs), we opted for the resampling method, based on the procedure use on a Markov chain marginal bootstrap (*MCMB*) because, without hypothesis on random distributions, this method gives robust empirical confidence intervals in a reasonable calculation time (He and Hu, 2002).

II.4 Symbolic analysis of empirical distributions of specific costs

II.4.1 Principal component analysis of distributions

The PCA of the interval extrema

In the extension of the principal components analysis (PCA) to the interval data proposed by (Cazes, Chaouakria, Diday and Schektman, 1997), called PCA according to the vertices of the intervals (*V-PCA*[10]), a standard ACP of the centered Z-reduced array (standard ACP) is carried out. In this way, the vertices of the hyper-rectangles are vectors of $\mathbb{R}^p$, while the estimates of the conditional quantiles are elements of $\mathbb{R}^N$. Thus, the V-PCA provides a dual representation of the specific empirical cost distributions represented by their estimation intervals, which are the symbolic objects, and conditional quantiles which are the descriptors.

As in classical ACP, the proper subspace (optimal for the dual representation) is structured by orthonormal axes v_m ($1 \leq m \leq p$, maximizing the sum of squares of vertex coordinates $\psi_m = Zv_m$ and satisfying in $\mathbb{R}^N$ to the characteristic eigenvector equation v_m and eigenvalues λ_m of the matrix $\frac{1}{N}Z'Z$:

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

The dual analysis $\mathbb{R}^p$ leads to a similar equation

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

[10] *Vertex Principal Component Analysis*

having the same non-zero eigenvalues but eigenvectors w_m such as:

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

The interpretation of the axes of the V-PCA is based on the conditional quantiles (variables of the V-PCA) presenting the strongest contributions. In normalized PCA, the contributions to the inertia of the variables *j* to the axis *m* are calculated as the square of the correlation between the factorial axis and the variable (factorial coordinates). The coordinates of the projections of the extremes of the estimation intervals of the conditional quantiles (vertices *s (i)*) of the empirical distribution ω_i specific costs (symbolic object) on the main factorial axes are provided by the relation:

$$\psi_{i,m} = Z_i v_m$$

The representation of the empirical distribution ω_i on the factorial axis m is provided by the projections of the ends of the estimated quantile intervals (hyper-rectangle of maximum inaccuracy , HRIM). The projection of the HRIM on a factorial plane provides a maximum imprecision rectangle (RIM) for the empirical distribution represented by the symbolic object ω_i.

In order to avoid the over-dimensioning projection bias of projected rectangles, (Chouakria, Cazes and Diday, 1998) propose to retain among the representations those whose vertices are best represented by using as a criterion the relative contribution (CTR) defined in cosine terms, that is, for an end of the estimation interval *s (i)* of the quantile distribution i :

$$CRT_{s(i),m} = \sum_{j=1}^{p} \left(z_{s(i),j} v_m\right)^2 \Big/ \sum_{j=1}^{p} {z_{s(i),j}}^2$$

The problems of representation of symbolic objects are studied by (Verde and De Angelis, 1997) in terms of better adjustment of convex envelopes.

The PCA's estimate interval centers

Let be $\delta_i^j = \left[\underline{x_i^j}; \overline{x_i^j}\right]$, the estimate interval of the conditional quantile j for the empirical distribution i of specific costs, this estimation interval can be represented by the data of the couple $(m_{ij}; r_{ij})$ where $m_{ij} = \frac{\underline{x_i^j} + \overline{x_i^j}}{2}$ is the middle of the interval and $r_{ij} = \frac{\overline{x_i^j} - \underline{x_i^j}}{2}$ its radius. The T matrix of the interval data is then constituted by the concatenation of the matrix of the estimation interval centers with the matrix of the estimation interval radii.

The PCA of the interval centers (PCA-IC) in $\mathbb{R}^N$, the space of the specific cost distributions, corresponds to the following characteristic equation:

$$\frac{1}{n} \widetilde{M}' \widetilde{M} v_m = \lambda_m v_m$$

where $\widetilde{M}$ is the matrix M standardized by the standard deviation of the interval centers, v_m and λ_m are respectively the eigenvectors and the eigenvalues associated with the inertia operator $\frac{1}{n} \widetilde{M}' \widetilde{M}$. It is therefore the diagonalization of the correlation matrix $\frac{1}{n} \widetilde{M}' \widetilde{M}$ conditional quantile estimators across all specific cost distributions. The centers of quantile estimation intervals [11] are projected on the factorial planes, with in additional projection the ends of the estimation intervals (vertices).

In the vertice-based analysis (PCA-V), these are considered as independent statistical units. In order not to lose information on the size and shape of the hyper-rectangles, (Lauro and Palumbo, 2000) introduce a constraint of cohesion between the vertices. The method is based on maximizing the variance between symbolic inter-objects. Let be *A*, the Boolean matrix indicating the membership of the *N* estimation interval ends to the *n* empirical distributions. The expression of the variance between symbolic objects is given by:

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

If all the empirical distributions have the same number of estimation intervals, then the same number of vertices $\frac{N}{n} = 2^p$ (our case study where the same number of conditional quantiles were estimated for each empirical distribution), then $A'A = 2^p I_n$.

Let be, the orthogonal projector P_A associated with the *A* matrice on the sub-space of reference :

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

In the space $\mathbb{R}^N$, the factorial axes of inertia are obtained as a solution of the following eigenvalue equation:

$$\frac{1}{N}Z'P_A Z\tilde{v}_m = \tilde{\lambda}_m\tilde{v}_m$$

where $\tilde{\lambda}_m$ et $\tilde{v}_m$ are the eigenvalues and the eigenvectors associated with the inertia operator $\frac{1}{N}Z'P_A Z$.

The coordinates of the hyper-rectangle associated with the empirical distribution are then computed as follows:

$$\tilde{\psi}_{i,m} = Z_i\tilde{v}_m$$

The analysis in $\mathbb{R}^p$ is equivalent to solving the following equation for eigenvalues:

$$(A'A)^{-1/2}(A'ZZ'A)(A'A)^{-1/2}\widetilde{w}_m = \tilde{\lambda}_m\widetilde{w}_m$$

where $\widetilde{w}_m = (A'A)^{-1/2}A'Z\tilde{v}_m$ modulo the $\frac{1}{N}$ constant.

The relative contributions of the variables (CRT) are defined in the same way as for the ACP-IS. These CRTs are also used to select the empirical distributions to be represented on the factorial graphs. As in V-PCA, representations of empirical distributions as a symbolic object are constructed by the Maximum Covering Area Rectangular (MCAR). If one compares the V-PCA with the IC-PCA proposed by (Cazes

[11] It should be noted that Markov chain marginal estimation intervals (MBMC) are not symmetrical in contrast to asymptotic estimation intervals, whereas interval-based PCA-IC assumes in its representation the symmetry with respect to center. Nevertheless, we propose to use the point estimate as a center and to introduce the concept of lower radius and upper radius to locate the ends of the MBMC interval.

and *al*., 1997), where the PCA is performed only on the interval data centers standardized with the matrix of the correlations of the centers of the variables, several improvements were introduced: first, the data are standardized by the standard deviation of the vertices considered as active units in the analysis, whereas they are considered in the IC-PCA as additional units; more generally, this can be considered an improvement of the IC-PCA because it can be applied to data constrained by logical or hierarchical relationships.

PCA ranges of estimation intervals

Partial PCA can also be used to better emphasize the differences between symbolic objects. The following section shows a partial PCA where the vertices of the hyper-rectangles are centered relative to the *Inf* value.

In order to take into account only the sizes and shapes of the hyper-rectangles associated with the descriptions, (Lauro and Palumbo, 2000) proposed a PCA based on scaling interval data that summarizes useful information to describe the size and shape of symbolic objects, with the following transformation $\mu(z_i^j) = \overline{x_i^j} - \underline{x_i^j}$. The *Description Potential – DP,* (De Carvalho, 1992, 1997) is the hyper-volume associated to the description of the i^{th} empirical distribution, domain defined by the cartesian product of the ranges $Z_i = z_i^1 \times \cdots \times z_i^j \cdots z_i^p$ of p parameters associated with the description of the symbolic object ω_i. Its measurement is defined by: $\pi(\omega_i) = \prod_{j=1}^{p} \mu(z_i^j)$

where $\mu(z_i^j) = z_i^j / s_j$ is the normalized range relative to the domain $D^j = \{z_i^j; i \in I\}$ of the descriptor Z^j. However, if the measure of one of the descriptors tends to zero, then the description potential tends to zero. To overcome this drawback, we use the linear measure of the potential of description (Carvalho, 1997), or for the instance a_i of the symbolic object ω_i: $\sigma(a_i) = \sum_{j=1}^{p} \mu(z_i^j)$.

Let all the instances of the empirical distributions of specific costs $\{a_1, \cdots, a_i, \cdots, a_n\}$ and $\boldsymbol{X}$ the matrix $n \times p$ of general term $x_i^j = \sqrt{z_i^j}$, then the PCA in range of estimation ranges is defined by the factor decomposition of the total linear description potential $LDP = \sum_{i=1}^{n} \sigma(a_i)$, allowing a different geometrical representation of the vertices than in the V-PCA.

The transformation of the data into a range with an affine translation where the minima $\left\{\underline{x_i^1}, \cdots, \underline{x_i^j}, \cdots, \underline{x_i^p}\right\}$ are all located at the origin. Thus, the search for an optimal representation subspace for the size and shape of each symbolic object is made from a non-centered PCA of maxima $\left\{\overline{x_i^1}, \cdots, \overline{x_i^j}, \cdots, \overline{x_i^p}\right\}$.

This Range Transformation PCA (RT-PCA) breaks down the criterion

$$LDP = tr(X'X) = tr(XX') = \sum_{i=1}^{n} \sigma(a_i)$$

according to the following characteristic eigenvector equations:

$$X'Xt_m = \mu_m t_m$$

and

$$XX'u_m = \mu_m u_m$$

Thus, the sum of the eigenvalues μ_m associated with the eigenvectors t_m in $\mathbb{R}^n$and u_m in $\mathbb{R}^p$ corresponds to the factorial decomposition of the linear description potential:

$\sum_{m=1}^{p} \lambda_m = \sum_{i=1}^{n} \sigma(a_i)$.

The factorial coordinates of the representation of the specific cost distributions in the optimal subspace are given by: $\phi_m = Xt_m$

The absolute contribution (CTA), as the ratio between the factorial coordinate and the eigenvalue, measures the contribution of the empirical distribution of specific costs to the potential of description of the *m*[th] factorial axis; it is defined by: $CTA_{i,m} = \frac{\phi_{i,m}^2}{\mu_m}$

The relative contribution (CTR) measures the representation quality of the empirical distribution in the chosen representation factorial subspace: $CTR_{i,m} = \frac{\sum_{m=1}^{M*} \phi_{i,m}^2}{\sum_{j=1}^{p} x_{i,j}^2}$

The interpretation of the factorial axes is performed according to the contributions (factorial coordinates) of the estimated quantiles estimated for the empirical distributions of specific costs, as descriptors of the symbolic objects : $CTA_{j,m} = t_{j,m}^2$.

The range ACP can be represented by the projection of the factorial coordinates of the maxima. The distributions described by conditional quantile estimates, share representations in hyper-rectangles similar in size and shape if they are projected in the same neighborhood.

If all the terms of matrix X are positive then the first eigenvectoru_1 and the associated factor

$\phi_1 = Xt_1$

are positive (Lauro and Palumbo, 2000). The first major component can therefore be interpreted as a size factor, while the higher order factors order the empirical distributions according to their shape characteristics.

Mixed Strategy PCA of Estimation Intervals

The mixed strategy in principal component analysis of symbolic objects (*SO-PCA)* combines the vertex PCA (V-PCA) and the range PCA (RT-PCA) in a three-step approach to account for differences in scale and shape between empirical distributions of specific costs:

i) PCA ranges to extract the main axes that best represent the scales and forms of empirical distributions of conditional quantiles;
ii) Projection from Z to $\hat{Z} = P_A Z$ in order to take into account the relations between the different extrema, given the order relationships between the different conditional quantiles of the distribution of specific costs ;
iii) PCA line projections $\hat{Z}_i$ on the sub-space of optimal representation $\Phi = \{\phi_1, \cdots, \phi_m, \cdots, \phi_{M^*}\}$ on the sub-space of optimal representation $P_\Phi = \Phi(\Phi'\Phi)^{-1}\Phi'$.

The mixed analysis strategy is therefore based on the solution of the following characteristic equation :

$$\hat{Z}'P_{\Phi}\hat{Z} = Z'A(A'A)^{-1/2}P_{\Phi}(A'A)^{-1/2}A'Zs_m = \rho_m s_m$$

where the diagonal matrix $(A'A)^{-1}$ s broken down in the product $(A'A)^{-1/2}(A'A)^{-1/2}$

for reasons of symmetry, with respectively ρ_m and s_m , eigenvalues and eigenvectors associated with the usual orthonormality conditions.

The interpretation of the results of the analysis depends on the choice of the projection operator P_{Φ}, whose diagonal term, interpretable as a normalized weight, is equal to:

$$\phi_i(\phi'_i\phi_i)^{-1}\phi'_i = \sum_{m=1}^{M^*} \phi_{i,m}^2/\mu_m$$

II.4.2 Automatic clustering of empirical distributions of specific costs

For all empirical distributions of specific costs $\Omega = \{\omega_1, \cdots, \omega_i, \cdots \omega_n\}$ described as symbolic objects by a set of *p=6* descriptors[12] which are the conditional quantiles $X = \{\tilde{Q}_{0,10}, \tilde{Q}_{0,25}, \tilde{Q}_{0,50}, Q_{0,75}, \tilde{Q}_{0,90}\} = \{x_1, \dots, x_j, \dots, x_p\}$. The dissimilarities associated with interval estimates of conditional quantiles $\delta_i^j = \left[Inf = \underline{x_i^j}; Sup = \overline{x_i^j}\right]$ can be calculated between two closed intervals of the *j^th^* conditional quantile $\delta_i^j = \left[\underbrace{x_i^j}; \overbrace{x_i^j}\right]$ and $\delta_k^j = \left[\underline{x_k^j}; \overline{x_k^j}\right]$ respectively associated with the distributions characterizing the country *i* and the country *k*, according to the following three standards:

Metric L_1 (sum of absolute differences)[13] : $\delta_1(\delta_i^j, \delta_k^j) = \left|\underline{x_i^j} - \underline{x_k^j}\right| + \left|\overline{x_i^j} - \overline{x_k^j}\right|$

Metric L_2 (sum of quadratic differences)[14] : $\delta_2(\delta_i^j, \delta_k^j) = \sqrt{\left(\underline{x_i^j} - \underline{x_k^j}\right)^2 + \left(\overline{x_i^j} - \underline{x_k^j}\right)^2}$

---

[12] This choice of a small number of descriptors was made for comparative convenience with some more classical graphic approaches (Desbois, Butault and Surry, 2013) and (Desbois, Butault and Surry, 2015); however, like these earlier works, it could be extended without disadvantage to sets of cardinality descriptors p = 9 (deciles), or even p = 99 (percentiles) if the analysis objectives required it.

[13] Labeled « Type L1 » in SCLUST.

[14] Labeled « Euclidean » in SCLUST.

Metric L_∞ (distance from Chebyshev)[15] : $\delta_\infty(\delta_i^j, \delta_k^j) = Sup\left\{\left(\underline{x_i^j} - \underline{x_k^j}\right); \left(\overline{x_i^j} - \overline{x_k^j}\right)\right\}$

For each of these metrics M on $\mathbb{R}$, a dissimilarity between empirical distributions based on the differences between estimation intervals of the conditional quantiles can be calculated according to a quadratic criterion: $d(\omega_i, \omega_k) = \left(\sum_{j=1}^{p} \delta_M^2(\delta_i^j, \delta_k^j)\right)^{1/2}$.

The data of a matrix of dissimilarities between national empirical distributions of specific costs makes it possible to directly apply the classical methods of automatic classification based on dissimilarities: minimum ultrametric methods (*single linkage*), maximum ultrametric (*complete linkage*), centroid method , and Ward's method.

Several automatic classification procedures have been developed for interval data. Among the first procedures, (Chavent, 1998) proposes a divisive hierarchical classification procedure (DIV procedure) on interval data.

III Data collection and distributional analysis of specific agricultural costs in the EU

III.1 European RICA, the model, the aggregates and the countries studied

Since its establishment in 1965[16], the European RICA has been defined by European regulations specifying the implementation modalities and their revisions, the most recent being the EC Regulation n ° 1217/2009 published in JOE L328 of 15/12/2009 for a entered into force on 01/04/2010. Together with the Census of Agriculture and the Structural Surveys, it completes the tripod of the Community acquis on agricultural statistics, which makes it possible to define the population of agricultural holdings, to follow the evolution of their productive structures, and finally to evaluate variations in their income. Focused from the outset on monitoring the income of so-called "professional" farmers and analyzing the economic functioning of their farms, it has gradually established itself as a vital database for ex ante and ex-post analysis of the impact of agricultural policy measures, in particular those related to the reforms of the Common Agricultural Policy (CAP). As underlined (Chantry, 1998 and 2003), the European FADN is the result of a process of adaptation and harmonization of pre-existing national arrangements within the Member States. European FADN as a database is fed by national FADN which despite the harmonization of accounting and technical-economic concepts carried out [17] under the auspices of the Directorate-General for Agriculture (DG Agri), presents a certain number of specificities relating mainly to the selection of the sample (sampling plan, selection method, economic size thresholds) and the conduct of the survey. For each Member State, the data for each holding ("record") collected at European level ("Community record") are derived from data collected at national level ("national record"). For some Member States such as Belgium and the Netherlands, the national FADN survey questionnaire collects more information than the European FADN questionnaire. Conversely, for other Member States, the "European file" incorporates as missing data the possibilities of exemption provided for by the Regulation due to limitations or constraints related to national FADN. However, as regards the accounting aggregates used in our work (gross products and specific expenses), the definitions are harmonized in both plant and animal production; the elements of differentiation that can influence estimates via weighting are mainly in the reference

[15] Labeled « Hausdorff » in SCLUST.
[16] European regulation n°79/65/CEE, of 15 June 1965.
[17] by the unit L3 in charge of the relations with the national operators of the FADN.

population of farms defined as professional (economic size thresholds) and in the sampling methodology (random selection versus quota selection).

Equation (5) is the basic model for estimating conditional quantiles of the direct costs specific to the products studied, including the dairy milk, our product of interest. Endogenous variable of the empirical modelling noted X, the specific costs[18] are defined as the sum:

- Crop-specific inputs, i.e. items of expenditure on seeds and seedlings, fertilizers and amendments, plant protection products, and other crop-specific costs ;
- Livestock-specific inputs, which include herbivore, granivore, and other animal-specific expense items ;
- inputs specific to forestry activities.

Exogenous variables of the empirical modelling, for each of the speculations implemented by the multi-product farm, the raw products [19] (denoted Y) relate to all plant, animal and animal products, or even forest products, where appropriate, with the following breakdown into fifteen aggregates: wheat, other cereals, industrial crops, protein crops, oilseeds, horticultural productions, fruit, wine, other vegetable or forest products, cattle, swine, poultry, dairy milk, other animal products, other raw products.

The sub-populations of farms selected as the basis of estimation are those corresponding to the following European FADN samples: for 2006, the following twelve Member States were selected: Austria, Belgium, Denmark, France, Germany, Hungary , Italy, Netherlands, Poland, United Kingdom and Sweden, together noted EU 12.

The weighted conditional quantile estimation is carried out using the SAS software, by the QUANTREG procedure associated with the WEIGHT instruction, for each of the countries but also for each of the dimension classes.

III.2 Distributional characteristics of specific agricultural costs

According to (Angrist & Pischke, 2009), "*For better or worse, 95% of estimates in econometrics are provided by averages*" however "*applied economists want more and more to know what's going on, not just on average , but for the whole distribution, the losers as the winners*“. Thus, in many evaluative and prospective studies, it is often useful to be able to compare results across a large number of sub-populations, to reflect the heterogeneity of the populations studied, and to be able to propose more realistic adjustments.

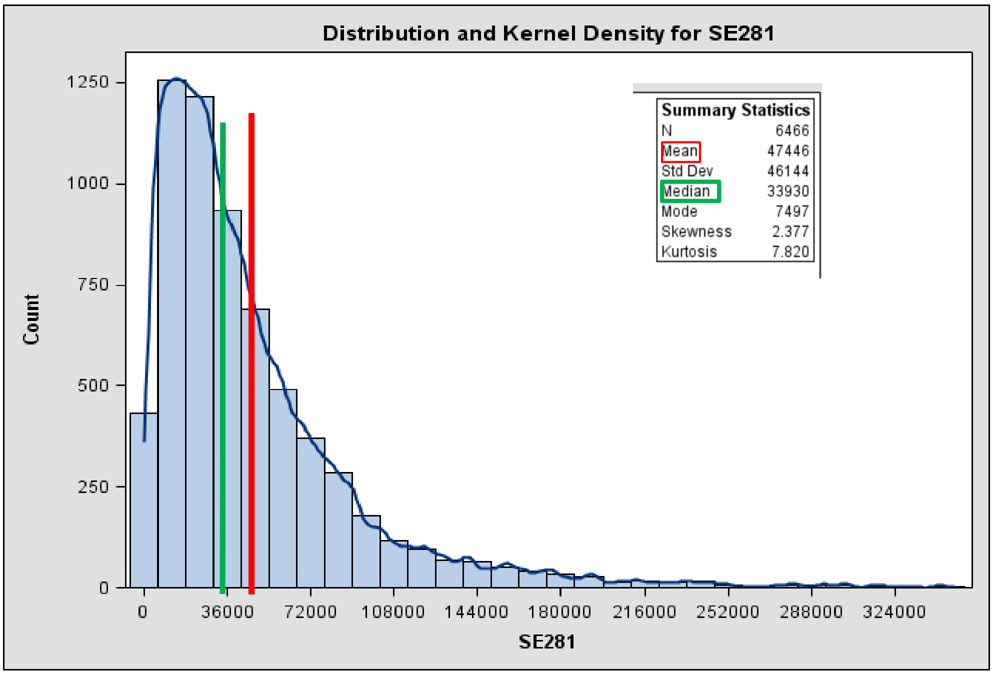
The non-parametric estimate of the density of specific costs by the kernel method highlights the asymmetry (2,377), indicated by the difference between the median (€ 33,930) and the average (€ 47,446) eccentric by the weight of extrema by higher values (see Figure 1, below representing the empirical distribution of French FADN for the calendar year 2006). This asymmetry, in addition to the

[18] The specific costs are recorded by the European FADN under the variable label SE281.

[19] The gross product is defined, with the variations of stock, as the total gross production from which the total of the intra-consumptions is subtracted. .

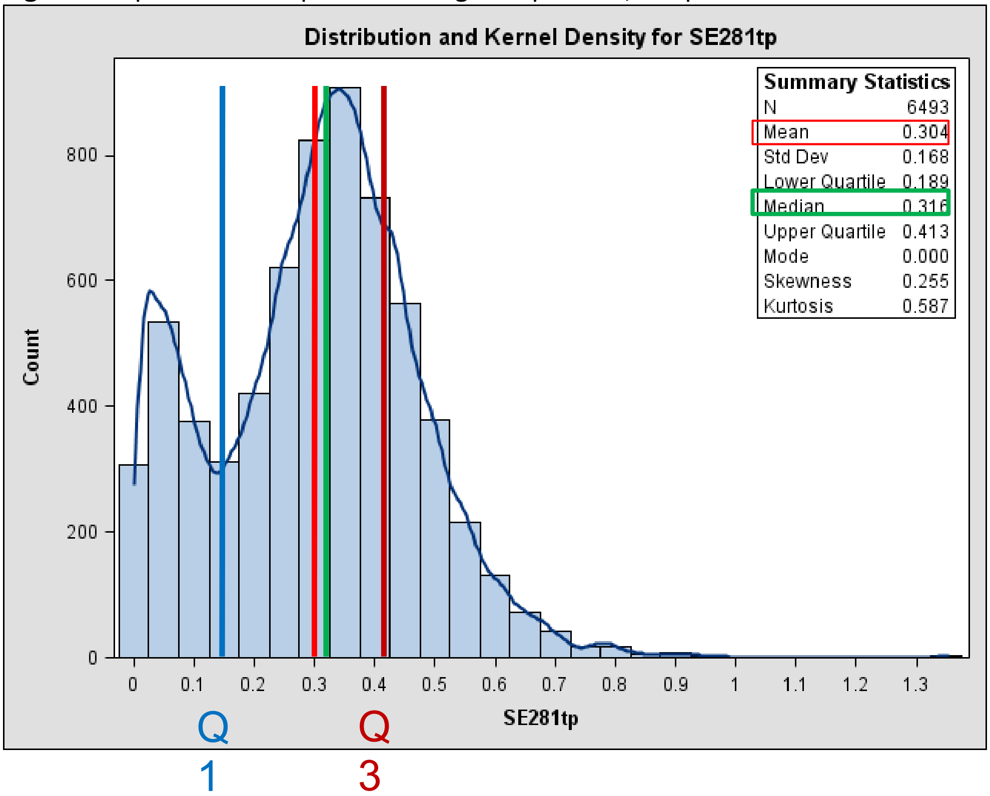
dispersion of cultivated areas, reveals the underlying heterogeneity in the mobilization of specific factors of production..

Figure 1: Specific costs, empirical distribution of French FADN, 2006.



Source: author's processing, from French FADN 2006.

Figure 2: specific costs per unit of gross product, empirical distribution of French FADN, 2006.



Source: author's processing, from French FADN 2006.

For such asymmetric distributions, it is well known that the median is a better estimator of central tendency than the arithmetic mean. However, very often specific cost distributions are not unimodal, as is the case for example for the distribution of specific costs per unit of gross product (see Figure 2) that we define as **specific costs**. Other values may then be needed to better characterize the form of the empirical distribution of specific costs, such as those of the lower quartile (Q1 = 0.189) and higher quartile (Q3 = 0.413) providing more precise information on the empirical distribution of specific costs. than that given by the single estimate provided by the average.

Tableau 1 : national distributions of specific costs per farm, EU 12.

Country	Sample	Mean	CoV	Skewness	Kurtosis	D1	Q1	Medain	Q3	D9	IRD*
Austria	1 790	16 870	139%	6,5	86,7	3 500	5 840	10 430	19 700	37 350	133%
Belgium	1 040	74 150	134%	4,5	37,1	11 270	21 980	43 660	90 370	166 150	157%
Denmark	1 690	112 200	241%	3,4	23,1	4 670	10 810	34 620	155 180	314 640	417%
France	6 510	39 310	160%	5,9	63,1	5 620	12 290	24 910	47 500	83 000	141%
Germany	6 750	63 420	261%	6,6	67,1	11 080	19 590	38 170	75 730	137 550	147%
Hungary	1 690	14 850	1023%	7,5	85,7	1 070	1 880	4 350	10 240	25 460	192%
Italia	13 200	12 180	939%	14,4	314,5	700	1 320	2 670	7 200	20 860	220%
Netherlands	1 340	124 330	218%	3,4	17,1	9 870	25 350	56 100	138 300	294 040	201%
Poland	11 000	7 010	383%	10,6	209,5	1 470	2 220	3 660	7 180	14 300	136%
United-Kingdom	2 590	82 620	210%	7,9	97,8	14 300	23 090	44 220	93 150	177 050	158%
Sweden	850	53 970	187%	8,5	111,6	7 300	15 840	28 850	67 030	122 760	177%
United-Kingdom	2 590	82 620	210%	7,9	97,8	14 300	23 090	44 220	93 150	177 050	158%
Total	56 180	22 250	570%	9,0	135,9	1 010	2 010	5 050	18 070	51 490	318%

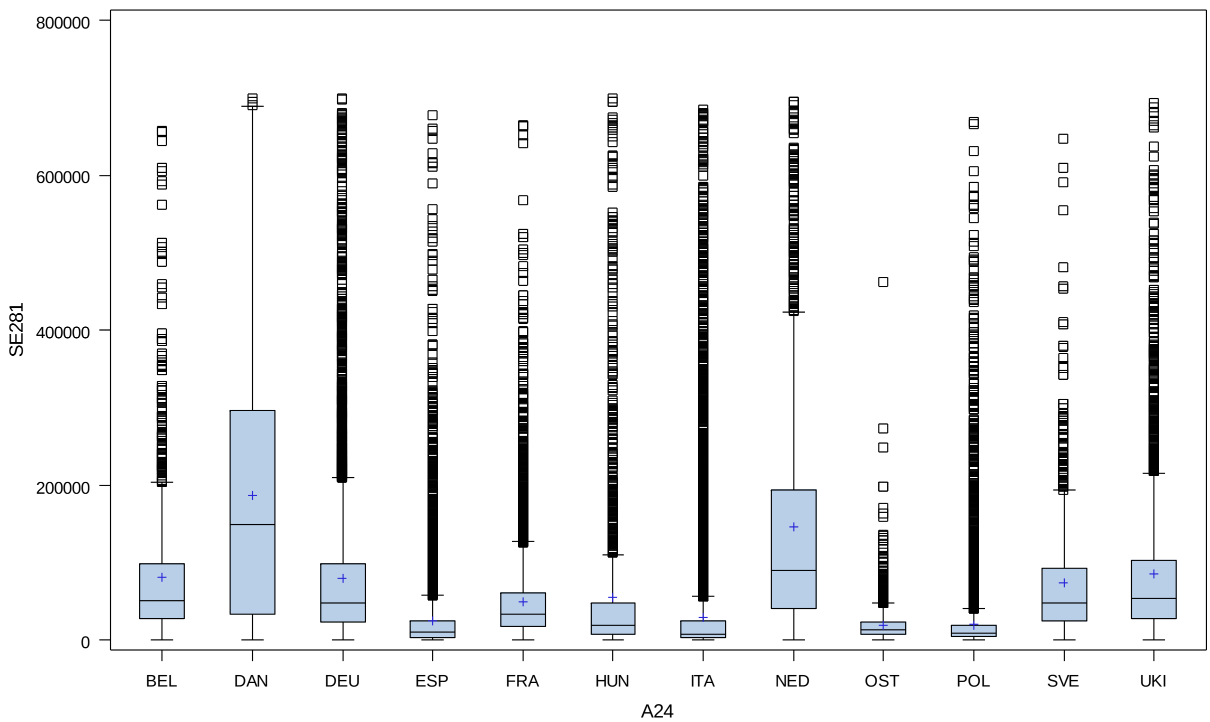
Source: author's processing, from EU-FADN 2006; (*) interquartile ratio of dispersion.

First of all, if we compare the countries of the EU12 set in terms of central tendency, the median varies from € 2,800 (Spain) to € 10,500 (Austria) for a first group also comprising Poland and Portugal. Italy; then, we distinguish an intermediate group with France and Sweden between € 25,000 and € 29,000, respectively; finally, the third group comprising the other countries with median costs per farm varies from € 34,500 for Denmark to € 56,000 for the Netherlands, and also including Germany, Belgium and the United Kingdom.

Secondly, if we compare these EU12 countries in terms of dispersion, the coefficient of variation (CoV)[20] is between 134% for Belgium, a country with the most homogeneous production structures (followed by Austria, France and Sweden) and 1023% for Hungary which appears with Italy as the most heterogeneous country.

Although these distributions are all asymmetrical by positive values with a large number of extreme values, they differ however in their form: more asymmetric by positive values (greater dispersion of the values above the median) as in Italy, Spain, Poland or Hungary, with a skewness between 5 and 6; or less asymmetrical as in Denmark, Austria, the Netherlands, Belgium or France (with a skewness of 1.5 to 2.5), with the United Kingdom having an intermediate skewness of a factor of 4. In addition, the kurtosis varies from minima to 17 for the Netherlands, or 23 for Denmark or 37 for Belgium, to a maximum of 210 to 315 for Poland, Spain or Italy.

Figure 3 : distribution of the specific costs of the operation (SE 281 <€ 750,000) by country, EU12.



Source: author's processing, from EU-FADN 2006.

[20] Expressed as a percentage, the coefficient of variation reports the value of the standard deviation to the mean: $CoV = \hat{\sigma}/\hat{\mu}$

For very asymmetric distributions with extreme values, the interquartile ratio of dispersion (IRD)[21] is preferable to the CV for measuring relative dispersion: Austria, France and Germany (between 130 and 150%) have the lowest relative dispersion, while Denmark has the highest relative dispersion of specific costs (420%). By ignoring the extreme values of inherited farms from increasingly marginal collective structures in its production, the IRD reduces the relative dispersion of Hungary to that of Italy.

Thus, the lowest levels and dispersions per farm are found in southern and eastern European countries, with more asymmetrical and flattened distributions, while the highest levels and dispersions are observed in countries in the North and West of Europe, with less asymmetrical and more concentrated distributions than the previous ones. Since specific cost distributions have many extreme values, it is more appropriate to use quartiles to locate the scale of specific costs, as well as the interquartile range or interquartile ratio of dispersion to measure dispersion rather than the mean, standard deviation and coefficient of variation are weight-sensitive to these extreme values.
The ratio of the specific costs to the raw products makes it possible to analyze the productivity of the inputs and to compare it with that of the other factors of production. Therefore, it is interesting to be able to describe by structural type the structural differences from the point of view of the specific costs between countries where the production is located: this angle of analysis is therefore developed in the rest of the presentation.

IV Econometric results: national estimates of specific costs

As we have shown in the methodological section, the estimation according to the conditional quantiles makes it possible to carry out a conditional allocation of the specific costs by products, allowing the comparison of the different workshops within the framework of a multi-product exploitation based on gross margin, its complement to the gross product. We use this conditional allocation to provide specific cost estimates to answer farm competitiveness measurement questions, which are posed by ex-ante or ex-post design and evaluation of different agricultural policy options. In the framework of the FACEPA research project, the choice chosen by the managers in charge of the Knowledge Based Bio-Economy program (KBBE) of the 7th PCRD was made for feasibility reasons on the three main agricultural commodities that are wheat, milk and pig, produced at a level sufficiently broad at the European level to allow cross-country comparisons. Quantile estimates are therefore made for each of the EU12 Member States in order to test the national differentiation of the productive framework at European level. we have chosen to analyze the estimates obtained for the year 2006[22], in order to compare the results of the conditional quantile approach later with those of the Seemingly Unrelated Regressions Equations (SURE) approach. Initially proposed by (Zellner, 1962), the latter approach is the standard procedure for estimating the GECOM model of the FACEPA project.

Thus, we analyze the results obtained in particular for the pig, one of the conveniences selected in the framework of the FACEPA project[23].

---

[21] As a ratio of interquartile dispersion at the median level, the quartile dispersion coefficient $IRD = (Q_3 - Q_1)/Q_2$ provides a non-parametric measure of relative dispersion.

[22] The analysis over the entire period is the subject of work in progress to adapt the quantile estimates approach to a panel data structure.

[23] The results obtained simultaneously on the other productions are the subject of analyzes in progress, conducted in parallel.

IV.1 Pig, comparative analysis of raw products between twelve EU countries

In 2010, according to EuroStat estimates[24], the EU-27 accounts for 24.4% of world pig production. The European Union is behind China the second largest producer in the world with 26 million tonnes in 2010. The number of pigs slaughtered in 2010 was 302.6 million head, or 23% of the herd. The countries studied are among the main producing countries in terms of tonnes of carcasses produced, in descending order: Germany (21.58%), Spain (13.33%), France (8.29%), Poland (7.44%), Denmark (6.61%), Italy (6.15%), the Netherlands (5.18%), Belgium (4.26%), the United Kingdom United Kingdom (2.91%), Austria (2.09%), Hungary (1.67%) and Sweden (0.98%), or 80.5% of European production.

Even if the correlation with national statistics is less good[25], the hierarchy of raw products observed within the European FADN (table 2) remains in line with the hierarchy of national statistics on pig production[26], ranking differences exist for the Netherlands (overestimation of 6%), France (overestimation of 4%) and Spain (underestimation of 3%).

Table 2: pig, distribution of gross product by country, EU12.

Country	Population	D1	Q1	Median	Q3	D9	Mean
Austria	29 040	64	126	506	33 031	81 340	24 698
Belgium	5 140	19 243	85 255	184 581	319 312	512 674	228 803
Denmark	6 950	5 544	65 231	213 944	506 288	902 508	354 530
France	13 370	1 620	41 845	123 884	271 555	450 627	194 908
Germany	52 980	1 341	11 946	51 488	135 958	230 833	91 407
Hungary	19 330	761	1 318	2 483	6 843	16 751	16 008
Italia	17 310	400	650	1 530	6 897	132 600	95 450
Netherlands	6 530	45 457	101 343	251 186	460 617	751 500	359 503
Poland	422 190	402	980	2 100	4 624	10 928	5 275
United-Kingdon	2 980	2 739	23 329	133 288	310 707	658 173	228 966
Sweden	3 100	3 209	13 388	54 656	173 638	319 639	124 786
United-Kingdon	2 980	2 739	23 329	133 288	310 707	658 173	228 966
Total	56 180	606 610	368	1 088	10 765	77 316	35 625

Source: author's processing, from EU-FADN 2006.

IV .2 Factor Analysis of Estimated Range Distributions

Table 3 presents the main estimates of conditional quantiles (lower decile D1, lower quartile Q1, median Q2, upper quartile Q3, upper decile D9) for pig, derived from quantile regression and ordinary least squares regression (OLS) for the specific costs of agricultural production (accounting aggregate SE281 of the European FADN) from a breakdown of the gross product into fifteen aggregates (cf. III.1), for the subset of 12 European countries selected in 2006. Among the results that may be encountered in estimates of conditional specific cost quantiles for pig (table 3), the estimated gross product shares for pig from the standard FACEPA model, cf. table A3.3 in (Kleinhanss, Offerman, Butault and Surry,

[24] According to Focus on the Common Agricultural Policy, Eurostat 2012.

[25] Coefficient of correlation: r = 0.92.

[26] Coefficient of correlation r = 0.98.

2011), show consistent ranking: in fact, in 11 EU Member States[27], the first (Q1) and second (Q2) conditional quartile estimators are significantly correlated with the linearly constrained estimator of FACEPA (the levels are close to the level reached by the OLS estimator[28]).

Table 3: pig, specific costs for 1,000 € of gross product, EU12.

Pig	D1 [Min ; Max]	Q1 [Min ; Max]	Q2 [Min ; Max]	Q3 [Min ; Max]	D9 [Min ; Max]	OLS [Min ; Max]
Austria	[347.2 ; 369.2]	[397.3 ; 409.1]	[425.6 ; 447.4]	[463.5 ; 485.3]	[523.1 ; 562.5]	[433.7 ; 442.1]
Belgium	[539.9 ; 566.1]	[561.2 ; 579.8]	[591.5 ; 608.7]	[642.7 ; 674.7]	[684.6 ; 707.4]	[630.9 ; 641.8]
Denmark	[445.2 ; 458.6]	[503.0 ; 515.4]	[558.9 ; 570.9]	[617.7 ;632.1]	[654.7 ; 671.3]	[535.2 ; 542.6]
France	[470.9 ; 493.5]	[509.9 ; 527.7]	[547.9 ; 561.7]	[577.6 ; 594.2]	[610.8 ; 644.6]	[541.6 ; 547.4]
Germany	[444.3 ; 454.7]	[475.6 ; 485.4]	[514.8 ; 526.4]	[567.7 ; 582.7]	[593.4 ; 618.8]	[493.6 ; 502.7]
Hungary	[369.0 ; 451.8]	[459.3 ; 568.1]	[589.2 ; 662.2]	[633.3 ; 681.5]	[647.8 ; 737.0]	[605.1 ; 620.7]
Italy	[116.5 ; 170.1]	[162.2 ; 245.2]	[325.1 ; 386.5]	[559.6 ; 633.0]	[627.9 ; 718.3]	[300.7 ; 307.8]
Netherlands	[487.4 ; 506.2]	[528.2 ; 550.4]	[584.6 ;602.4]	[639.9 ; 661.5]	[676.0 ; 721.6]	[573.2 ; 595.1]
Poland	[471.3 ; 483.3]	[541.8 ; 552.2]	[603.0 ; 618.0]	[655.1 ; 674.7]	[704.5 ; 727.3]	[641.7 ; 648.1]
Spain	[191.5 ; 285.3]	[369.8 ; 441.6]	[552.3 ; 638.5]	[743.8 ; 802.2]	[824.6 ; 893.4]	[449.9 ; 456.7]
Sweden	[396.3 ; 443.9]	[507.2 ; 533.1]	[533.1 ; 578.1]	[547.5 ; 619.5]	[641.7 ; 722.9]	[528.1 ; 543.2]
United-Kingdom	[376.8 ; 559.2]	[548.4 ; 596.0]	[599.2 ; 629.6]	[641.3 ; 712.9]	[723.2 ; 805.4]	[565.7 ; 588.4]

Source: author's processing, from EU-FADN 2006.

The visualization of the specific cost estimates is done according to the graph in Figure 3, showing the conditional quantile estimates in ascending order for each country. For 2006, this graph of the point estimates of conditional quantiles of specific pig costs by country identifies four types of distributional scales: first type, Italy (ITA) and Spain (ESP) having a similar form (high inter-quantile growth with an inter-decile difference D1-D9 greater than 400 €) despite distinct locations (the minimum of the differentials between respective quantiles is greater than 200 €); second type, Austria (OST) opposes the previous model with an inter-decile gap D1-D9 of 100 €; third type, the United Kingdom presenting a distributional scale with significant inter-quantile growth (inter-decile gap D1-D9 greater than 300 €); and fourth, a subset of countries with moderate inter-quantile growth (inter-decile range of between € 200 and € 300). The conditional median (Q2) estimation levels are also a second criterion for distinguishing between these different distributional scales with two subsets: on the one hand, Italy (ITA) and Austria (OST) on the median estimates. less than 450 €; on the other hand, all the other countries whose median conditional estimates are between € 500 and € 600.

Among the differences that can be identified in 2006, let us first note the significant difference between two similar distributional scales with heterogeneous slopes, Spain (ES)[29] and Italy (IT)[30], figure 3 confirming this separation of distributional scales for all conditional quantiles; this is an illustration of the linear model of conditional quantile with heterogeneous slope (cf. above § II.2.ii). Less easy to identify, we secondly note the absence of overlapping distributional scales of Belgium (BE), Denmark (DA) and Austria (OS) whose separation of confidence intervals can be seen on the figure 3. Apart from certain differences in precision for the estimation of the upper conditional decile (D9) between Belgium

---

[27] Austria was not included in the FACEPA report.

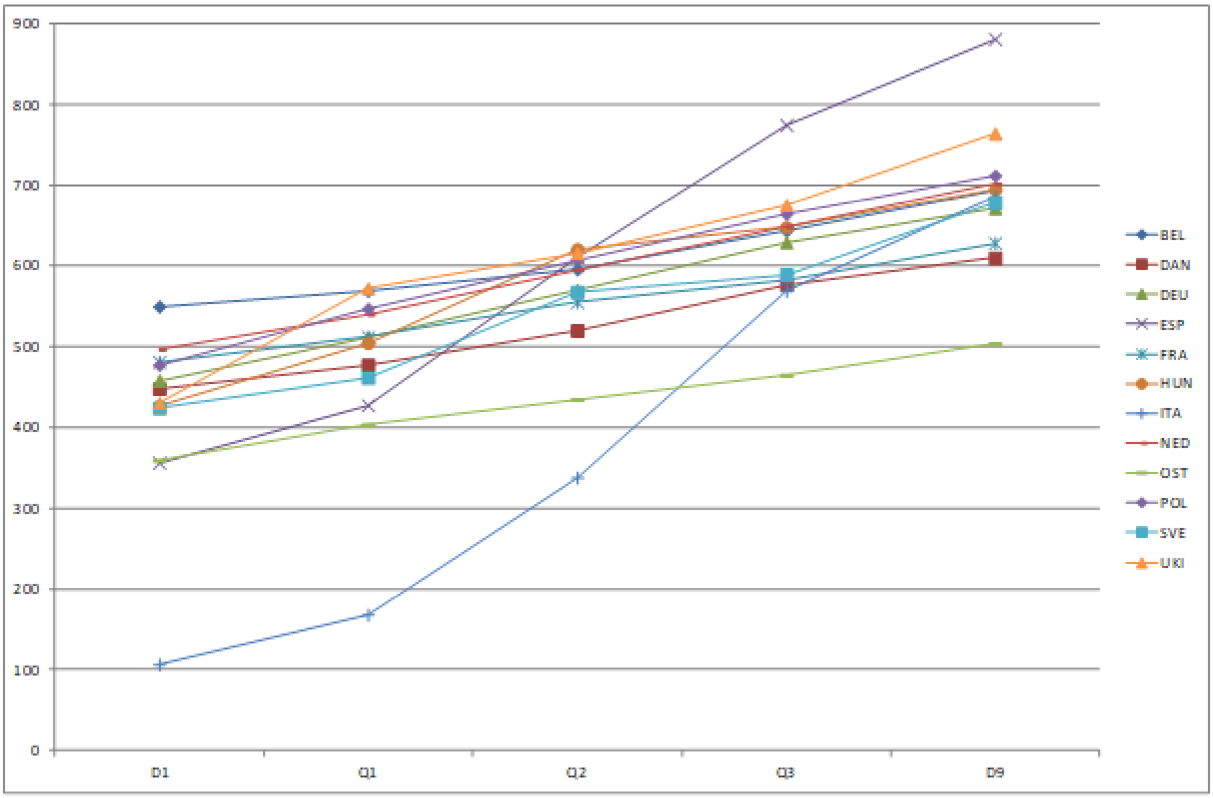
[28] The rank correlation levels of Spearman are increasing from $corr(SURE, D1) = 0{,}62$ to $corr(SURE, D9) = 0{,}69$, comparable to $corr(SURE, MCO) = 0{,}72$.

[29] For which, the differences between extreme conditional quantiles overlap those between *Comunitat Valenciana* (PDO *Jamon de Teruel*) at the highest costs and *Extremadura* at the lowest costs.

[30] Whose highest quantile estimates correspond to those recorded in *Emilia-Romagna* (PDO *Prosciutto di Parma*) or *Venetto* (PDO *Veneto Berico-Eugeano*), which oppose the lowest quartile and decile estimates in *Lombardia*.

and Austria on the one hand, and Denmark on the other hand; this illustrates the linear model of conditional quantile with homogeneous slope (cf. above § II.2.i).

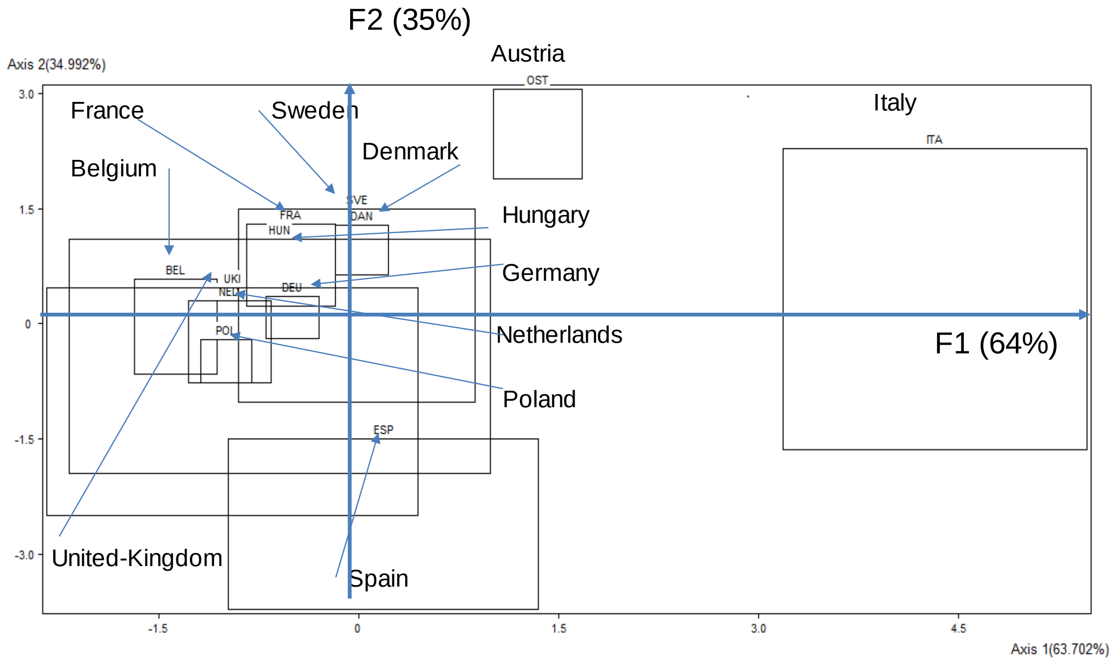
Figure 4: pig, estimation of conditional quantiles for 12 EU member states (2006).



Source: author's processing, from EU-FADN 2006.

Figure 5 : pig, quantile estimation interval SO-PCA, factorial plane F1xF2 of EU12 countries (2006).

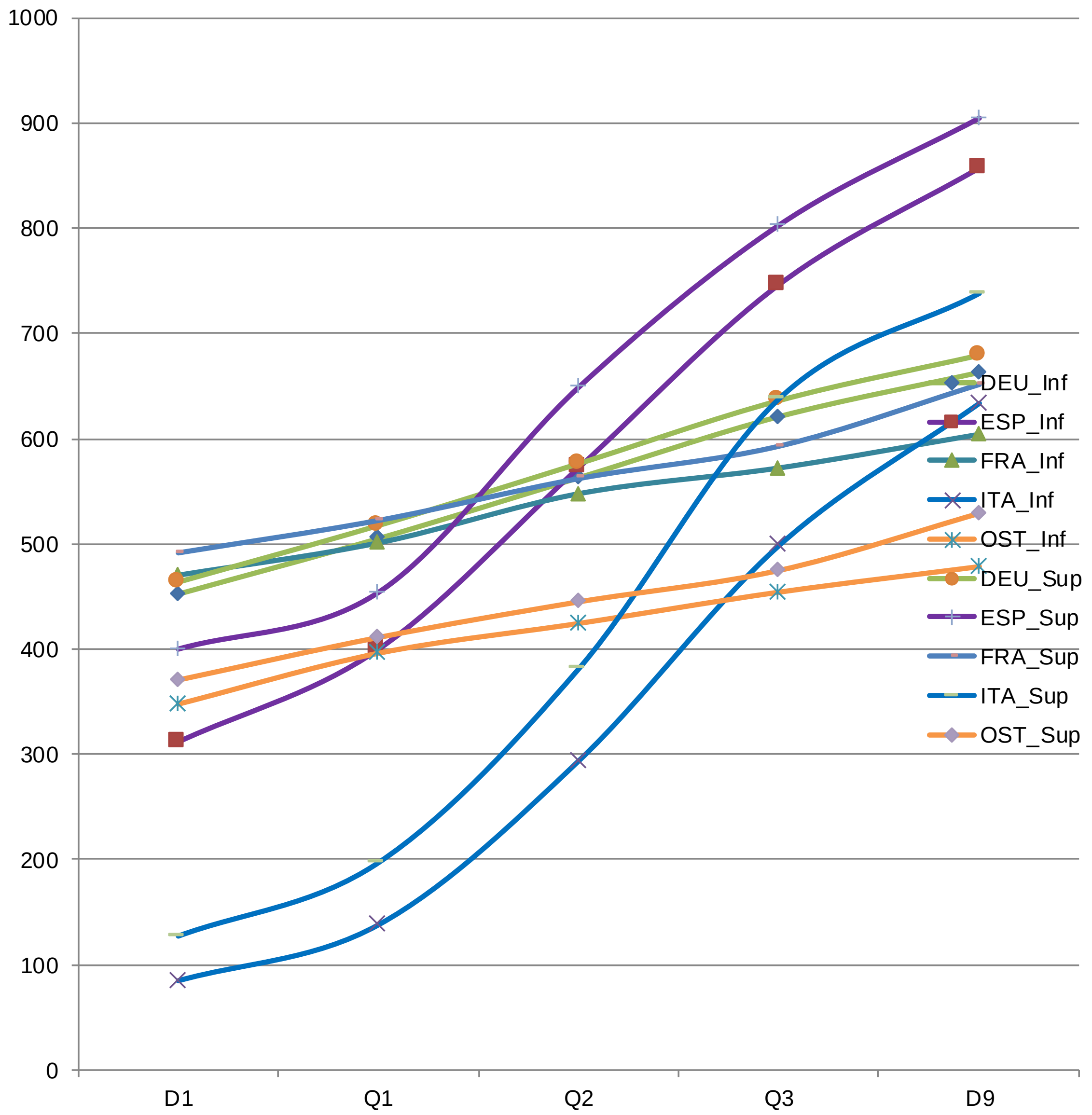
SPCA - SOs Interval Coordinates



Source: author's processing, from EU-FADN 2006.

The SO-PCA interval estimates of conditional quantiles allow to specify this distributional structure. The first principal component F1 (Axis 1) representing 64% of the inertia, is negatively correlated with the first conditional quantiles (decile D1 and quartile Q1 highly correlated). The second principal component F2 (Axis 2), representing 35% of the inertia, is positively correlated with the upper decile (D9) and the third quartile (Q3). The median Q2 is also correlated with the first two major components. the ACPEIQ F1xF2 first factorial factor, representing 99% of the country variability, makes it possible to identify two distinct groups of countries differentiating according to the level of the conditional estimate of the first quantiles (D1 and Q1): on the other hand, in F1> 0, Italy with the first quantiles lower than 205 € and, on the other hand in F1 <0, all the other countries for which the first quantiles are greater than 235 €. The second main component makes it possible to distinguish three groups: on the one hand, Austria in F2> 0 to the most homogeneous quantile estimates situated between € 350 and € 430; on the other hand, Spain with the highest estimates (from € 770 for Q3 to € 860 for D9); and all other countries in the quadrant (F1 <0, F2 <0) or close to it.

Figure 6: pig, location shift model (FRA-DEU/OST) versus location-scale shift model (ITA/ESP).



Source: author's processing, from EU-FADN 2006.

Thus, does the SO-PCA identify Austria's cost homogeneity model and distinguish two models of cost heterogeneity, one for lower cost quantiles (d1 and Q1) by Italy, the second by the higher costs (D9 and Q3) represented by Spain. Finally, taking into account additional parameters could make it possible to better separate two putative subgroups: on the one hand, Denmark, France, Hungary and Sweden; on the other hand, Germany, Belgium, Poland and the United Kingdom.

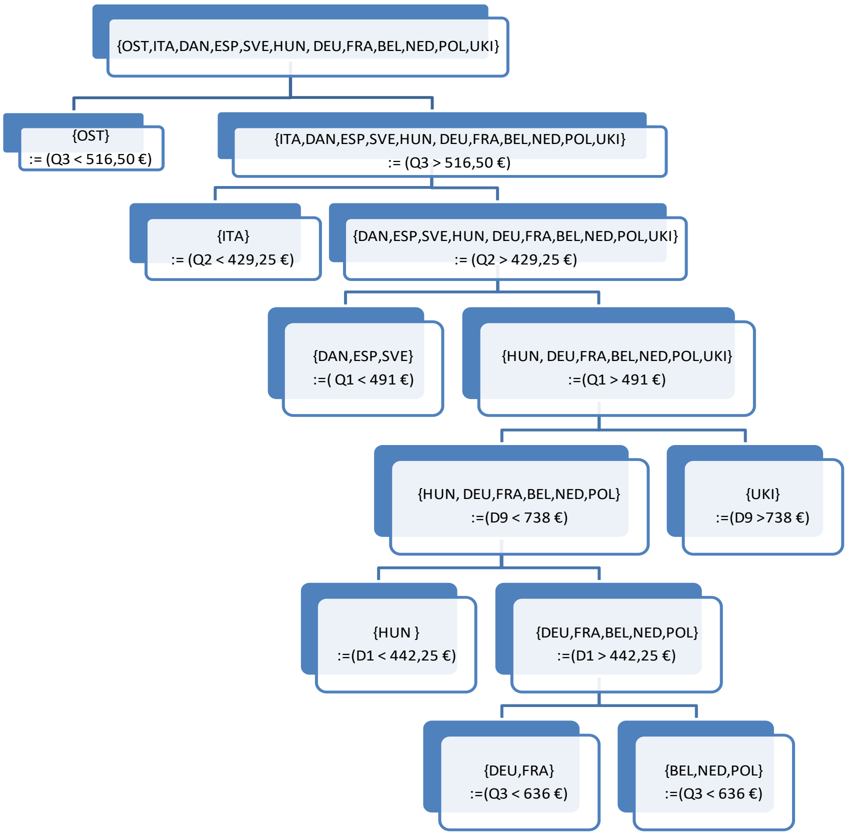
Hierarchical descending clustering (DIV)[31] allows the cost structure to be specified by country class (figure 7). First, there is a major distinction in the location of distributional scales: Austria (OST) is separated from other countries by an upper quartile estimate Q3 <516.50 €; Italy stands out with a median estimate Q2 <429.75 €; Spain, Denmark and Sweden are distinguished by an estimate of the first quartile Q1 <€ 491; the United Kingdom is characterized by a higher decile estimate .is less than this value; on the other hand, among these other countries, a supplementary distinction must be made between those whose estimate of the last decile (D9) exceeds € 738; Hungary is distinguished by an

[31] Unsupervised classification algorithm on the MBMC confidence interval table at 95% quantile estimates (SODAS 2.5 software).

estimate of the first decile (D1) less than 442.25 €. The other countries are divided into two sub-groups: on the one hand, France and Germany are characterized by an estimate of the upper quartile (Q3) of less than € 636; on the other hand, Belgium and the Netherlands, and Poland, which are distinguished by a quartile estimate (Q3) of over € 636.

This descending hierarchy shows that the set of quantile estimates is mobilized by the discriminant values, which implies keeping all the parameters describing the distribution, and possibly extending it by a finer quantile scale allowing some of the national distributions to be better distinguished.

Figure 7: pig, specific costs for € 1,000 gross product, country classification, EU12.



Source: author's processing, according to EU-FADN 2006.

V Discussion of results

The heterogeneity of national distributions of specific costs covers the combined effect of different dispersal factors, including the economic dimension of farms that should be analyzed. In fact, the European countries studied have neither the same composition in terms of the economic dimension of the farms, nor the same thresholds to define a professional exploitation. Thus, the heterogeneity of the quantile estimates of specific costs within national distributions, either in Italy either in Spain, probably covers those of very different production structures both in their economic dimension and in the production technology used.

Regionalised estimates make it possible to specify national situations that are not all homogeneous: the Spanish region *Communitat Valencian*a is distinguished by the maximum estimate of the median quantile of specific costs; in contrast, the Spanish region of *Extremadura* and the Italian region of *Venetto* are distinguished by the lowest overall levels of quantile estimates, especially for the median quantile; the Spanish region *Andalucia* and the Italian region *Emilia-Romagna* are characterized by higher quantile estimates (Q3 and D9); the central Swedish region *Skogs-och mellanbygdslan*, the central Hungarian region *Kozép-Magyarorszàg*, the French region *Basse-Normandie*, and the German region *Sachsen-Anhalt* are associated with lower quantile estimates (D1 and Q1) among the highest.

It cannot be ruled out that the high values of some estimates may in some cases come from artifacts related to the estimation methodology for countries where pig production correlates with other production facilities on the farm. Indeed, the size of the pig workshop compared to the other workshops, according to the more or less pronounced productive specialization of the farms, can produce artifacts resulting from the productive correlations at the level of mixed technical-economic orientations to the extent that the weight of the costs Specifically related to the other workshops would lead, depending on the hierarchy of specific costs, either to an underestimation bias for minority pig workshops compared to other products with a smaller production detour or conversely to an overestimation bias for productions presenting with the detour of more important production such as pig production.

However, the existence of very high specific costs may also signal the maintenance of technically less efficient producers in less favorable areas because of the existence of comprehensive income support measures (Barkaoui, Daniel and Butault, 2009), or even agri-environmental measures specific to certain productive contexts, in particular those aimed at maintaining agricultural production in certain territories. On the other hand, the lower estimates can point to either the presence of intensive farms that perform better technically, such as for pig producers in western France, or the presence of productive systems based on less demanding input and output techniques as in piedmont and mountain areas.

VI Conclusions

On the basis of European FADN, we have tested the feasibility of the micro-econometric estimation methodology of the specific production costs according to the conditional quantiles, and we have illustrated its relevance to take into account the intrinsic heteroscedasticity of these distributions fornone of the major commodities of the European market, the pig. The lessons learned from these analyzes are relatively consistent for the pig: the lower quantiles (D1 and Q1) and, respectively, the higher quantiles (Q3 and D9) are the specific cost parameters that can differentiate national productions according to their cost distributions based on regional differences observed.

The analysis of these estimates makes it possible to identify types of national distributions of specific costs. The main producing countries are located in a two-dimensional repository based on a principal component analysis of the conditional quantile interval estimates that provides a test of significance for the differences found between national distribution scales according to their respective conditional quantiles. Differences and similarities between countries are exploited using hierarchical top-down classification to produce country classes with comparable costs. The differences between these groups of countries are delimited by thresholds expressed according to the conditional quantiles in terms of the gross product. These thresholds can be mobilized to segment farm populations to analyze the differential effects of agricultural policy measurement. These analyzes therefore make it possible to identify different models of distributional scale, notably that of the location shift one opposite that of the location-scale shift one.

We hypothesize that the differentiation of these national distributions takes place on the one hand between specialized and input-intensive farms and on the other hand mixed or extensive farms in inputs. We also trace the prospects of pursuing and valuing the estimation methodology according to the conditional quantiles in the context of an input-output analysis of European agriculture. The unit estimates given in terms of the share of the specific costs in the gross product that we have privileged in this paper can be used in the context of the calculation of standard gross margins, whether at the normative level to provide a statistical basis for the estimation of but also feed the input-output matrices of the particular agri-food sector to a set of EU countries, or even certain groups of European regions, to implement sensitivity analyzes for possible options. agricultural policy through social and environmental accounts matrices (Léon and Surry, 2009). In the current context of the greening of the Pac, the proposed national typology for the pig could be mobilized to carry out simulations aimed at exploring the relocation of pig production in mountain areas or in intermediate regions.

Références bibliographiques

Angrist, J. et Pischke, J. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton Univ 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., janvier.

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. (2002) *Symbolic Data Analysis: Definitions and Examples*, Technical Report. http://www.stat.uga.edu/faculty/LYNNE/Lynne.html.

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.

Barkaoui A., Daniel K. et 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.

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., et Reignier, E. (1988). *Les coûts de production des principaux produits agricoles dans la CEE*. Luxembourg: Office of Official Publications of the European Communities.

Camaret D., Le Stum H., Murphy M., Petersen A. (1990) Les coûts de production du blé tendre en East-Anglia, au Schlesswig-Holstein et en France. In: *Économie rurale*. n°195, pp. 7-12.

Cameron A. C., Trivedi P. K. (2005). *Microeconometrics. Methods and Applications*. Cambridge, Cambridge University Press.

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. (1997) *Une méthode divisive de classification*, Thèse, Université Paris IX-Dauphine.

Chavent, M. (1998). A monothetic clustering method, *Pattern Recognition Letters*, vol. 19, pp. 989-996.

Chavent M. (2000) Citerion-Based Divisive Clustering for Symbolic Objects, *Analysis of Symbolic Data*, eds. H.H.Bock, E.Diday, Springer, 2000.

Dantzig G.B. (1949) Programming in a linear structure, *Econometrica,* n°17, pp. 73–74.

Diday E. (1971) Une nouvelle méthode en classification automatique et reconnaissance des formes, la méthode des nuées dynamiques, *Revue de statistique appliquée*, tome 19, n°2, pp. 19-33.

Dixon, B. L., M. T. Batte and S. T. Sonka. 1984. Random Coefficients Estimation of Average Total Product Costs For Multiproduct Firms. *Journal of Business Econonomics and Statistics* n°2, pp. 360-366.

Desbois D. (2015). *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., 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., 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.
EU-DGAgri (2009) *Production costs and margins of pig fattening farms – 2008 Report*. European Commission, Directorate General for Agriculture and Rural Development, 59 p.
Givord P. et D'Haultfoeuille (2013) La régression quantile, Document de travail « Méthodologie statistique », M 2013/01, Institut national de la Statistique et des Etudes économiques, 39 p.
Hallam, D., Bailey, A., Jones, Ph., and Errington, A. (1999). Estimating input use and production costs from farm survey panel data. *Journal of Agricultural Economics*, 50, 440-449.
He, X. and Hu, F. (2002), Markov Chain Marginal Bootstrap, *Journal of the American Statistical Association*, 97, 783–795.
Karmakar R. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, n°4, pp. 373-395.
Kleinhanss W, Offerman F., Butault J-P.et 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, p. 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, p. 223-235.
Laisney C., Portet F., Vert J. (2011) Prospective Agriculture énergie 2030. Comment l'agriculture s'adptera-telle aux futures défis énergétiques? *Futuribles*, n°376, pp.99-116.
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. and Nielsen, H. B. (1993), A Finite Smoothing Algorithm for Linear Estimation, *SIAM Journal on Optimization*, n°**3**, pp. 223–235.
OCDE (2010). *Agricultural Policies in OECD Countries: At a Glance 2010*. *Highlights,* Paris, OCDE, 42 p.
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, p. 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.