

Applying interval PCA and clustering to :empirical distributions of fertilizer cost quantile estimates for yearly crops in European Countries

Dominique Desbois

▶ **To cite this version:**

Dominique Desbois. Applying interval PCA and clustering to :empirical distributions of fertilizer cost quantile estimates for yearly crops in European Countries. CLIMAE Seminar, Nov 2022, Versailles, France. 2022. hal-03868997

HAL Id: hal-03868997
https://hal.inrae.fr/hal-03868997

Submitted on 24 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Applying interval PCA and clustering to quantile estimates: empirical distributions of fertilizer cost for yearly crops in European Countries

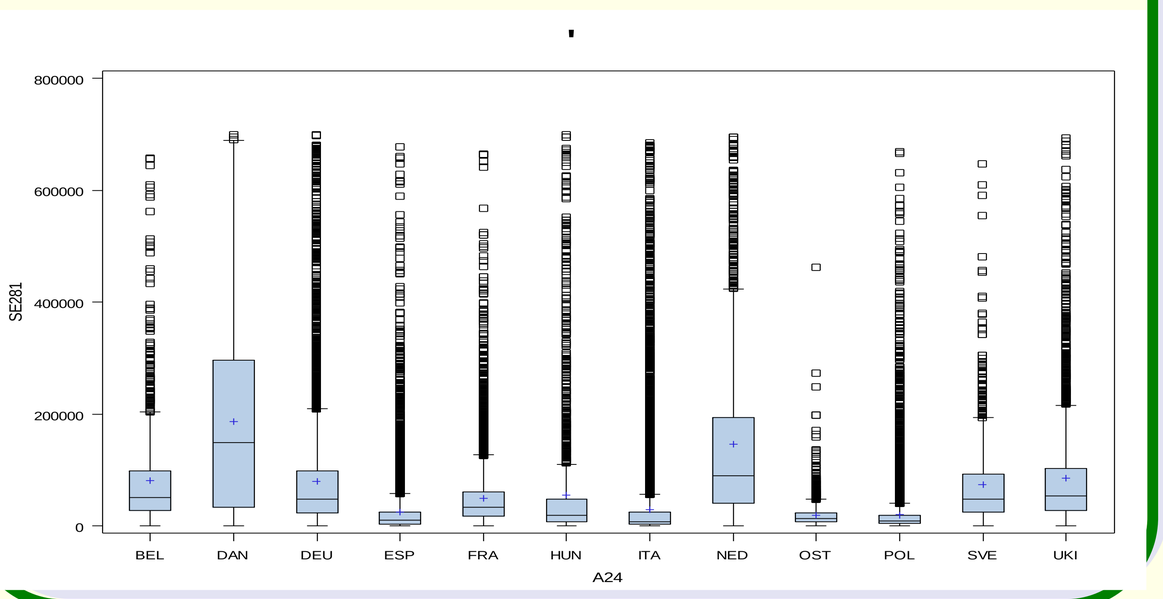
Dominique Desbois

Paris Saclay Applied Economics, INRAE-AgroParisTech, Université Paris-Saclay

Supported by the **ASSESS** project (ANR-16-NME1-0008) of the ***OTE-Med*** Eranet

After recalling the conceptual framework of the estimation of agricultural production costs, we present the empirical data model, the quantile regression approach and the interval principal component analysis and clustering tools used to obtain typologies of European countries on the basis of the conditional quantile distributions of fertilizer cost empirical estimates issued from FADN data.

1. Distribution of Production Costs in Agriculture



2. Input Allocation Model

$$X_{ih} = \sum_{k=1}^{K} \alpha_{ih}^{k} Y_{kh} + \varepsilon_{ih}$$

Produits / Charges	Y_{1h}	...	Y_{kh}	...	Y_{Kh}	Total charge
X_{1h}	a_{1h}^{1}	...	a_{1h}^{k}	...	a_{1h}^{K}	$\sum X_{1h}$
⋮	⋮		⋮		⋮	⋮
X_{ih}	a_{ih}^{1}	...	a_{ih}^{k}	...	a_{ih}^{K}	$\sum X_{ih}$
⋮	⋮		⋮		⋮	⋮
X_{Ih}	a_{Ih}^{1}	...	a_{Ih}^{k}	...	a_{Ih}^{K}	$\sum X_{Ih}$
Total produit	$\sum Y_{1h}$	...	$\sum Y_{kh}$	...	$\sum Y_{Kh}$	$\sum_k Y_{kh} = \sum_i X_{ih}$

3. Estimation Method: Quantile Regression (Koenker, 2005)

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

4. Estimation Intervals, Yield Crops (UE12, 2006)

Unit Costs for € 1 of gross product

Country	D1	Q1	Q2	Q3	D9
Austria	[0.000 ; 0.029]	[0.043 ; 0.057]	[0.068 ; 0.086]	[0.106 ; 0.127]	[0.155 ; 0.179]
Belgium	[0.009 ; 0.019]	[0.023 ; 0.030]	[0,038 ; 0.047]	[0.056 ; 0.080]	[0.082 ; 0.110]
Denmark	[0.018 ; 0.024]	[0.035 ; 0.035]	[0.056 ; 0.056]	[0.094 ; 0.094]	[0.140 ; 0.140]
France	[0.023 ; 0.028]	[0.053 ; 0.065]	[0.125 ; 0.125]	[0.182 ; 0.182]	[0.232 ; 0.232]
Germany	[0.004 ; 0.009]	[0.025 ; 0.033]	[0.082 ; 0.082]	[0.140 ; 0.140]	[0.181 ; 0.181]
Hungary	[0.020 ; 0.038]	[0.056 ; 0.071]	[0.093 ; 0.110]	[0.138 ; 0.164]	[0.197 ; 0.197]
Italy	[0.007 ; 0.011]	[0.019 ; 0.022]	[0.041 ; 0.041]	[0.078 ; 0.078]	[0.121 ; 0.121]
Netherlands	[0.001 ; 0.004]	[0.004 ; 0.006]	[0.009 ; 0.012]	[0.017 ; 0.022]	[0.026 ; 0.029]
Poland	[0.024 ; 0.032]	[0.052 ; 0.059]	[0.088 ; 0.099]	[0.146 ; 0.165]	[0.215 ; 0.228]
Spain	[0.013 ; 0.017]	[0.025 ; 0.033]	[0.058 ; 0.058]	[0.103 ; 0.103]	[0.169 ; 0.169]
Sweden	[-0.007 ; 0.016]	[0.003 ; 0.038]	[0.100 ; 0.100]	[0.215 ; 0.215]	[0.293 ; 0.293]
United-Kingdom	[0.006 ; 0.029]	[0.036 ; 0.047]	[0.088 ; 0.088]	[0.137 ; 0.137]	[0.171 ; 0.171]

5. Distribution Analysis of Interval Estimates

$$z_l^q = \left[Inf_\hat{\gamma}_l^{j_0}(q); Sup_\hat{\gamma}_l^{j_0}(q) \right] = \left[\underline{z_l^q}; \overline{z_l^q} \right]$$

PCA of Estimation Intervals in Mixed Strategy

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

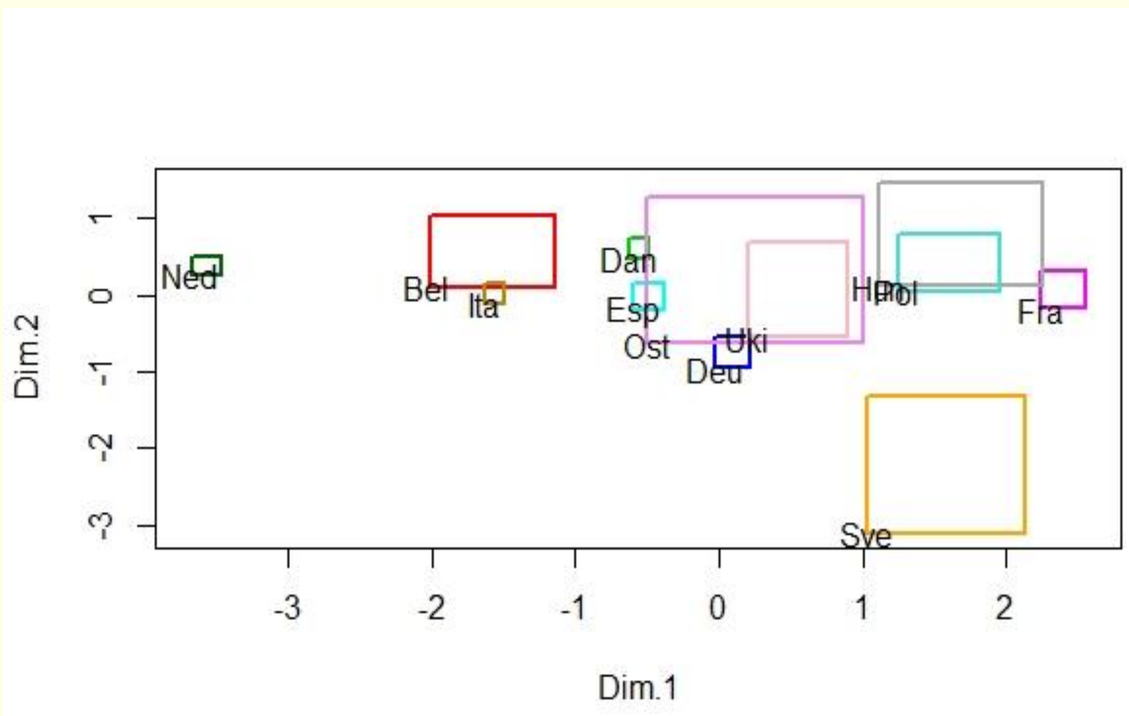
(Lauro & Palumbo, 2000)

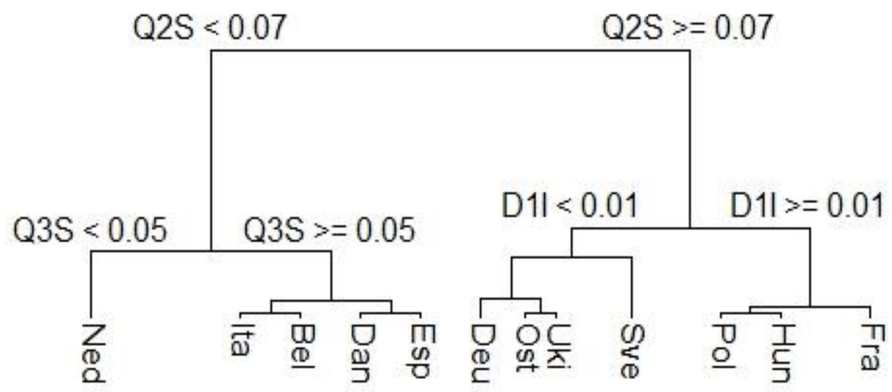
Hierarchical Divisive Classification on Interval Data

$$d(\omega_l, \omega_{l'}) = \left(\sum_{q=1}^{Q} \delta_M^2 \left(z_l^q, z_{l'}^q \right) \right)^{1/2}$$

(Chavent, 1998)

6. Factor Plane (Interval PCA) and Divisive Segmentation Tree





Excerpts from: Dominique Desbois (2021) Applying interval PCA and clustering to quantile estimates: empirical distributions of fertilizer cost estimates for yearly crops in European Countries, *Communications in Statistics: Case Studies, Data Analysis and Applications*, 7:4, 695-716, DOI:10.1080/23737484.2021.1972875

CLIMAE, November 23-24 , 2022, Versailles.