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
  - \textit{a priori}: specify a fixed scale in advance
  - \textit{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 \textit{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.
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)

\[ y_{ij} \sim \text{Neg Binomial}(\mu_{ij}, \phi_j) \]

\[ \log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_j \theta_i \]

Interpretation:
- \( \alpha_i \) the variable length of the document (total length)
- \( \zeta_j \) the baseline frequency of a category (issue coverage)
- \( \lambda_j \) responsiveness of the category to the latent variable
- \( \theta_i \) the latent variable (e.g. left-right position)
- \( 1/\phi_j \) overdispersion parameter
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 $i = 1, \ldots, N$ documents, falling into category $j = 1, \ldots, J$.

$$y_{ij} \sim \text{Neg.Binomial}(\mu_{ij}, \phi_j)$$

$$\log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_j \theta_i$$

- Interpretation:
  - $\alpha_i$: the variable length of the document (total length)
  - $\zeta_j$: the baseline frequency of a category (issue coverage)
  - $\lambda_j$: responsiveness of the category to the latent variable
  - $\theta_i$: the latent variable (e.g. left-right position)
  - $1/\phi_j$: overdispersion parameter
Parameters and Outcomes

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

- $\lambda = -1$
- $\lambda = -0.5$
- $\lambda = 0$
- $\lambda = 0.5$
- $\lambda = 1$

Latent position $\theta$

Expected number of items

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

- $\lambda = -1$
- $\lambda = -0.5$
- $\lambda = 0$
- $\lambda = 0.5$
- $\lambda = 1$

Expected probability/proportion

Latent position $\theta$
Identification

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

\[
\begin{align*}
\zeta_j &= 0 \\
\bar{\lambda}_j &= 0 \\
\theta_i &\sim N(0, 1) \\
\lambda_j &> \lambda_{j'}
\end{align*}
\]
Inference

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

\[
\begin{align*}
\alpha_i & \sim N(\mu_\alpha, \sigma_\alpha) \\
\zeta_j & \sim N(\mu_\zeta, \sigma_\zeta) \\
\lambda_j & \sim N(0, \sigma_\lambda) \\
\theta_i & \sim N(0, 1) \\
\mu_\alpha & \sim N(0, 5) \\
\mu_\zeta & \sim N(0, 5) \\
\sigma_\alpha & \sim \text{Half-Cauchy}(0, 5) \\
\sigma_\zeta & \sim \text{Half-Cauchy}(0, 5) \\
\sigma_\lambda & \sim \text{Half-Cauchy}(0, 5) \\
1/\phi & \sim \text{Uniform}(0, 200)
\end{align*}
\]
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

<table>
<thead>
<tr>
<th>Left categories</th>
<th>Right categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>103 Anti-imperialism</td>
<td>104 Military: positive</td>
</tr>
<tr>
<td>105 Military: negative</td>
<td>201 Freedom and human rights</td>
</tr>
<tr>
<td>106 Peace</td>
<td>203 Constitutionalism: positive</td>
</tr>
<tr>
<td>107 Internationalism: positive</td>
<td>305 Political authority</td>
</tr>
<tr>
<td>202 Democracy</td>
<td>401 Free enterprise</td>
</tr>
<tr>
<td>403 Market regulation</td>
<td>402 Incentives</td>
</tr>
<tr>
<td>404 Economic planning</td>
<td>407 Protectionism: negative</td>
</tr>
<tr>
<td>406 Protectionism: positive</td>
<td>414 Economic orthodoxy</td>
</tr>
<tr>
<td>412 Controlled economy</td>
<td>501 Welfare state limitation</td>
</tr>
<tr>
<td>413 Nationalisation</td>
<td>601 National way of life: positive</td>
</tr>
<tr>
<td>504 Welfare state expansion</td>
<td>603 Traditional morality: positive</td>
</tr>
<tr>
<td>506 Education expansion</td>
<td>605 Law and order</td>
</tr>
<tr>
<td>701 Labour groups: positive</td>
<td>606 Social harmony</td>
</tr>
</tbody>
</table>

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 ($\zeta$ values):

<table>
<thead>
<tr>
<th>Top 10 Categories</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare State Expansion +</td>
<td>2.20</td>
</tr>
<tr>
<td>Environm. Protection +</td>
<td>1.72</td>
</tr>
<tr>
<td>Technology and Infrastr.</td>
<td>1.70</td>
</tr>
<tr>
<td>Education Expansion +</td>
<td>1.66</td>
</tr>
<tr>
<td>Political Authority +</td>
<td>1.63</td>
</tr>
<tr>
<td>Social Justice +</td>
<td>1.53</td>
</tr>
<tr>
<td>Gov. and Admin. Effic. +</td>
<td>1.50</td>
</tr>
<tr>
<td>Non-Econ. Demogr. Gr. +</td>
<td>1.37</td>
</tr>
<tr>
<td>Democracy +</td>
<td>1.35</td>
</tr>
<tr>
<td>Farmers +</td>
<td>1.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bottom 10 Categories</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marxist Analysis +</td>
<td>-3.37</td>
</tr>
<tr>
<td>Education Limitation +</td>
<td>-3.22</td>
</tr>
<tr>
<td>Labour Groups -</td>
<td>-2.71</td>
</tr>
<tr>
<td>National Way of Life -</td>
<td>-2.18</td>
</tr>
<tr>
<td>Centralisation +</td>
<td>-1.86</td>
</tr>
<tr>
<td>Foreign Special Rel. -</td>
<td>-1.46</td>
</tr>
<tr>
<td>Protectionism -</td>
<td>-1.45</td>
</tr>
<tr>
<td>Constitutionalism -</td>
<td>-1.29</td>
</tr>
<tr>
<td>Corporatism +</td>
<td>-1.23</td>
</tr>
<tr>
<td>Keynesian Dem. Man. +</td>
<td>-1.15</td>
</tr>
</tbody>
</table>
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

Kenneth Benoit and Thomas Däubler
Scaling Manifesto Data
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

Kenneth Benoit and Thomas Däubler
Scaling Manifesto Data
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

Scaling Estimates of Theta
Validation with Expert Survey Placements

$r = 0.82$ W Europe
$r = 0.55$ E Europe
$r = 0.89$ Pacific & N America
$r = 0.75$ All
Country-by-Country Validation (vs. CMP Rile)

Romania not shown (r for rile -.55, r for IRT theta -.44)
\( \hat{\lambda} \) (post-1972 democracies)

- Western Europe, \( \hat{\lambda}_j \)
- E Europe, \( \hat{\lambda}_j \)
- Pacific, \( \hat{\lambda}_j \)
- Far East, \( \hat{\lambda}_j \)
- N America, \( \hat{\lambda}_j \)

- Rile LEFT
- Rile RIGHT

R = 0.37
R = 0.64
R = 0.49
R = -0.097
In the two-dimensional case, we model the mean of the counts as:

$$\log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_1 j \theta_1 i + \lambda_2 j \theta_2 i$$

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 $\lambda_{dj}$
- Reference category for $\zeta_j$
$\lambda_d$ for 2D-model (post-1972 democracies)

(a) Economic Dimension
(b) Social Dimension
Validation with Expert Survey Placements - 2D

**Economic**

- Scaling Estimates of Theta (Economic Dimension) vs Expert Survey Estimates of Taxes vs Spending
- $r = 0.73$

**Social**

- Scaling Estimates of Theta (Social Dimension) vs Expert Survey Estimates of Social Dimension
- $r = 0.53$
Using Alternative Items - Belgian CAP data

Scaling Estimates of Theta

Kenneth Benoit and Thomas Däubler
Scaling Manifesto Data
Validation of Belgian CAP Data Results

Expert Surveys

-1.5 -0.5 0.5 1.0

0.83 0.84 0.79

Agendas Project IRT

-20 -15 -10 -5 0 5

-1.5 -0.5 0.0 0.5

0.46 0.43

CMP IRT

-20 -15 -10 -5 0 5

-1.5 -0.5 0.5 1.0

0.84

RILE

-20 -15 -10 -5 0 5

-20 -15 -10 -5 0 5

0.79
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
  - …