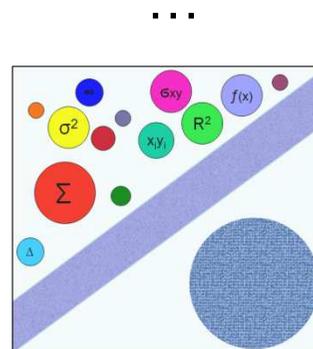# Making subjectivity explicit

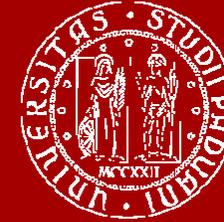## A measure of Social Inclusion for European administrative regions

Ludovico Carrino[1], Silvio Giove[2]

[1] Department of Economics - University of Venice, University of Trieste

[2] Department of Economics - University of Venice

…

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
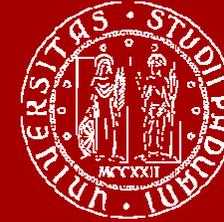
The paper in a nutshell

- Main topic: implicit trade-offs resulting from
  - Normalization function (re-scaling)
  - Aggregation function (preference structure)
- Case-study: measure of social inclusion for European regions
- Method: compare strategies for normalization and aggregation
  - Top-down approach: parameters are set by the researcher
  - Expert-based approach: parameters are elicited by experts/stakeholders
- Results
  - Data-driven vs experts-driven strategies lead to very different results and economic interpretations (positive vs normative)

UNIVERSITÀ DEGLI STUDI DI PADOVA
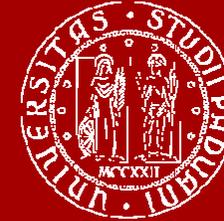
Case study: 4 variables of social inclusion in Europe

- Theoretical framework on social inclusion starts with Atkinson et al. (2002).
- EUROSTAT DATABASE 2014 for 58 administrative regions, 2004-2012

Average 2004-2012	Health	Education	Labour market	Economic Resources
	Longevity	Early school-leaving	Long-term unemployment	At-risk-of-poverty rate
Belgium	79.8	12.10%	3.80%	14.80%
Germany	80.2	12.30%	4.10%	14.50%
Italy	82	18.40%	3.60%	16.90%
Spain	81.4	29.30%	4.80%	20.80%
maximum	84.2 (Madrid 2012)	54.2% (Ceuta 2005)	18.2% (Ceuta 2012)	44.3% (Sicilia 2011)
minimum	77.5 (Wallonie 2004)	5.4% (Thüringen 2009)	0.5% (Trent.A.A. 2004)	5.2% (V.d'Aosta 2006)

Case study: social inclusion in Europe
Normalization
Aggregation
Conclusions

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

The baseline model

- Generic aggregation model of Social Inclusion

$$F(\mathbf{x}) = \left[w_1 v_1 (x_1)^{\beta} + \cdots + w_4 v_4 (x_4)^{\beta} \right]^{1/\beta}$$

- Min-max normalization function

For «positive» variables

$$\nu_+^{i,t}(x_+^{i,t}) = 100 \cdot \frac{x^{i,t} - b\min(x_+)}{b\max(x_+) - b\min(x_+)}$$

$$\nu_+^{i,t}(x_+^{i,t}) = 0 \text{ if } x_+^{i,t} \leq b\min(x_+)$$

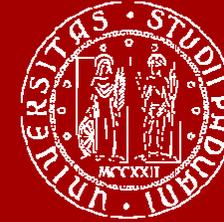$$\nu_+^{i,t}(x_+^{i,t}) = 100 \text{ if } x_+^{i,t} \geq b\max(x_+)$$

For «negative» variables

$$\nu_-^{i,t}(x_-^{i,t}) = 100 \cdot \frac{b\max(x_-) - x_-^{i,t}}{b\max(x_-) - b\min(x_-)}$$

$$\nu_-^{i,t}(x_-^{i,t}) = 0 \text{ if } x_-^{i,t} \geq b\max(x_-)$$

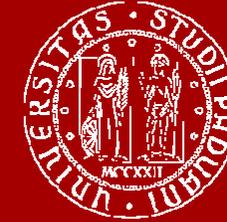$$\nu_-^{i,t}(x_-^{i,t}) = 100 \text{ if } x_-^{i,t} \leq b\min(x_-)$$

- Normalization benchmarks bmin and bmax can be:
  - Data-driven (i.e., correspond to observed performance in data)
  - Based on preferences of experts/stakeholders

UNIVERSITÀ DEGLI STUDI DI PADOVA
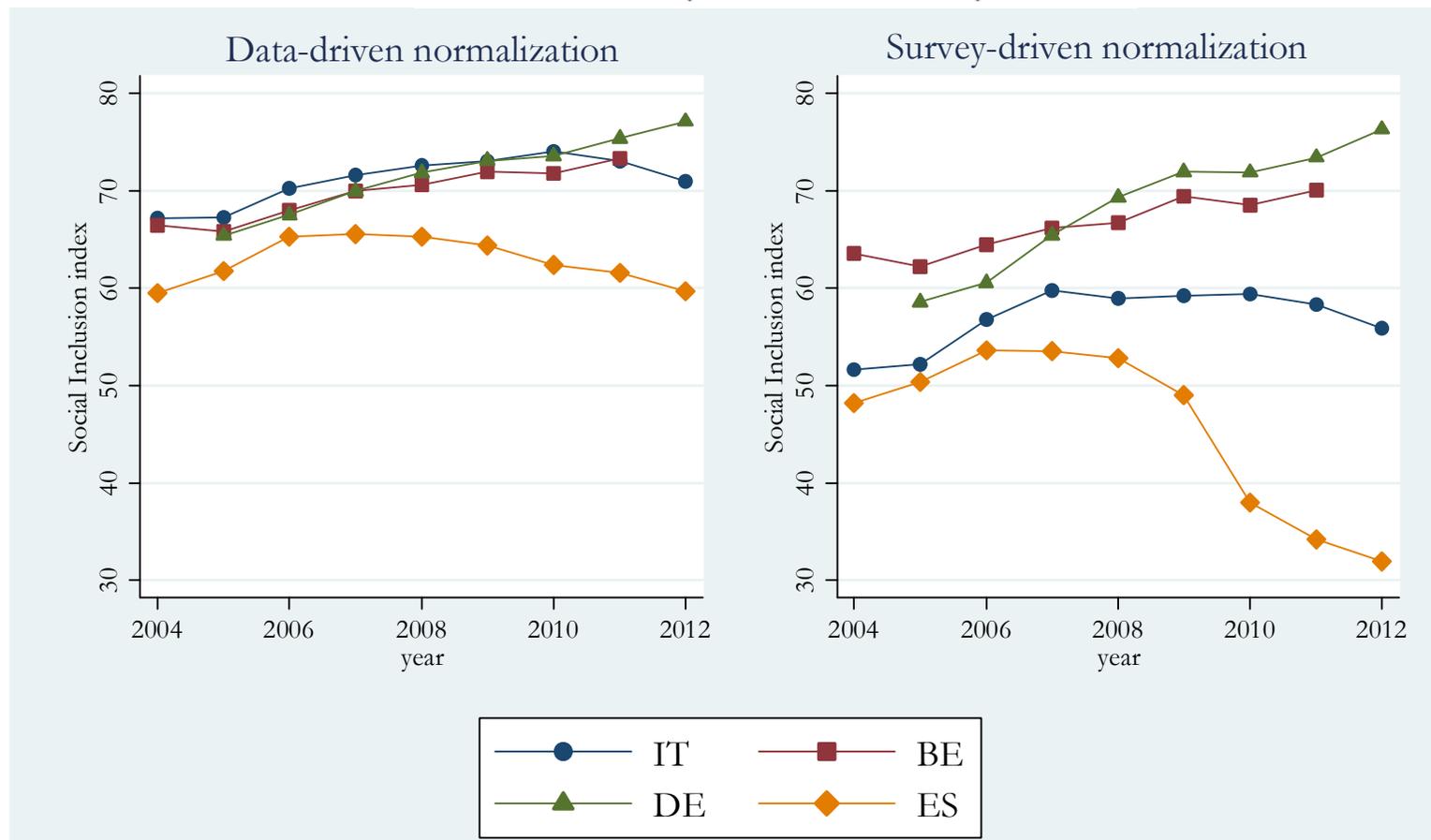
Elicitation of experts benchmarks

- Population: 143 professors in Economics and Management at the Ca' Foscari University of Venice (88 respondents)
- For each variable, we ask for two thresholds that would represent:
  - a certainly **undesirable** social condition
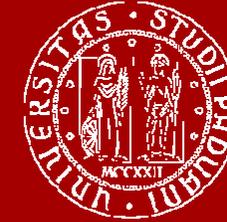 - a certainly **desirable** social condition

Variable	SURVEY benchmarks median responses *(interquartile range)*		DATA-DRIVEN benchmarks	
	Minimum	Maximum	Minimum	Maximum
Early school leaving	10% [5% - 10%]	20% [15% - 25%]	5.4%	54.2%
Life expectancy	73 years [70 - 75]	*83 years* [80 - 85]	77.5	84.2
L.T. Unemployment	3% [2% - 4%]	9% [6% - 10%]	0.5%	18.2%
Poverty rate	5% [3% - 7%]	20% [17% - 21.5%]	5.2%	44.3%

Case study: social inclusion in Europe
Normalization
Aggregation
Conclusions

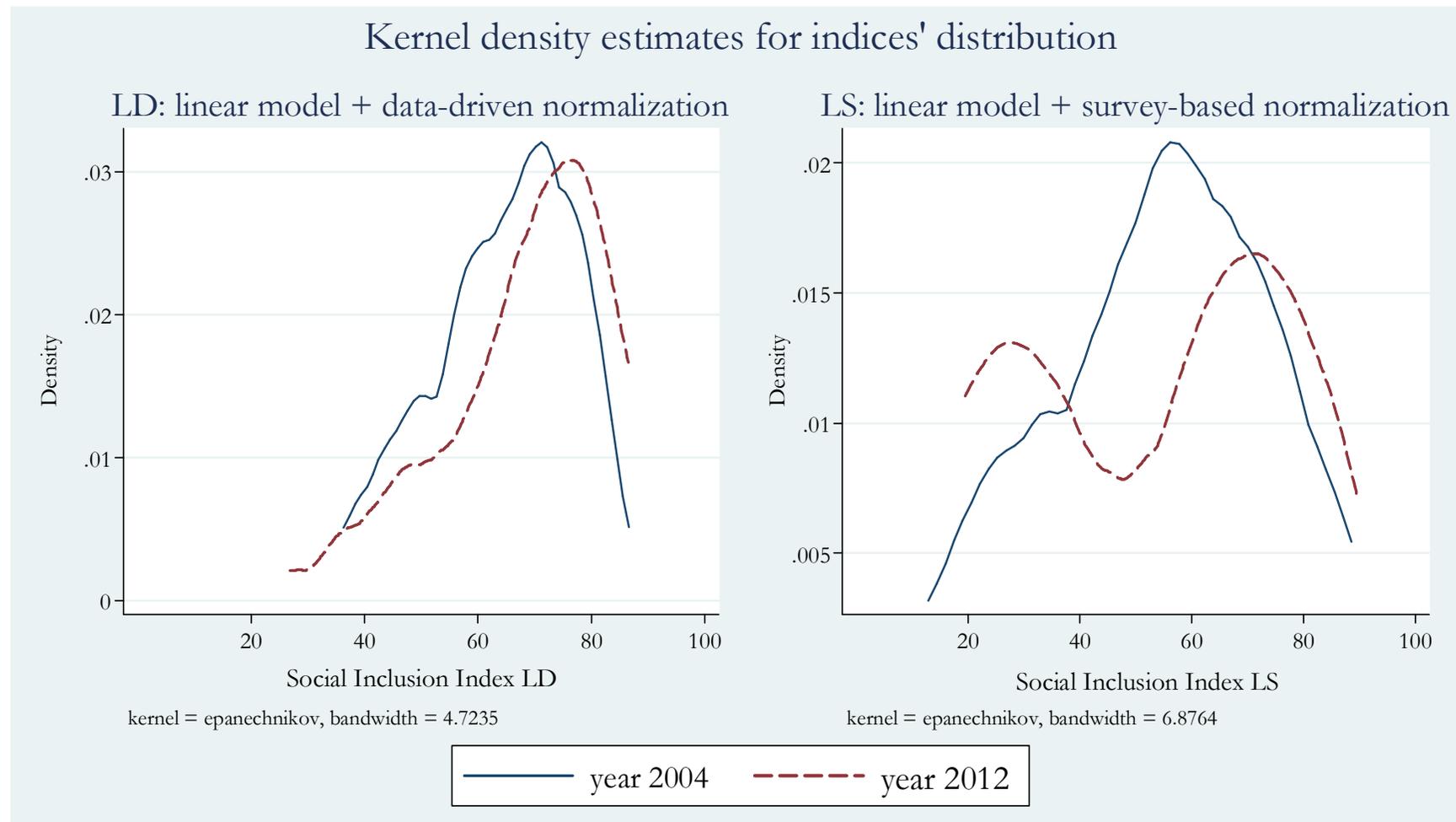UNIVERSITÀ
DEGLI STUDI
DI PADOVA

A simple average model with alternative normalizations

$$F^i\left(\nu(\mathbf{x}^i)\right) = \frac{1}{4}\nu_1(x_1^i) + \cdots + \frac{1}{4}\nu_4(x_4^i)$$

Case study: social inclusion in Europe
Normalization
Aggregation
Conclusions

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

- ## Normalization affects the distribution of the index

Kernel density estimates for indices' distribution

LD: linear model + data-driven normalization

LS: linear model + survey-based normalization

kernel = epanechnikov, bandwidth = 4.7235

kernel = epanechnikov, bandwidth = 6.8764

year 2004 — — — year 2012

UNIVERSITÀ
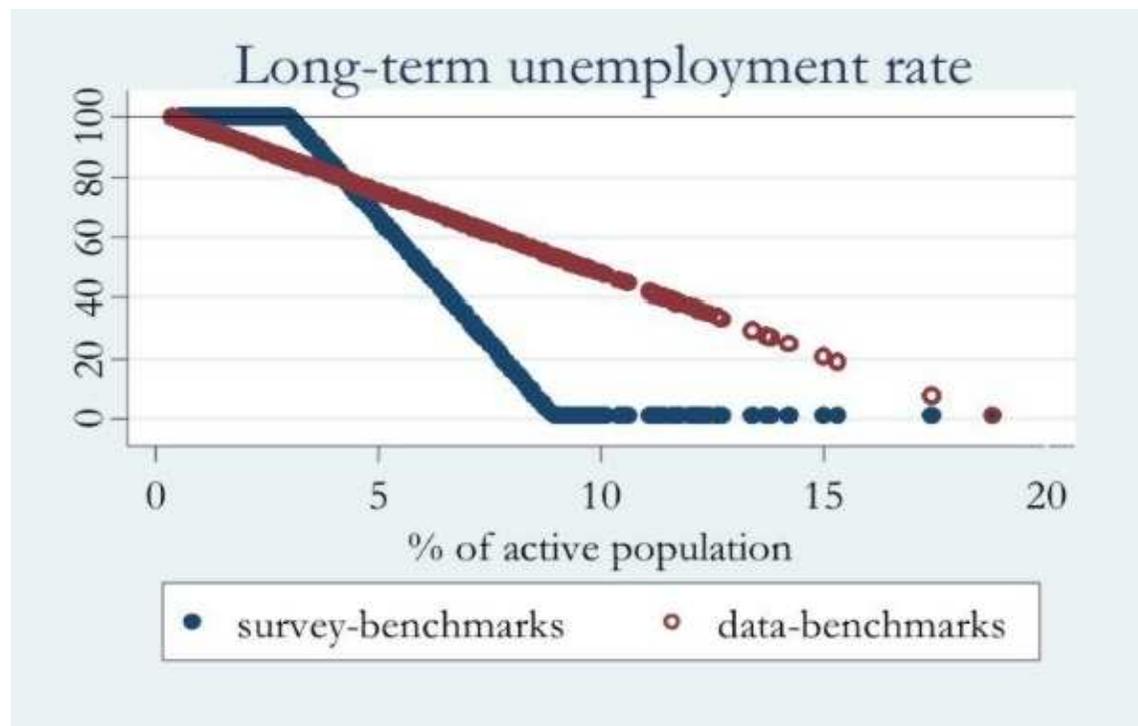DEGLI STUDI
DI PADOVA
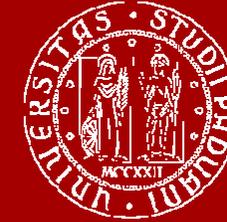
- ## Partial derivatives identify variables' relative relevance

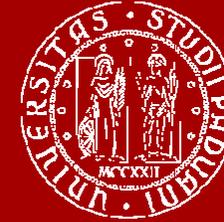| DATA-DRIVEN normalization	longevity	early school leavers	long-term unemployment	poverty-rate
$\partial F / \partial x_j$	3.72	-0,512	-1,412	-0,57
Relative importance	59.8%	8.2%	22.7%	9.1%

SURVEY normalization	longevity	early school leavers	long-term unemployment	poverty-rate
$\partial F / \partial x_j$	2.5	-2.5	-4	-1.65
Relative importance	23.5%	23.5%	37.6%	15.4%

- Rank reversal between Italy and Germany because of different relative weights, especially on longevity and unemployment

- These characterizations have different economic justifications. Positive vs normative.
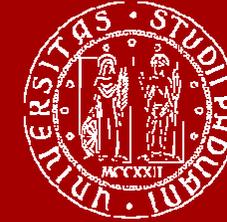
Case study: social inclusion in Europe
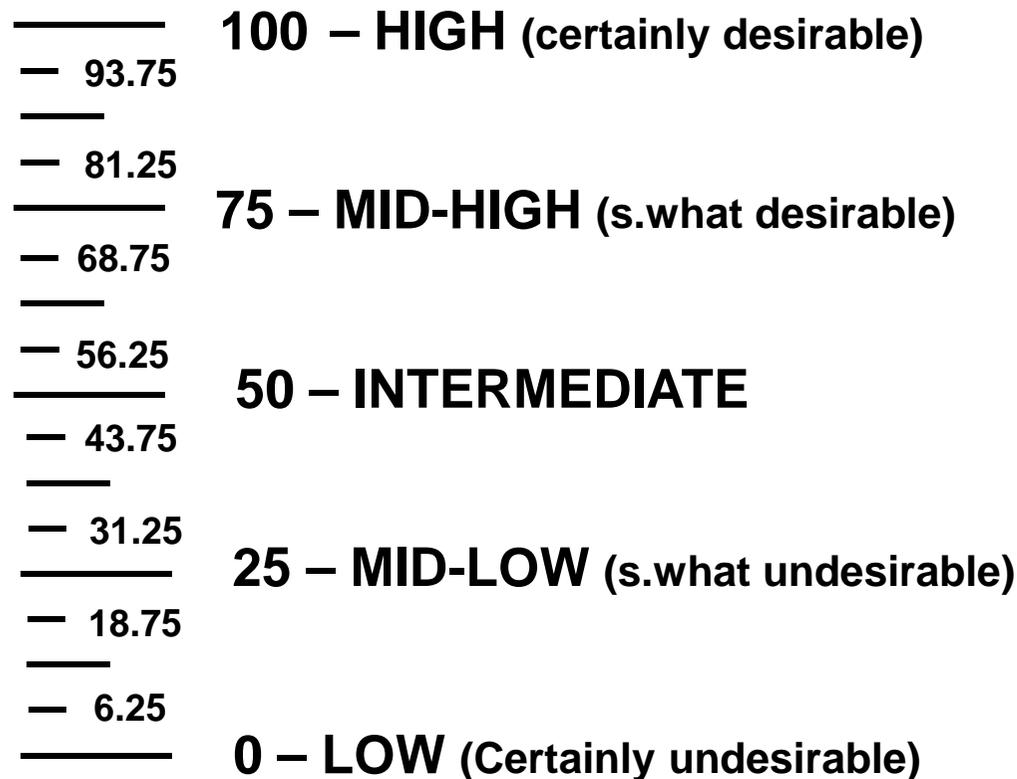**Normalization**
Aggregation
Conclusions

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

- # Normative vs positive

Long-term unemployment rate

survey-benchmarks   data-benchmarks

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Estimate the aggregation function through experts' panel

- The general CES model is $F(\mathbf{x}) = \left[w_1 v_1 \left(x_1 \right)^\beta + \cdots + w_4 v_4 \left(x_4 \right)^\beta \right]^{1/\beta}$

  - Set of parameters to be estimated: P = (β, **w**)

- Importance of eliciting expert/institutional preferences

  - Kim et al. (SIR, 2015), Decancq, Lugo (Econometric Reviews, 2013)

- Population: 20 Regional Directors General of Social Policy

- Elicitation strategy: Scenarios evaluation (individual interviews)

- Scenario: random combination of normalized values for the 4 variables

  - Assumption: each dimension can take three normalized levels: Certainly desirable (100), Intermediate condition (50), Certainly undesirable (0)

  - *Min-max* normalization function with expert-based benchmarks

  - Set of 27 scenarios, same for every Decision-Maker

Case study: social inclusion in Europe
Normalization
Aggregation
Conclusions

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Scenarios' evaluation

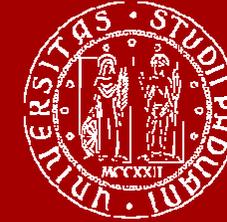- Evaluation on a 0-100 scale, using 5 trivial scenarios as guidelines.

Istruzione – – Lavoro – – Reddito = Salute + +			Scenario 10
ISTRUZIONE condizione **BASSA**	**LAVORO** condizione **BASSA**	**REDDITO** condizione **INTERMEDIA**	**SALUTE** condizione **ALTA**
abbandono scolastico 20 %	disocc. lunga durata 9 %	povertà relativa 12,5 %	speranza di vita 82 anni

Catalunia (ES)

Istruzione + Lavoro + Reddito + Salute +			Scenario MA
ISTRUZIONE	LAVORO	REDDITO	SALUTE
condizione **MEDIO-ALTA** abbandono scolastico 12,5 %	condizione **MEDIO-ALTA** disocc. lunga durata 4,5 %	condizione **MEDIO-ALTA** povertà relativa 9 %	condizione **MEDIO-ALTA** speranza di vita 79,5 anni

Trivial mid-high (75) scenario

100 – HIGH (certainly desirable)

— 93.75

— 81.25

75 – MID-HIGH (s.what desirable)

— 68.75

— 56.25

50 – INTERMEDIATE

— 43.75

— 31.25

25 – MID-LOW (s.what undesirable)

— 18.75

— 6.25

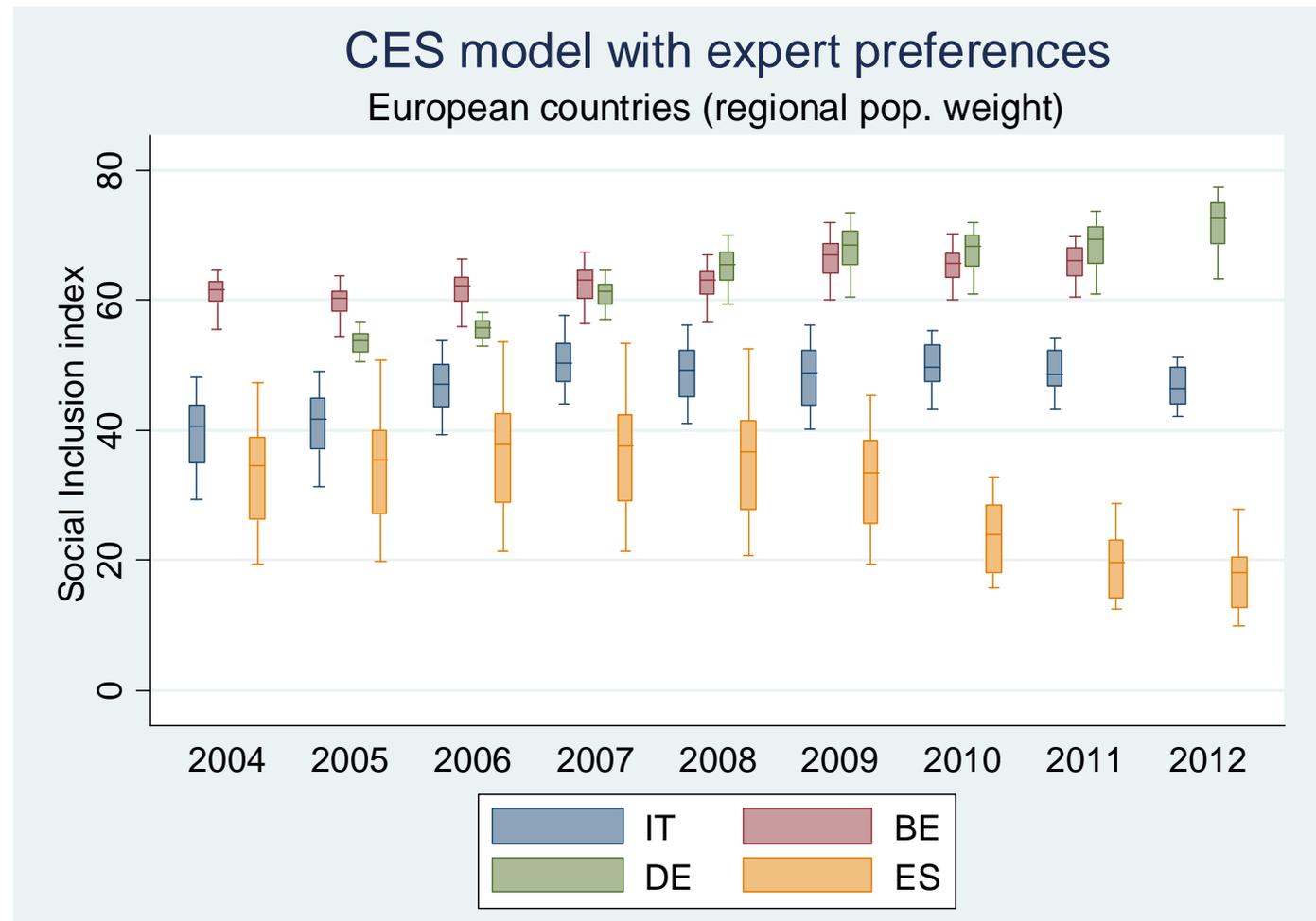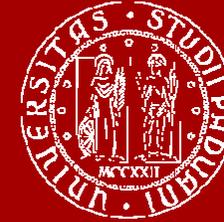**0 – LOW** (Certainly undesirable)

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

NLS estimation results per decision-maker		β	w_1 Education	w_2 Lab. mkt	w_3 Econ. Res.	w_4 Health	R^2
1	DM-Piemonte	0.95	0.38	0.20	0.30	0.13	0.95
2	DM-Lombardia	0.66	0.11	0.2	0.37	0.33	0.93
3	DM-Liguria	0.71	0.27	0.23	0.25	0.26	0.84
4	DM-Veneto	0.53	0.32	0.27	0.19	0.22	0.9
5	DM-TAA	0.34	0.27	0.27	0.28	0.18	0.89
6	DM-Friuli – V. Giulia	0.77	0.31	0.27	0.22	0.20	0.9
7	DM-Emilia-Romagna	0.27	0.41	0.21	0.13	0.24	0.92
8	DM-Toscana	0.29	0.30	0.36	0.20	0.14	0.94
9	DM-Marche	0.71	0.52	0.21	0.15	0.10	0.88
10	DM-Abruzzo	1.00	0.22	0.36	0.28	0.14	0.82
11	DM-Campania	0.67	0.28	0.21	0.29	0.21	0.91
12	DM-Puglia	1.00	0.27	0.26	0.39	0.08	0.84

UNIVERSITÀ DEGLI STUDI DI PADOVA

Results: average national Social Inclusion

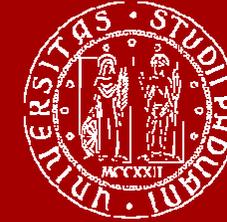- We do not aggregate experts' preferences

- Each set of preference generate an index

- We graph the indices distribution for each region / nation

CES model with expert preferences

European countries (regional pop. weight)
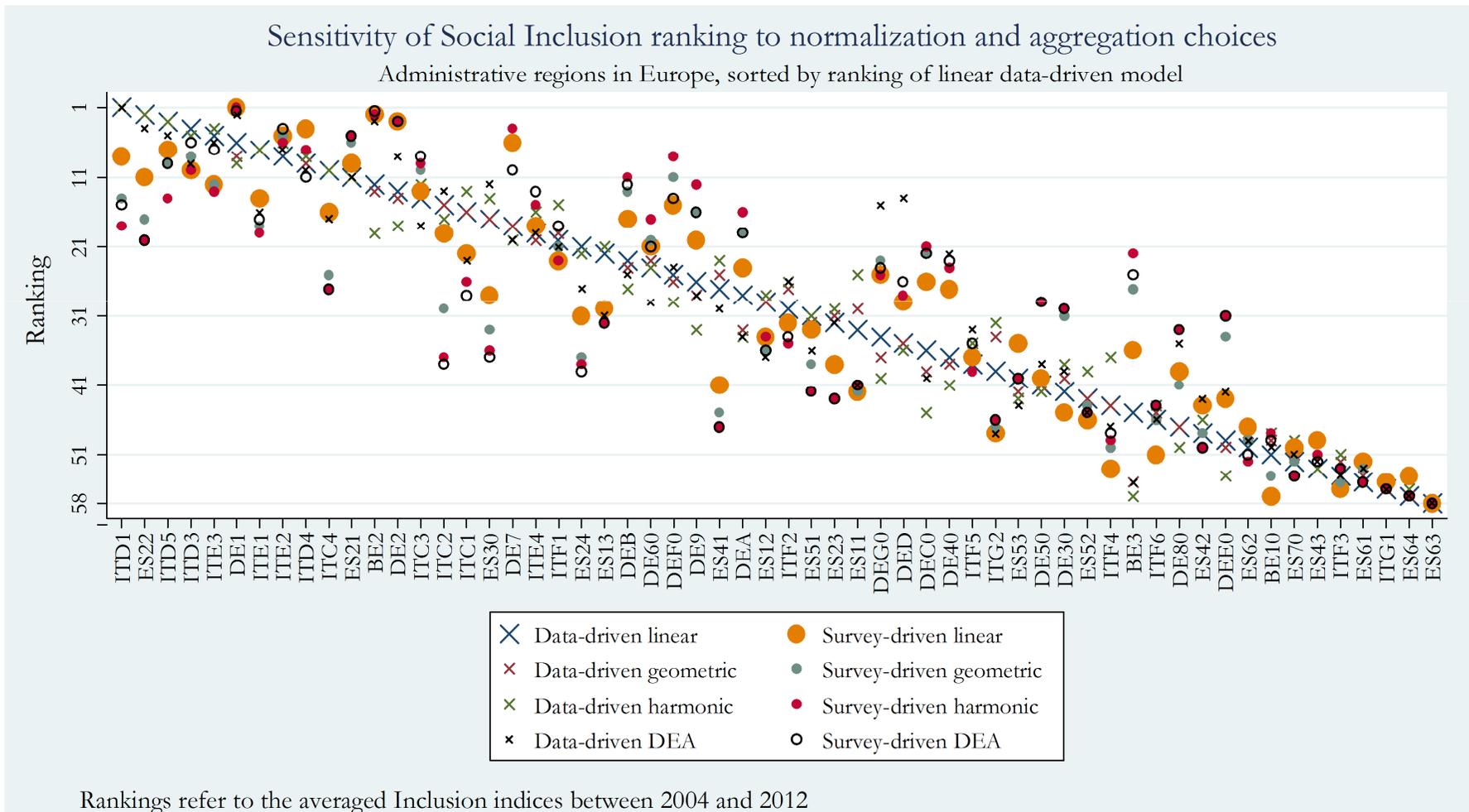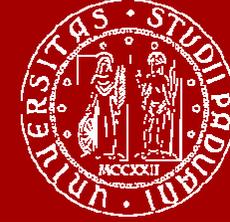
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Conclusions

- There is hardly a neutral method to build synthetic measures of abstract phenomena

- Normalization is a partial weighting stage which can strongly affect results

- Implicit trade-offs can be made transparent

- Economic premises and interpretation of results may largely differ depending on the nature of parameters' selection

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Changing aggregation functions (geometric, harmonic)

Sensitivity of Social Inclusion ranking to normalization and aggregation choices

Administrative regions in Europe, sorted by ranking of linear data-driven model

Rankings refer to the averaged Inclusion indices between 2004 and 2012

FOLLOW-UP SLIDES

SOGLIA DI RISCHIO: qual è un livello di speranza di vita alla nascita che Lei valuterebbe come sicuramente negativo, non desiderabile e problematico in un generico contesto territoriale?
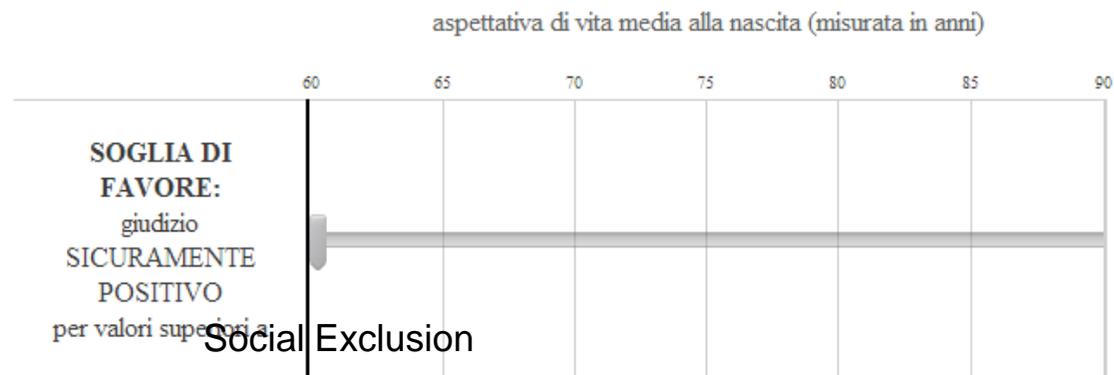
Per individuare la soglia nella tabella seguente occorrerà trascinare il cursore verticale (che assume colore blu non appena lo si sfiori) fino a raggiungere il numero di anni desiderato.

aspettativa di vita media alla nascita (misurata in anni)

SOGLIA DI RISCHIO:
giudizio
SICURAMENTE
NEGATIVO
per valori inferiori a

... I FAVORE: un livello che Lei valuterebbe
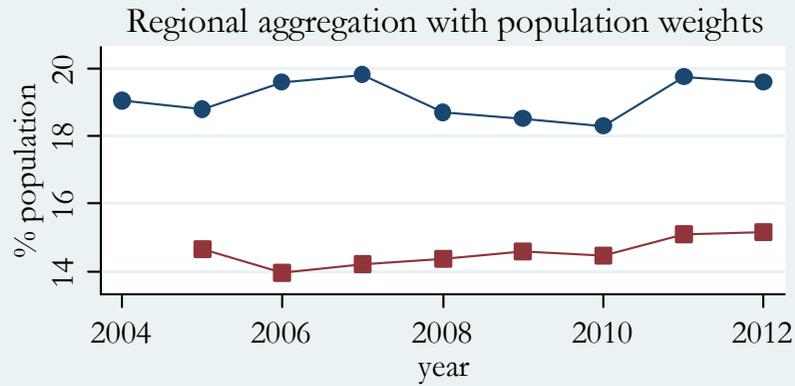... una condizione virtuosa in un generico

La *soglia di favore* deve essere **maggiore** (o uguale) rispetto alla *soglia di rischio* selezionata nella domanda precedente: una longevità più elevata è preferibile ad una meno elevata.

Per individuare la soglia nella tabella seguente occorrerà trascinare il cursore verticale (che assume colore blu non appena lo si sfiori) fino a raggiungere il numero di anni desiderato.
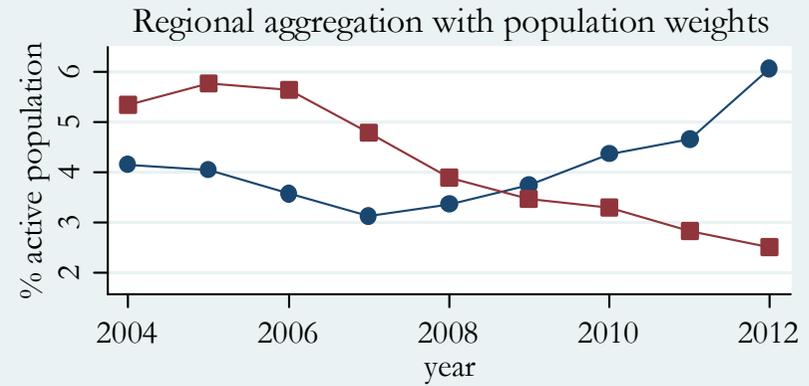
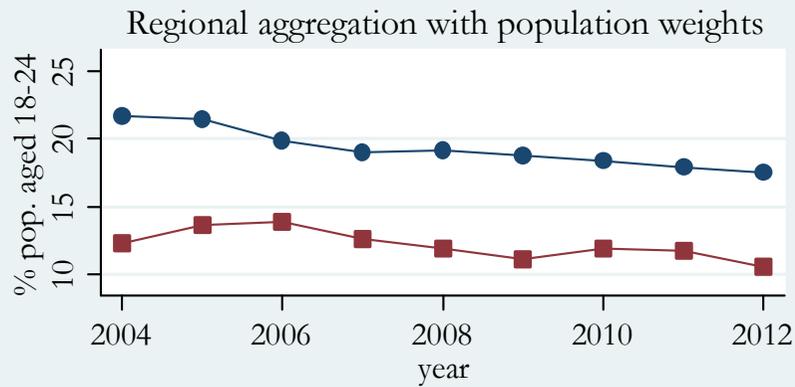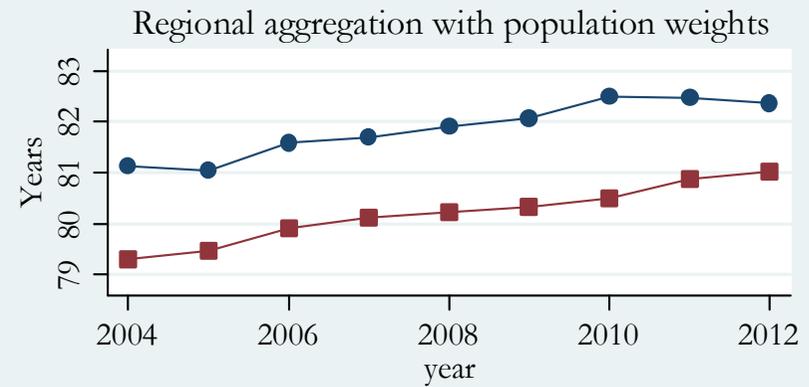aspettativa di vita media alla nascita (misurata in anni)

SOGLIA DI FAVORE:
giudizio
SICURAMENTE
POSITIVO
per valori superiori a

Social Exclusion

Time trend of indicators

Poverty rate
Regional aggregation with population weights

Long-term unemployment
Regional aggregation with population weights

Early school-leavers
Regional aggregation with population weights

Life expectancy at birth
Regional aggregation with population weights

Italy   Germany

- Recall the CES function and fix a beta<1

$$F^i(v(x)) = \left[w_1 v_1 (x_1^i)^b + L + w_m v_m (x_m^i)^b \right]^{1/b}$$

- Suppose β=0, we get a geometric mean

$$G^i(n(\mathbf{x}^i)) = n_1 (x_1^i)^w * L * n_m (x_m^i)^w$$

- Higher penalization for bad performances, Extreme case: if one normalized-attribute is zero, the Index collapses to zero. (Ravallion 2012 for a discussion)

- Now an attribute's relevance depends also on its relative performance with respect to the others. Worst performances have higher relevance.

$$\frac{\P F(v(\mathbf{x}))}{\P x_j} = w_j v'(x_j) \left(\frac{\left[w_1 v_1(x_1)^b + L + w_m v_m(x_m)^b \right]^{1/b}}{v_j(x_j)} \right)^{1-b} = w_j v'(x_j) \left(\frac{F(v(\mathbf{x}))}{v_j(x_j)} \right)^{1-b}$$
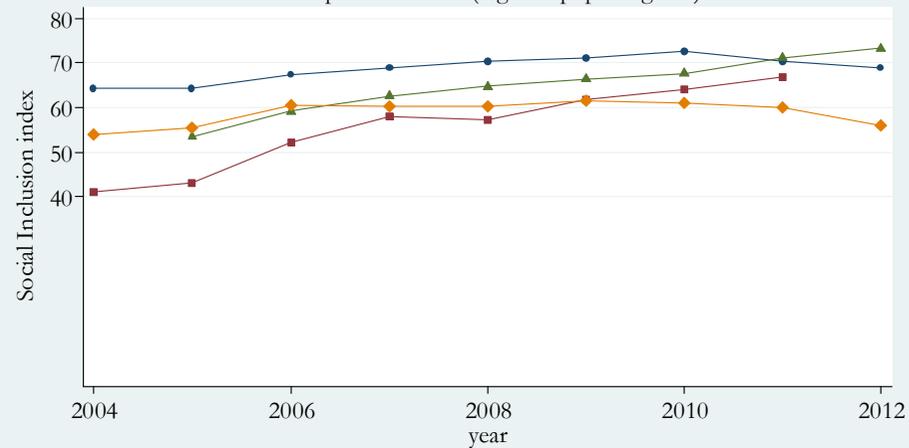
Geometric model with data-driven normalization — European countries (regional pop. weighted)
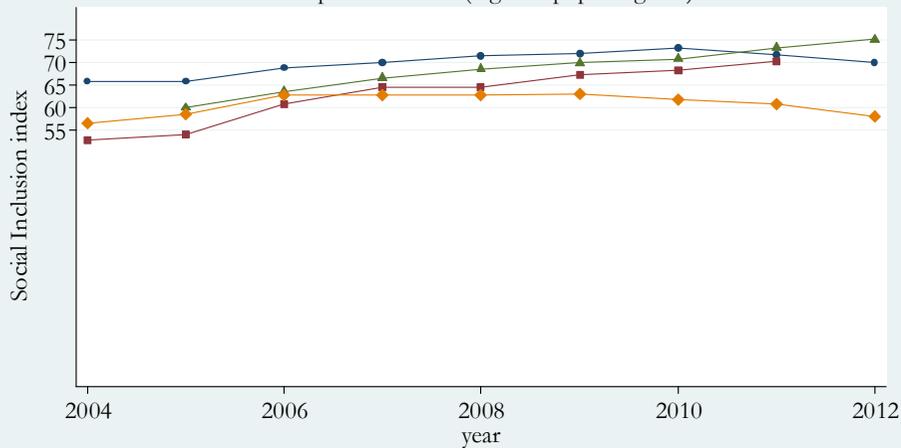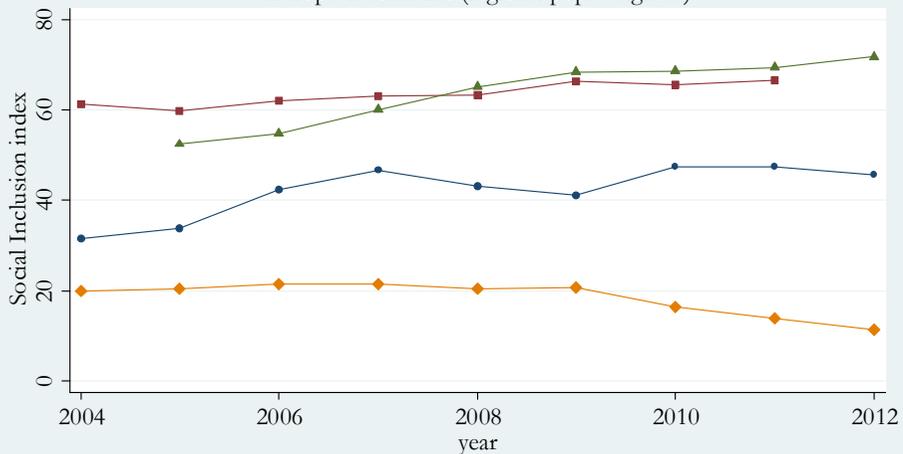
Armonic model with data-driven normalization — European countries (regional pop. weighted)
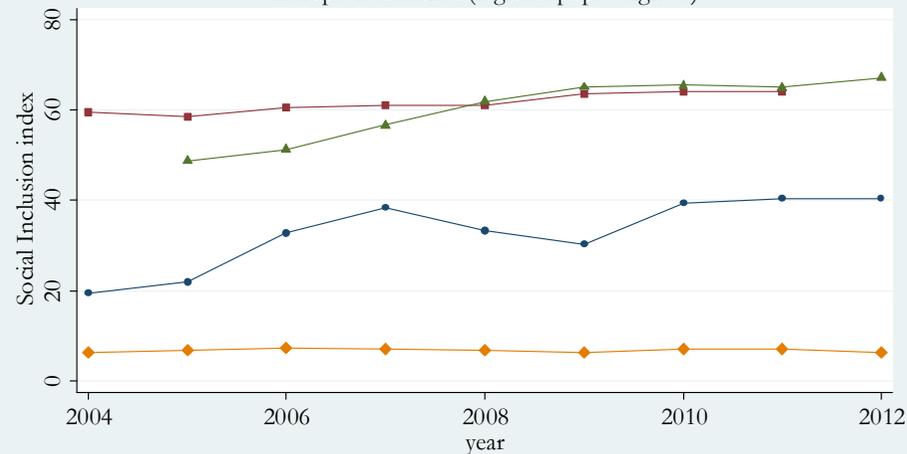
Geometric model with survey-driven normalization — European countries (regional pop. weighted)

Armonic model with survey-driven normalization — European countries (regional pop. weighted)

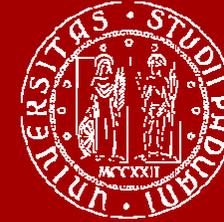Legend: IT, BE, DE, ES

Number of scenarios

- $3^4 = 81$ possible scenarios

- Reduce them to 27 with fractional factorial (orthogonal arrays)
  - Chen J, Sun DX, Wu CFJ, 1993, International Statistical Review
  - Ferrini, Scarpa, 2007, Journal of Environmental Economics and Management
  - Wu, Amada, 2009, Wiley
  - Street et al, 2005, International Journal of Research in Marketing

- Orthogonal array of strength t *(integer number)*:
  - a "table" whose entries come from a fixed finite set of symbols (ex: 0,1)
  - **for every selection of t** columns, **all ordered t-tuples** of the symbols, formed by taking the entries in each row restricted to these columns, **appear the same number of times**.

- We create an orthogonal array of strength 2, with $3^3 = 27$ lines

Interview' stages

- **Preliminary**: introduction to the phenomena of social exclusion, description of the 4 dimensions and indicators, details on the normalization, familiarisation with scenario-cards, clearing the desk:

- **First**: manually allocate scenarios on a 5-levels scale (corresponding to 5 areas on the desk), with option to use $\pm$ 6.25 steps

- **Second**: test for validity of the weak axiom of revealed preferences
  - take a random subset of 10 scenarios and repeat the experiment to check for coherence and independence of irrelevant alternatives

- **Third**: explicit question on relative weights w_1, w_2, w_3, w_4.
  - Budget allocation: «how would you allocate 100 points…?»

- **Last**: self-assessment of answers' reliability and other questions
  - 5 points scale, from "no confidence" to "very high confidence";
  - Respondent's education level, years worked in public sector, past experiences,…
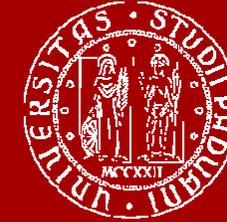
Example of responses from an expert

# scenario	RESPONSE	Education $v(x_A)$	Labour Market $v(x_B)$	Econ. resources $v(x_C)$	Health $v(x_D)$
1	**100**	100	100	100	100
3	**81.25**	50	100	50	100
11	**81.25**	100	50	50	100
15	**68.75**	50	1	100	100
21	**68.75**	50	50	100	50
23	**68.75**	100	100	50	50
9	**56.25**	100	50	1	50
17	**56.25**	100	50	100	1
13	**56.25**	1	50	100	100
14	**56.25**	100	1	100	50
20	**43.75**	1	100	100	50
12	**31.25**	1	50	50	50
22	**25**	100	1	50	1
10	**25**	1	100	50	1
16	**25**	100	1	1	100
24	**25**	1	100	1	100
27	**1**	1	50	1	1
25	**1**	50	1	1	1
19	**1**	1	1	1	50

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Estimation of β and w_i

- Henningsen & Henningsen, 2012, Economic Letters

1. Grid search on β
    a. Fix a range of values for β
        - example: from 1 to -3 with steps of 0,01
    b. For each β perform a constrained OLS
        - Weights should be non-negative
        - Sum of weights should be 1
        - Get optimal coefficients **w** conditioned on β
    c. Choose the β (and related **w**|β) that minimize sum of squared residuals

2. Perform a NLS estimation with starting values coming from grid search.

- Obtain a set of parameters (β, **w**) for each expert

Case study: social inclusion in Europe
Normalization
Aggregation
Conclusions

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Results: Social Inclusion for specific regions

CES model with expert preferences
Selected European regions

Legend: Veneto, Friuli - V. Giulia, Flanders, Valencia