

Putting Text in Context: How to Estimate Better Left-Right Positions by Scaling Party Manifesto Data

Kenneth Benoit[§] Thomas Däubler*

[§]LSE and Trinity College Dublin

*University of Mannheim and MZES

Questions

- ▶ How should we understand the “left-right dimension”?
- ▶ How should we construct a left-right index using content analysis results from political documents (here: manifestos)?
- ▶ What does this say about the past, present, and future of measuring left-right?

Locating Parties on a Left-Right Dimension

- ▶ Two approaches
 - ▶ *a priori*: specify a fixed scale in advance

Locating Parties on a Left-Right Dimension

- ▶ Two approaches
 - ▶ *a priori*: specify a fixed scale in advance
 - ▶ *a posteriori*: estimate a scale using inductive methods from data (higher-dimensional measurements)

Locating Parties on a Left-Right Dimension

- ▶ Two approaches
 - ▶ *a priori*: specify a fixed scale in advance
 - ▶ *a posteriori*: estimate a scale using inductive methods from data (higher-dimensional measurements)
- ▶ One common application: estimating left-right from category counts based on human-coding of policy documents

Locating Parties on a Left-Right Dimension

- ▶ Two approaches
 - ▶ *a priori*: specify a fixed scale in advance
 - ▶ *a posteriori*: estimate a scale using inductive methods from data (higher-dimensional measurements)
- ▶ One common application: estimating left-right from category counts based on human-coding of policy documents
- ▶ The CMP/MARPOR's main deliverable is its fixed scale: RILE

“Its a priori, deductive nature is important in allowing its application **in all places at all times** without the qualifications about content or context which apply to inductive scales. **It is a substantively invariant measure whose numeric values always carry the same meaning.**” (Budge and Meyer 2013: 88)

An IRT Model for Counts of Unordered Categorical Outcomes

- ▶ IRT modelling (Rasch 1960, Bock 1972)

An IRT Model for Counts of Unordered Categorical Outcomes

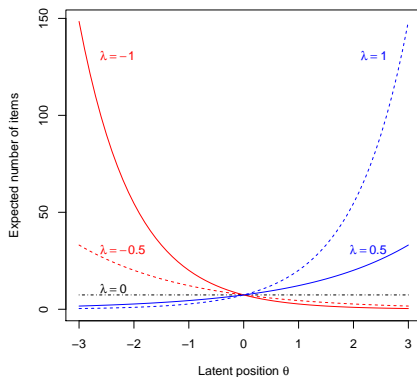
- ▶ IRT modelling (Rasch 1960, Bock 1972)
- ▶ Let y_{ij} represent the counts of coded text units from $i = 1, \dots, N$ documents, falling into category $j = 1, \dots, J$.

$$y_{ij} \sim \text{Neg. Binomial}(\mu_{ij}, \phi_j)$$
$$\log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_j \theta_i$$

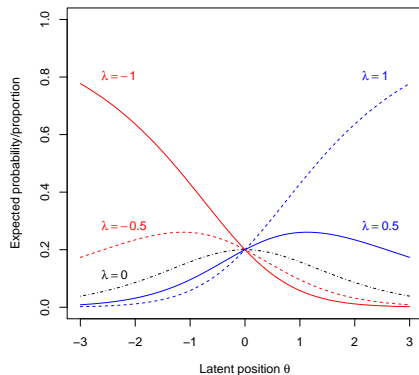
- ▶ Interpretation:
 - α_i the variable length of the document (total length)
 - ζ_j the baseline frequency of a category (issue coverage)
 - λ_j responsiveness of the category to the latent variable
 - θ_i the latent variable (e.g. left-right position)
 - $1/\phi_j$ overdispersion parameter

Parameters and Outcomes

Expected category counts ($\alpha = 1$ and all $\zeta = 1$)



Item response category functions (any α , all $\zeta = 1$)



Identification

- ▶ There are five fundamental indeterminacies
- ▶ The following constraints constitute one way of identifying the model:

$$\zeta_j = 0$$

$$\bar{\lambda}_j = 0$$

$$\theta_i \sim N(0, 1)$$

$$\lambda_j > \lambda_{j'}$$

Inference

- ▶ Bayesian approach (using HMC in Stan) with priors as follows:

$$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha)$$

$$\zeta_j \sim N(\mu_\zeta, \sigma_\zeta)$$

$$\lambda_j \sim N(0, \sigma_\lambda)$$

$$\theta_i \sim N(0, 1)$$

$$\mu_\alpha \sim N(0, 5)$$

$$\mu_\zeta \sim N(0, 5)$$

$$\sigma_\alpha \sim \text{Half-Cauchy}(0, 5)$$

$$\sigma_\zeta \sim \text{Half-Cauchy}(0, 5)$$

$$\sigma_\lambda \sim \text{Half-Cauchy}(0, 5)$$

$$1/\phi \sim \text{Uniform}(0, 200)$$

The CMP/MARPOR Data

- ▶ Human coding of 3000+ election manifestos from 50+ countries
- ▶ Standard coding scheme has 56 categories, some directed, others not
- ▶ Frequently used “canned” RILE index

Table 1

The components of the CMP left-right scale

Left categories	Right categories
103 Anti-imperialism	104 Military: positive
105 Military: negative	201 Freedom and human rights
106 Peace	203 Constitutionalism: positive
107 Internationalism: positive	305 Political authority
202 Democracy	401 Free enterprise
403 Market regulation	402 Incentives
404 Economic planning	407 Protectionism: negative
406 Protectionism: positive	414 Economic orthodoxy
412 Controlled economy	505 Welfare state limitation
413 Nationalisation	601 National way of life: positive
504 Welfare state expansion	603 Traditional morality: positive
506 Education expansion	605 Law and order
701 Labour groups: positive	606 Social harmony

Source: Budge et al. (2001), Mapping Policy Preferences, Appendix III. Left-right score = proportion (right – left) × 100.

And Now for Some Results...

- ▶ Sample of 2288 manifestos from post-1972 democracies
- ▶ The most and least frequent categories (ζ values):

Top 10 Categories	
Category	ζ
Welfare State Expansion +	2.20
Environm. Protection +	1.72
Technology and Infrastr. +	1.70
Education Expansion +	1.66
Political Authority +	1.63
Social Justice +	1.53
Gov. and Admin. Effic. +	1.50
Non-Econ. Demogr. Gr. +	1.37
Democracy +	1.35
Farmers +	1.25

Bottom 10 Categories	
Category	ζ
Marxist Analysis +	-3.37
Education Limitation +	-3.22
Labour Groups -	-2.71
National Way of Life -	-2.18
Centralisation +	-1.86
Foreign Special Rel. -	-1.46
Protectionism -	-1.45
Constitutionalism -	-1.29
Corporatism +	-1.23
Keynesian Dem. Man. +	-1.15

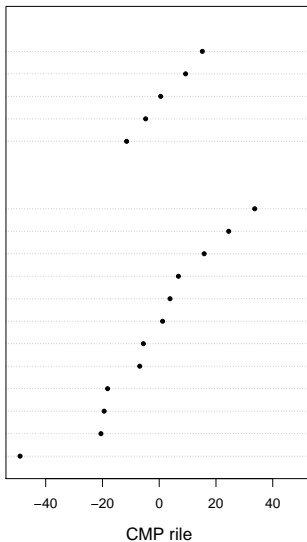
Left-Right Positions in Italy: CMP Rile

2008

PdL People of Freedom
IdV List Di Pietro – Italy of Values
PD Democratic Party
LN Northern League
UdC Union of the Center

2013

UdC Union of the Center
CD Democratic Centre
FDI-CDN Brothers of Italy
SVP South Tyrolean People's Party
PdL People of Freedom
SC Civic Choice
3L Labour and Freedom List
PD Democratic Party
Autonomy Progress Federalism
SEL Left Ecology Freedom
RC Civil Revolution
M5S Five Star Movement



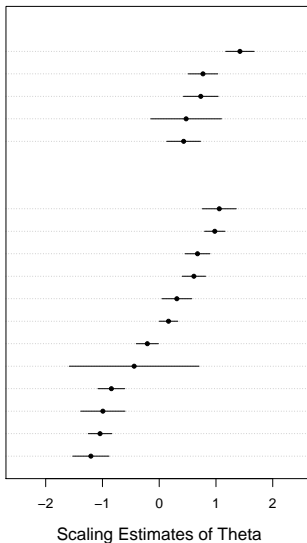
Left-Right Positions in Italy: Poisson IRT

2008

PdL People of Freedom
UdC Union of the Center
PD Democratic Party
IdV List Di Pietro – Italy of Values
LN Northern League

2013

PdL People of Freedom
FDI-CDN Brothers of Italy
SC Civic Choice
CD Democratic Centre
3L Labour and Freedom List
SVP South Tyrolean People's Party
UdC Union of the Center
Autonomy Progress Federalism
RC Civil Revolution
M5S Five Star Movement
SEL Left Ecology Freedom
PD Democratic Party



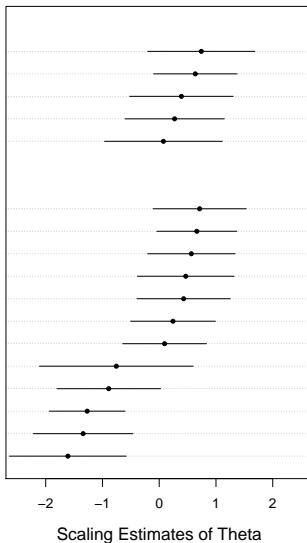
Left-Right Positions in Italy: NegBin IRT

2008

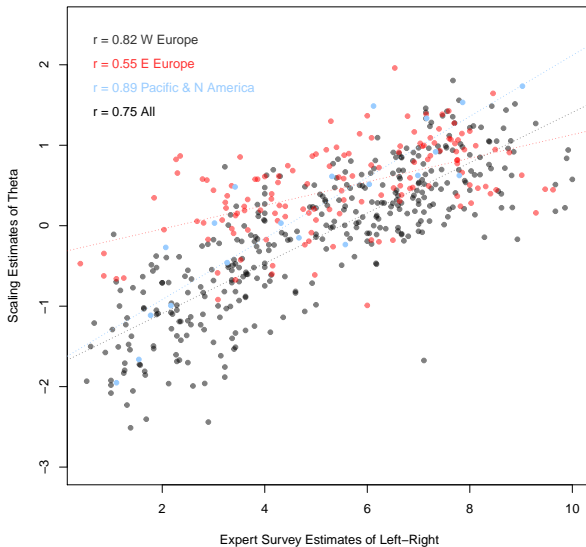
PdL People of Freedom
PD Democratic Party
UdC Union of the Center
IdV List Di Pietro – Italy of Values
LN Northern League

2013

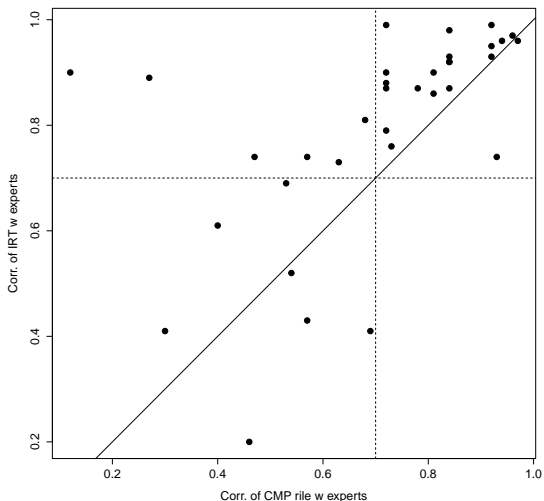
FDI-CDN Brothers of Italy
UdC Union of the Center
CD Democratic Centre
PdL People of Freedom
SC Civic Choice
SVP South Tyrolean People's Party
3L Labour and Freedom List
Autonomy Progress Federalism
PD Democratic Party
RC Civil Revolution
SEL Left Ecology Freedom
M5S Five Star Movement



Validation with Expert Survey Placements

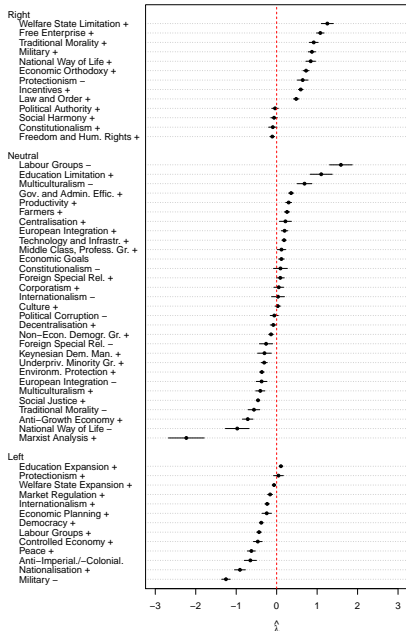


Country-by-Country Validation (vs. CMP Rile)

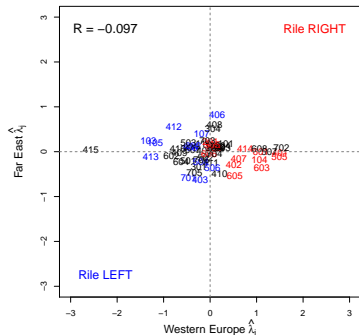
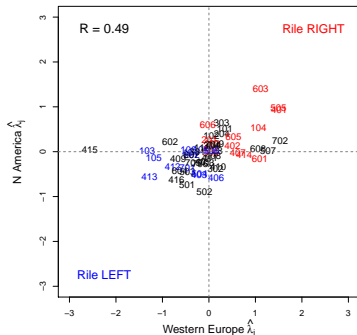
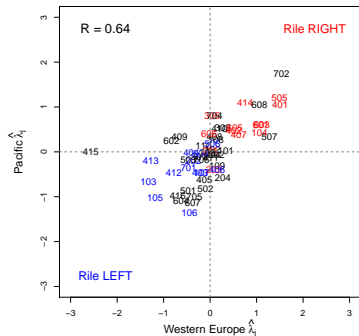
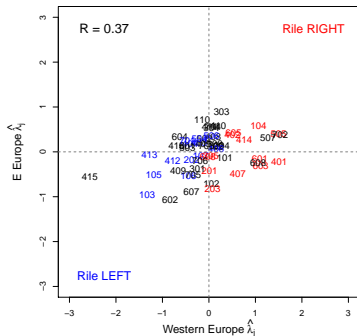


Romania not shown (r for rile -.55, r for IRT theta -.44)

$\hat{\lambda}$ (post-1972 democracies)



$\hat{\lambda}$ (post-1972 democracies)



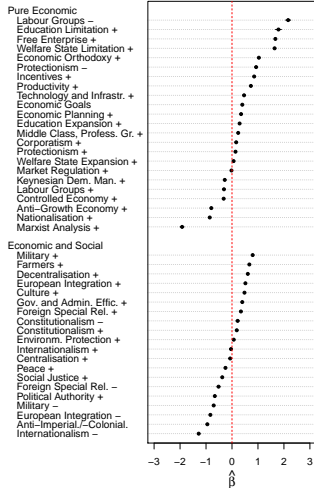
Take Me to Another Dimension

- ▶ In the two-dimensional case, we model the mean of the counts as:

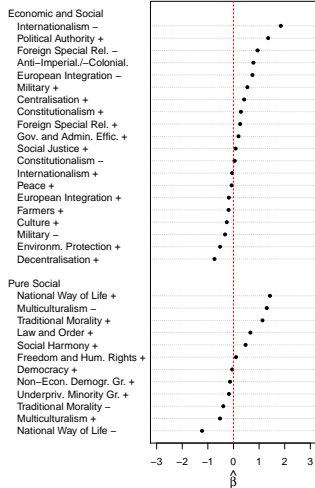
$$\log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_{1j}\theta_{1i} + \lambda_{2j}\theta_{2i}$$

- ▶ The model requires 13 constraints for identification
 - ▶ Mean of positions zero in each dimension
 - ▶ “Dimension founders”
 - ▶ 1 single, 1 single, 1 double, 2 further zero constraints on λ_{dj}
 - ▶ Reference category for ζ_j

$\hat{\lambda}_d$ for 2D-model (post-1972 democracies)

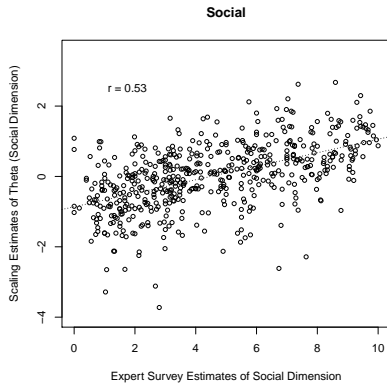
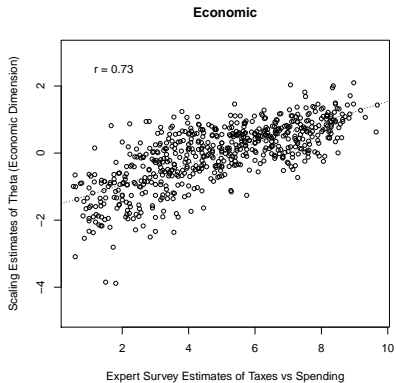


(a) Economic Dimension

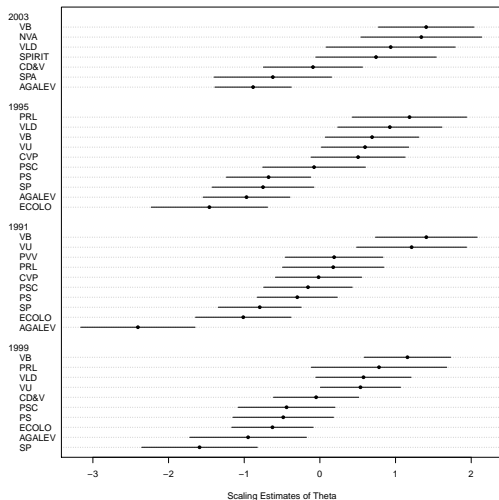


(b) Social Dimension

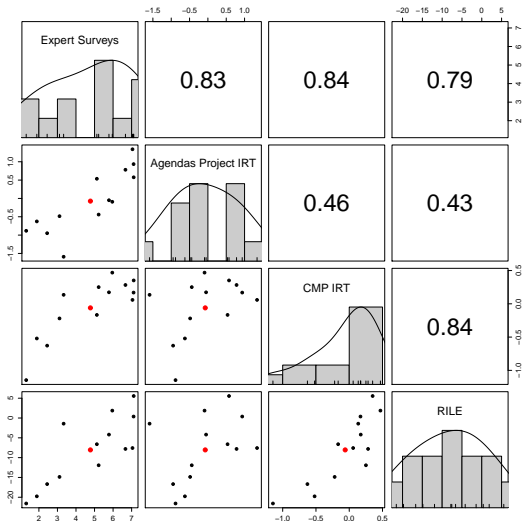
Validation with Expert Survey Placements - 2D



Using Alternative Items - Belgian CAP data



Validation of Belgian CAP Data Results



Conclusion

- ▶ IRT model as inductive approach for inferring a left-right dimension from category counts
- ▶ Left-right as a super-issue rather than defined by content
- ▶ Superior to fixed content-based index (RILE)
- ▶ Can incorporate a priori information explicitly
- ▶ Flexible in terms of extensions:
 - ▶ Differential item functioning
 - ▶ Model the parameters with covariates
 - ▶ ...