

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

Dominique DESBOIS

UMR Economie publique, INRA-AgroParisTech, Université Paris Saclay

Session Interval Data
ADVANCES IN DATA SCIENCE

2019, January 10th



CEREMADE LAMSADE

Problem:

How to keep the maximum amount of information for analyzing estimated parameter distributions?

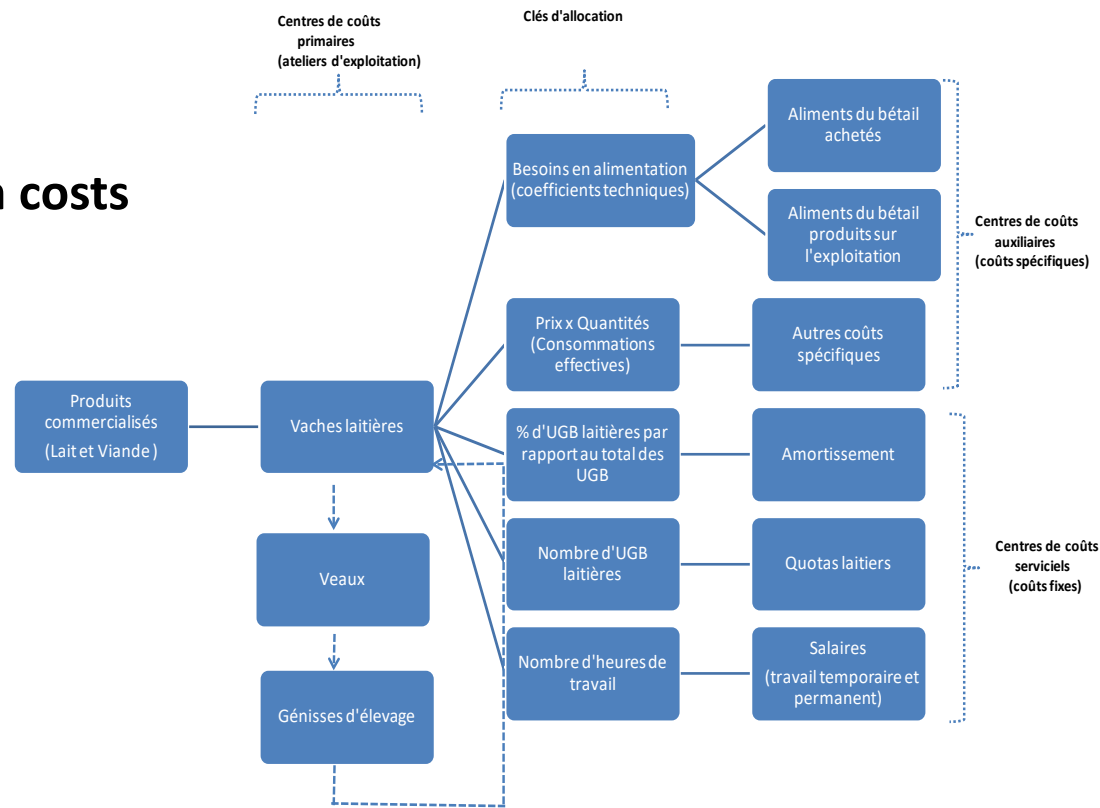
Outline :

- 1. The estimation by intervals of conditional quantiles;**
- 2. Principal component analysis on intervals;**
- 3. Viewing estimated conditional quantiles distributions;**
- 4. Detailed analyses of the results.**

1.1 Evaluation of Agricultural production costs

1. Technical accounting of Production costs

Choice of a method of Technical-Economic Accounting



2. Econometric Modeling of production costs

Choosing a Model with coefficients

$$X_{ih} = \sum_{k=1}^K \alpha_{ih}^k Y_{kh} + \varepsilon_{ih} \text{ with } \varepsilon_{ih} \text{ i.i.d.}$$

CHARGES	PRODUITS					TOTAL CHARGE
	Y_{1h}	...	Y_{kh}	...	Y_{Kh}	
X_{1h}	a_{1h}^1	...	a_{1h}^k	...	a_{1h}^K	$\sum X_{1h}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_{ih}	a_{ih}^1	...	a_{ih}^k	...	a_{ih}^K	$\sum X_{ih}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_{Ih}	a_{Ih}^1	...	a_{Ih}^k	...	a_{Ih}^K	$\sum X_{Ih}$
TOTAL PRODUIT	$\sum Y_{1h}$...	$\sum Y_{kh}$...	$\sum Y_{Kh}$	$\sum_k Y_{kh} = \sum_l X_{lh}$

I.2 Quantitative estimation of specific cost of agricultural production

Problem : In the face of the heterogeneity of agricultural production structures and production behaviours in Europe, how to retain the maximum useful information when estimating production costs ?

Conceptual model : The Input-Output table

Method :

- i) The L1 standard regression allows, by estimating conditional quantiles, to generate a distribution of specific production costs and gross margins ;
- ii) The analysis of symbolic data allows an optimum visualization of the scale and the form of the distributions of estimates by interval.

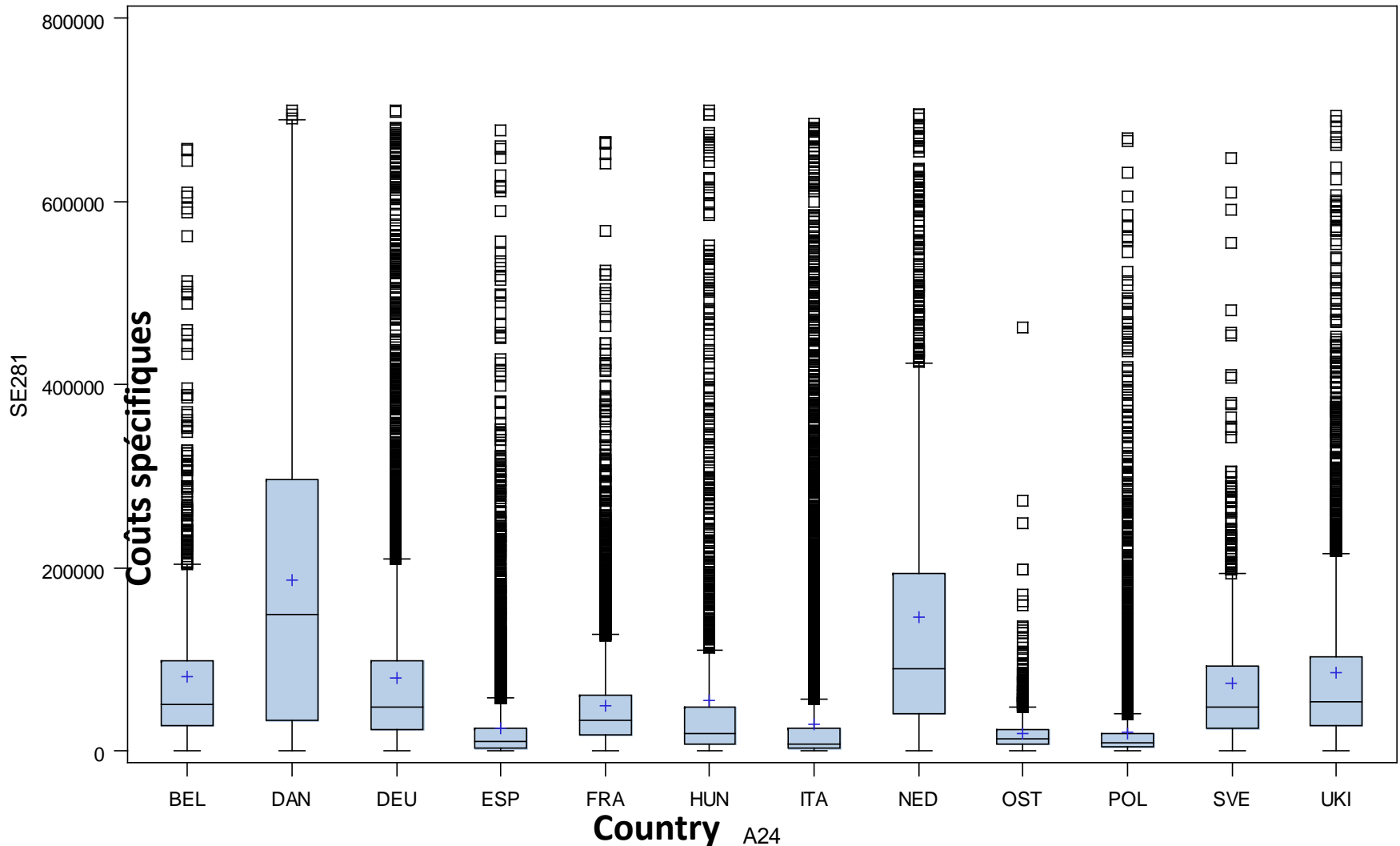
Data : The Farm Accounting Data Network (FADN) in France and Europe

Application :

Estimation of specific costs to deduct gross margin (important for insurance and solidarity funds)

I.3 Distribution of specific inputs of agricultural production in the EU

Distribution of specific inputs (< 750 €0) by country, RICA 2006, Eu12



I.4 Estimation of conditional quantiles.

The estimation of conditional quantiles is obtained by solving a problem of minimizing the following loss function for each quantile Q , according to the mean absolute deviation (MAD): criterion

$$\sum_{i: y_i \geq x_i' \beta} q |x_i - y_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - q) |x_i - y_i' \beta|$$

For a data generator process

$$X = Y' \beta + u \quad \text{as} \quad u = Y' \alpha * \varepsilon$$

Following a linear model with multiplicative heteroscedasticity (i.e. $u = Y * e$), under the assumption $Y' \alpha > 0$

The conditional quantile q^{th} of cost X conditionally to Y is equal to:

$$\mu_{q(Y, \beta, \alpha)} = Y' [\beta + \alpha * F_{\varepsilon}^{-1}(q)]$$

So linear in Y .

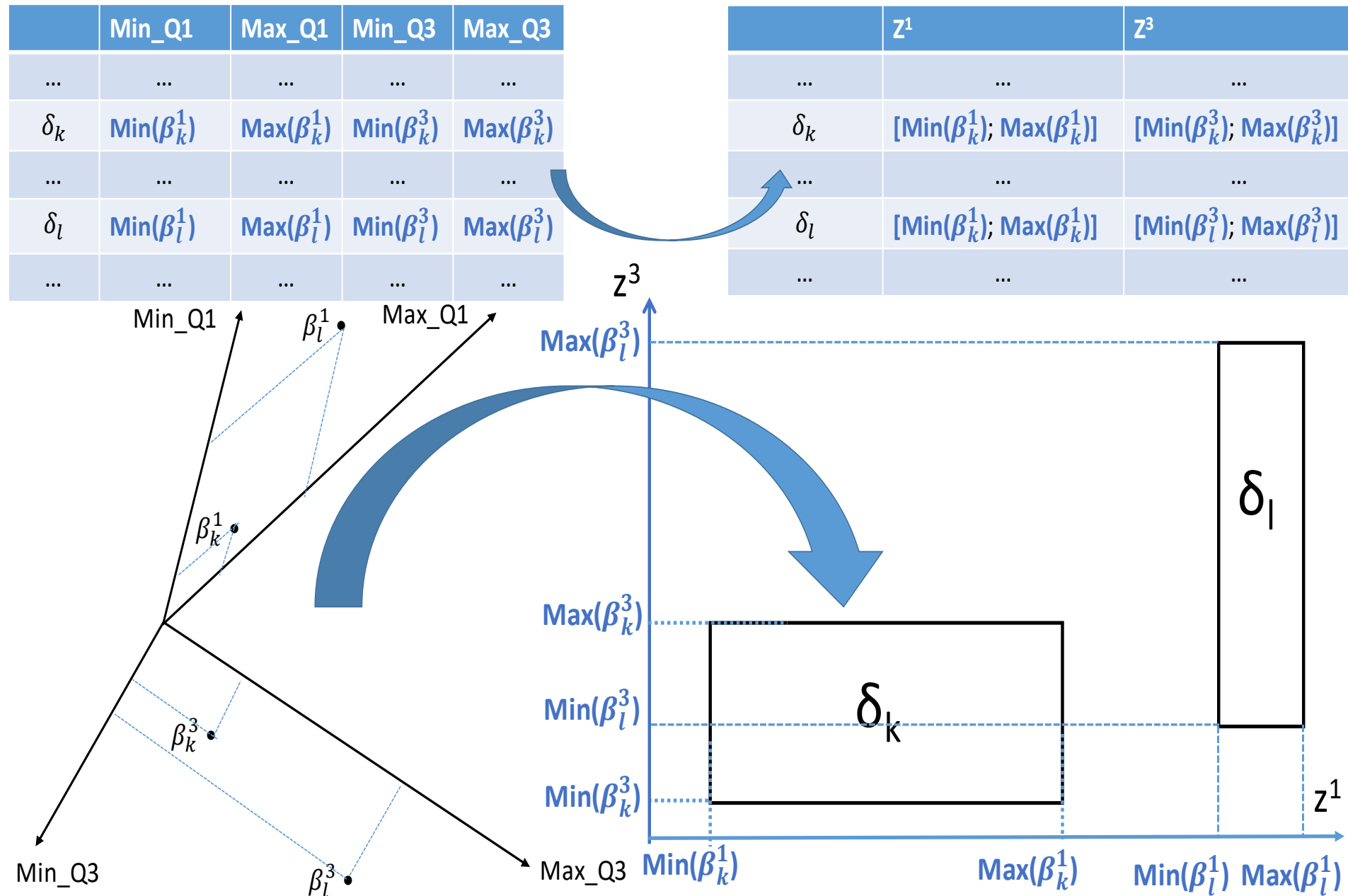
(Cameron & Trivedi, 2005)

II.1 Distribution of the specific costs of agricultural production (EU12)

Pig : specific cost (€) for 1,000 € of gross product, 2006

Specific Costs (€ for 1,000 € of gross product)						
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]

II.2 Distribution of estimated intervals: recoding in 2 dimensions



II.3 PCA of the estimation intervals in mixed strategy (SO-PCA)

The mixed strategy (Lauro & Palumbo, 2000) of main component analysis of symbolic objects (SO-PCA) combines the PCA of the Vertice PCA (V-PCA) with the PCA of the Ranges (RT-PCA) in a three-step approach to take into account the differences in scale and shape between Empirical distributions of specific costs:

- i) PCA of the ranges to extract the main axes that best render the scales and forms of the empirical distributions of conditional quantiles ;
- ii) Z Projection en $\hat{\mathbf{Z}} = \mathbf{P}_A \mathbf{Z}$ in order to take into account the relations between The different extrema, given the relationship of order between the different conditional quantiles on the distribution of specific costs ;
- iii) PCA projections of the lines $\hat{\mathbf{Z}}_i$ on the optimum representation sub-space $\Phi = \{\phi_1, \dots, \phi_m, \dots, \phi_{M^*}\}$ for the representation of scales and forms of distributions, with the projection operator

$$\hat{\mathbf{Z}}' \mathbf{P}_\Phi \hat{\mathbf{Z}} = \mathbf{Z}' \mathbf{A} (\mathbf{A}' \mathbf{A})^{-1/2} \mathbf{P}_\Phi (\mathbf{A}' \mathbf{A})^{-1/2} \mathbf{A}' \mathbf{Z} \mathbf{s}_m = \rho_m \mathbf{s}_m$$

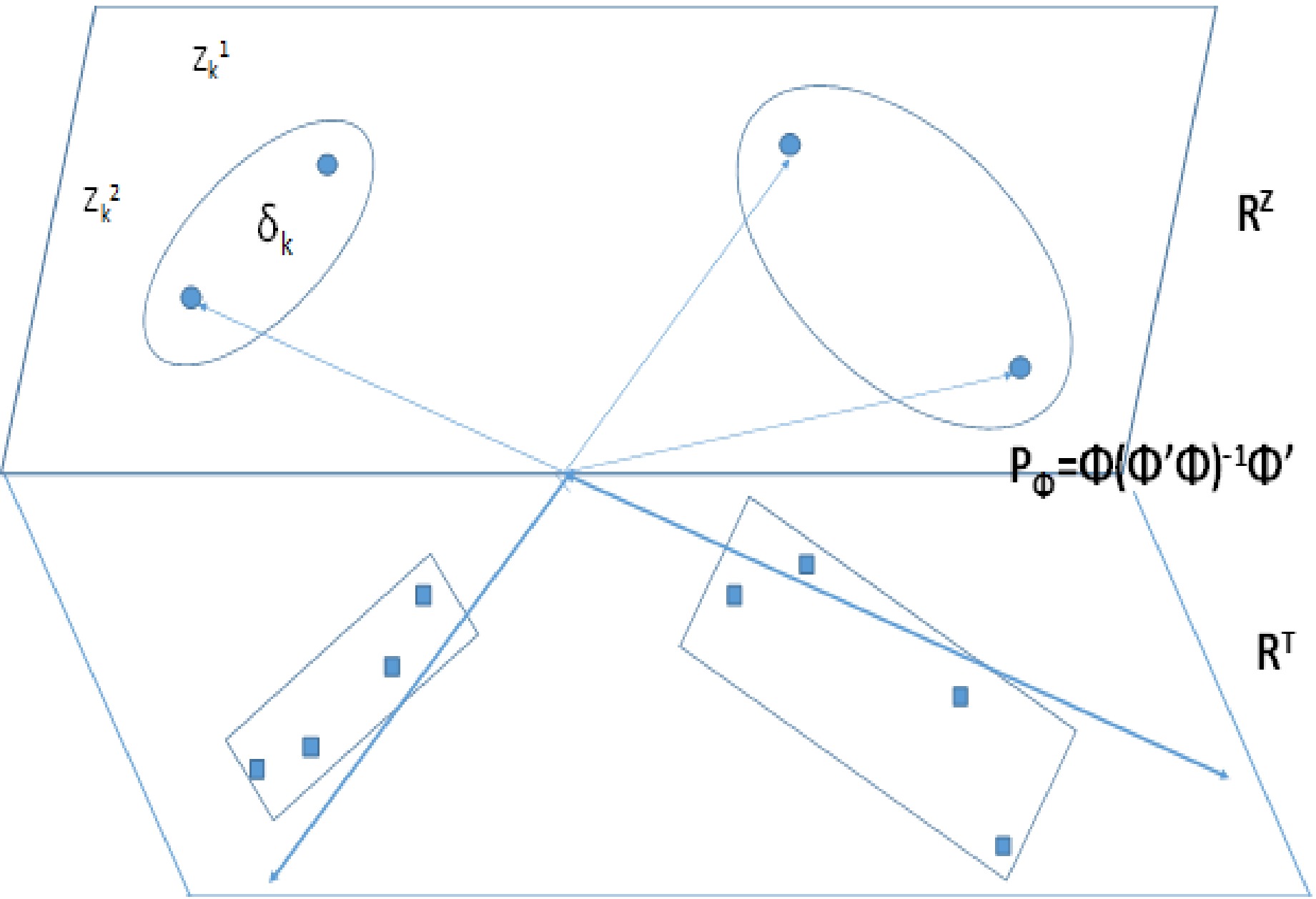
The mixed analysis strategy is therefore based on the solution of the following characteristic equation:

$$\hat{\mathbf{Z}}' \mathbf{P}_\Phi \hat{\mathbf{Z}} = \mathbf{Z}' \mathbf{A} (\mathbf{A}' \mathbf{A})^{-1/2} \mathbf{P}_\Phi (\mathbf{A}' \mathbf{A})^{-1/2} \mathbf{A}' \mathbf{Z} \mathbf{s}_m = \rho_m \mathbf{s}_m$$

The interpretation of the results of the analysis depends on the choice of the projection operator \mathbf{P}_Φ , whose diagonal term, interpreted as a standardized weight, is equal to:

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

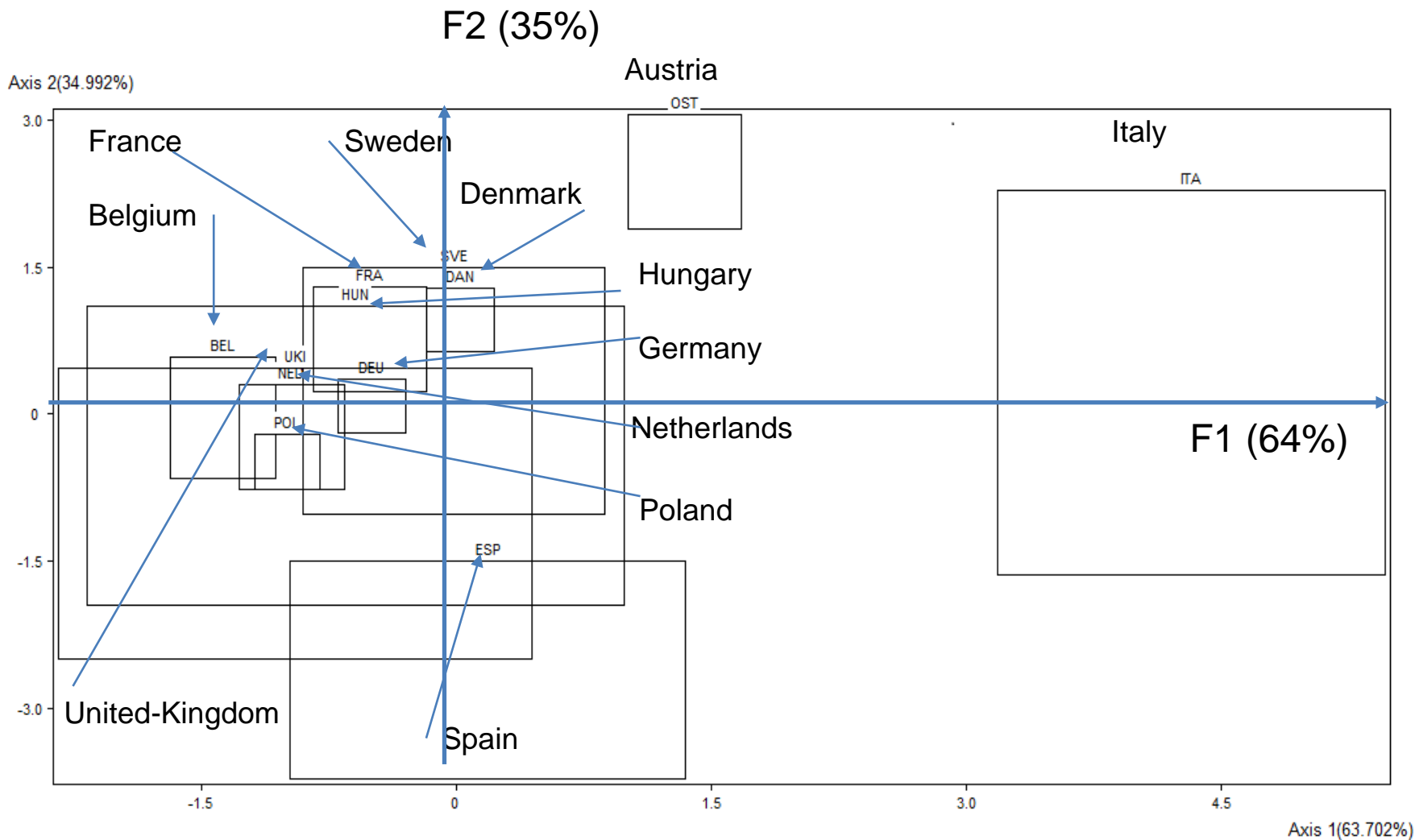
II.4 Symbolic PCA : projection on the referent sub-space



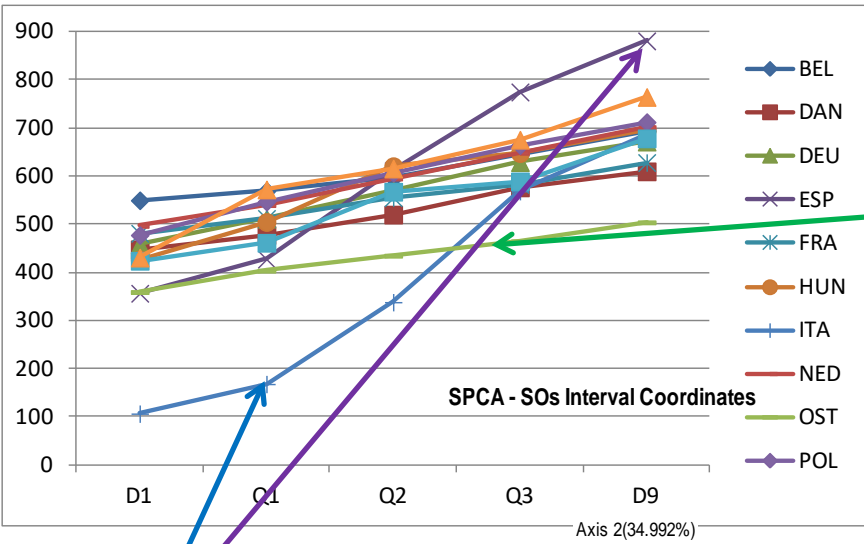
III.1 Quantile Estimates of production costs in agriculture

Fig: PCA on the estimated intervals

SPCA - SOs Interval Coordinates



IV.1 Pig-2006: Conditional Quantile Estimates of 12 EU-Members, SPCA

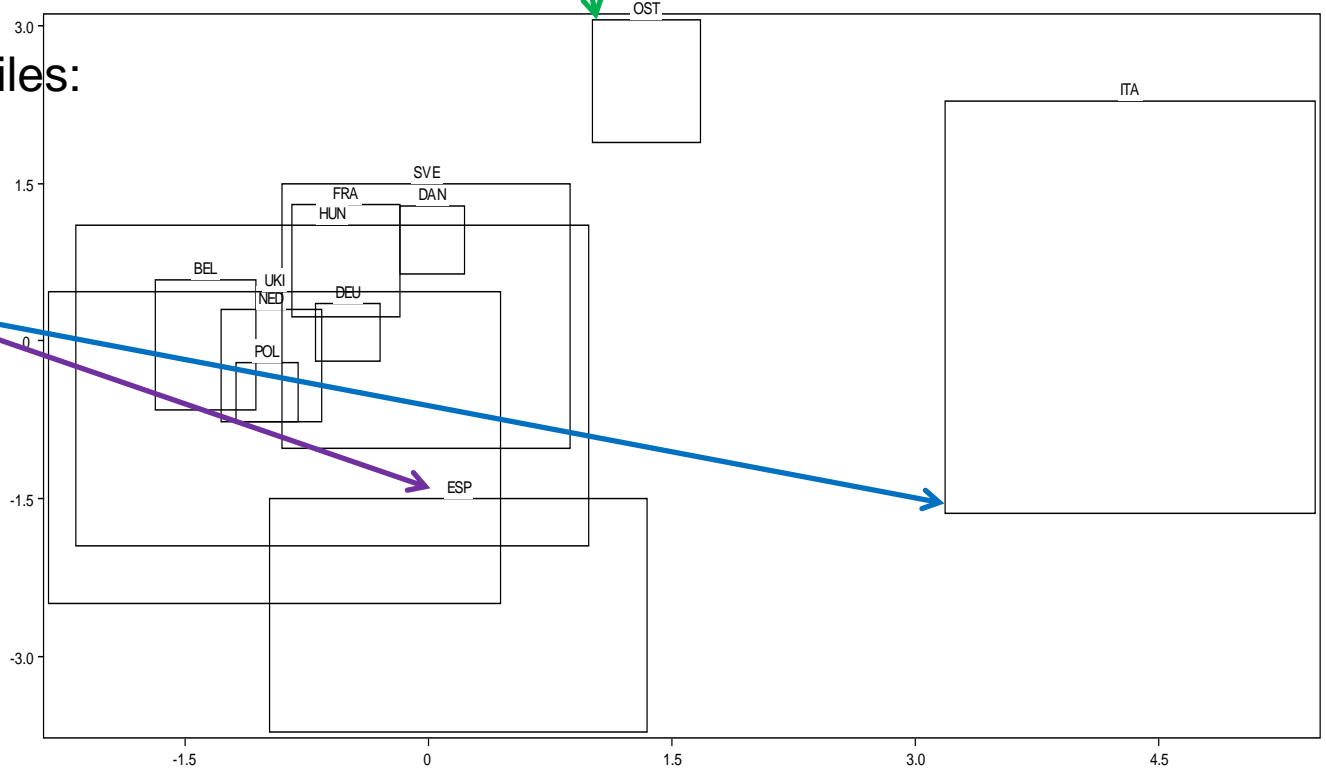


Homogeneous slope profiles:

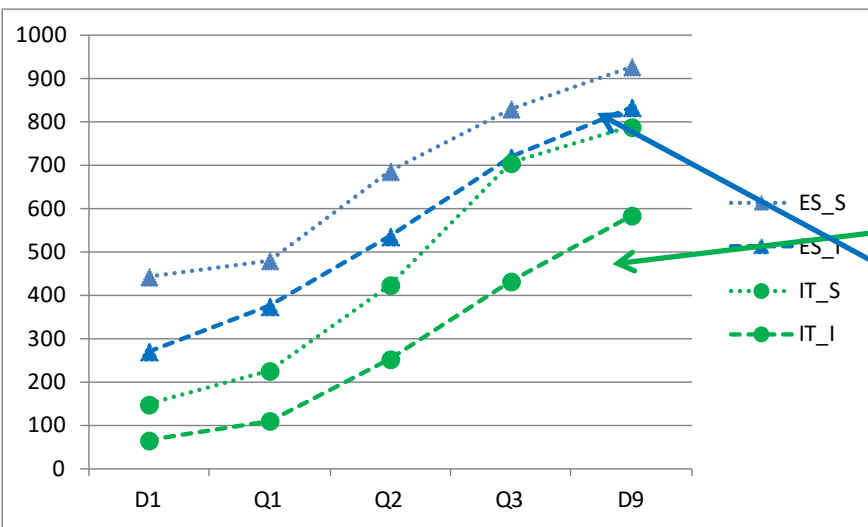
- DEU Deutschland
- BEL Belgium
- OST Österreich
- FRA France
- HUN Hungary
- NED Nederland
- POL Poland

Heterogeneous slope profiles:

- ESP Espana
- ITA Italia



IV.2) Pig-2006: Conditional Quantile Estimates of 12 EU-Members, SPCA

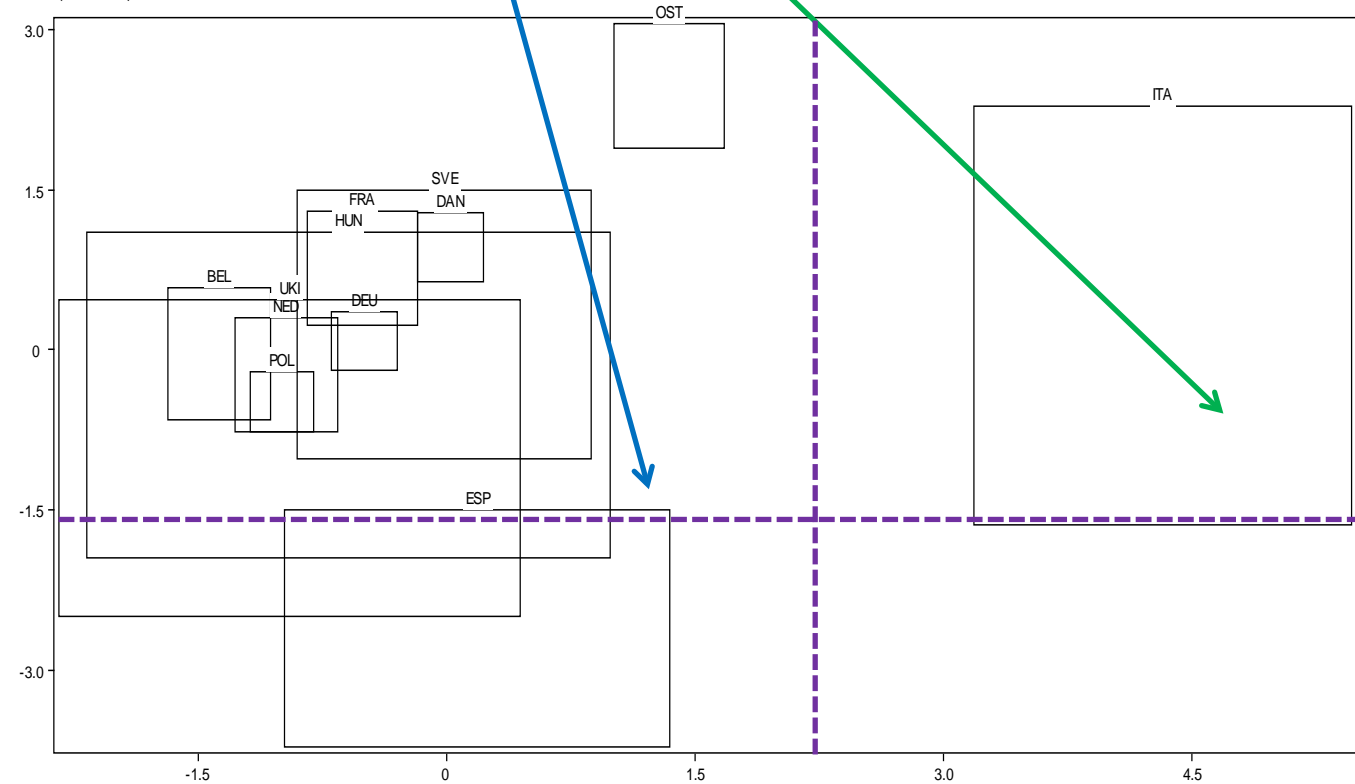


Graphical test of significant differences between :

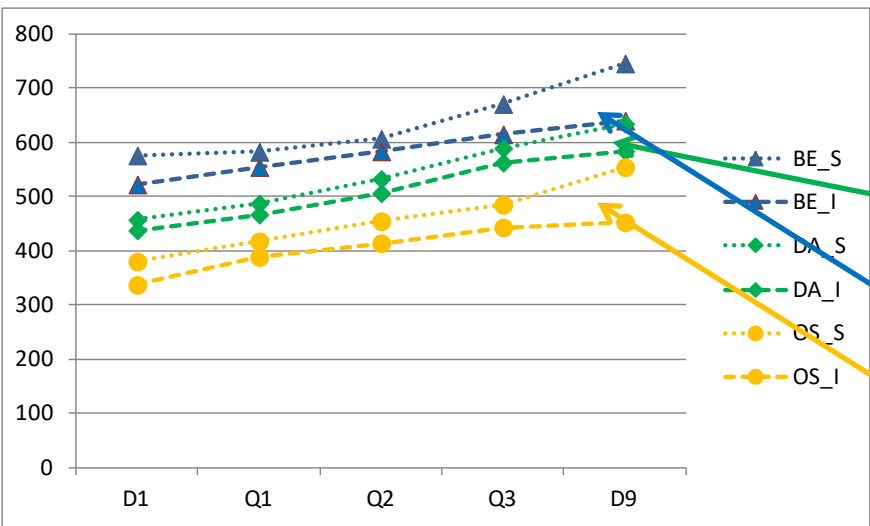
ITA Italia (heterogeneous)

ESP Espana (heterogeneous)

Axis 2(34.992%)



IV.3 Pig-2006: Conditional Quantile Estimates of 12 EU-Members, SPCA

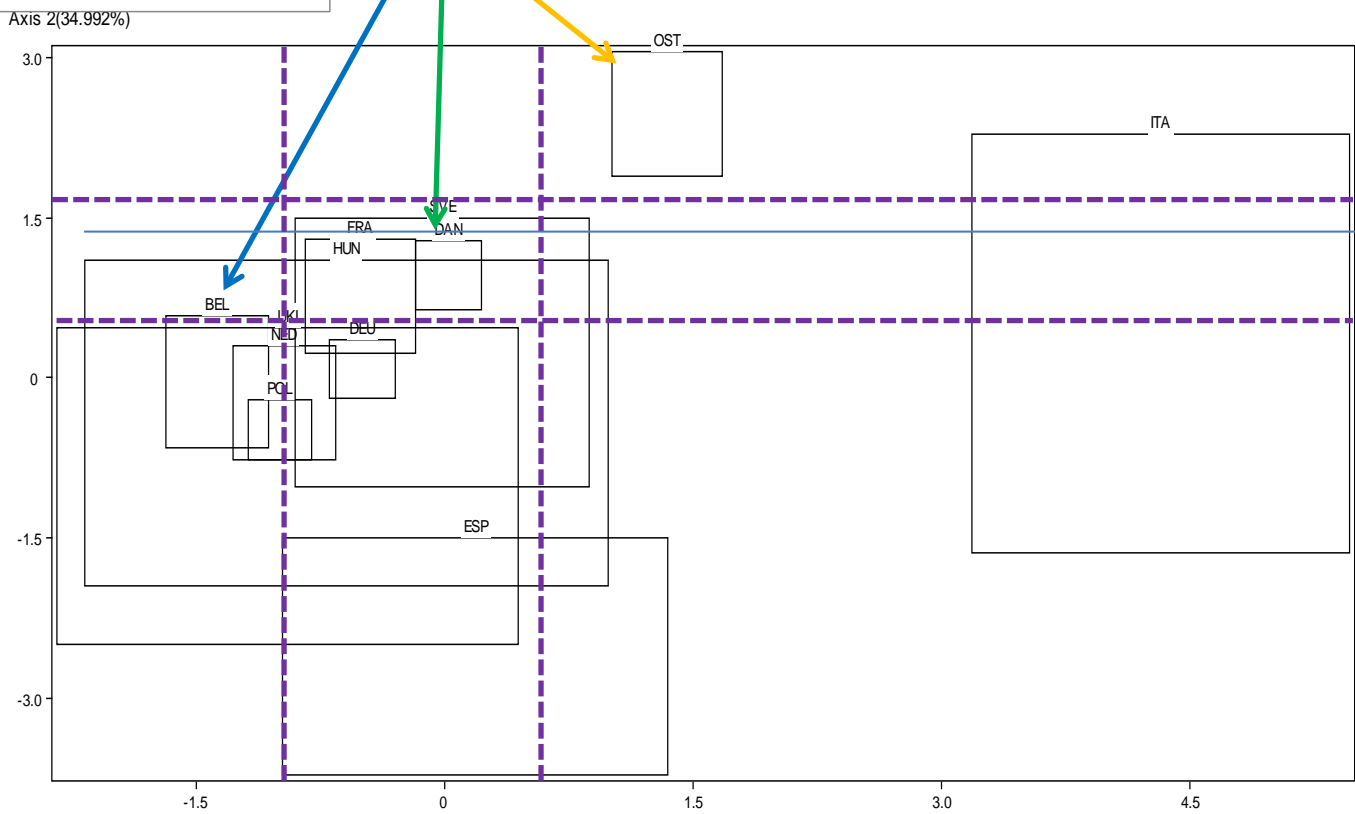


Graphical test of significant differences between :

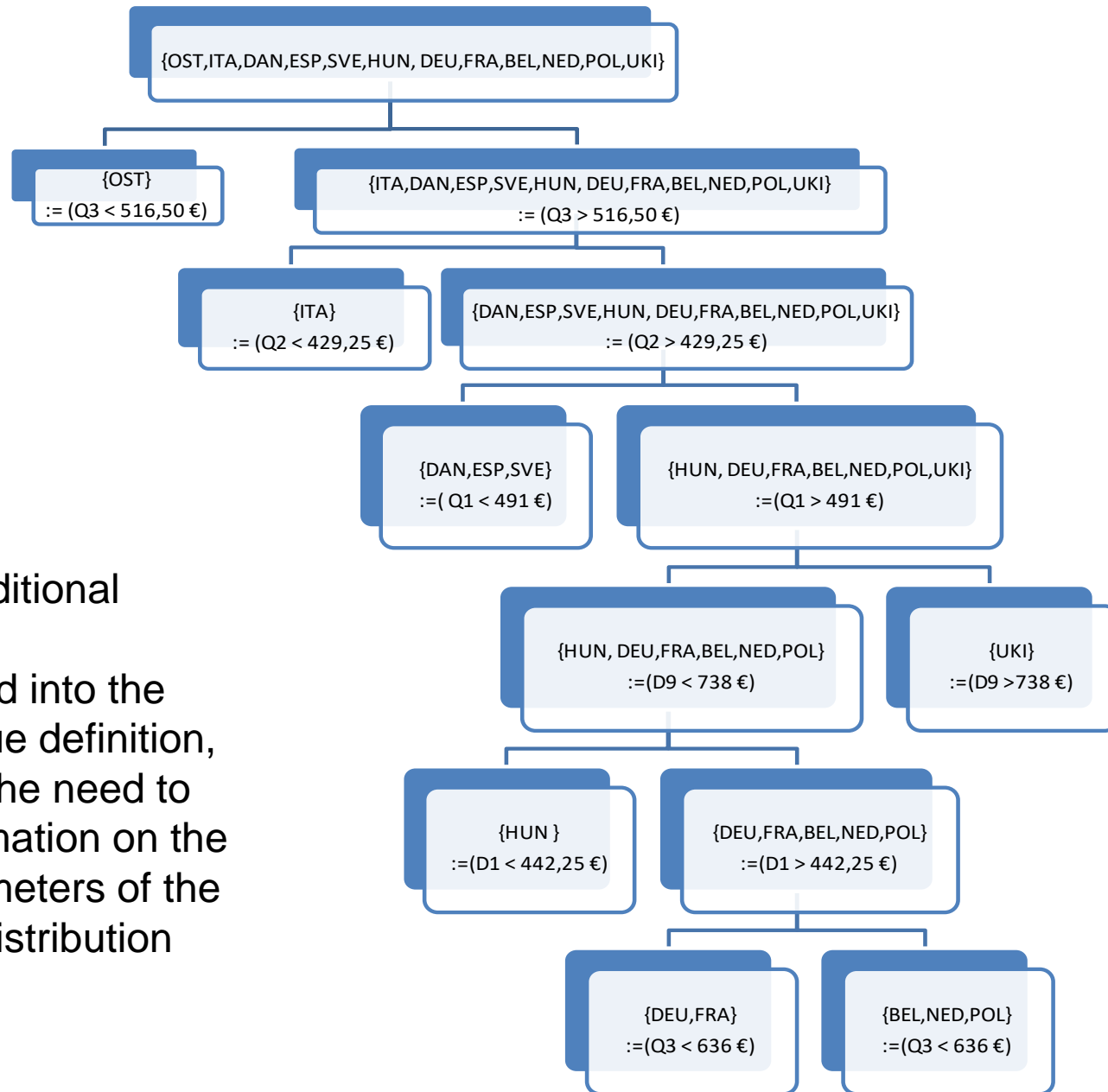
DAN Danemark (homogeneous)

BEL Belgium (homogeneous)

OST Österreich (homogeneous)



IV.5 Quantile estimates of production costs: pig hierarchical divisive clustering on the estimation intervals



All the conditional
quantiles
are involved into the
cutting value definition,
stressing the need to
keep information on the
main parameters of the
quantiles distribution

Conditionnal Quantiles of Specific Costs for Pig Production

Conclusions

Specific costs: in UE12, for the pig

- The national context is a significant factor of heterogeneity distinguishing northern Europe (homogeneous production systems) from Southern Europe (heterogeneous production systems) ;
- The specific costs of the pig workshop seem to be lower at large for small specialized granivorous farms, whereas they are generally higher when the pig workshop is located on mixed farms of medium or large economic size;
- There are significant differences between the southern regions of the EU12 European countries (Italy, Spain) due to regional specializations in certain types of pig production..



Collaborative research platform

Sign up for a **FREE** account



Search

Search

authors titles abstracts



Numéro 45

MODULAD 2018. *Modulad - Le Monde des Utilisateurs de L'Analyse de Données* vol.Modulad 45



PDF



Bib



Editors : Dominique Desbois, Edwin Diday

Dominique Desbois, Edwin Diday

Éditorial : analyse des données symboliques aujourd'hui, pp.1-2

Edwin Diday

Pouvoir explicatif et discriminant de variables et de tableaux de données symboliques, pp.3-18

Oldemar Rodriguez Rojas

Shrinkage linear regression for symbolic interval-valued variables, pp.19-38

Stéphanie Bougeard, Carole Toque

Symbolic Covariance ACP et régression pour variables à valeurs d'intervalles. Application en épidémiologie vétérinaire. , pp.39-54

Sun Makosso-Kallyth

Analyse en composante principales dun tableau de distributions macroéconomiques. , pp.55-74

Dominique Desbois

Explorer la distribution des intervalles destinations quantiles conditionnels : une application à l'estimation de coûts spécifiques de production du lait de vache dans l'Union européenne, pp.75-100

Frédéric Lebaron

La « complexité » du social. Quelques réflexions sur l'usage de l'analyse des données symboliques en sociologie. , pp.101-114

Daniel Defays

Appariement de matrices de dissimilarités, pp.115-128

<https://editions-rnti.fr/>