Specific costs and gross margins: estimation practices

Dominique Desbois

Specific costs and gross margins: estimation practices

Accédez aux articles scientifique
publiés par les Éditions RNT

Recherche | Saisir votre recherche | Cherch
auteurs | titres | résumés

Revue des Nouvelles Technologies de l'Information

Le Monde des Utilisateurs de l'Analyse de Données

Numéro 45

MODULAD 2018. *Modulad - Le Monde des Utilisateurs de L'Analyse de Données*
vol.Modulad 45
Rédacteurs invités : Dominique Desbois, Edwin Diday

Dominique Desbois, Edwin Diday
Éditorial : l'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

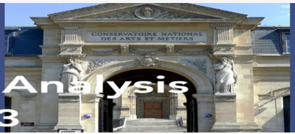
Stéphanie Bougeard, Carole Toque
Symbolic Covariance ACP et régression pour variables à valeurs d'intervalles. Application en épidémiologie vétérinaire. , pp.39-54

Sun Makosso-Kallyth
Analyse en composante principales d'un tableau de distributions macroéconomiques., pp.55-74

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



Estimating costs and margins: input-output methodology

Econometric model of specific production costs

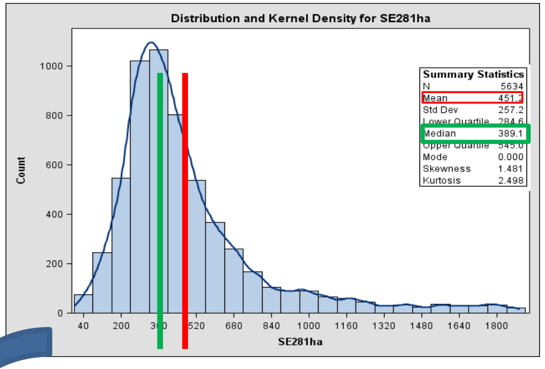
$$X_{ih} = \sum_{k=1}^{K} \alpha_{ih}^{k} Y_{kh} + \varepsilon_{ih} \quad with \quad \varepsilon_{ih} \quad i.i.d.$$

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	$\sum Y_{1h}$	…	$\sum Y_{kh}$	…	$\sum Y_{Kh}$	$\sum_k Y_{kh} = \sum_i X_{ih}$

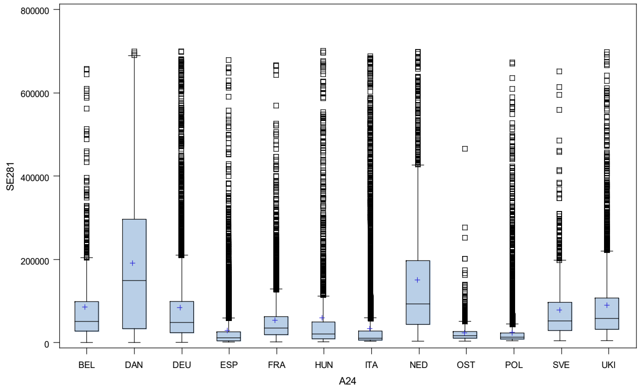
(D.Desbois, 2015), FACEPA Project, FP7 European program

Quantile estimates of costs and margins in agricultural production

1. Specific costs



2. Gross margins

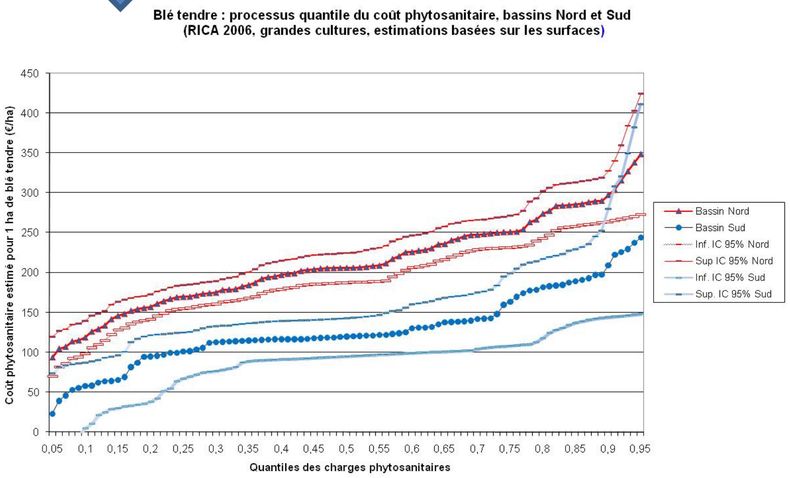


Method: quantile regression

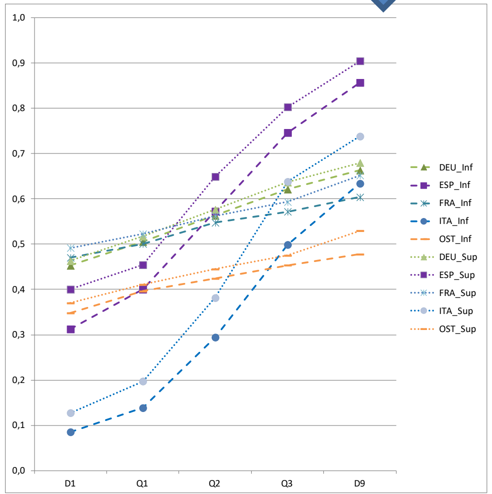
$$\underset{\beta}{Min}\left\{\sum_{\vec{i}\,x_i \geq y_i^{'}\beta} q\left|x_i - y_i^{'}\beta\right| + \sum_{\vec{i}\,x_i \leq y_i^{'}\beta} (1-q)\left|x_i - y_i^{'}\beta\right|\right\}$$

Estimation: conditional quantiles

Quantile process: wheat, France, 2006



Distribution in European countries: wheat, 2006



(D.Desbois, 2015)

Equivariance properties of conditional quantiles under monotonic transformation

The order statistics are equivariant by monotonic transformation:

If $\lambda \in [0; \infty]$ then $\mu_q(\lambda * X + C|Y)$=C+ $\lambda * \mu_{(q)}(X|Y)$

If $\lambda \in]-\infty; 0]$ then $\mu_q(\lambda * X + C|Y)$=C+ $\lambda * \mu_{(1-q)}(X|Y)$

By reparametrisation in X of $M = Y - X$, the result is:

$$\mu_q\widehat{(M \mid Y)} = \mu_q(Y\widehat{-X} \mid Y) = 1 - \mu_{\widehat{(1-q)}}(X \mid Y)$$

(D.Desbois, NOV'AE, 8, 2023)

Specific costs of wheat

$$\hat{\mu}_{0.25}^{mb} = 646$$

$$= 1000 - 354$$

$$= 1000 - \hat{\mu}_{(0.75)}^{cs}$$

Blé		Coûts spécifiques (€) pour 1 000 € de produit brut						Marges brutes (€) pour 1 000 € de produit brut					
Code	Régions RICA	D1cs	Q1cs	Q2cs	Q3cs	D9cs	MCOcs	D1mb	Q1mb	Q2mb	Q3mb	D9mb	MCOmb
A010	Schleswig-Holstein	389	407	364	354	542	464	458	646	636	593	611	536
A030	Niedersachsen	214	229	295	351	327	305	673	649	705	771	786	695
A050	Nordrhein-Westfalen	325	245	222	224	233	204	767	776	778	755	675	796
A060	Hessen	383	410	343	356	333	428	667	644	657	590	617	572
A090	Bayern	283	354	418	561	669	487	331	439	582	646	717	513
A115	Sachsen-Anhalt	245	214	290	284	289	188	711	716	710	786	755	812
F131	Champagne-Ardenne	379	514	537	530	752	492	248	470	463	486	621	508
F132	Picardie	314	299	376	465	465	332	535	535	624	701	686	668
F134	Centre	303	317	376	490	568	367	432	510	624	683	697	633
F135	Basse-Normandie	699	646	617	525	687	507	313	475	383	354	301	493
F136	Bourgogne	475	442	543	657	779	491	221	343	457	558	525	509
F141	Nord-Pas-de-Calais	321	370	452	508	687	498	313	492	548	630	679	502
F151	Lorraine	524	508	488	398	531	452	469	602	512	492	476	548
F152	Alsace	589	721	817	886	1001	864	-1	114	183	279	411	136
F153	Franche-Comté	689	759	680	625	832	706	168	375	320	241	311	294
F162	Pays de la Loire	428	421	435	562	555	403	445	438	565	579	572	597
F163	Bretagne	368	400	456	421	483	396	517	579	544	600	632	604
F192	Rhône-Alpes	609	640	904	935	911	903	89	65	96	360	391	97
F193	Auvergne	336	384	528	792	804	434	196	208	472	616	664	566
I260	Emilia-Romagna	187	270	374	437	481	346	519	563	626	730	813	654
I282	Umbria	316	305	327	358	359	376	641	642	673	695	684	624
B341	Flandre	305	363	446	496	1085	552	-85	504	554	637	695	448
B343	Wallonie	397	306	342	407	407	404	593	593	658	694	603	596
D370	Danemark	213	261	363	462	552	426	448	538	637	739	787	574
U411	England-North	332	348	346	471	564	375	436	529	654	652	668	625
U412	England-East	264	302	339	421	475	323	525	579	661	698	736	677
U413	England-West	363	343	294	308	406	349	594	692	706	657	637	651
U431	Scotland	338	405	522	443	490	478	510	557	478	595	662	522
E570	Extremadura	182	206	218	373	498	299	502	627	782	794	818	701
O660	Österreich	201	240	266	261	344	200	656	739	734	760	799	800
S710	Slattbygdslan	341	331	388	522	567	323	433	478	612	669	659	677
H761	Kozép-Dunàntùl	382	465	419	348	638	563	362	652	581	535	618	437
H762	Nyugat-Dunàntùl	361	357	393	388	509	759	491	612	607	643	639	241
H765	Eszak-Alfoid	302	238	339	371	341	197	659	629	661	762	698	803
P785	Pormoze & Mazury	300	353	395	505	562	418	438	495	605	647	700	582
P790	WiekoPolska &Slask	252	280	344	430	536	424	464	570	656	720	748	576
P795	Mazowsze & Podlasie	266	263	302	349	458	336	542	651	698	737	734	664
P800	Malopolska & Pogörze	254	290	372	331	383	347	617	669	628	710	746	653

Gross margins of wheat

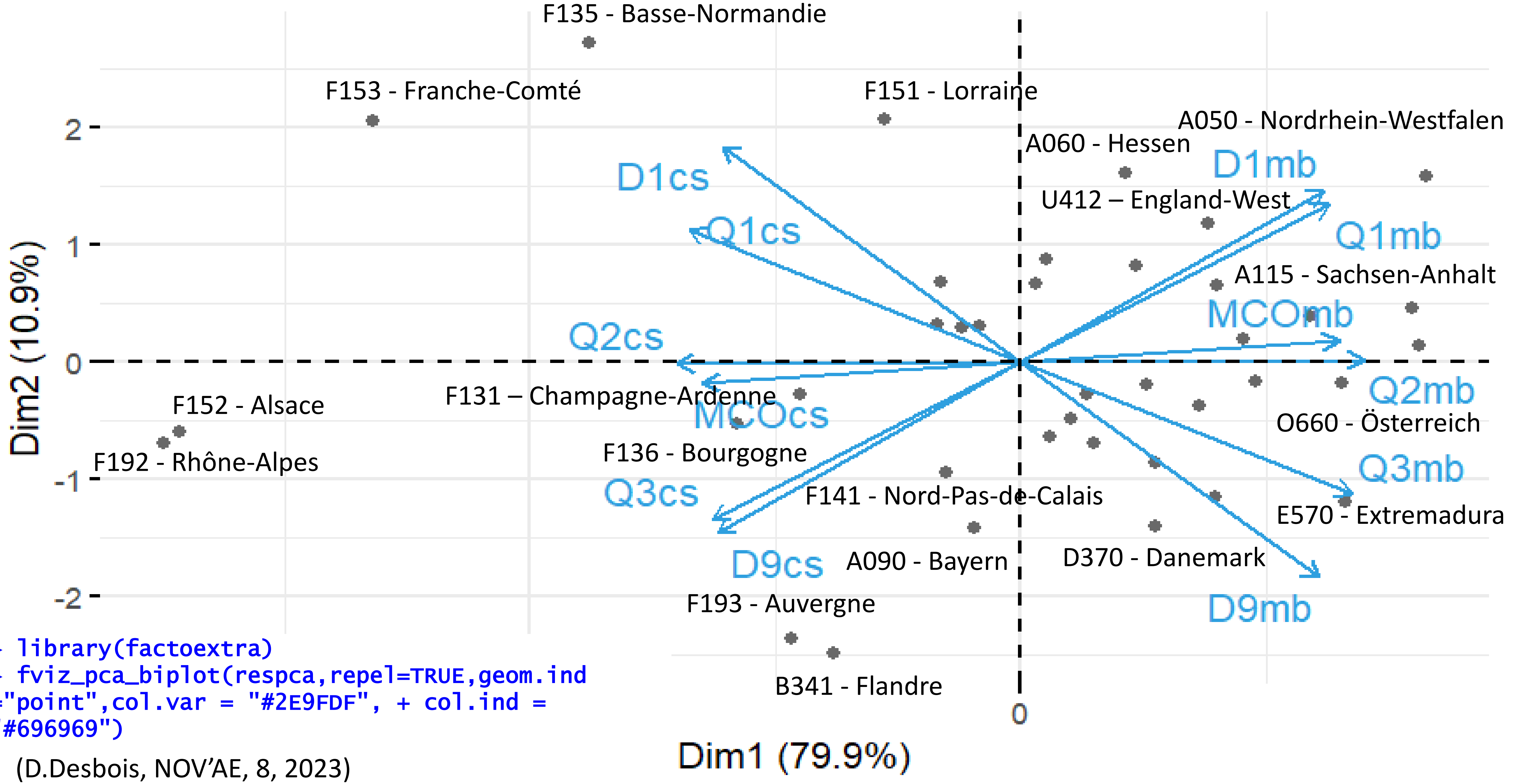
$$\hat{\mu}_{0.1}^{mb} = 458$$

$$= 1000 - 542$$

$$= 1000 - \hat{\mu}_{(0.9)}^{cs}$$

(D.Desbois, NOV'AE, 8, 2023)

nPCA Biplot




```
> library(factoextra)
> fviz_pca_biplot(respca,repel=TRUE,geom.ind
="point",col.var = "#2E9FDF", + col.ind =
"#696969")
```

(D.Desbois, NOV'AE, 8, 2023)

Divisive Classification Algorithm

Chavent's Method for Divisive Classification:

Generated by the binary response (yes/no) to the question

$$\Psi = [\, z^q \ \leq c\,?\,],$$

Let be $\{A_k, \overline{A_k}\}$ the bipartition induced on the C_k clusters formed by the n_k objects ;

As in Ward's method, the upper "hierarchy" of the P_K partition is sorted by the h index of the C_K, clusters, defined by their inter−inertia (B):

$$h(C_k) = B\left(A_k, \overline{A_k}\right) = \frac{\mu(A_k)\mu\left(\overline{A_k}\right)}{\mu(A_k) + \mu\left(\overline{A_k}\right)} d^2\left(g(A_k), g\left(\overline{A_k}\right)\right)$$

the DIVCLUS−T algorithm splits the C_K^* cluster maximising $h(C_K)$, such as P_{K+1}, the new partition

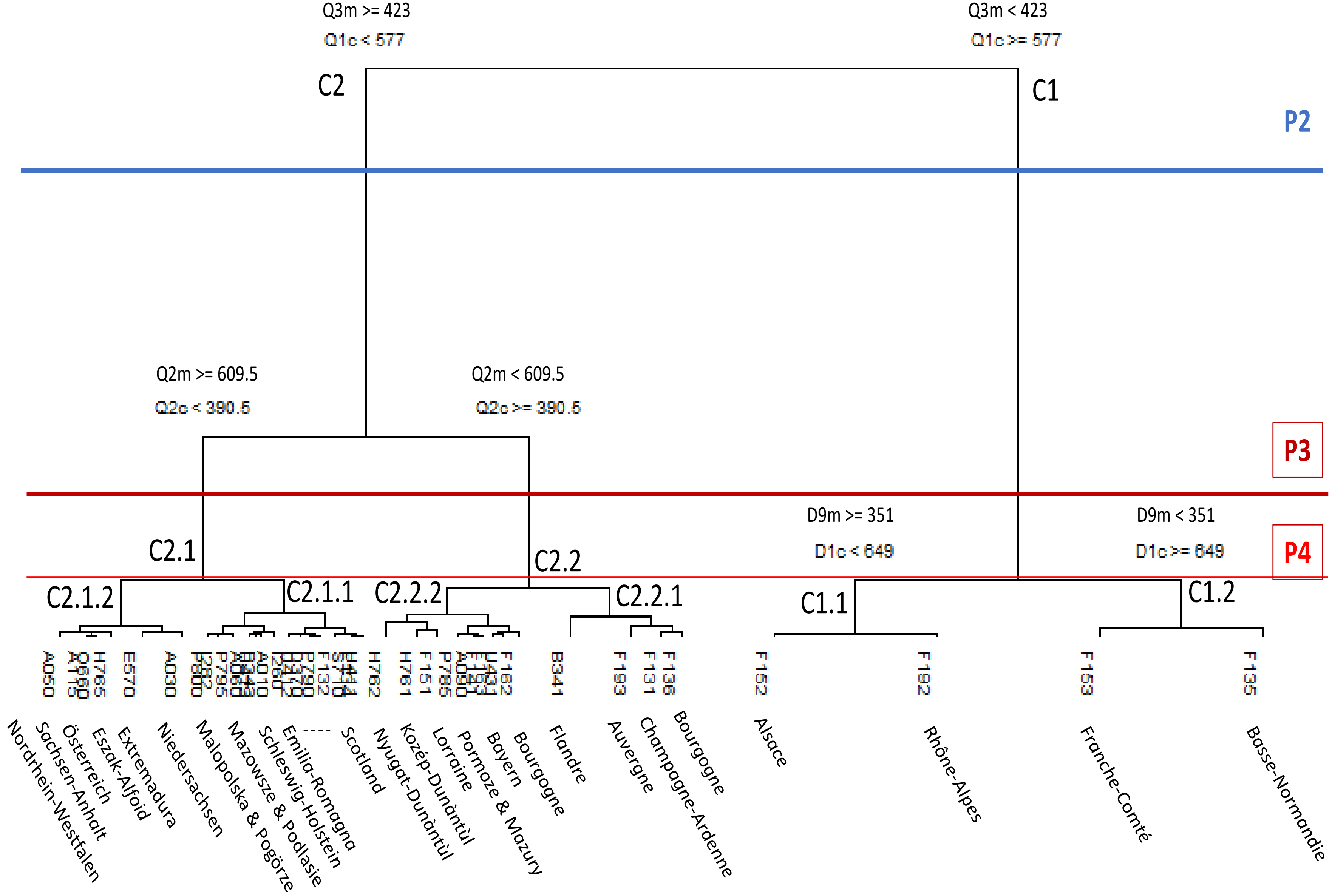
$$P_{K+1} = P_K \bigcup \{A_K, \overline{A_K}\} - C_K^*$$

has the lowest intra inertia, according to the following rule:

$$W(P_{K+1}) = W(P_K) - \ h(C_K^*).$$

Partitions into two (P2), three (P3) and four (P4) classes from the divisive class hierarchy of the DIVCLUST procedure

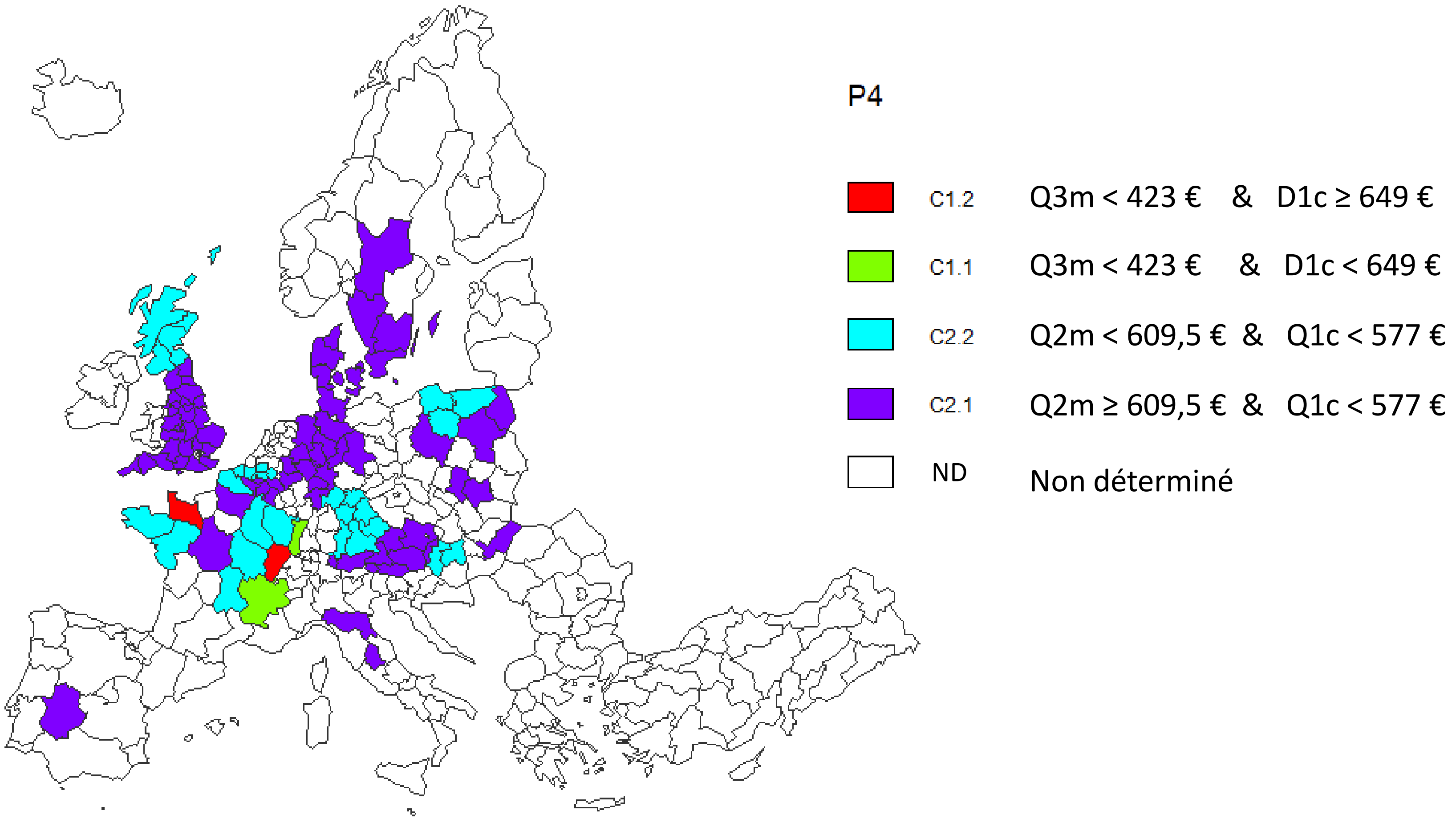
Typology of costs and margins




```
> resdivP4<-divclust(reg06ble[,3:14],K=4)
```

(D.Desbois, NOV'AE, 8, 2023)

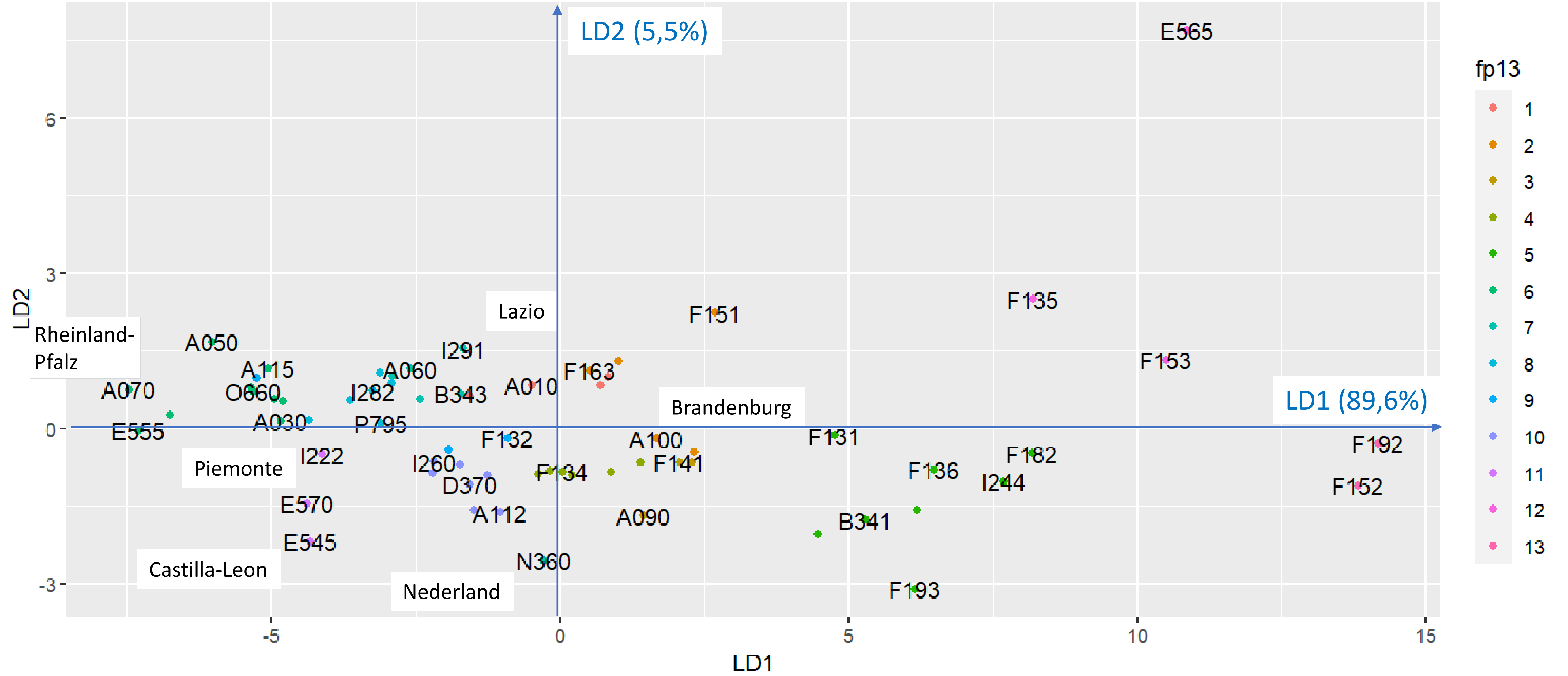
Map projection of partition P4 of the divisive clustering




```
➤ typoLayer(spdf = nuts2.spdf, df = nuts2P4n3.df, var = "P4", legend.pos ="topright", col=c(rainbow(4),
  "white"), legend.values.order=c("C1.2","C1.1","C2.2","C2.1","NULL"))
```

(D.Desbois, NOV'AE, 8, 2023)

Linear discrimination P13: first linear discriminating plane

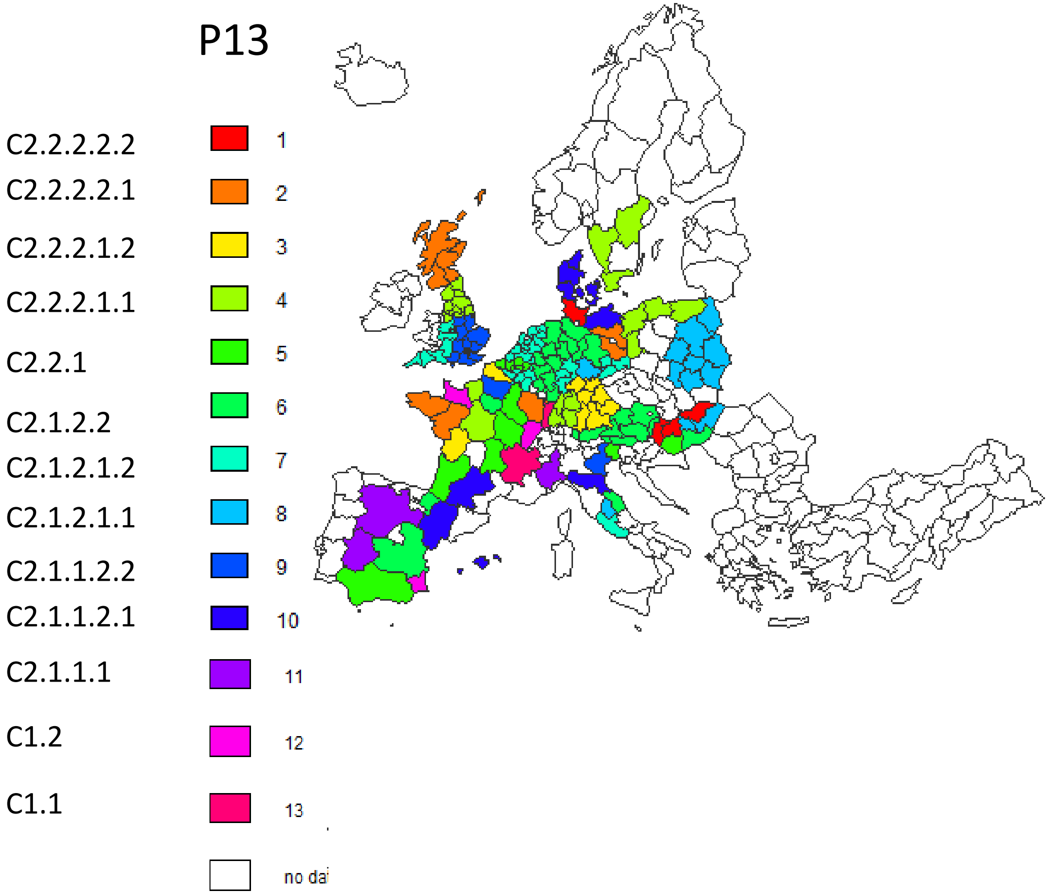


Learning sample: fully estimated regions **Test sample: partially estimated regions**

```
> fp13<-as.factor(r3828p13[,2])
> reslda<-lda(x=r3828p13[,3:8],grouping=fp13)
> predlda<-as.data.frame(predict(reslda)$x)
> ggplot(predlda,aes(LD1, LD2))+geom_point(aes(color=fp13))+geom_text(label=r3828p13[,1],check_overlap=TRUE)
```

(D.Desbois, NOV'AE, 8, 2023)

Map projection of partition P13 (full and partial estimates)



(D.Desbois, NOV’AE, 8, 2023)

```
> typoLayer(spdf = nuts2.spdf, df = nuts2P13n2.df, var = "p13", legend.pos ="bottomleft",
col=c(rainbow(13)),legend.values.order=c("1","2","3","4","5","6","7","8","9","10","11","12","13"),
legend.title.cex = 0.5,legend.values.cex = 0.5, lwd=0.05)
```

From margins to costs: invariance of distances between quantile processes symmetrically constructed

- **Supremum Distance**

For any symmetrically constructed quantile process, i.e. $\mathrm{T} = \{\tau_1 = \varepsilon, \dots, \tau_j = q, \ \dots, \tau_{j+p} = 1-q, \dots, \tau_{2p} = 1-\varepsilon\}$ pour $J = \{1, \dots, 2p\}$, by re-parametrizing M into $X - Y$, the supremum distance is written as follows:

$$d_\infty^M(i', i) = d_\infty^M(i, i') = Sup_{j\in J}\left(\left|\widehat{\mu_M\ (q)}_i^j - \widehat{\mu_M(q)}_{i'}^j\right|\right) = Sup_{j\in J}\left(\left|\left[1-\mu_X\ \widehat{(1-}q)_i^j\right] - \left[1-\mu_X\widehat{(1-}q)_{i'}^j\right]\right|\right)$$

$$= Sup_{j\in J}\left(\left|\mu_X\widehat{(1-}q)_{i'}^j - \mu_X\ \widehat{(1-}q)_i^j\right|\right) = d_\infty^X(i', i) = d_\infty^X(i, i')$$

- **Manhattan Distance**

For any symmetrically constructed quantile process, i.e. $\mathrm{T} = \{\tau_1 = \varepsilon, \dots, \tau_j = q, \ \dots, \tau_{j+p} = 1-q, \dots, \tau_{2p} = 1-\varepsilon\}$ pour $J = \{1, \dots, 2p\}$, by re-parametrizing M into $X - Y$, the Manhattan distance is written as follows:

$$d_1^M(i', i) = d_1^M(i, i') = \sum_{j\in J}\left|\widehat{\mu_M\ (q)}_i^j - \widehat{\mu_M(q)}_{i'}^j\right| = \sum_{j\in J}\left|\left[1-\mu_X\ \widehat{(1-}q)_i^j\right] - \left[1-\mu_X\widehat{(1-}q)_{i'}^j\right]\right|$$

$$= \sum_{j\in J}\left|\left[\mu_X\widehat{(1-}q)_{i'}^j\right] - \left[\mu_X\ \widehat{(1-}q)_i^j\right]\right| = \delta_1^X(i', i) = \delta_1^X(i, i')$$

- **Minkowski Distance of *k* order (Euclidean distance for k=2)**

For any symmetrically constructed quantile process, i.e. $\mathrm{T} = \{\tau_1 = \varepsilon, \dots, \tau_j = q, \ \dots, \tau_{j+p} = 1-q, \dots, \tau_{2p} = 1-\varepsilon\}$ pour $J = \{1, \dots, 2p\}$, by re-parametrizing M into $X - Y$, the Minkowski distance of k order is written as follows:

$$d_k^M(i', i) = d_k^M(i, i') = \sqrt[k]{\sum_{j\in J}\left(\widehat{\mu_M\ (q)}_i^j - \widehat{\mu_M(q)}_{i'}^j\right)^k} = \sqrt[k]{\sum_{j\in J}\left(\left[1-\mu_X\ \widehat{(1-}q)_i^j\right] - \left[1-\mu_X\widehat{(1-}q)_{i'}^j\right]\right)^k}$$

$$= \sqrt[k]{\sum_{j\in J}\left(\left[\mu_X\widehat{(1-}q)_{i'}^j\right] - \left[\mu_X\ \widehat{(1-}q)_i^j\right]\right)^k} = d_k^X(i', i) = d_k^X(i, i')$$

Distance Invariance with Duality: Tchebichev Distance

$$d(i,i') = Sup_j\left(\left|x_i^j - x_{i'}^j\right|\right)$$

Margins

```
> dman_cm06<-dist(cm06[,c(13:23,25)],method="maximum")
> dman_cm06
```

	BEL	DAN	DEU	ESP	FRA	HUN	ITA	NED	OST	POL	SVE
DAN	0.058										
DEU	0.098	0.046									
ESP	0.087	0.029	0.037								
FRA	0.150	0.092	0.052	0.079							
HUN	0.115	0.078	0.039	0.061	0.044						
ITA	0.039	0.019	0.062	0.048	0.111	0.086					
NED	0.058	0.114	0.154	0.143	0.206	0.171	0.095				
OST	0.073	0.046	0.035	0.028	0.077	0.042	0.049	0.129			
POL	0.132	0.074	0.034	0.062	0.037	0.024	0.093	0.188	0.059		
SVE	0.211	0.153	0.113	0.124	0.063	0.096	0.172	0.267	0.138	0.079	
UKI	0.089	0.043	0.026	0.034	0.061	0.065	0.059	0.145	0.033	0.043	0.122

Costs

```
> dman_cm06<-dist(cm06[,c(2:12,24)],method="maximum")
> dman_cm06
```

	BEL	DAN	DEU	ESP	FRA	HUN	ITA	NED	OST	POL	SVE
DAN	0.058										
DEU	0.099	0.046									
ESP	0.087	0.029	0.037								
FRA	0.150	0.092	0.052	0.079							
HUN	0.115	0.078	0.039	0.061	0.044						
ITA	0.039	0.019	0.062	0.048	0.111	0.086					
NED	0.081	0.114	0.155	0.143	0.206	0.171	0.095				
OST	0.073	0.046	0.034	0.028	0.077	0.042	0.058	0.150			
POL	0.133	0.088	0.048	0.062	0.037	0.031	0.107	0.199	0.060		
SVE	0.211	0.153	0.113	0.124	0.061	0.096	0.172	0.267	0.138	0.078	
UKI	0.089	0.043	0.026	0.034	0.061	0.065	0.059	0.145	0.033	0.057	0.122

Margins & Costs

```
> dmax_cm06<-dist(cm06[,2:25],method="maximum")
> dmax_cm06
```

	BEL	DAN	DEU	ESP	FRA	HUN	ITA	NED	OST	POL	SVE
DAN	0.058										
DEU	0.099	0.046									
ESP	0.087	0.029	0.037								
FRA	0.150	0.092	0.052	0.079							
HUN	0.115	0.078	0.039	0.061	0.044						
ITA	0.039	0.019	0.062	0.048	0.111	0.086					
NED	0.081	0.114	0.155	0.143	0.206	0.171	0.095				
OST	0.073	0.046	0.035	0.028	0.077	0.042	0.058	0.150			
POL	0.133	0.088	0.048	0.062	0.037	0.031	0.107	0.199	0.060		
SVE	0.211	0.153	0.113	0.124	0.063	0.096	0.172	0.267	0.138	0.079	
UKI	0.089	0.043	0.026	0.034	0.061	0.065	0.059	0.145	0.033	0.057	0.122

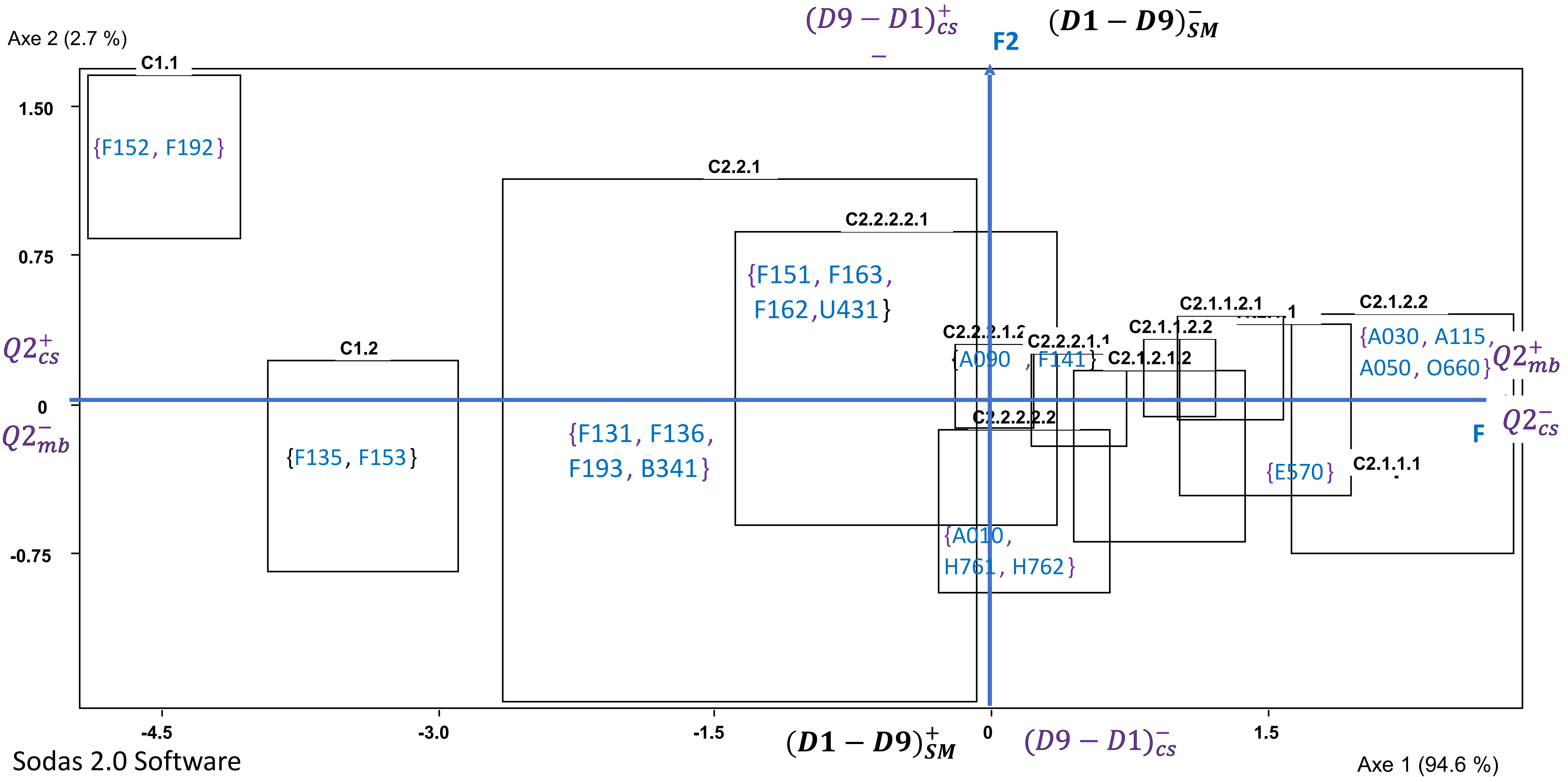
What contribution can the analysis of symbolic data make to the study of costs and margins?

- **What about symbolic mean, symbolic variances, symbolic covariances, & symbolic PCA ?**

- **What about symbolic distances & symbolic clustering ?**

- **What about symbolic discrimination ?**

Standardized interval PCA of the typological reference frame



Sodas 2.0 Software

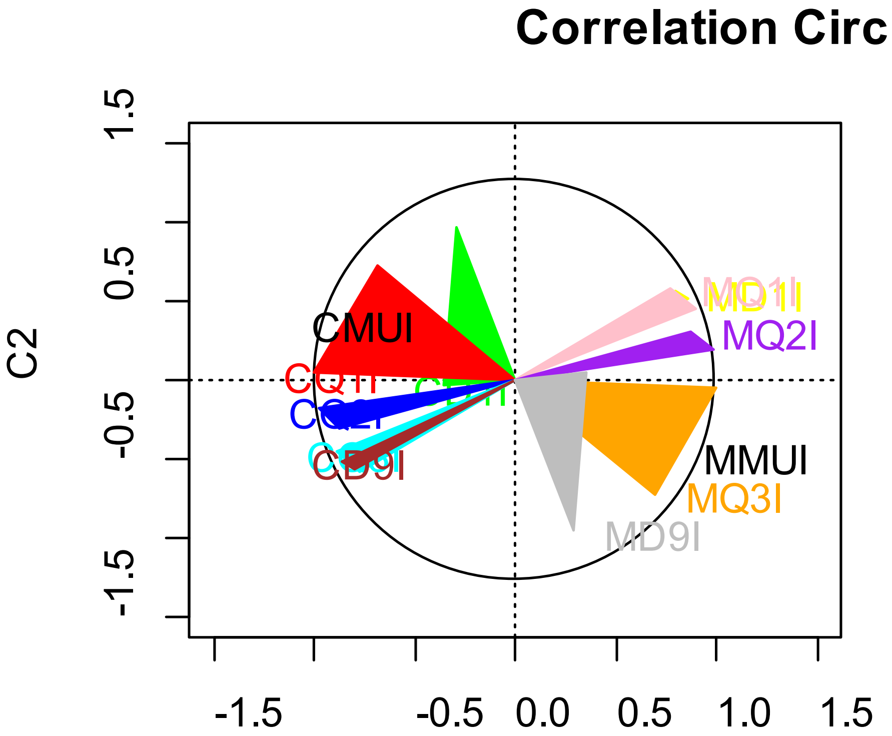
Fertilizer costs & associated gross margins

Using this duality of conditional quantile estimates, we can study other variations of the concepts of costs and margin
For example, the costs of fertilizers and the associated gross margin to compare the efficiency of production systems in relation to different inputs.

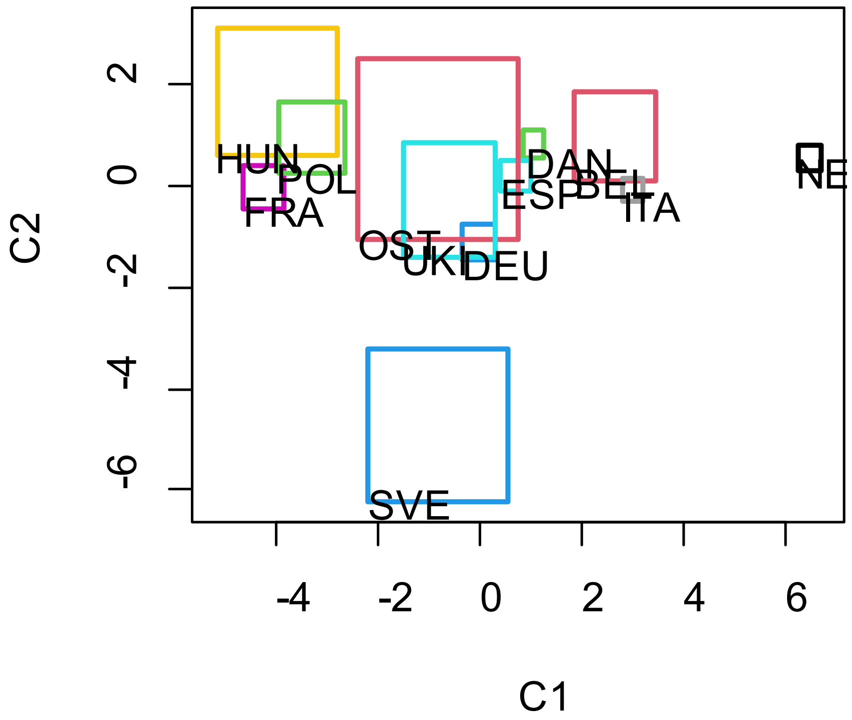
	CD1I	CQ1I	CQ2I	CQ3I	CD9I	MD1I	MQ1I	MQ2I	MQ3I	MD9I	CMUI	MMUI
BEL	[0.01 : 0.02]	[0.02 : 0.03]	[0.04 : 0.05]	[0.06 : 0.08]	[0.08 : 0.11]	[0.89 : 0.92]	[0.92 : 0.94]	[0.95 : 0.96]	[0.97 : 0.98]	[0.98 : 0.99]	[0.04 : 0.04]	[0.96 : 0.96]
DAN	[0.02 : 0.02]	[0.04 : 0.04]	[0.06 : 0.06]	[0.09 : 0.09]	[0.14 : 0.14]	[0.86 : 0.86]	[0.91 : 0.91]	[0.94 : 0.94]	[0.96 : 0.96]	[0.98 : 0.98]	[0.03 : 0.04]	[0.96 : 0.97]
DEU	[0.00 : 0.01]	[0.02 : 0.03]	[0.08 : 0.08]	[0.14 : 0.14]	[0.18 : 0.18]	[0.82 : 0.82]	[0.86 : 0.86]	[0.92 : 0.92]	[0.97 : 0.98]	[0.99 : 1.00]	[0.07 : 0.08]	[0.92 : 0.93]
ESP	[0.01 : 0.02]	[0.02 : 0.03]	[0.06 : 0.06]	[0.10 : 0.10]	[0.17 : 0.17]	[0.83 : 0.83]	[0.90 : 0.90]	[0.94 : 0.94]	[0.97 : 0.98]	[0.98 : 0.99]	[0.06 : 0.06]	[0.94 : 0.94]
FRA	[0.02 : 0.03]	[0.05 : 0.06]	[0.12 : 0.12]	[0.18 : 0.18]	[0.23 : 0.23]	[0.77 : 0.77]	[0.82 : 0.82]	[0.88 : 0.88]	[0.94 : 0.95]	[0.97 : 0.98]	[0.08 : 0.09]	[0.91 : 0.92]
HUN	[0.02 : 0.04]	[0.06 : 0.07]	[0.09 : 0.11]	[0.14 : 0.16]	[0.20 : 0.20]	[0.80 : 0.80]	[0.84 : 0.86]	[0.89 : 0.91]	[0.93 : 0.94]	[0.96 : 0.98]	[0.11 : 0.11]	[0.89 : 0.89]
ITA	[0.01 : 0.01]	[0.02 : 0.02]	[0.04 : 0.04]	[0.08 : 0.08]	[0.12 : 0.12]	[0.88 : 0.88]	[0.92 : 0.92]	[0.96 : 0.96]	[0.98 : 0.98]	[0.99 : 0.99]	[0.03 : 0.03]	[0.97 : 0.97]
NED	[0.00 : 0.00]	[0.00 : 0.01]	[0.01 : 0.01]	[0.02 : 0.02]	[0.03 : 0.03]	[0.97 : 0.97]	[0.98 : 0.98]	[0.99 : 0.99]	[0.99 : 1.00]	[1.00 : 1.00]	[0.01 : 0.01]	[0.99 : 0.99]
OST	[0.00 : 0.03]	[0.04 : 0.06]	[0.07 : 0.09]	[0.11 : 0.13]	[0.16 : 0.18]	[0.82 : 0.84]	[0.87 : 0.90]	[0.91 : 0.93]	[0.94 : 0.96]	[0.97 : 1.00]	[0.08 : 0.08]	[0.92 : 0.92]
POL	[0.02 : 0.03]	[0.05 : 0.06]	[0.09 : 0.10]	[0.15 : 0.16]	[0.22 : 0.23]	[0.77 : 0.79]	[0.84 : 0.85]	[0.90 : 0.91]	[0.94 : 0.95]	[0.97 : 0.98]	[0.09 : 0.09]	[0.91 : 0.91]
SVE	[-0.01 : 0.02]	[0.00 : 0.04]	[0.10 : 0.10]	[0.22 : 0.22]	[0.29 : 0.29]	[0.71 : 0.71]	[0.78 : 0.78]	[0.90 : 0.90]	[0.96 : 1.00]	[0.98 : 1.01]	[0.02 : 0.03]	[0.97 : 0.98]
UKI	[0.01 : 0.03]	[0.04 : 0.05]	[0.09 : 0.09]	[0.14 : 0.14]	[0.17 : 0.17]	[0.83 : 0.83]	[0.86 : 0.86]	[0.91 : 0.91]	[0.95 : 0.96]	[0.97 : 0.99]	[0.04 : 0.05]	[0.95 : 0.96]

```
> library(RSDA)
> iscm06<-read.sym.table("C:/Users/Desbois/Documents/SDA/SDA_2023/iscm06.txt", header=TRUE, sep='\t',dec=',', row.names=1)
> row.names(iscm06)<-c("BEL","DAN","DEU","ESP","FRA","HUN","ITA","NED","OST","POL","SVE","UKI")
> View(iscm06)
```

Centers sPCA



CD1+
CMU+
CQ1+
CQ2+
CQ3+
CD9+



MQ1+
MQ2+
CM2+
MMU+
MQ3+
MD9+

```
>library(rsda)
>res_centers<-sym.pca(cmis06,method='centers')
>plot(res_centers,choix="var")
>plot(res_centers,choix="ind")
```

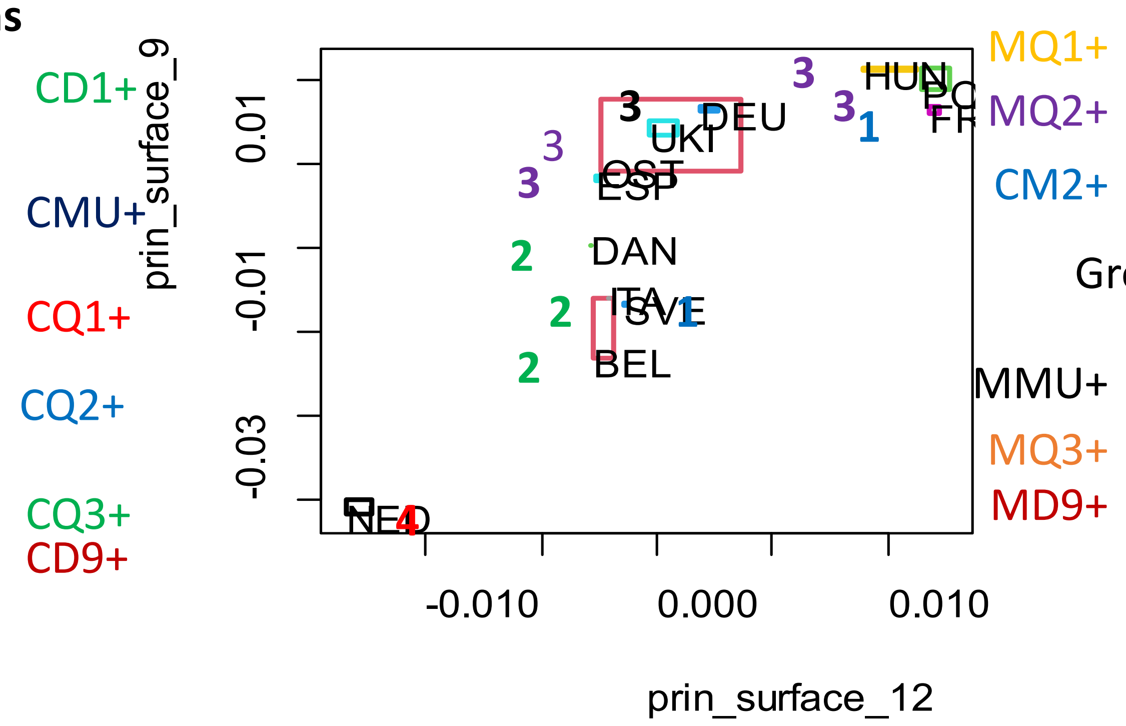
Duality of costs and margins: sym.pca & sym.kmeans

Specific Costs & Gross Margins

```
>library(rsda)
> res_kmeans4<-sym.kmeans(cmis06,k=4)
> res_kmeans4
Clustering vector:
BEL DAN DEU ESP FRA HUN ITA NED OST POL SVE UKI
 2   2   3   3   1   3   2   4   3   3   1   3
```

Specific Costs

```
> res_kmeans4<-sym.kmeans(cmis06[,c(6,7,8,9,10,12)],k=4)
>  res_kmeans4
Clustering vector:
BEL DAN DEU ESP FRA HUN ITA NED OST POL SVE UKI
 2   2   3   3   1   3   2   4   3   3   1   3
```



Groups : 1, 2, 3, 4

Gross Margins

```
> res_kmeans4<-sym.kmeans(cmis06[,c(1,2,3,4,5,11)],k=4)
> res_kmeans4
Clustering vector:
BEL DAN DEU ESP FRA HUN ITA NED OST POL SVE UKI
 1   1   4   4   3   4   1   2   4   4   3   4
```

with the permutation s=(1->2; 2->4; 3->1; 4->3)

```
 2   2   3   3   1   3   2   4   3   3   1   3
```

- `>res_princurv<-sym.pca(cmis06,method='principal.curves')`
- `>plot(res_princurv,choix="ind")`

DIV (Sodas 2.5): Margin Divisive Hierarchy

```
      +----   Classe 1  (Ng=1)
      !
  !----2- [D1 Margin <= 0.740000]
  !   !
  !   !          +----   Classe 3  (Ng=1)
  !   !          !
  !   !     !---7- [Q1 Margin <= 0.832500]
  !   !     !    !
  !   !     !    !     +----   Classe 8  (Ng=1)
  !   !     !    !     !
  !   !     !    !---10- [D1 Margin <= 0.790000]
  !   !     !          !
  !   !     !          +----   Classe 11  (Nd=1)
  !   !     !
  !   !----4- [D1 Margin <= 0.810000]
  !         !
  !         !          +----   Classe 5  (Ng=1)
  !         !          !
  !         !    !---9- [D1 Margin <= 0.825000]
  !         !    !     !
  !         !    !     +----   Classe 10  (Nd=1)
  !         !    !
  !         !---6- [Q1 Margin <= 0.872500]
  !              !
  !              !     +----   Classe 7  (Ng=1)
  !              !     !
  !              !---8- [Q1 Margin <= 0.892500]
  !                    !
  !                    +----   Classe 9  (Nd=1)
  !
!----1- [D1 Margin <= 0.845000]
  !
  !         +----   Classe 2  (Ng=2)
  !         !
  !   !----5- [D1 Margin <= 0.892500]
  !   !     !
  !   !     +----   Classe 6  (Nd=1)
  !   !
  !----3- [Q1 Margin <= 0.955000]
      !
      +----   Classe 4  (Nd=1)
```

{SVE} IF [D1 Margin <= 0.74] & [D1 Margin <= 0.845]
< IF [D9 Cost > 0.155] & [D9 Cost > 0.26] >

{FRA} IF [Q1 Margin <= 0.8325] & [D1 Margin <= 0.81] & [D1 Margin > 0.74] &[D1 Margin <= 0.845]

{POL} IF [D1 Margin <= 0.79] & [Q1 Margin > 0.8325] & [D1 Margin <= 0.81] & [D1 Margin > 0.74] & [D1 Margin <= 0.845]

{HUN} IF [D1 Margin > 0.79] & [Q1 Margin > 0.8325] & [D1 Margin <= 0.81 & [D1 Margin > 0.74] & [D1 Margin <= 0.845]

{DEU} IF [D1 Margin <= 0.825] & [Q1 Margin <= 0.8725] &[D1 Margin > 0.81] & [D1 Margin > 0.74] & [D1 Margin <= 0.845]

{UKI} IF [D1 Margin > 0.825] &[Q1 Margin <= 0.8725] & [D1 Margin > 0.81] & [D1 Margin > 0.74] & [D1 Margin <= 0.845]

{OST} IF [Q1 Margin <= 0.8925] & [Q1 Margin > 0.8725] & [D1 Margin > 0.81] & [D1 Margin > 0.74] & [D1 Margin <= 0.845]

{ESP} IF [Q1 Margin > 0.8925] & [Q1 Margin > 0.8725] & [D1 Margin > 0.81] & [D1 Margin > 0.74] & [D1 Margin <= 0.845] IS TRUE

{DAN,ITA} IF [D1 Margin <= 0.8925] & [Q1 Margin <= 0.955] & [D1 Margin > 0.845]

{BEL} IF [D1 Margin > 0.8925] & [Q1 Margin <= 0.955] & [D1 Margin > 0.845]

{NED} < IF [D9 Cost <= 0.155] & [Q3 Cost <= 0.045]) >
IF [Q1 Margin > 0.955] & [D1 Margin > 0.845]

DIV (Sodas 2.5): Cost Divisive Hierarchy

```
          +----   Classe 1 (Ng=1)
          !
     !----3- [Q3 Cost <= 0.045000]
     !    !
     !    !    +----   Classe 4 (Ng=1)
     !    !    !
     !    !----5- [D9 Cost <= 0.107500]
     !         !
     !         +----   Classe 6 (Nd=2)
     !
!----1- [D9 Cost <= 0.155000]
     !
     !              +----   Classe 2 (Ng=1)
     !              !
     !         !---7- [Q2 Cost <= 0.070000]
     !         !    !
     !         !    !    +----   Classe 8 (Ng=1)
     !         !    !    !
     !         !    !---8- [Q3 Cost <= 0.130000]
     !         !         !
     !         !         !    +----   Classe 9 (Ng=1)
     !         !         !    !
     !         !         !---9- [D1Cost <= 0.012500]
     !         !              !
     !         !              +----   Classe 10 (Nd=1)
     !         !
     !    !----4- [D9 Cost <= 0.190000]
     !    !    !
     !    !    !         +----   Classe 5 (Ng=1)
     !    !    !         !
     !    !    !    !---10- [D1Cost <= 0.027500]
     !    !    !    !    !
     !    !    !    !    +----   Classe 11 (Nd=1)
     !    !    !    !
     !    !    !---6- [Q3 Cost <= 0.167500]
     !    !         !
     !    !         +----   Classe 7 (Nd=1)
     !    !
     !----2- [D9 Cost <= 0.260000]
          !
          +----   Classe 3 (Nd=1)
```

{NED: Nederland} **IF [D9 Cost <= 0.155] & [Q3 Cost <= 0.045]**
< IF [Q1 Margin > 0.955] & [D1 Margin > 0.845]>

{BEL: Belgium} **IF [D9 Cost <= 0.155] & [Q3 Cost > 0.045] & [D9 Cost <= 0.1075]**

{DAN: Danemark; ITA: Italy } **IF [D9 Cost <= 0.155] & [Q3 Cost > 0.045] & [D9 Cost > 0.1075]**

{ESP: Spain} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [D9 Cost <= 0.19] & [Q2 Cost <= 0.07]**

{OST: Austria} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [D9 Cost <= 0.19] & [Q2 Cost > 0.07] & [Q3 Cost <= 0.13]**

{DEU: Germany} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [D9 Cost <= 0.19] & [Q2 Cost > 0.07] & [Q3 Cost > 0.13] & [D1Cost <= 0.0125]**

{UKI: United Kingdom} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [Q2 Cost > 0.07] & [Q3 Cost > 0.13] & [D1 Cost > 0.0125] & [D9 Cost <= 0.19]**

{POL: Poland} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [D9Cost > 0.19] & [Q3 Cost <= 0.1675] & [D1Cost <= 0.0275]**

{HUN: Hungary} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [D9 Cost > 0.19] & [Q3 Cost <= 0.1675] & [D1 Cost > 0.0275]**

{FRA: France} **IF [D9 Cost > 0.155] & [D9 Cost <= 0.26] & [D9 Cost > 0.19] & [Q3 Cost > 0.1675]**
<IF [D1 Margin <= 0.74] & [D1 Margin <= 0.845]>

{SVE: Sweden} **IF [D9 Cost > 0.155] & [D9 Cost > 0.26]**

Symbolic Mean & Symbolic Variances

- Interval data $(i = 1, \ldots, n)$ $([a_{i1}, b_{i1}], \ldots, [a_{ip}, b_{ip}])'$

- Symbolic mean:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} \frac{a_{ij} + b_{ij}}{2}$$

- Symbolic variance:

$$s_{jj}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{a_{ij} + b_{ij}}{2} - \bar{x}_j^{(1)} \right)^2 .$$

de Carvalho & al. (2006)

$$s_{jj}^{(2)} = \sum_{i=1}^{n} \frac{(a_{ij} - \bar{x}_j)^2 + (b_{ij} - \bar{x}_j)^2}{2n}$$

Bertrand & Goupil (2000)

$$s_{jj}^{(3)} = \sum_{i=1}^{n} \frac{b_{ij}^2 + b_{ij}a_{ij} + a_{ij}^2}{3n} - \left[\sum_{i=1}^{n} \frac{b_{ij} + a_{ij}}{2n} \right]^2$$

Symbolic Covariances

- Billard & Diday (2003)

$$s_{jl}^{(1)} = \sum_{i=1}^{n} \frac{(b_{ij} + a_{ij})(b_{il} + a_{il})}{4n} - \bar{x}_j \, \bar{x}_l,$$

- Billard (2008)

$$s_{jl}^{(3)} = \frac{1}{6n} \sum_{i=1}^{n} \Big[(a_{ij} - \bar{x}_j)(b_{il} - \bar{x}_l) + (b_{ij} - \bar{x}_j)(a_{il} - \bar{x}_l) + 2(a_{ij} - \bar{x}_j)(a_{il} - \bar{x}_l) + 2(b_{ij} - \bar{x}_j)(b_{il} - \bar{x}_l) \Big].$$

What properties for various definitions, algorithms & representations in …

- **Adding a constant?**
- **Multiplication by a constant ?**
- **Mixing intervals ?**

So, as a user of such methods, it is our responsibility to check the relevance of these properties to our analytical objectives

… and let me introduce one last personal request

Edwin, could you inspire us in the analysis of our environmental footprint, please ?

Bibliographical References

Butault J.P. (1991). Coûts, prix et revenus selon les produits dans les agricultures européennes en 1984, 1985 et 1986 : résultats généraux du modèle. *Actes et Communications n°5*, INRA-INSEE, Paris, pp. 1331.

Butault J.P., Hassan C.R., Reignier E. (1988). *Les coûts de production des principaux produits agricoles dans la CEE*. Luxembourg, Office of Official Publications of the European Communities.

Commission européenne (2013). Présentation de la réforme de la PAC 2014-2020, *Brief : Les perspectives de la politique agricole*, n°5, décembre, 11 p.

Commission européenne (2017). *The Future of Food and Farming*, COM 713, 29 novembre, 26 p.

Cameron A.C., Trivedi P.K. (2006). *Microeconometrics. Methods and Applications*. Cambridge University Press, New-York.

Cazes, P., Chouakria, A., Diday, E., Schektman Y. (1997) Extensions de l'analyse en composantes principales à des données de type intervalle. *Revue de Statistique Appliquée*, n°24, pp. 5-24.

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. (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.

Desbois D. (2023) « Coûts spécifiques et marges brutes du blé en Europe: pratiques innovantes d'estimations », NOV'AE, https://doi.org/10.17180/novae-2023-NO-art08.

Desgraupes B. (2017). Clustering Indices, *Vignette R*, CRAN.

D'Haultfoeuille X., Givord P. (2014). La régression quantile en pratique. *Economie et statistique*, n°471, pp. 85-111.

Eurostat (2012). *Agriculture, fishery and forestry statistics, Main results – 2010-11*, 221 p.

Fuentes M., Chavent M. (2015). Clustering divisif monothétique, *Vignette R*, 4e Rencontre R.

He X., Hu F. (2002). Markov Chain Marginal Bootstrap. *Journal of the American Statistical Association*, vol. 97, pp. 783-795.

Koenker R., Bassett G. (1978). Regression quantiles. *Econometrica,* n°46, pp. 33-50.

Koenker R., Bassett G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica,* vol. 50, n°1, pp. 43-61.

Koenker, R., Machado, J. A. F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association*, vol. 94, n°448, pp. 1296-1310.

Koenker R., Zhao Q. (1994). L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics*, n° 3, pp. 223-235.

SAS Institute (2008) *SAS/STAT 9.2 User's Guide. The QUANTREG Procedure, Chapter 72*, SAS, pp. 5352-5425, Institute Cary NC, USA.

SODAS (2004) Analysis System of Symbolic Official Data, release 2.5, www.info.fundp.ac.be/asso/sodaslink.htm.

Surry Y., Desbois D., Butault J.-P. (2013). *Quantile Estimation of Specific Costs of Production*. FACEPA, D8.2.

Trouvé A., Berriet-Solliec M. (2008). 2nd pilier de la Politique Agricole Commune et régionalisation : vers plus de cohésion ? *Revue d'Économie Régionale & Urbaine*, 2008/1, pp. 87-108.

Ward, J. H., Jr. (1963), Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, n° 58, pp. 236–244.