High Dimensional-Non Additive Municipalities' Performance Index in a Two Levels Approach

L. FARNIA[a,b,c], A. GIORGIO[c], F. RAVONI[c]

[a]Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy

[b]International Center for Climate Governance, Italy

[c]Fondazione Eni Enrico Mattei, Italy

The Role of National Statistical Offices in the production of Leading, Composite and Sentiment Indicators

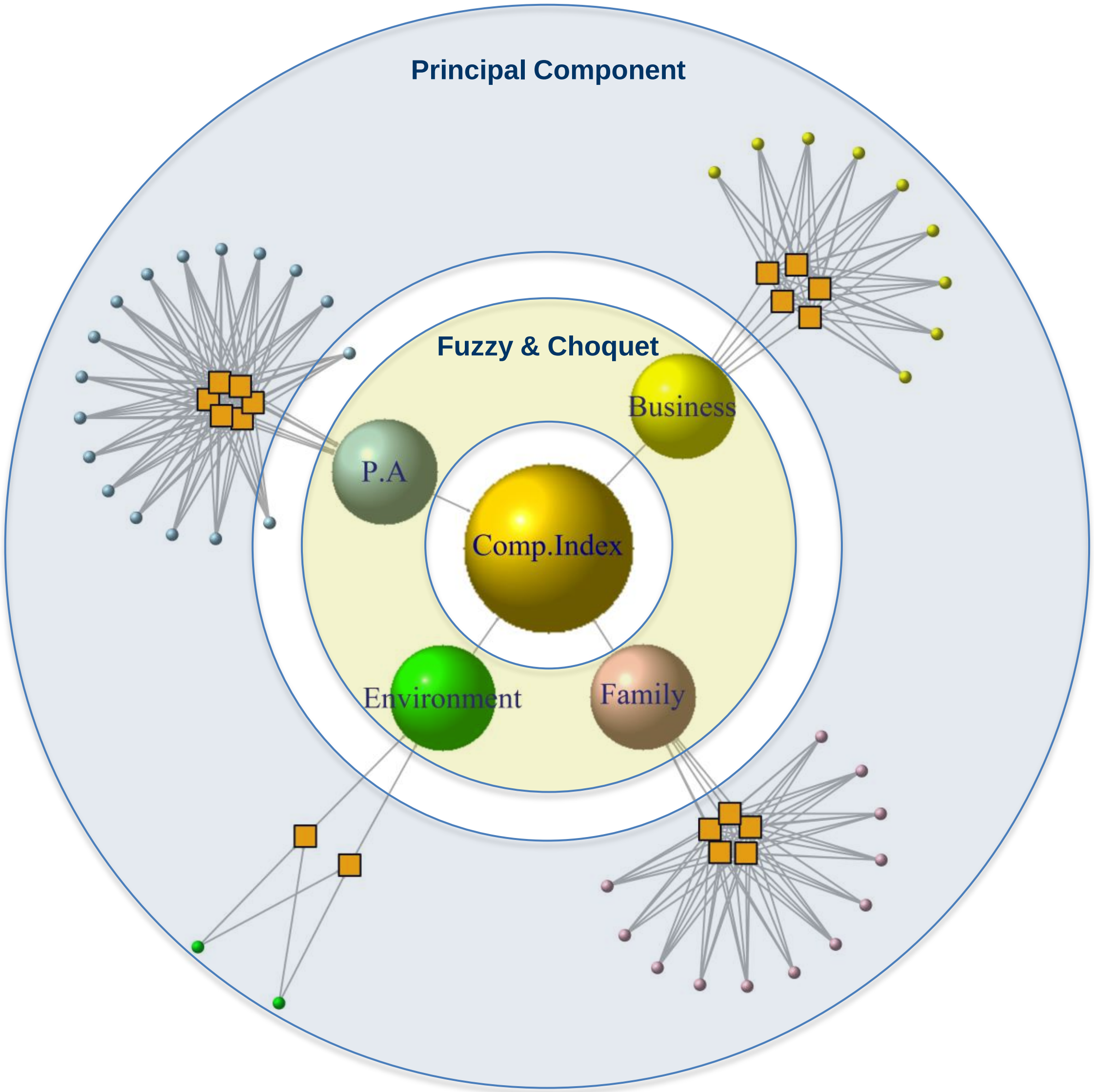
UNECE

Geneva, 6 - 7 July 2017

Outline

- High Dimension
  - Weighting technique
    - Principal Component
  - Non Additive Aggregation Function
    - Fuzzy Measures and Choquet Integral
    - Preferences' Elicitation

Composite Index Structure



Composite Index structure

- 4 Dimensions
- 17 Indicators for *P.A*.
- 2 Indicators for *Environment*
- 12 Indicators for *Family*
- 9 Indicators for *Business*
- 131 Municipalities evaluated

Computation techniques

- Two levels approach; first within, then between dimensions.
- Additive aggregation within Dimensions (PC technique)
- Non additive aggregation between dimensions (Fuzzy measures and Choquet Integral)

Data Management

Dimensions:

- Family: indicators related to demographic dimensions, households' income, internal and external migration flows;
- Business: indicators related to total number of business, business types, survival rate;
- P. A.: indicators related to different municipalities balance sheet items;
- Environment: two indicators related to recycled urban waste and total urban waste

Data have been:

- Transformed in relative terms to guarantee comparability among municipalities (by number of citizens, firms, taxpayers, etc.)
- Transformed to guarantee the same polarity of indicators (the higher the measure of an indicator the better it is)
- Standardized to ensure zero mean and unit variance (PCA)

Weighting and Aggregation

Principal component Technique

Indicators' Weight inside a Component

- The weight of each indicator in the *k*-th component is proportional to its absolute influence on that component (not as a function of the explained variance):

$$z_k = a_{k1}x_1 + \cdots + a_{kn}x_n$$

$$w_{x_j;z_k} = \frac{|a_{ki}|}{\sum_j^n |a_{kj}|} \quad \textit{not } w_{x_j;z_k} = {a_{kj}}^2$$

Components' Weight Inside the Dimension

- The weight of each component in the dimension is the same; it is not hence a function of the variance explained.

Two reasons:

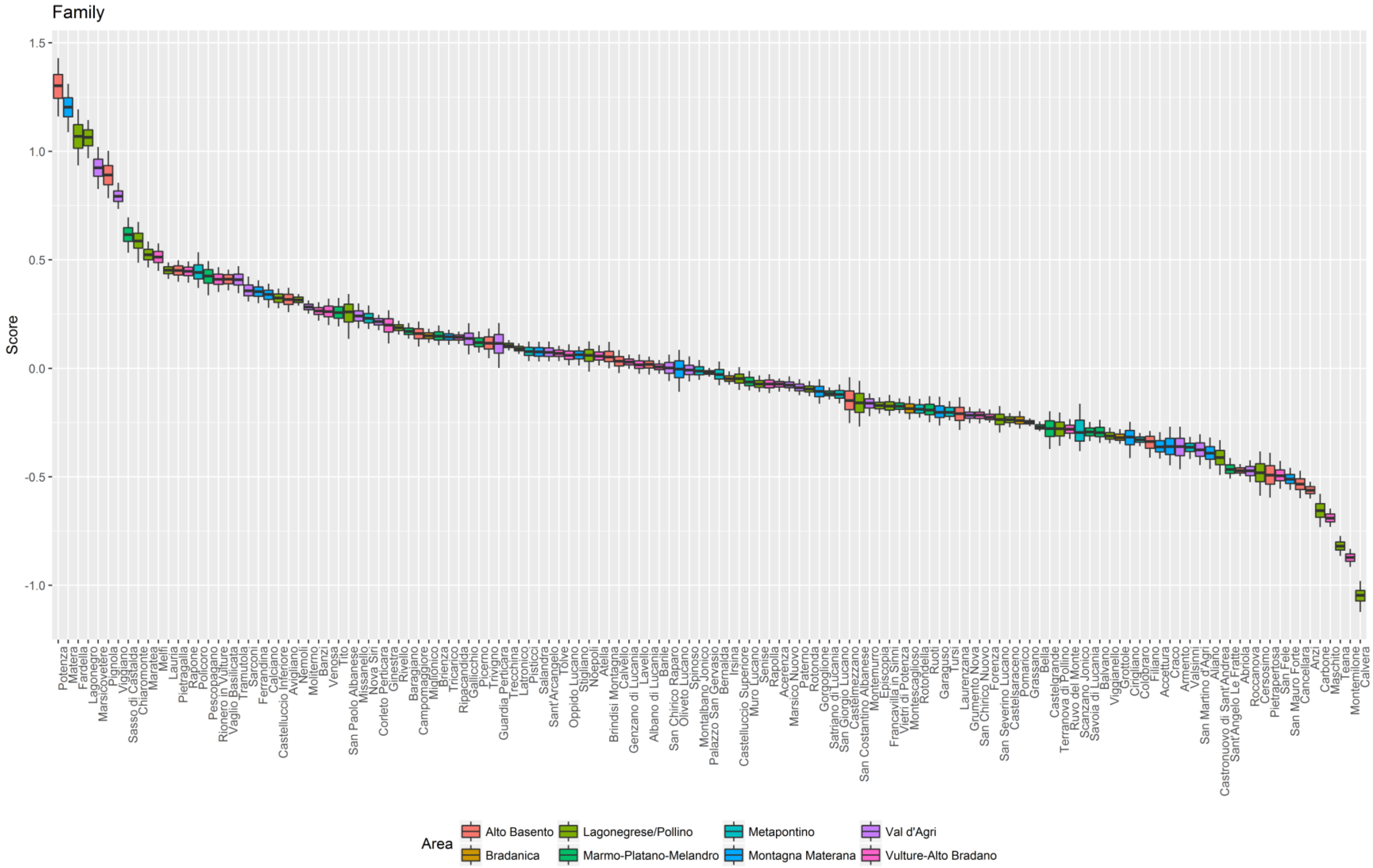
- In the context of C.I. a component cannot be more important than others;
- Mitigate the risk of overweighting correlated variables.

Example:

$$\mathbf{R_x} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -0.5 & 0 \\ -0.5 & 0 \\ -0.5 & 0 \\ -0.5 & 0 \end{bmatrix}; \boldsymbol{\lambda}' = \begin{bmatrix} 4 & 1 \end{bmatrix}$$

According to variance explained, the first component weights 80% and the second 20%

Family Dimension



Weighting and Aggregation

Fuzzy Measures and Choquet Integral

Weighting and Aggregation

Many composite indices existing in the literature assume implicitly the preferential independence among criteria assumption [D. Scott, P. Suppes, 1958; Fishburn, 1970; Marichal, Roubens 1998]

↔ *Arithmetic Mean, Weighted Mean*.

Depending on the objective of the composite indicator, this assumption could be limited or unrealistic because indicators may have some form of synergies and redundancies that should be taken into account in the aggregation phase

↔ *Geometric Mean, Constant Elasticity of Substitution*

However these aggregator functions are not flexible enough.

Fuzzy Measures and Choquet Integral

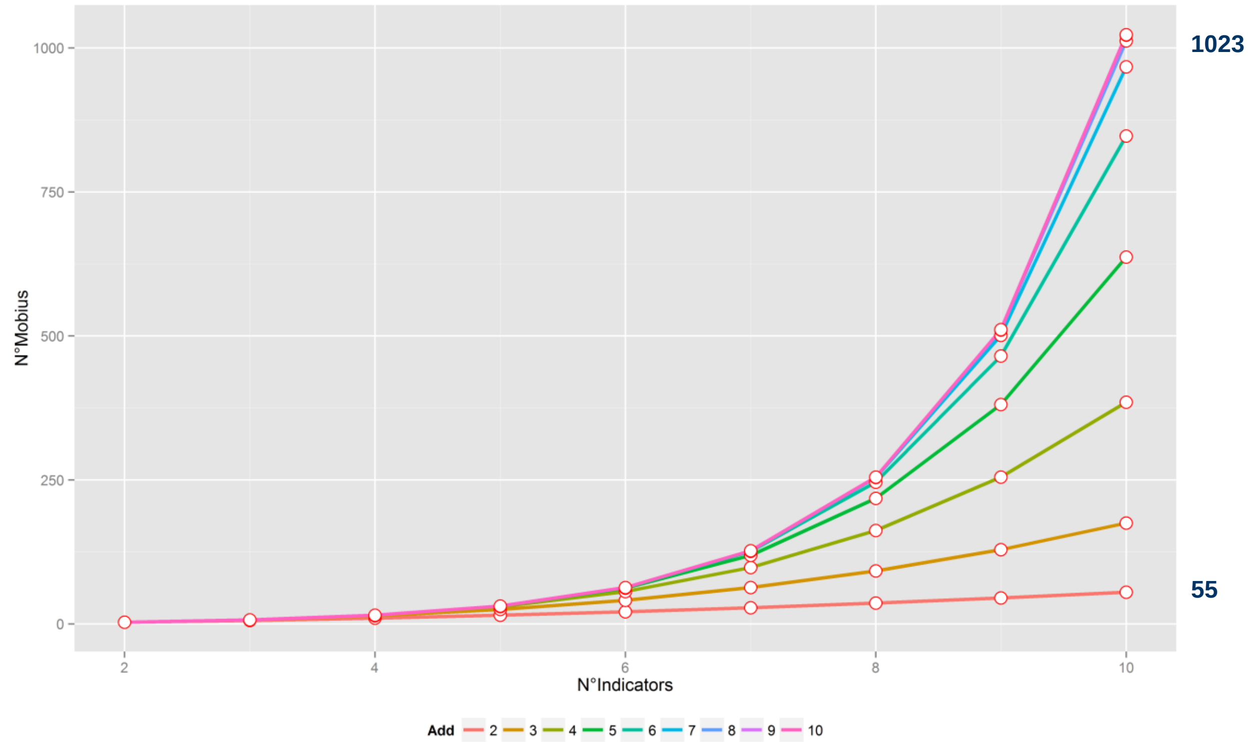
- Allow to model flexible interactions (from redundancy/substitutability to synergy/complementarity) and to consider hence complex dependencies among criteria, hence relaxing the *preferential independence assumption;*

- They can exactly replicate:
  - Weighted mean;
  - Ordered Weighted Averaging Operators (OWA);
  - Minimum and Maximum Operator.

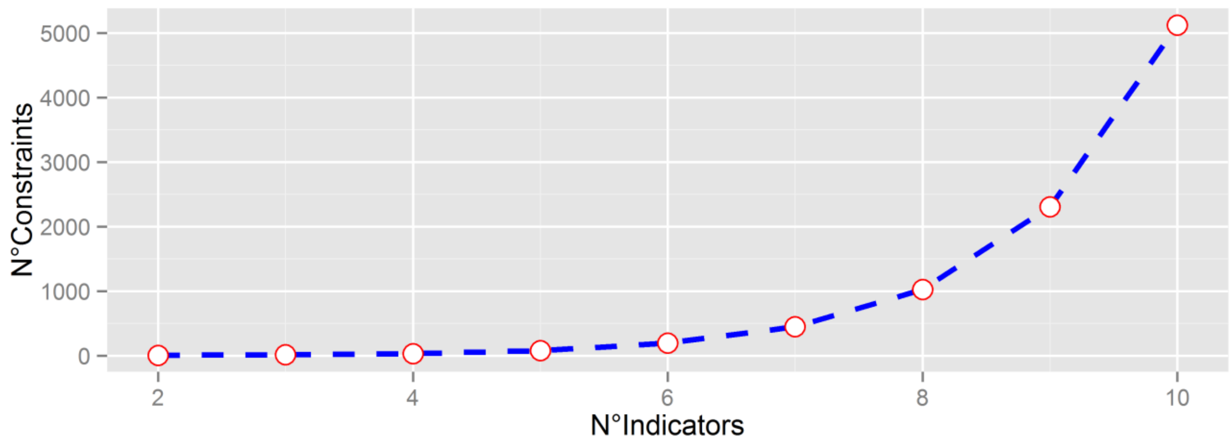
- They are a powerful tool to elicit Experts' preferences.

Exponential Complexity

$$N°Möbius = \sum_{j=1}^{k} \binom{n}{j}$$



$$N°Cons. = n \cdot 2^{(n-1)}$$



Preferences' Elicitation

LS Experts' Preference Elicitation

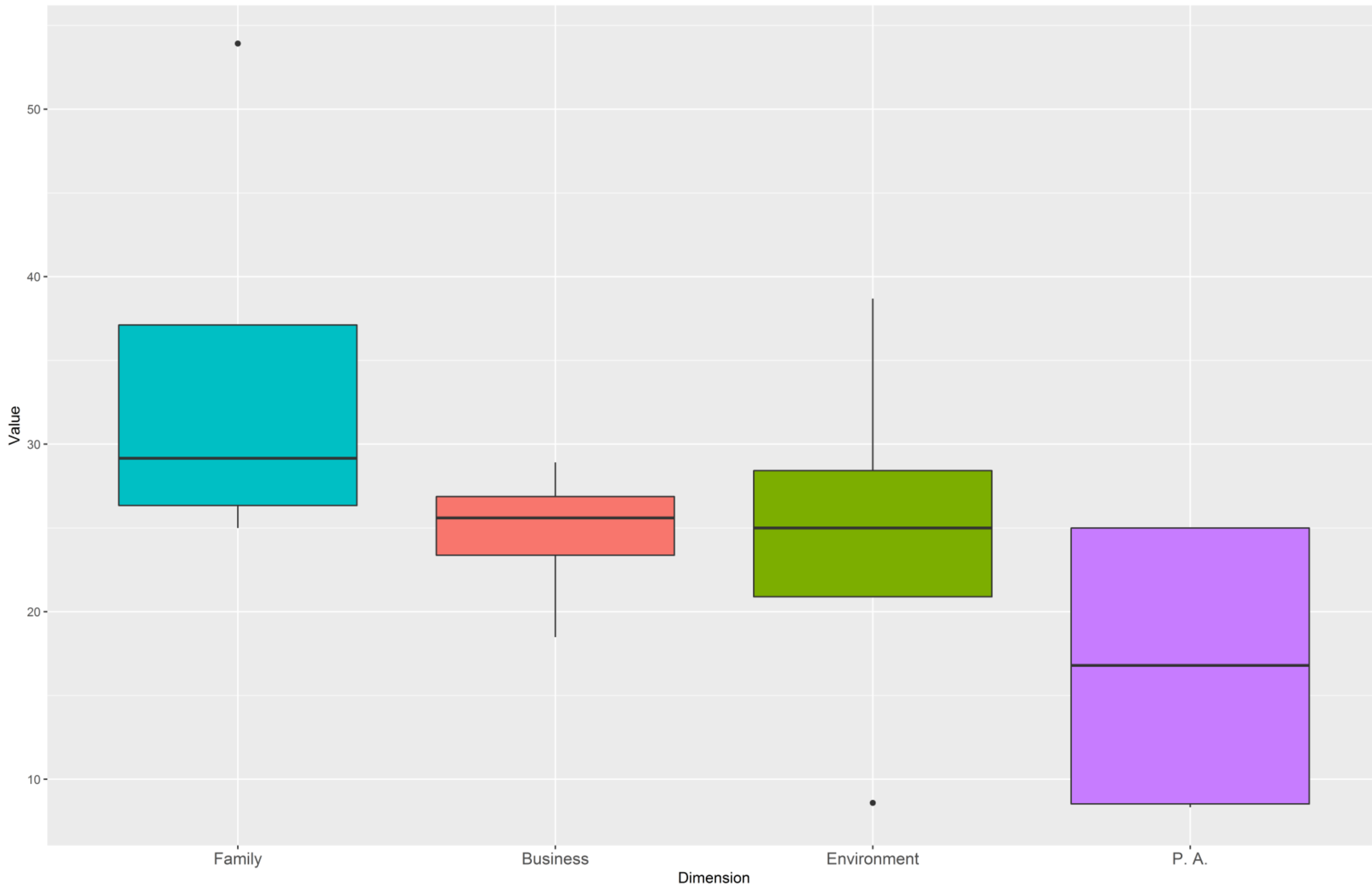
	Criteria				**Expert**
Alternative	*Family*	*Business*	*P.A.*	*Environment*	*Overall Evaluation*
1	Excellent	Very bad	Very bad	Very bad	-
2	Very bad	Excellent	Very bad	Very bad	-
3	Very bad	Very bad	Excellent	Very bad	-
4	Very bad	Very bad	Very bad	Excellent	-
5	Excellent	Excellent	Very bad	Very bad	-
6	Excellent	Very bad	Excellent	Very bad	-
7	Excellent	Very bad	Very bad	Excellent	-
8	Very bad	Excellent	Excellent	Very bad	-
9	Very bad	Excellent	Very bad	Excellent	-
10	Very bad	Very bad	Excellent	Excellent	-

$$Expert\ V. = \{Very\ Sat., Sat., Nor\ Sat\ or\ Disat., Disat., Very\ Dis.\}$$

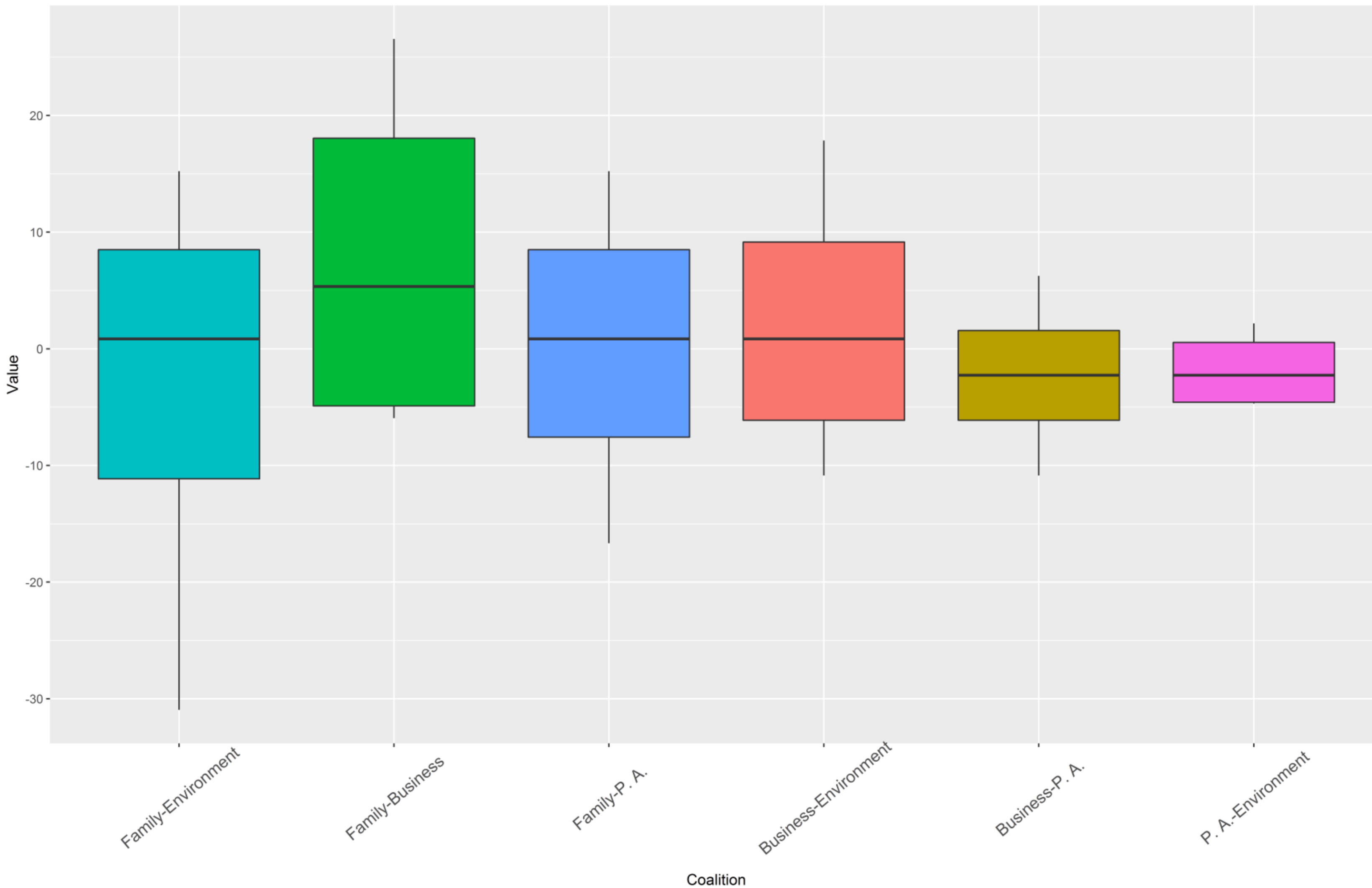
Test for answer’s consistency

Alternative	**Criteria**				**Expert** *Overall Evaluation*
	Family	*Business*	*P.A.*	*Environment*	
1	**Excellent**	Very bad	Very bad	Very bad	**Very Satisf.**
2	Very bad	Excellent	Very bad	Very bad	-
3	Very bad	Very bad	Excellent	Very bad	-
4	Very bad	Very bad	Very bad	Excellent	-
5	**Excellent**	**Excellent**	Very bad	Very bad	**Satisf.**
6	Excellent	Very bad	Excellent	Very bad	-
7	Excellent	Very bad	Very bad	Excellent	-
8	Very bad	Excellent	Excellent	Very bad	-
9	Very bad	Excellent	Very bad	Excellent	-
10	Very bad	Very bad	Excellent	Excellent	-

Shapley Value



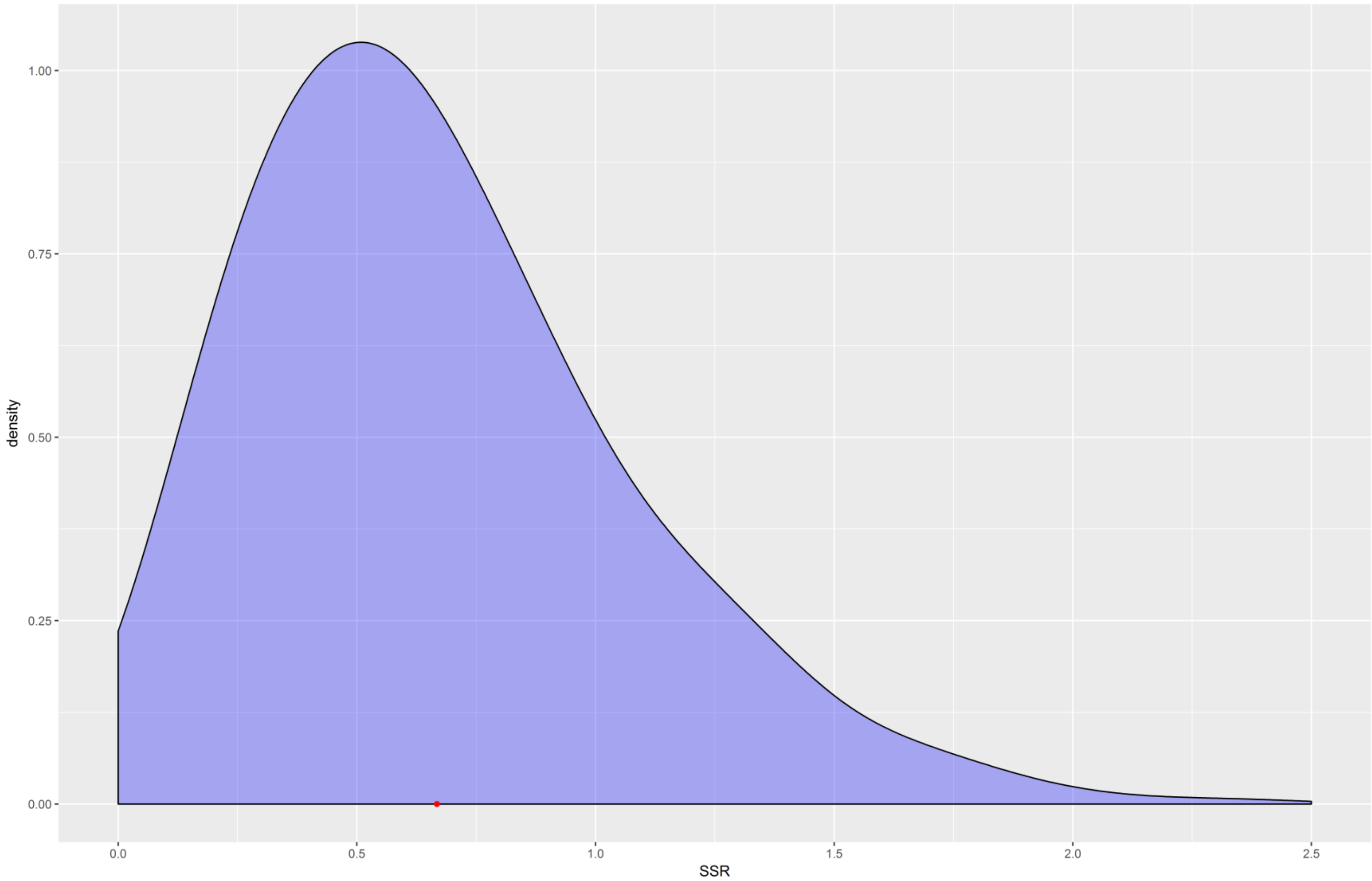
Interaction Index



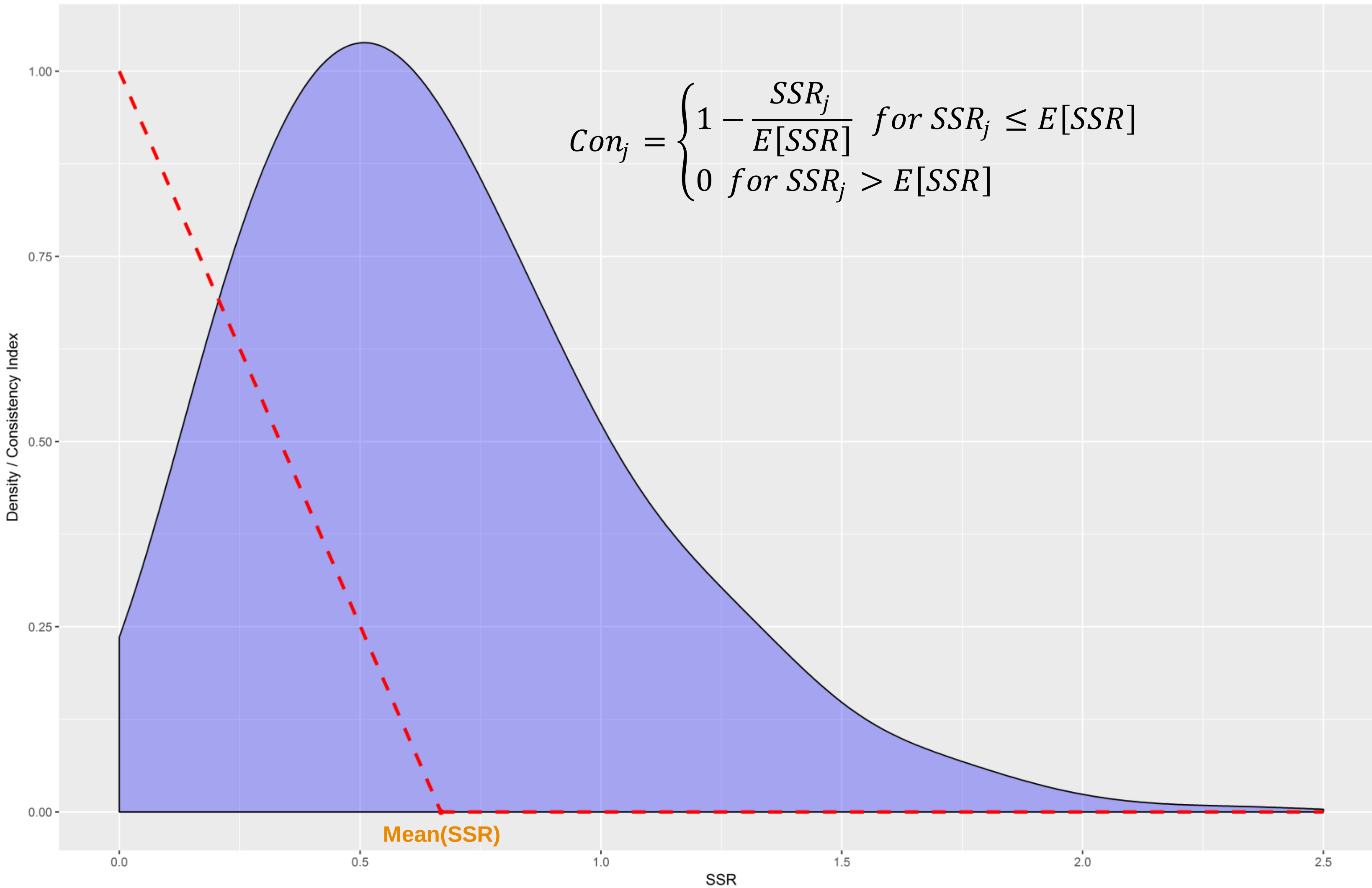
Aggregation of Individual Opinions

Consistency Index

SSR Density Distribution



Consistency Index



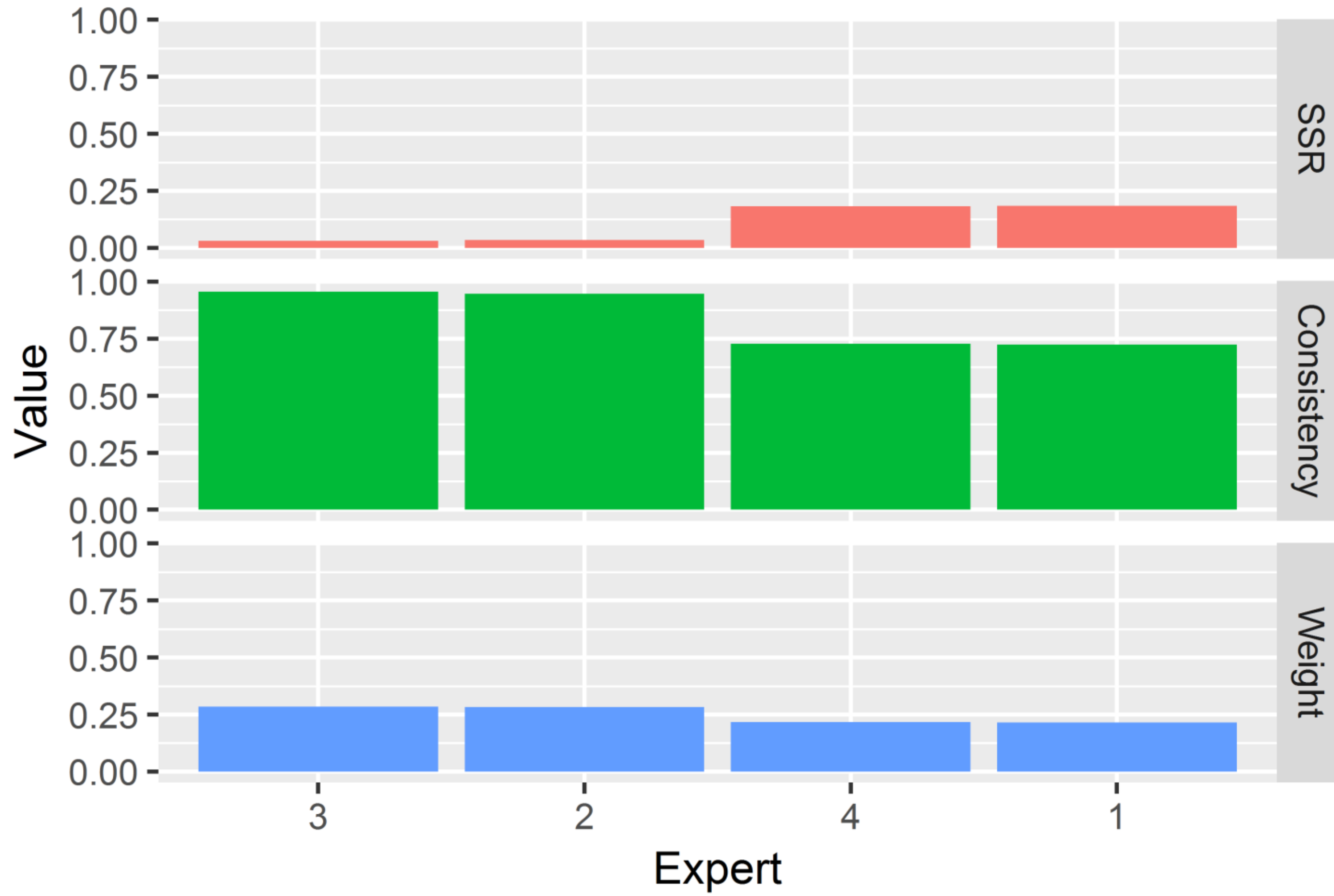
Aggregation of Individual Opinions

$$\mathbf{y}_j = \mathbf{C}\mathbf{h}_j + \boldsymbol{\varepsilon}_j$$

$$SSR_j = \boldsymbol{\varepsilon}_j{}'\boldsymbol{\varepsilon}_j$$

$$Con_j = f\left(SSR_j \middle| E[SSR]\right)$$

$$w_j = \frac{Con_j}{\sum_{j=1}^{d} Con_j}$$



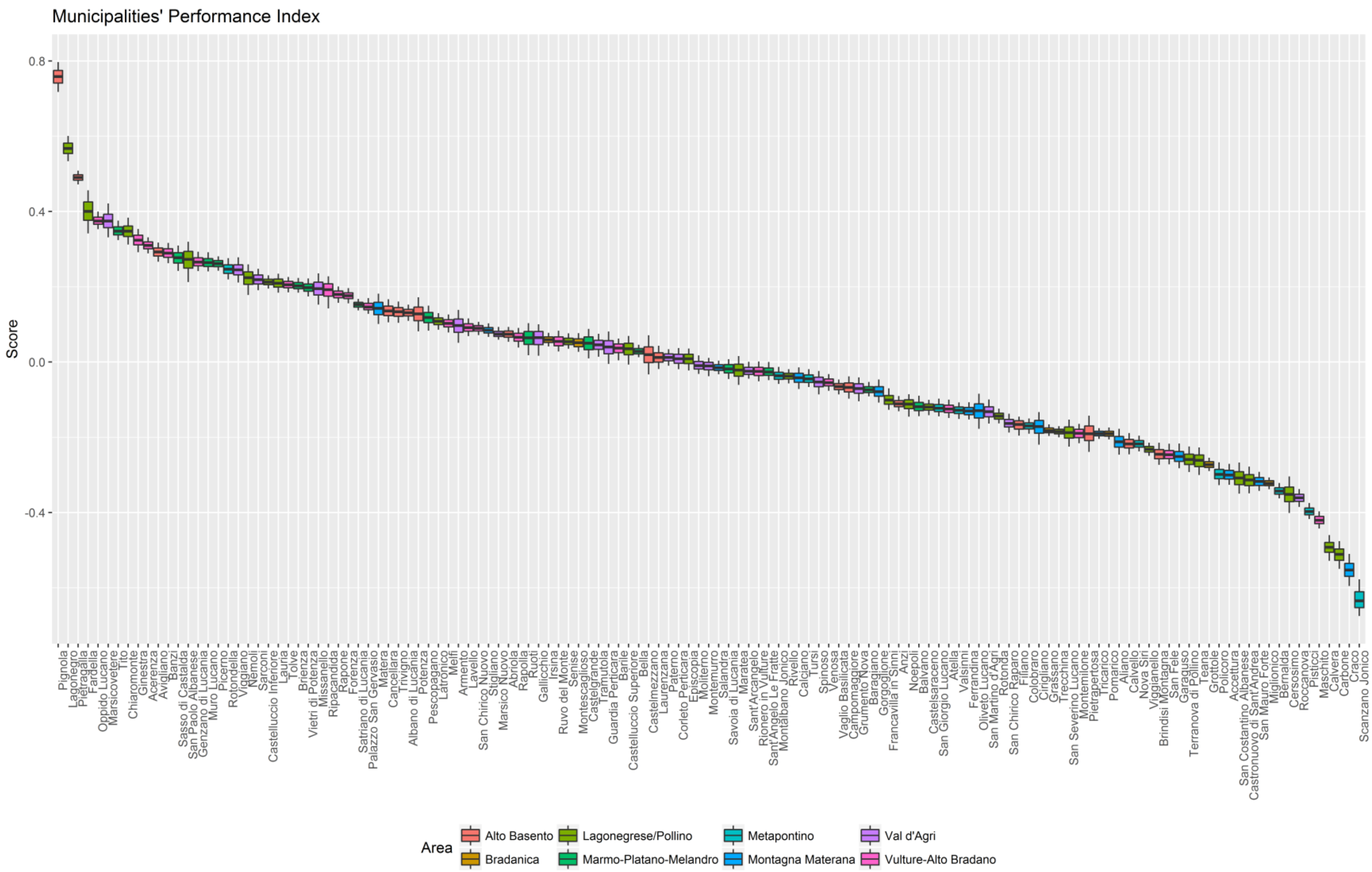
$$\Longrightarrow \quad m^*\{T\} = \sum_{j=1}^{d} w_j m_j\{T\}$$

Dimension	Shapley %
Family	35.40
Business	24.50
Environment	23.32
P.A.	16.76

Coalition	Interaction %
Family - Business	9.56
Family – P.A.	1.51
Family – Environm.	-1.56
Business – P.A.	-2.31
Business – Environm.	1.53
P.A. - Environment	-1.69

Non Additive Aggregation of Dimensions

Municipalities’ Performance Index



Conclusions

- In high dimension problems, the use of P. C. could be a good solution to take into account the correlation among variables and hence to reduce the risk of overweighting their influence on the C. I.
- More consensus among researchers is needed to derive the weights of individual indicators by means of P.C.
- Fuzzy Measures and Choquet integral are powerful tool to model interactions among indicators
- It is important to take into account the consistency of the respondents in evaluating a questionnaire

Thank you

luca.farnia@feem.it

Extra Slides

Fuzzy Measures and Choquet Integral

A Fuzzy measures or Non Additive Measure, defined over the set of criteria $N = \{1,2, \dots, n\}$, is a set function $\mu: 2^n \rightarrow [0,1]$ satisfying the following boundary and monotonicity conditions:

$$\begin{cases} \mu(\emptyset) = 0 \\ \mu(N) = 1 \\ \mu(S) \leq \mu(T) \leq 1, \qquad \forall\, S \subseteq T \subseteq N \end{cases}$$

Choquet Integral:

$$Ch(x_1, \dots, x_n) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mu\left(A_{(i)}\right)$$

$$x_{(0)} \equiv 0; \;\; x_{(1)} \leq \cdots \leq x_{(n)}; \;\; A_{(i)} = \left\{x_{(i)}, \dots, x_{(n)}\right\}$$

Fuzzy Measures

A Fuzzy measures or Non Additive Measure, defined over the set of criteria $N = \{1,2, \dots, n\}$, is a set function $\mu: 2^N \to [0,1]$ satisfying the following boundary and monotonicity conditions:

$$\begin{cases} \mu(\emptyset) = 0 \\ \mu(N) = 1 \\ \mu(S) \leq \mu(T) \leq 1, \qquad \forall\, S \subseteq T \subseteq N \end{cases}$$

Example: Two criteria $N = \{1,2\}$

$\mu\{\emptyset\} = 0;\ \mu\{1,2\} \geq \mu\{1\};\ \mu\{1,2\} \geq \mu\{2\};\ \mu\{1,2\} \equiv 1$

- If $\mu\{1\} + \mu\{2\} > \mu\{1,2\}$ sub-additive $\rightarrow \{1,2\}$ substitutes
- If $\mu\{1\} + \mu\{2\} < \mu\{1,2\}$ super-additive $\rightarrow \{1,2\}$ complements
- If $\mu\{1\} + \mu\{2\} = \mu\{1,2\}$ additive $\rightarrow \{1,2\}$ independendent

The Möbius representation of Fuzzy Measures allow to reduce the problem complexity (k-additivity); the monotonicity and boundary conditions are:

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{T \subseteq N} m(T) = 1 \\ \sum_{\substack{T \subseteq S \\ T \ni i}} m(T) \geq 0, \qquad \forall\, S \subseteq N, \qquad \forall\, i \in S \\ m(T) = 0 \;\; \forall\, t > k \end{cases}$$

The Choquet Integral with respect to Möbius representation:

$$C_m(x_1, \dots, x_n) = \sum_{T \subseteq N} m(T) \bigwedge_{i \in T} x_i$$

Example – Veto Effect

i) *Mathematics is a veto criterion*

ii) *Statistics $\succ$ Literature*

Student	***Math.***	***Stat.***	***Liter.***	***Choquet***
a	10	10	10	10
b	10	10	9	9.7
c	10	9	10	9.5
d	7	8	9	7
e	7	10	5	6.4
f	7	5	10	6
g	6	7	7	6
h	6	10	5	5.7
i	6	5	10	5.5
l	5	10	10	5