UNIVERSITY OF BRESCIA

BODAI – LAB

BIG&OPEN DATA INNOVATION LABORATORY

A mixture model for ordinal variables measured on semantic differential scales

5th International Conference on Econometrics and Statistics (EcoSta 2022)

4-6 June 2022, Ryukoku University, Kyoto, Japan

M. Manisera, P. Zuccolotto

University of Brescia

Kyoto, June 4-6, 2022

Motivation and aim

Subjective perceptions and attitudes are usually measured by questionnaires, with questions having **ordered response categories**

→ **Semantic differential scales**

Please rate your customer experience:

Sad	Neutral	Happy

Sad			Neutral			Happy
1	2	3	4	5	6	7

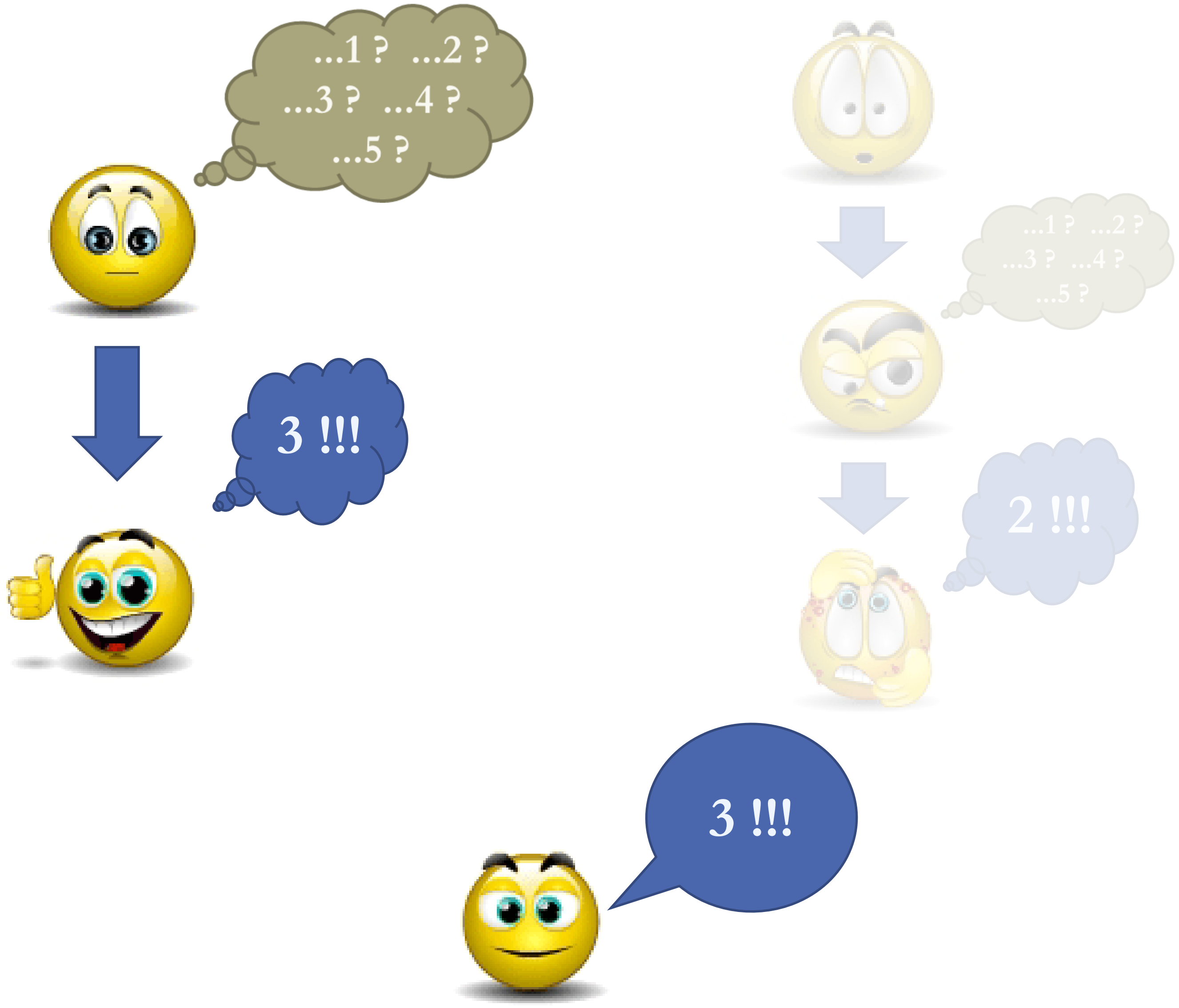
Motivation and aim

Aim: to propose a **new mixture model**

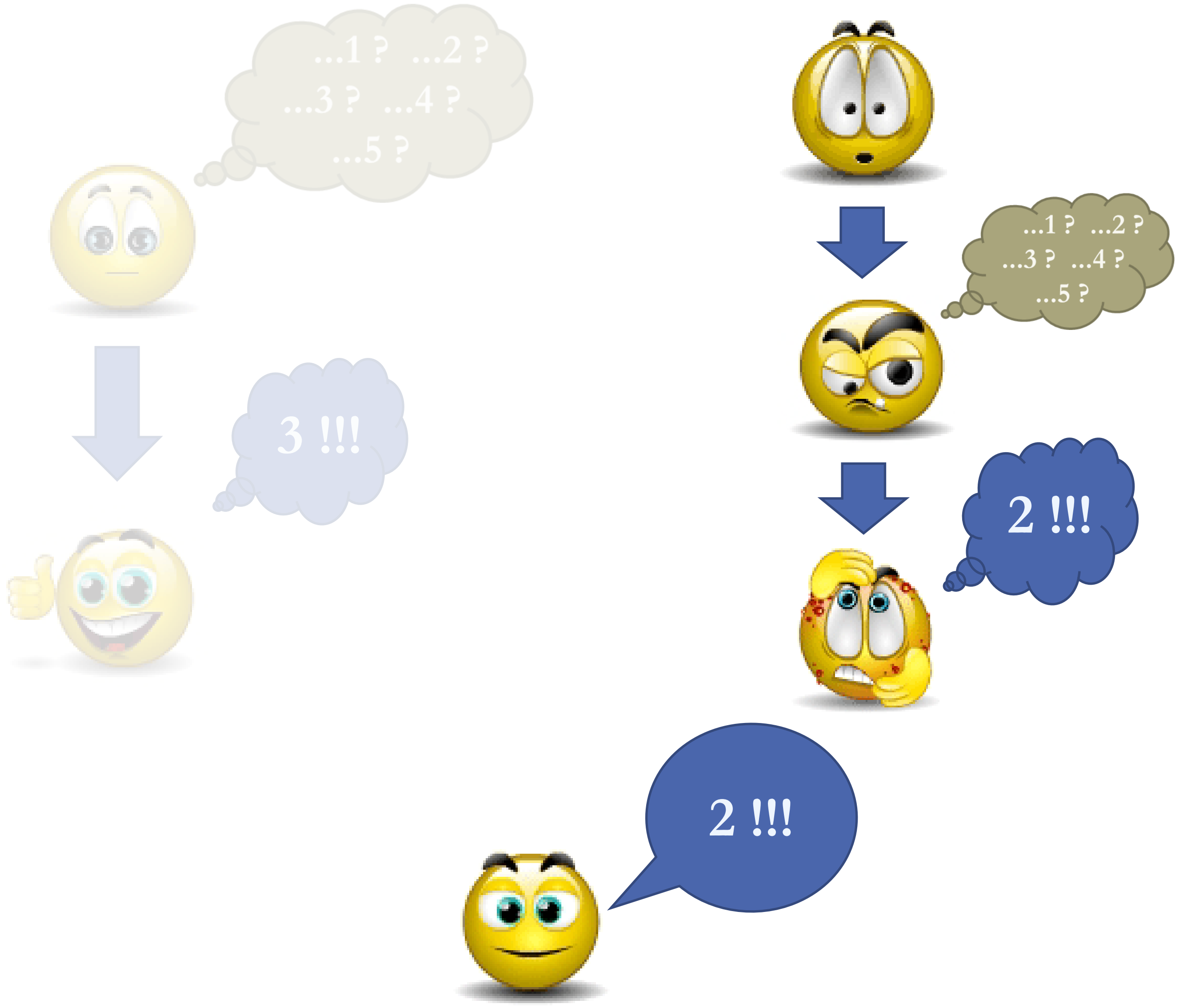
in the class of **CUB models**

starting from a general framework for the **Decision Process** that drives individuals to answer questions with ordered response categories

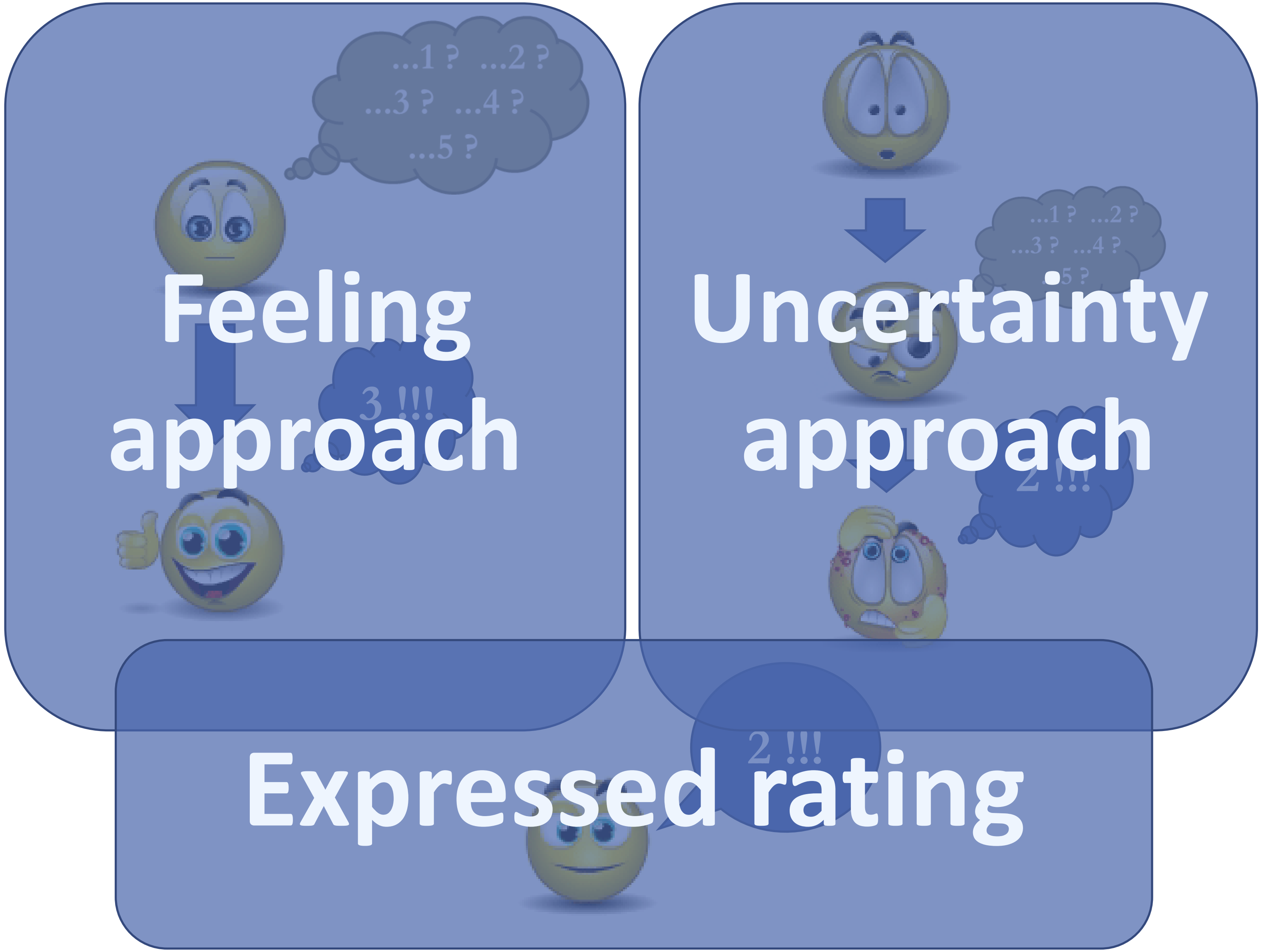
The unconscious DP: general framework



The unconscious DP: general framework



The unconscious DP: general framework



The unconscious DP: general framework

reasoned and logical thinking, the set of emotions, perceptions, subjective evaluations that individuals have with regard to the latent trait being evaluated

indecision inherently present in any human choice, not depending on the individuals' position on the latent variable

Expressed rating

The feeling path

- T consecutive steps
- at each step a **basic** judgment is formulated
- step-by-step, the basic judgments are **accumulated** and **transformed** into provisional ratings
- the final rating of the Feeling approach is the last provisional rating

The feelin

An assessment about the latent trait, but a simpler task than the full rating expression. Example: gather thoughts around a single aspect of the statement under evaluation and decide whether you agree or not (Yes/No)

- T consecutive steps
- at each step a **basic** judgment is formulated
- step-by-step, the basic judgments are **accumulated** and **transformed** into provisional ratings
- the final rating of the Feeling approach is the last provisional rating

UNIVERSITY OF BRESCIA

BODAI – LAB
Big&Open Data Innovation Laboratory

A) Feeling Approach

1. *Elementary judgments:* An *iid* sequence of random variables $X_1, \cdots, X_T$ with domains $\mathcal{D}_{X_1}, \cdots, \mathcal{D}_{X_T}$ generates T elementary judgments $x_1, \cdots, x_T$ progressively expressed along T steps.

2. *Accumulating function:* At each step t, a function $f : \mathcal{D}_{X_1} \times \cdots \times \mathcal{D}_{X_t} \to \Psi_t \subseteq \mathbb{R}$ summarizes the t past elementary judgments (for example, by summation). We say that f is an accumulating function, i.e. we require it obeys the following property: $\Psi_t \subseteq \Psi_{t+1}$, $\forall t$.

3. *Accumulated judgments:* A sequence of random variables $W_1, \cdots, W_T$, $W_t = f(X_1, \cdots, X_t)$, with domains $\mathcal{D}_{W_1} \equiv \Psi_1, \cdots, \mathcal{D}_{W_T} \equiv \Psi_T$ is then originated along the T steps of the DP with T corresponding realizations $w_1, \cdots, w_T$, $w_t = f(x_1, \cdots, x_t)$, called accumulated judgments.

4. *'Likertization' function:* At each step t, a non-decreasing function $d : \mathcal{D}_{W_T} \to (1, \cdots, m)$ transforms w_t into a provisional rating. Note that from the definition of accumulating function derives $\mathcal{D}_{W_1} \subseteq \cdots \subseteq \mathcal{D}_{W_T}$, so that d can always be computed on the domain of W_t, for all t.

5. *Provisional ratings:* A sequence of random variables $R_1, \cdots, R_T$, $R_t = d(W_t)$, with domains the space $(1, \cdots, m)$ is then originated along the T steps of the feeling path with T corresponding realizations $r_1, \cdots, r_T$, $r_t = d(w_t)$, called provisional ratings.

The feeling path

Several different models can be obtained, depending on the assumptions on:

- the **distribution** of the basic judgments
- the **accumulation function**
- the **transformation function**

✓ CUB class (D'Elia&Piccolo, 2005; Piccolo&Simone, 2019)

✓ Nonlinear CUB model (Manisera&Zuccolotto, 2014)

✓ CUM model (Manisera&Zuccolotto, forthcoming)

The CUM model

Response scale: $1, 2, \ldots, m$, with uneven number of options $m = 2k + 1$, middle point $k + 1$

Feeling approach: the respondent starts his reasoning focusing on **the middle option** and then moves downward or upward, according to different sensations

The number of steps in the **feeling path** is $T = k$

The CUM model
- Distribution of the basic judgments

The basic judgments can be modelled by a multivariate random variable $\mathbf{X}_t = [X_{t,D}, X_{t,U}, X_t]$ following a **Multinoulli distribution** with parameters $[\xi_D, \xi_U]$ $(\xi_D, \xi_U \geq 1,\ \xi_D + \xi_U \leq 1)$

- $X_{t,D} = 1$ (with probability ξ_D) when there is a perception that moves the respondent's position **downwards** along the response scale
- $X_{t,U} = 1$ (with probability ξ_U) when there is a perception that moves the respondent's position **upwards**
- $X_t = 1$ (with probability $1 - \xi_D - \xi_U$) when the respondent considers **staying still**

The CUM model
- Accumulation function

…is additive: $\mathbf{M}_t = \mathbf{X}_1 + \cdots + \mathbf{X}_t$

→ The accumulated judgments are modelled by

$\mathbf{M}_t = [M_{t,D}, M_{t,U}, M_t]$, a **Multinomial random variable** with parameters $[\xi_D, \xi_U]$ and trial parameter t

The CUM model
- Likertization function

$$l : (1, \cdots, t) \times (1, \cdots, t) \to (1, \cdots, 2t+1) = [-1 \;\; 1 \;\; 0] \begin{bmatrix} M_{t,D} \\ M_{t,U} \\ M_t \end{bmatrix} + k + 1.$$

- Provisional ratings

Follow a linearly transformed Multinomial distribution W_t

- Final ratings (feeling path)

Follow a linearly transformed Multinomial distribution $W_k\ (T = k)$

UNIVERSITY OF BRESCIA

BODAI – LAB
BIG&OPEN DATA INNOVATION LABORATORY

Econometrics and Statistics

The CUM model

The final expressed rating is a combination of the **feeling** approach and the **uncertainty** approach

$$p(R = r|\pi, \xi_D, \xi_U) = \pi W(r|\xi_D, \xi_U) + (1-\pi)U_m$$

$$\boxed{W(r|\xi_D, \xi_U) = p(W_k = r|\xi_D, \xi_U)} \qquad \boxed{U_m = 1/m}$$

Two feeling parameters!

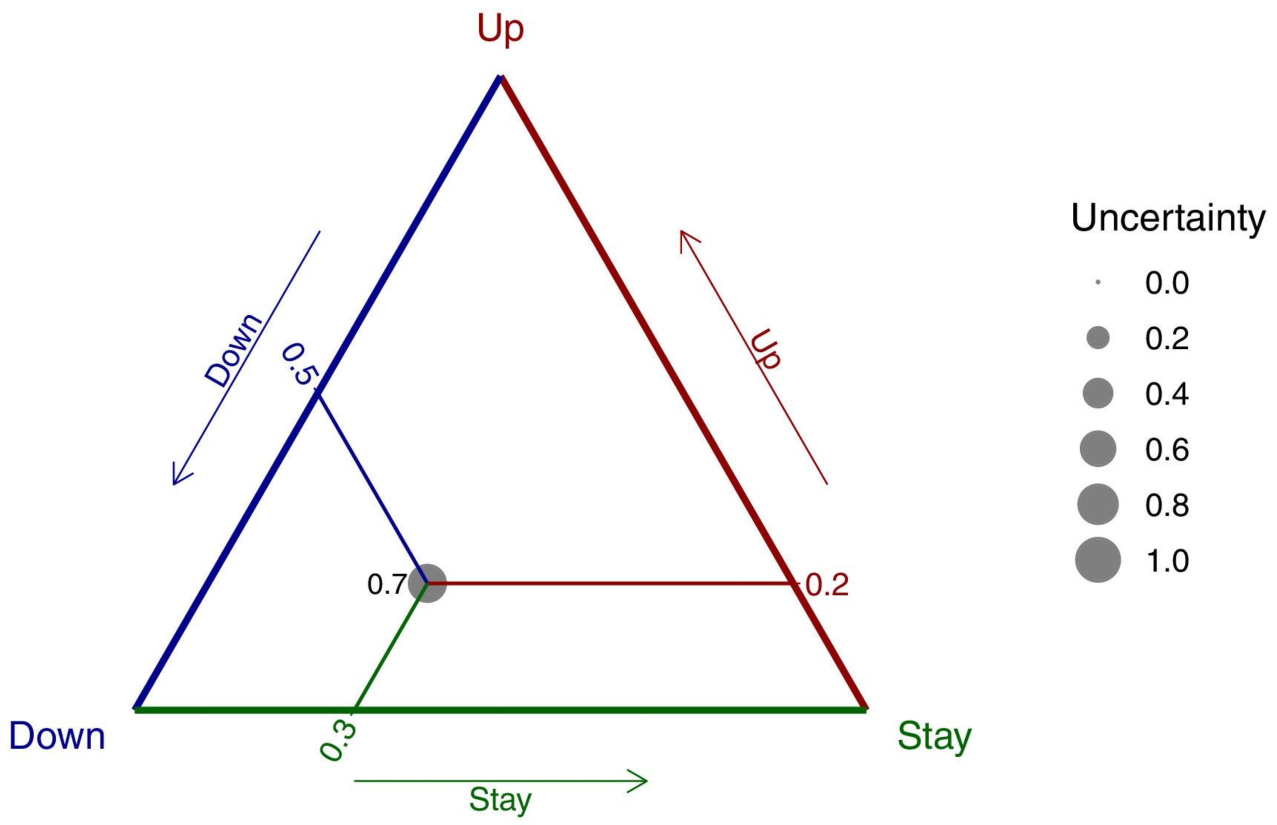
Combination of **U**niform and **M**ultinomial

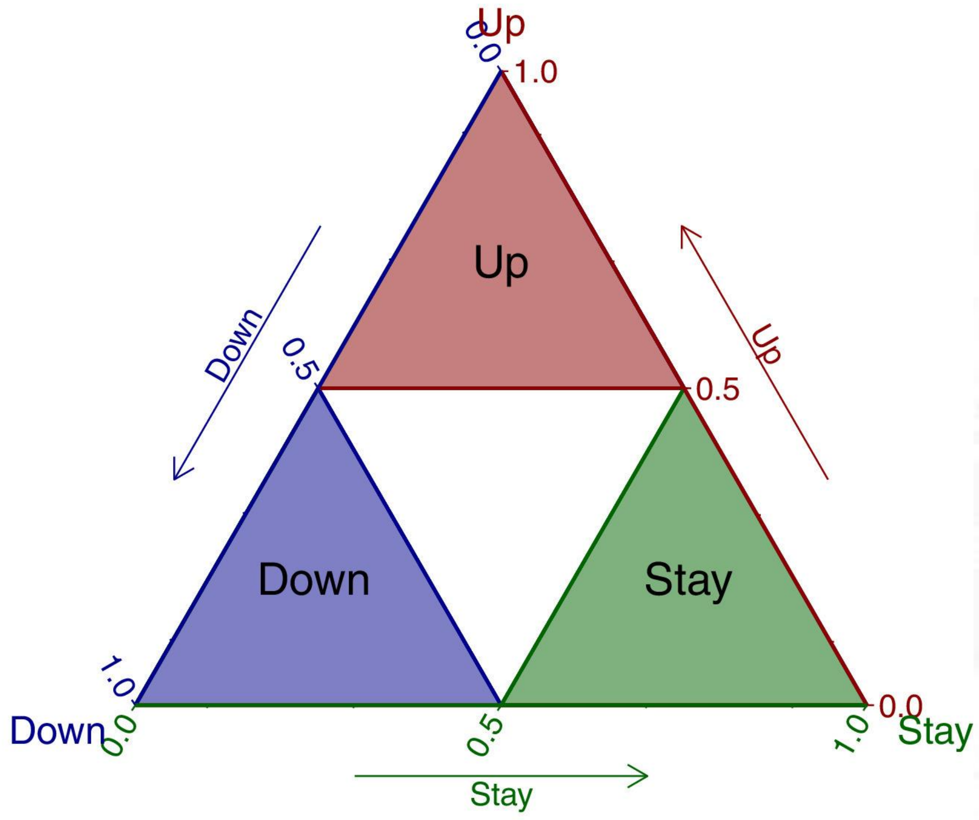
We derived the probability mass function of the variable W_k (based on the Multinomial distribution) and set up parameter estimation (via **EM algorithm**)

UNIVERSITY OF BRESCIA

BODAI – LAB
Big&Open Data Innovation Laboratory

Graphical representation of the parameter space





Case study: opinions of students about distance teaching

Please rate your opinion about the following questions between the two given extreme options

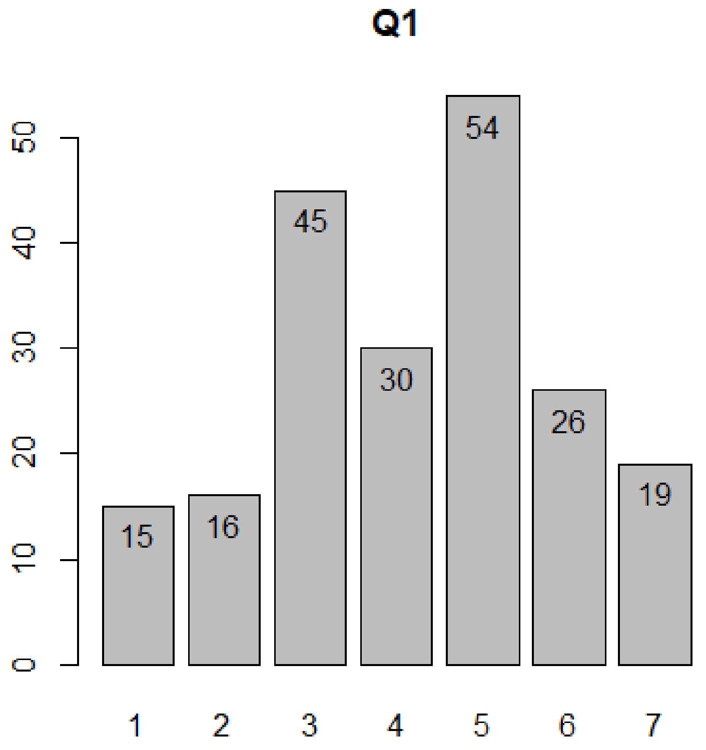
- **Q1**: Do you think that your learning was improved or worsened by distance teaching? (1=Much worsened, 4=Neither worsened nor improved, 7=Much improved)
- **Q2**: Do you think that the solutions arranged for distance teaching should be maintained or abandoned after the COVID-19 pandemic? (1=We should go completely back to the past, 4=We should select what to maintain and what to abandon, 7=We should maintain all the novelties introduced during the pandemic)

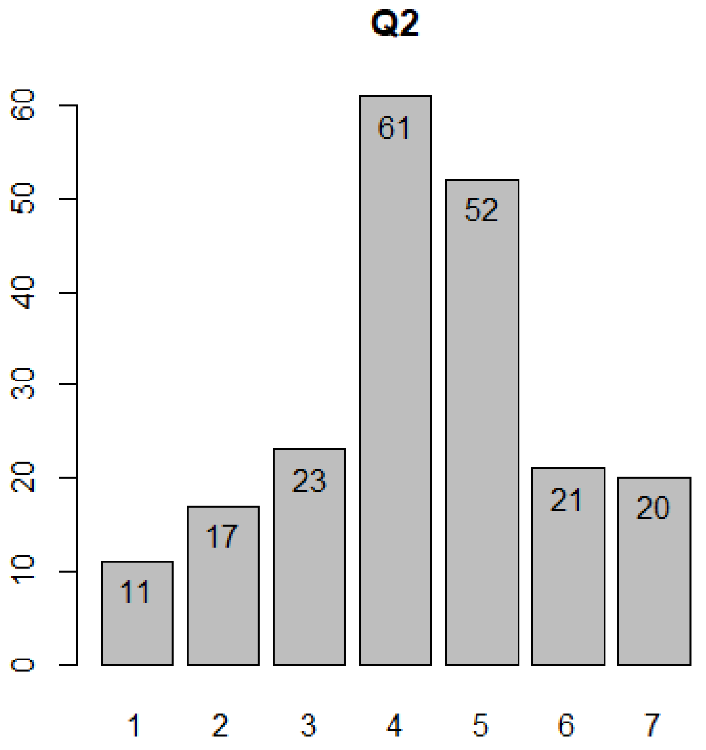
UNIVERSITY OF BRESCIA

BODAI – LAB
Big&Open Data Innovation Laboratory

Econometrics and Statistics

Case study: opinions of students about distance teaching

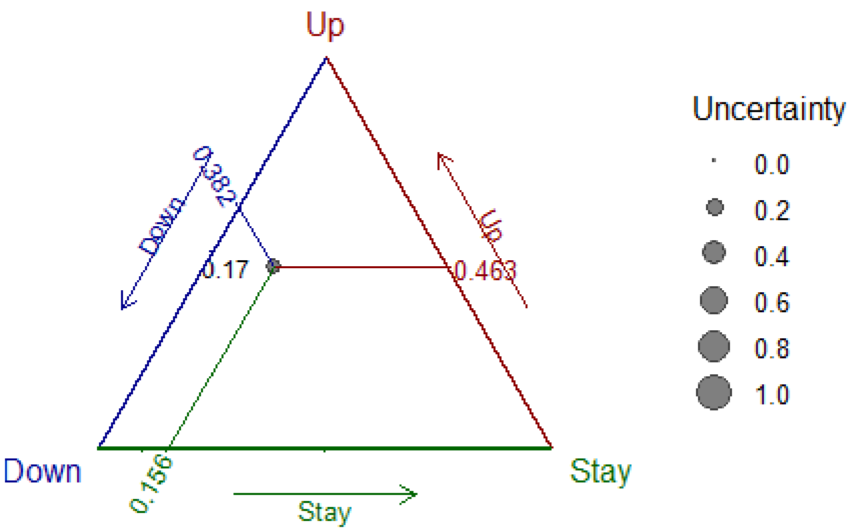


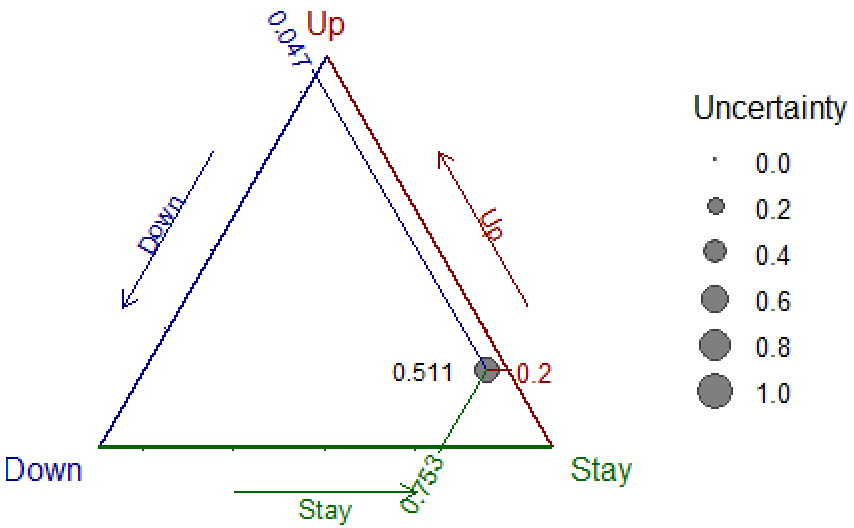


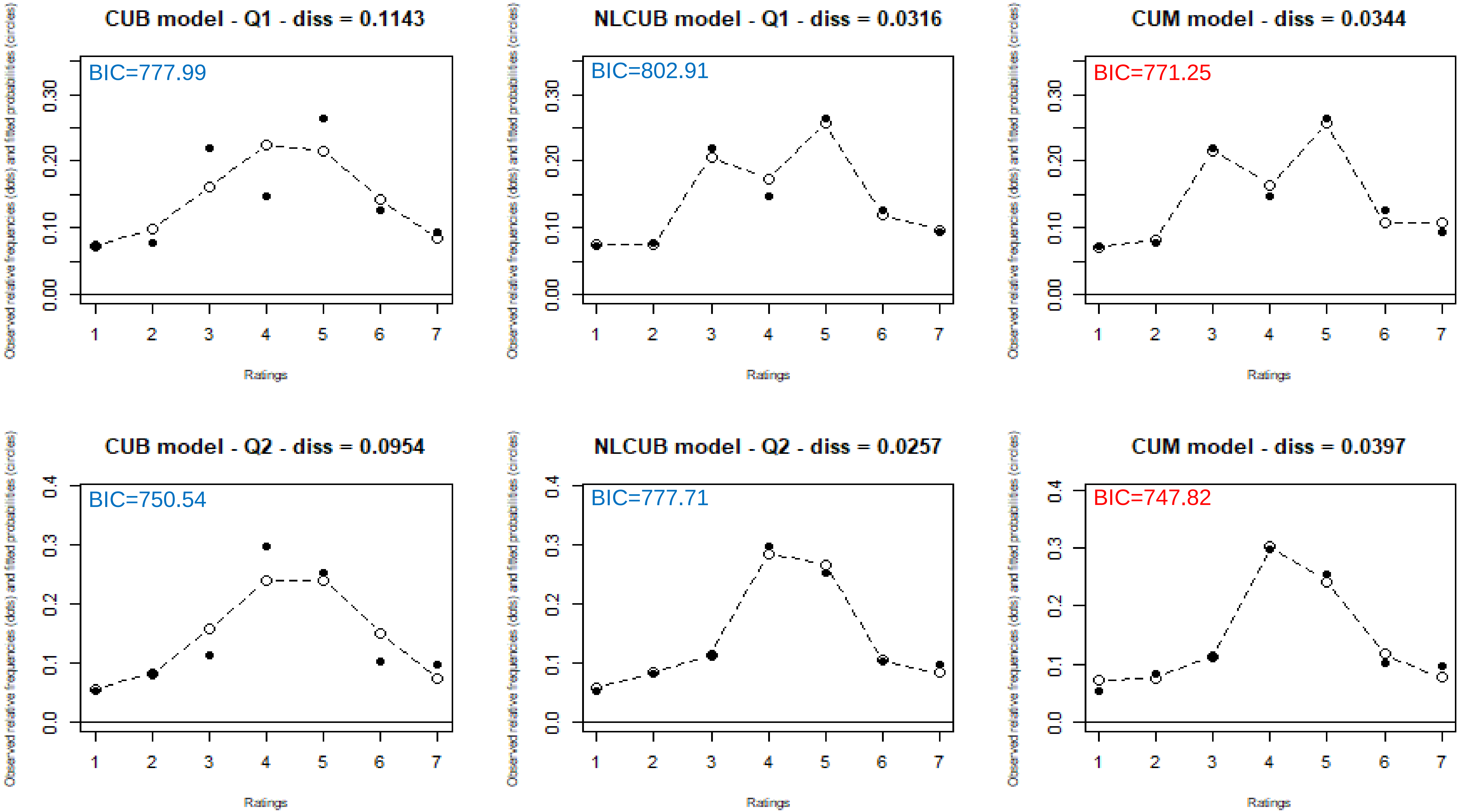
UNIVERSITY OF BRESCIA

BODAI – LAB
Big&Open Data Innovation Laboratory

Case study: opinions of students about distance teaching






CUB model - Q1 - diss = 0.1143
BIC=777.99
NLCUB model - Q1 - diss = 0.0316
BIC=802.91
CUM model - diss = 0.0344
BIC=771.25
CUB model - Q2 - diss = 0.0954
BIC=750.54
NLCUB model - Q2 - diss = 0.0257
BIC=777.71
CUM model - diss = 0.0397
BIC=747.82
Observed relative frequencies (dots) and fitted probabilities (circles)
Ratings

Conclusions

We proposed the **CUM model**, a mixture model for ordinal variables from semantic differential scales

It belongs to the **CUB class** and is based on a **mixture** of a **linearly transformed Multinomial** and a discrete **Uniform** random variables

It is derived as a **particular case of a general DP** that underlies the respondents' choice when facing questions with ordered response categories

We obtained parameter estimation by the **EM algorithm**.

Simulations and application to real data suggest that the model is promising and can be applied to get interesting insights in several fields.

Main References

D'Elia A., Piccolo D. (2005), A mixture model for preferences data analysis. *Computational Statistics & Data Analysis*, **49**(3), 917-934.

Manisera M., Zuccolotto P. (2014), Modeling rating data with Nonlinear CUB models. *Computational Statistics & Data Analysis*, **78**, 100-118.

Manisera M., Zuccolotto P. (2022), A mixture model for ordinal variables measured on semantic differential scales, Econometrics and Statistics, **22**, 98-123.

Piccolo, D., Simone, R. (2019). The class of cub models: statistical foundations, inferential issues and empirical evidence. *Statistical Methods & Applications*, **28**(3), 389-435.

UNIVERSITY OF BRESCIA

BODAI – LAB

BIG&OPEN DATA INNOVATION LABORATORY

Website and R codes

The main methodological advancements developed by the BODal-Lab research group on the topic of rating data analysis by means of statistical models in the CUB class are collected in the website

https://bodai.unibs.it/cub/

There, a page is devoted to the **CUM model**, where R codes and functions can be downloaded for the estimation of CUM with the EM algorithm in the case of $m = 7$. **An R package is being developed and will be released soon**.

Data for the case study discussed in this talk are also available, with the R script for obtaining the presented graphical representations.

UNIVERSITY OF BRESCIA

BODAI – LAB
BIG&OPEN DATA INNOVATION LABORATORY

Thank you!

marica.manisera@unibs.it

paola.zuccolotto@unibs.it

UNIVERSITY OF BRESCIA

BODAI – LAB
Big&Open Data Innovation Laboratory

Econometrics and Statistics