

A Spatial Model of Coded Political Texts

Martin Elff

E-mail: melff@essex.ac.uk

Department of Government
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom

Research Seminar, Nuffield College, University of
Oxford, 1st November 2011

- 1 The Problem
- 2 The Model
- 3 Estimation and Prediction
- 4 Applications
- 5 Conclusion

- 1 The Problem
- 2 The Model
- 3 Estimation and Prediction
- 4 Applications
- 5 Conclusion

Spatial models of politics

Spatial models of politics

- Formal spatial models of politics, spatial models of electoral behaviour, etc.

Spatial models of politics

- Formal spatial models of politics, spatial models of electoral behaviour, etc.
- Political options/policies etc. have a location in a political space

Spatial models of politics

- Formal spatial models of politics, spatial models of electoral behaviour, etc.
- Political options/policies etc. have a location in a political space
- Political space can be one- or multidimensional

Spatial models of politics

- Formal spatial models of politics, spatial models of electoral behaviour, etc.
- Political options/policies etc. have a location in a political space
- Political space can be one- or multidimensional
- 'Similar' political positions are less distant etc.

Spatial models of politics

- Formal spatial models of politics, spatial models of electoral behaviour, etc.
- Political options/policies etc. have a location in a political space
- Political space can be one- or multidimensional
- 'Similar' political positions are less distant etc.
- Actors have ideal points – of two policy options/policies actors will chose the one that is less distant to their ideal point

Spatial models of politics

- Formal spatial models of politics, spatial models of electoral behaviour, etc.
 - Political options/policies etc. have a location in a political space
 - Political space can be one- or multidimensional
 - 'Similar' political positions are less distant etc.
 - Actors have ideal points – of two policy options/policies actors will chose the one that is less distant to their ideal point
- ⇒ Political spaces are *Euclidean* spaces

Spatial models of politics

x
•

Spatial models of politics

a · ***x*** ·

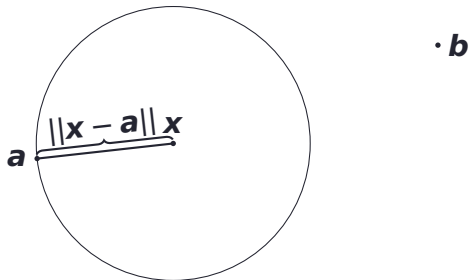
Spatial models of politics

a ·

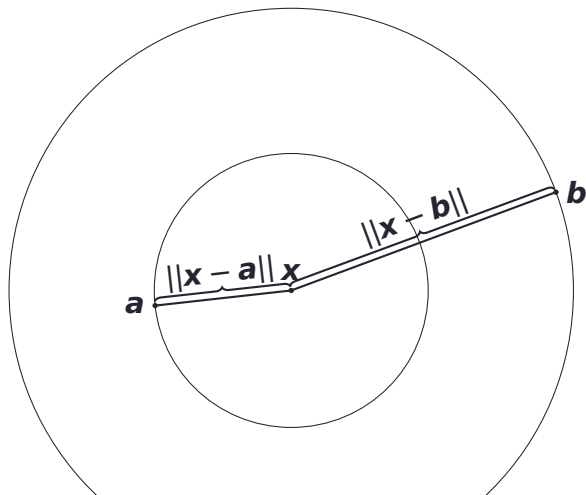
x
·

· ***b***

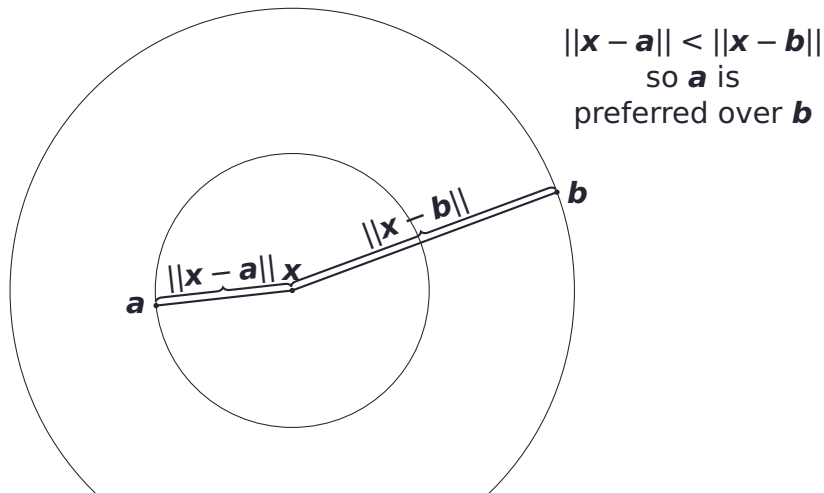
Spatial models of politics



Spatial models of politics



Spatial models of politics



Quantitative text analysis

Quantitative text analysis

- Usual procedure:

Quantitative text analysis

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)
 - 2 Assignment of linguistic units into categories (by trained coders or automatically)

Quantitative text analysis

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)
 - 2 Assignment of linguistic units into categories (by trained coders or automatically)
 - 3 Count of units within categories

Quantitative text analysis

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)
 - 2 Assignment of linguistic units into categories (by trained coders or automatically)
 - 3 Count of units within categories
- Issues:

Quantitative text analysis

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)
 - 2 Assignment of linguistic units into categories (by trained coders or automatically)
 - 3 Count of units within categories
- Issues:
 - Classification error when human coders are employed

Quantitative text analysis

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)
 - 2 Assignment of linguistic units into categories (by trained coders or automatically)
 - 3 Count of units within categories
- Issues:
 - Classification error when human coders are employed
 - Emphases of policy objectives reflect *both* the saliency of a substantial policy area *and* the position *within* a policy area

- Usual procedure:
 - 1 Segmentation of text into linguistic units (words, sentences, or quasi-sentences)
 - 2 Assignment of linguistic units into categories (by trained coders or automatically)
 - 3 Count of units within categories
- Issues:
 - Classification error when human coders are employed
 - Emphases of policy objectives reflect *both* the saliency of a substantial policy area *and* the position *within* a policy area
 - Emphasis counts are not directly compatible with locations/positions within an Euclidean policy space

- Salience of a policy area
 - Urgency of a policy area, e.g. economic policy
 - Varies according to political or external circumstances, e.g. economic problems – recession – or recovery/prosperity

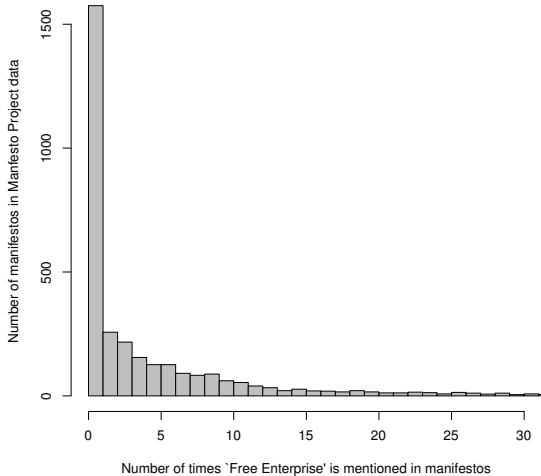
Salience and Position

- Salience of a policy area
 - Urgency of a policy area, e.g. economic policy
 - Varies according to political or external circumstances, e.g. economic problems – recession – or recovery/prosperity
- Position within a policy area
 - In terms of economic policy – e.g.: *laissez-faire* (economic ‘Right’) versus interventionism or state-controlled economy (economic ‘Left’)

Salience and Position

- Salience of a policy area
 - Urgency of a policy area, e.g. economic policy
 - Varies according to political or external circumstances, e.g. economic problems – recession – or recovery/prosperity
- Position within a policy area
 - In terms of economic policy – e.g.: *laissez-faire* (economic ‘Right’) versus interventionism or state-controlled economy (economic ‘Left’)
- $E = S \cdot P$
 - E a measure of emphasis
 - S related to salience of policy area
 - P related to position within policy area

Distribution of the emphasis of 'Free of enterprise' in the Manifesto Project data



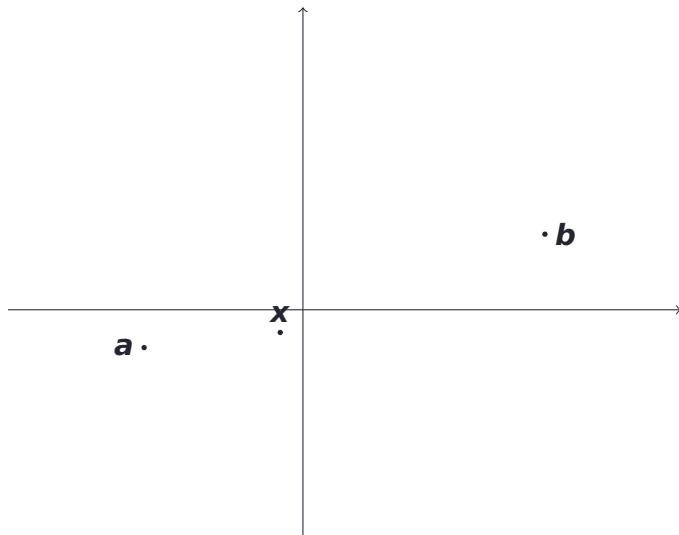
Locations, positions, and coordinates

a ·

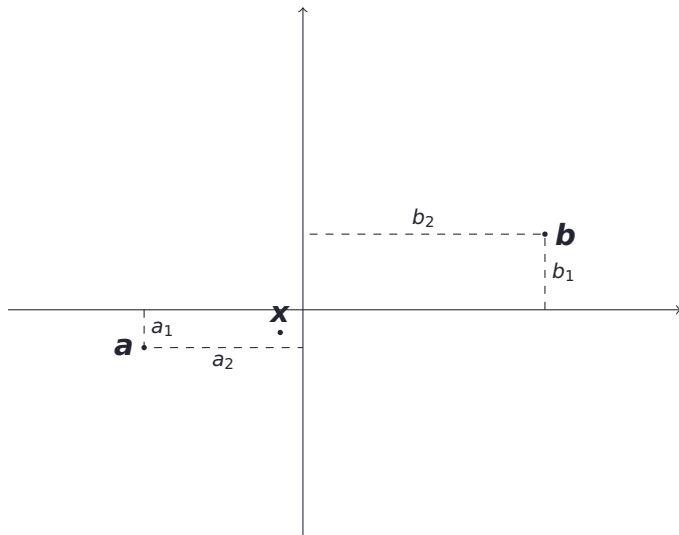
x
·

· ***b***

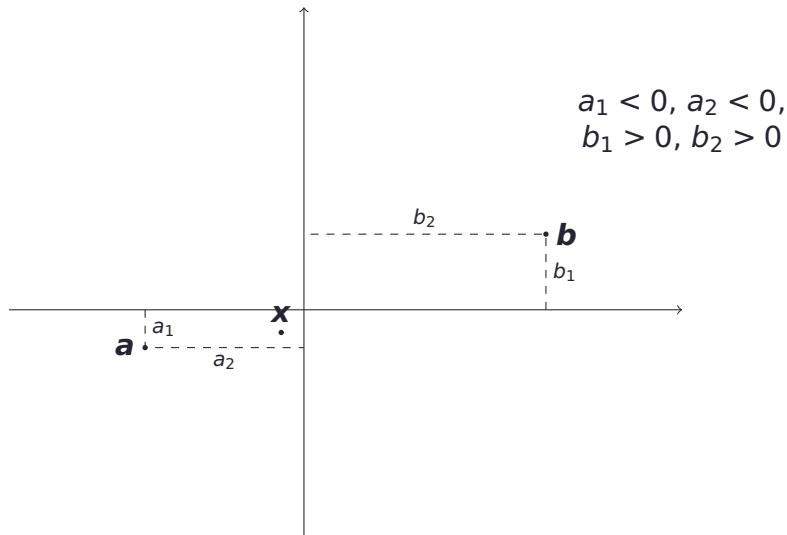
Locations, positions, and coordinates



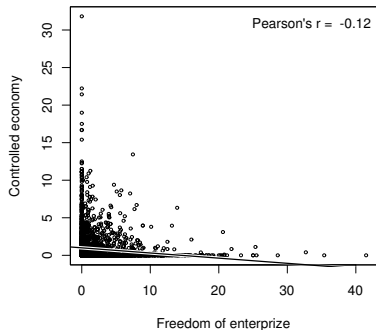
Locations, positions, and coordinates



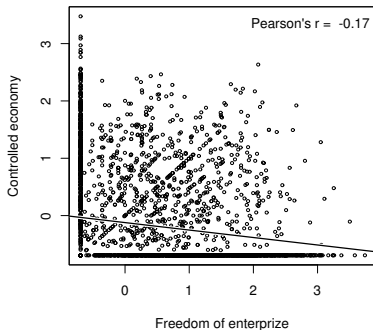
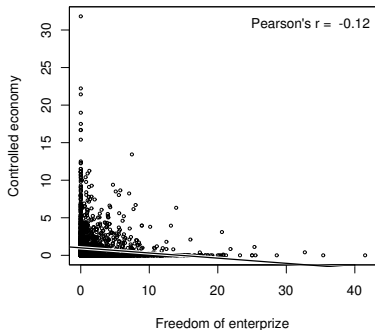
Locations, positions, and coordinates



Relation between the emphasis of the policy objective “Controlled economy” and of the policy objective “Freedom of enterprise”



Relation between the emphasis of the policy objective “Controlled economy” and of the policy objective “Freedom of enterprise”



The Model

1 The Problem

2 The Model

3 Estimation and Prediction

4 Applications

5 Conclusion

Assumptions about political positions (and their dynamics)

Assumptions about political positions (and their dynamics)

1 Policy area \leftrightarrow Euclidean space with D dimensions

Assumptions about political positions (and their dynamics)

- 1 Policy area \leftrightarrow Euclidean space with D dimensions
- 2 Policy objective $i \leftrightarrow$ location at α_i

Assumptions about political positions (and their dynamics)

- 1 Policy area \leftrightarrow Euclidean space with D dimensions
- 2 Policy objective $i \leftrightarrow$ location at α_i
- 3 Political actor j at time $t \leftrightarrow$ position at \mathbf{b}_{jt}

Assumptions about political positions (and their dynamics)

- 1 Policy area \leftrightarrow Euclidean space with D dimensions
- 2 Policy objective $i \leftrightarrow$ location at α_i
- 3 Political actor j at time $t \leftrightarrow$ position at \mathbf{b}_{jt}
- 4 Initial positions \mathbf{b}_{j0} have normal distribution with mean β and variance Σ_0

Assumptions about political positions (and their dynamics)

- 1 Policy area \leftrightarrow Euclidean space with D dimensions
- 2 Policy objective $i \leftrightarrow$ location at α_i
- 3 Political actor j at time $t \leftrightarrow$ position at \mathbf{b}_{jt}
- 4 Initial positions \mathbf{b}_{j0} have normal distribution with mean β and variance Σ_0
- 5 Conditional on prior positions, \mathbf{b}_{jt} have normal distribution with mean $\beta + \Gamma(\mathbf{b}_{jt-1} - \beta)$ and variance Σ_1

Assumptions about political positions (and their dynamics)

- 1 Policy area \leftrightarrow Euclidean space with D dimensions
- 2 Policy objective $i \leftrightarrow$ location at α_i
- 3 Political actor j at time $t \leftrightarrow$ position at \mathbf{b}_{jt}
- 4 Initial positions \mathbf{b}_{j0} have normal distribution with mean $\boldsymbol{\beta}$ and variance $\boldsymbol{\Sigma}_0$
- 5 Conditional on prior positions, \mathbf{b}_{jt} have normal distribution with mean $\boldsymbol{\beta} + \boldsymbol{\Gamma}(\mathbf{b}_{jt-1} - \boldsymbol{\beta})$ and variance $\boldsymbol{\Sigma}_1$ Put simpler: $\mathbf{u}_{jt} = \mathbf{b}_{jt} - \boldsymbol{\beta}$, and $E(\mathbf{u}_{jt}) = \boldsymbol{\Gamma}\mathbf{u}_{jt-1}$

Illustration of the dynamics of political positions

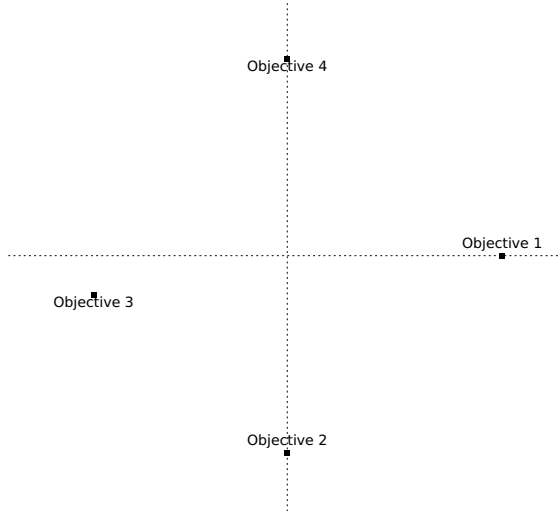


Illustration of the dynamics of political positions

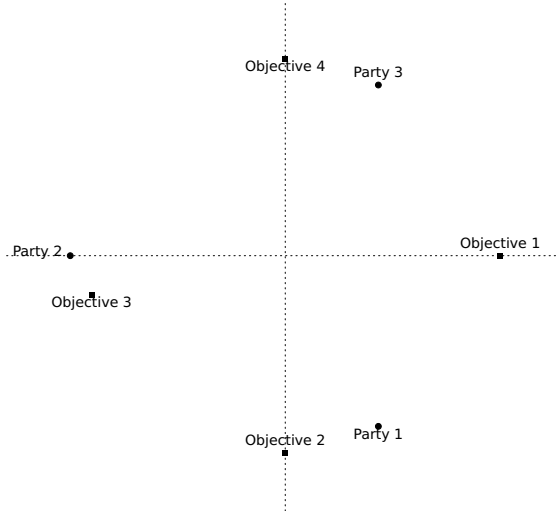


Illustration of the dynamics of political positions

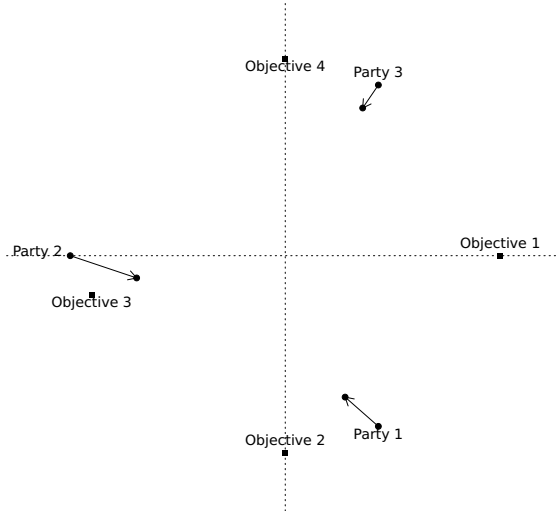


Illustration of the dynamics of political positions

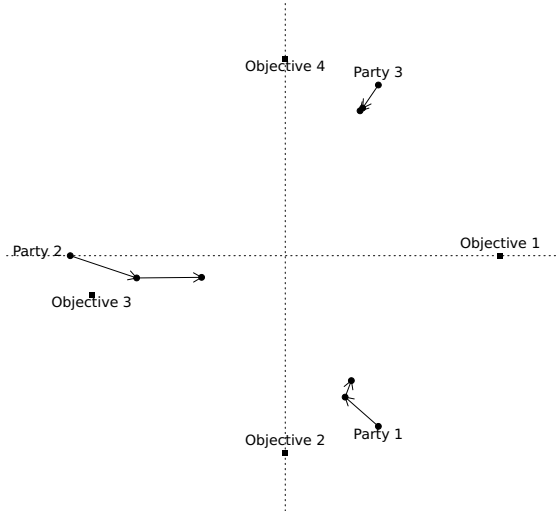


Illustration of the dynamics of political positions

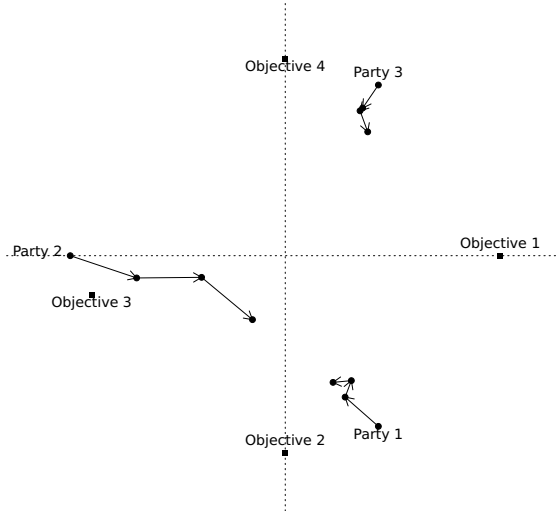


Illustration of the dynamics of political positions

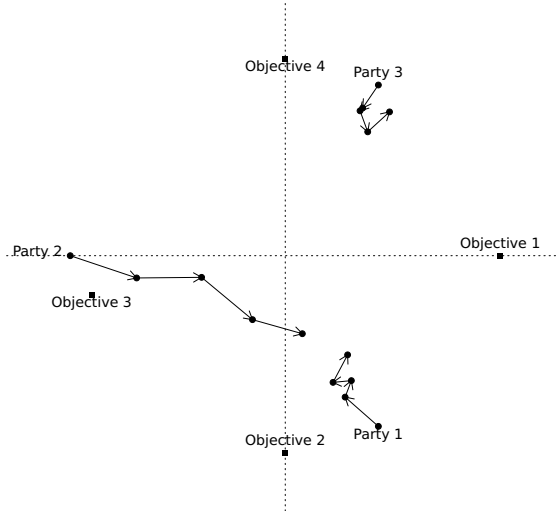


Illustration of the dynamics of political positions

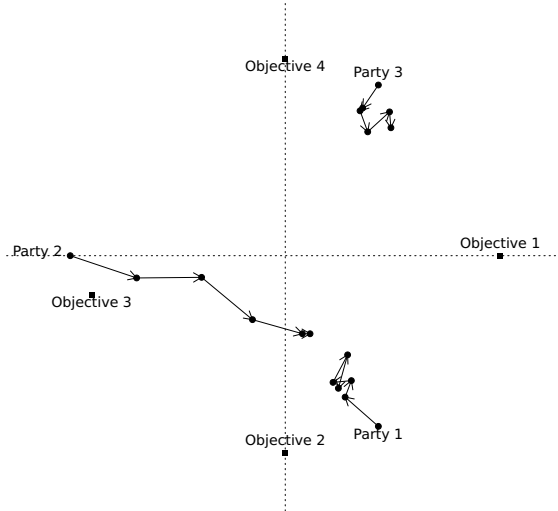


Illustration of the dynamics of political positions

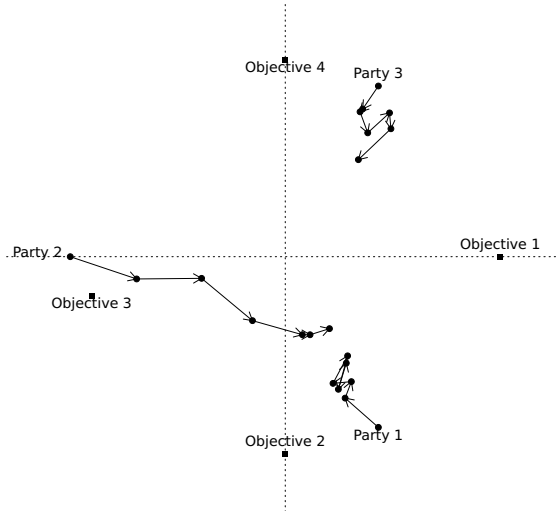


Illustration of the dynamics of political positions

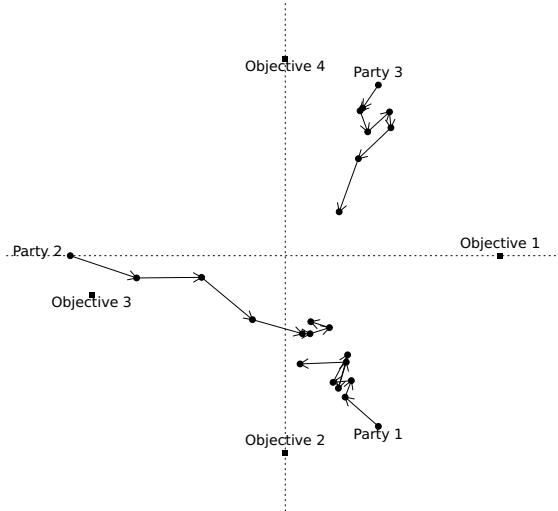


Illustration of the dynamics of political positions

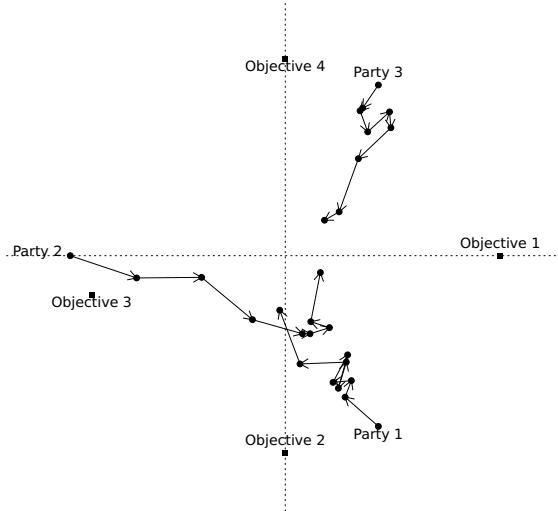
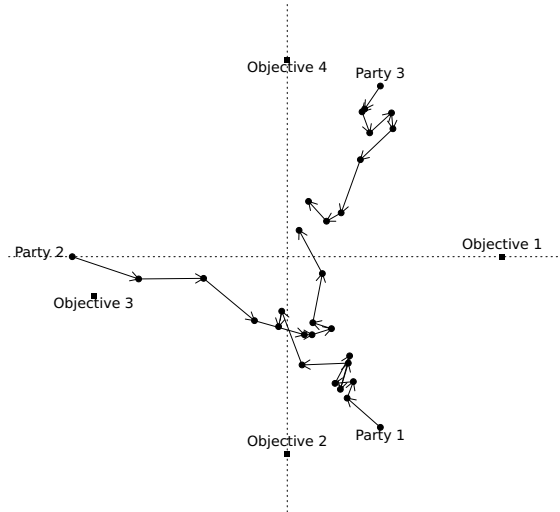


Illustration of the dynamics of political positions



Assumptions about the 'data-generating process'

Assumptions about the 'data-generating process'

- 1 Greater distance $\Delta_{ijt} = \|\alpha_i - \mathbf{b}_{jt}\|$ leads to lower expected emphasis $E(M_{ijt})$

Assumptions about the 'data-generating process'

- 1 Greater distance $\Delta_{ijt} = \|\alpha_i - \mathbf{b}_{jt}\|$ leads to lower expected emphasis $E(M_{ijt})$
- 2 Higher salience s_{jt} leads to higher expected emphasis $E(M_{ijt})$.

Assumptions about the 'data-generating process'

- 1 Greater distance $\Delta_{ijt} = \|\alpha_i - \mathbf{b}_{jt}\|$ leads to lower expected emphasis $E(M_{ijt})$
- 2 Higher salience s_{jt} leads to higher expected emphasis $E(M_{ijt})$.
- 3 Emphasis counts have Poisson distribution with mean parameter μ_{ijt} given by

$$E(M_{ijt}) = \mu_{ijt} = e^{\lambda_{jt}} e^{-\frac{1}{2}\Delta_{ijt}^2}$$

Assumptions about the 'data-generating process'

- 1 Greater distance $\Delta_{ijt} = \|\mathbf{\alpha}_i - \mathbf{b}_{jt}\|$ leads to lower expected emphasis $E(M_{ijt})$
- 2 Higher salience s_{jt} leads to higher expected emphasis $E(M_{ijt})$.
- 3 Emphasis counts have Poisson distribution with mean parameter μ_{ijt} given by

$$E(M_{ijt}) = \mu_{ijt} = e^{\lambda_{jt}} e^{-\frac{1}{2}\Delta_{ijt}^2}$$

Implication: *Conditional* on the sum $N_{jt} = \sum_i M_{ijt}$ the counts have a multinomial distribution with cell probabilities

$$\pi_{ijt} = \frac{e^{-\frac{1}{2}\Delta_{ijt}^2}}{\sum_h e^{-\frac{1}{2}\Delta_{hjt}^2}}$$

Illustration of the 'data-generating process'

Objective 1

α_1



Illustration of the 'data-generating process'

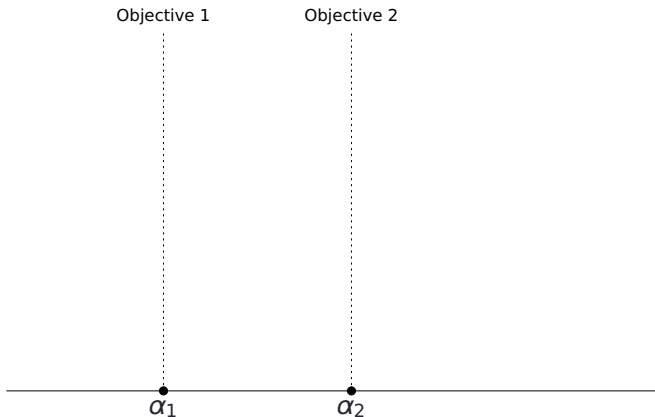


Illustration of the 'data-generating process'

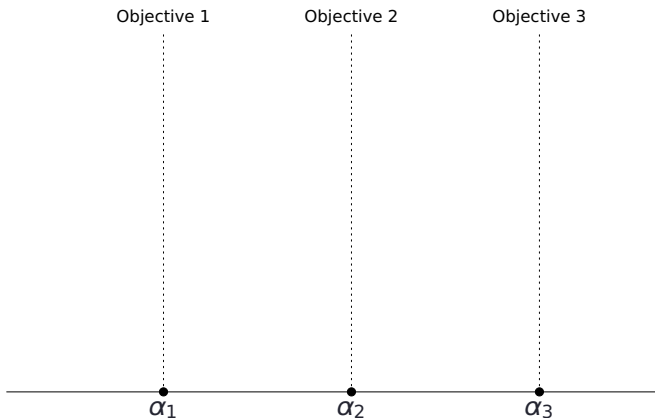


Illustration of the 'data-generating process'

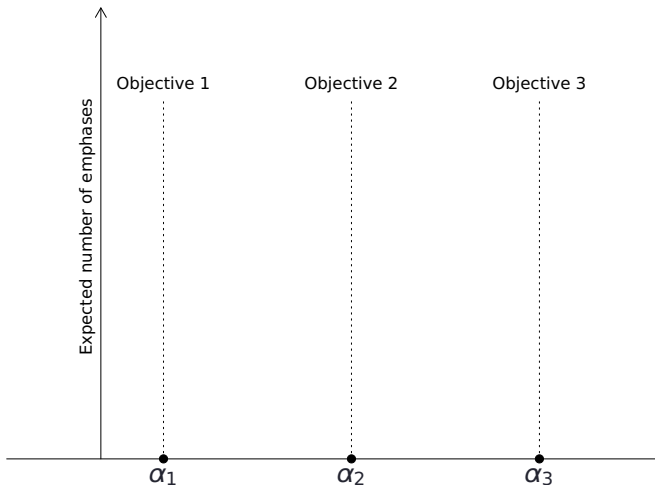


Illustration of the 'data-generating process'

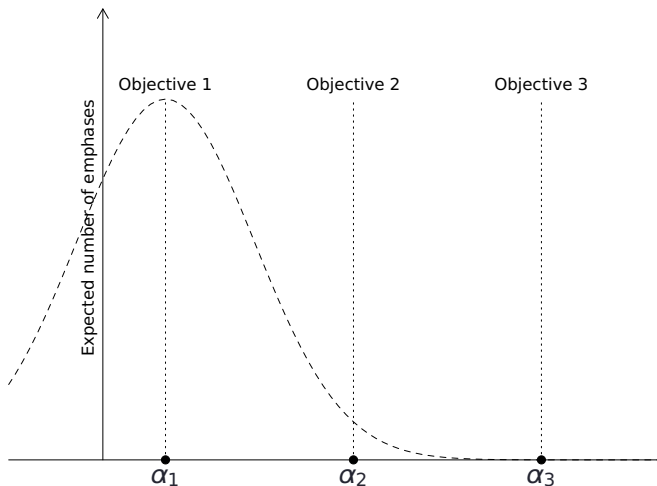


Illustration of the 'data-generating process'

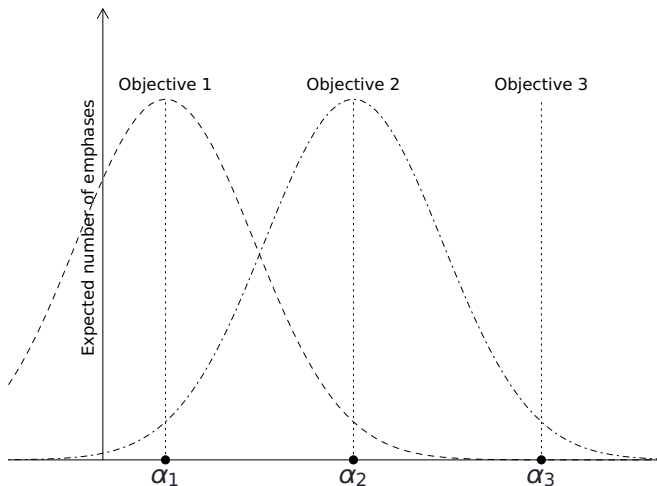


Illustration of the 'data-generating process'

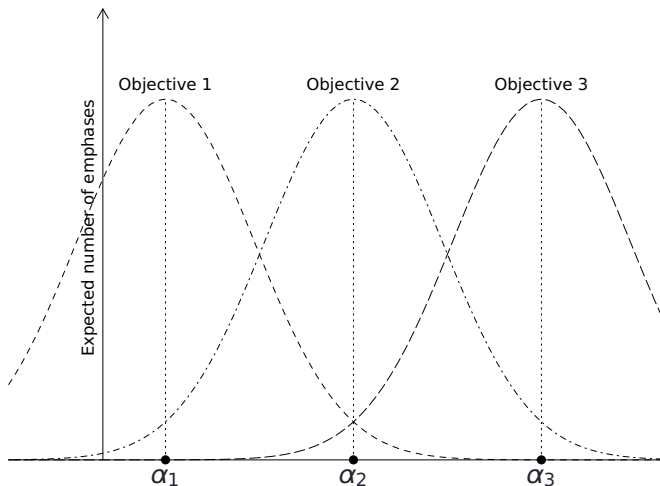


Illustration of the 'data-generating process'

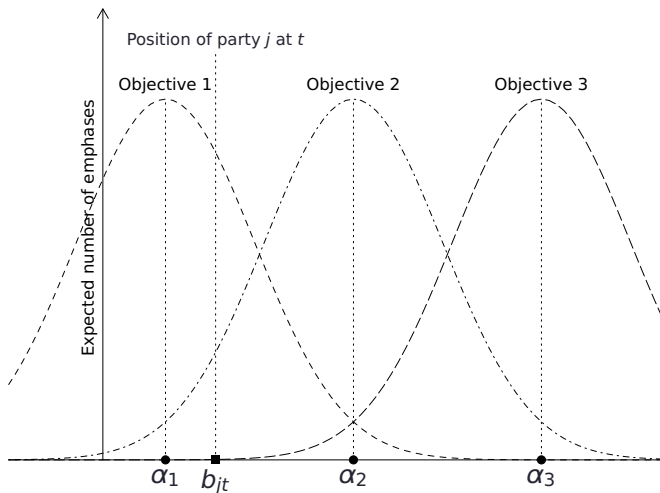


Illustration of the 'data-generating process'

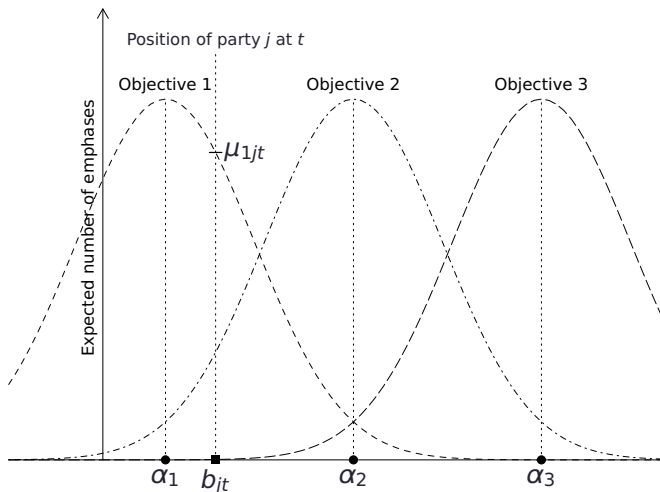


Illustration of the 'data-generating process'

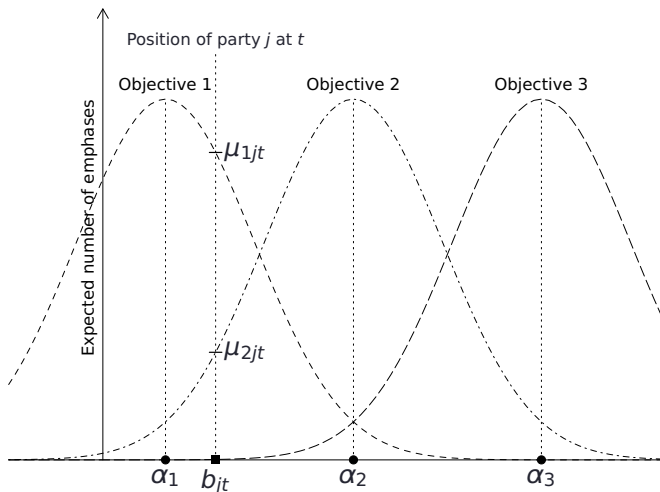
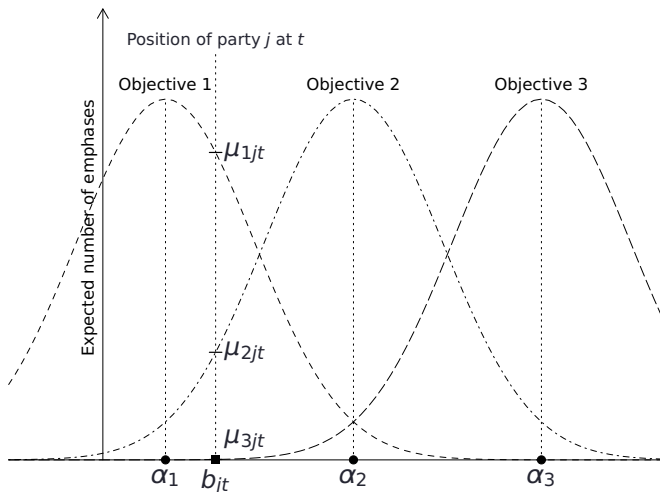
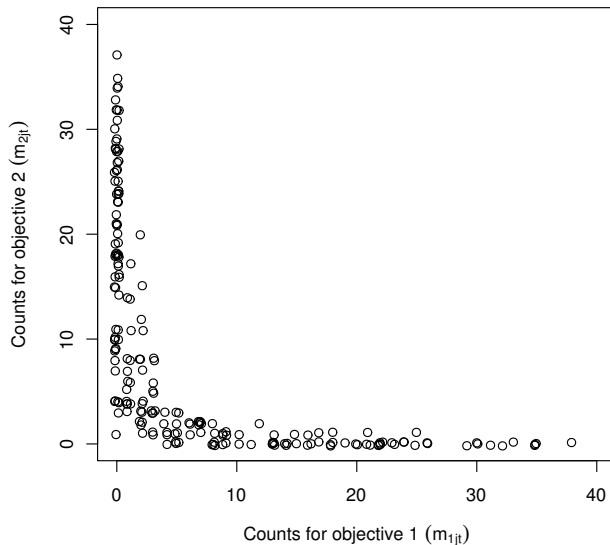


Illustration of the 'data-generating process'



Hypothetical distribution of emphasis counts



Estimation and Prediction

- 1 The Problem
- 2 The Model
- 3 Estimation and Prediction**
- 4 Applications
- 5 Conclusion

Estimation of model parameters

Estimation of model parameters

- Parameters:

Estimation of model parameters

- Parameters:
 - locations α_j of objectives

Estimation of model parameters

- Parameters:
 - locations α_j of objectives
 - overall mean β of positions,

Estimation of model parameters

- Parameters:
 - locations α_j of objectives
 - overall mean β of positions,
 - variance Σ_0 of initial positions
 - variance Σ_1 of 'updated' positions

Estimation of model parameters

- Parameters:
 - locations α_j of objectives
 - overall mean β of positions,
 - variance Σ_0 of initial positions
 - variance Σ_1 of 'updated' positions
- Estimation:

Estimation of model parameters

- Parameters:
 - locations α_j of objectives
 - overall mean β of positions,
 - variance Σ_0 of initial positions
 - variance Σ_1 of 'updated' positions
- Estimation:
 - Estimator is maximum (marginal) likelihood

Estimation of model parameters

■ Parameters:

- locations α_j of objectives
- overall mean β of positions,
- variance Σ_0 of initial positions
- variance Σ_1 of 'updated' positions

■ Estimation:

- Estimator is maximum (marginal) likelihood

$$\mathcal{L}(\mathbf{m}_j, \mathbf{b}_j; \boldsymbol{\psi}) = \prod_j \int f(\mathbf{b}_j; \boldsymbol{\psi}) f(\mathbf{m}_j | \mathbf{b}_j; \boldsymbol{\psi}) d\mathbf{b}_j$$

Estimation of model parameters

■ Parameters:

- locations α_j of objectives
- overall mean β of positions,
- variance Σ_0 of initial positions
- variance Σ_1 of 'updated' positions

■ Estimation:

- Estimator is maximum (marginal) likelihood

$$\mathcal{L}(\mathbf{m}_j, \mathbf{b}_j; \boldsymbol{\psi}) = \prod_j \int f(\mathbf{b}_j; \boldsymbol{\psi}) f(\mathbf{m}_j | \mathbf{b}_j; \boldsymbol{\psi}) d\mathbf{b}_j$$

- Algorithm is Expectation-Maximisation (EM)

Estimation of model parameters

■ Parameters:

- locations α_j of objectives
- overall mean β of positions,
- variance Σ_0 of initial positions
- variance Σ_1 of 'updated' positions

■ Estimation:

- Estimator is maximum (marginal) likelihood

$$\mathcal{L}(\mathbf{m}_j, \mathbf{b}_j; \boldsymbol{\psi}) = \prod_j \int f(\mathbf{b}_j; \boldsymbol{\psi}) f(\mathbf{m}_j | \mathbf{b}_j; \boldsymbol{\psi}) d\mathbf{b}_j$$

- Algorithm is Expectation-Maximisation (EM)
- Expectations approximation using Monte-Carlo integration (with [independent] importance sampling)

Prediction of actors' political positions

Prediction of actors' political positions

- Prediction of actors' positions rests on (EB) posterior distribution

$$f(\mathbf{b}_j | \mathbf{m}_j; \hat{\boldsymbol{\psi}}) = \frac{f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}})}{\int f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}}) d \mathbf{b}_j}$$

Prediction of actors' political positions

- Prediction of actors' positions rests on (EB) posterior distribution

$$f(\mathbf{b}_j | \mathbf{m}_j; \hat{\boldsymbol{\psi}}) = \frac{f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}})}{\int f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}}) d \mathbf{b}_j}$$

- Point predictions: (simulated) posterior means

Prediction of actors' political positions

- Prediction of actors' positions rests on (EB) posterior distribution

$$f(\mathbf{b}_j | \mathbf{m}_j; \hat{\boldsymbol{\psi}}) = \frac{f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}})}{\int f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}}) d \mathbf{b}_j}$$

- Point predictions: (simulated) posterior means
- Prediction uncertainty: (simulated) posterior variances

Prediction of actors' political positions

- Prediction of actors' positions rests on (EB) posterior distribution

$$f(\mathbf{b}_j | \mathbf{m}_j; \hat{\boldsymbol{\psi}}) = \frac{f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}})}{\int f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}}) d \mathbf{b}_j}$$

- Point predictions: (simulated) posterior means
- Prediction uncertainty: (simulated) posterior variances
- Interval predictions: (simulated) quantiles of posterior distribution

Prediction of actors' political positions

- Prediction of actors' positions rests on (EB) posterior distribution

$$f(\mathbf{b}_j | \mathbf{m}_j; \hat{\boldsymbol{\psi}}) = \frac{f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}})}{\int f(\mathbf{b}_j; \hat{\boldsymbol{\psi}})f(\mathbf{m}_j | \mathbf{b}_j; \hat{\boldsymbol{\psi}}) d \mathbf{b}_j}$$

- Point predictions: (simulated) posterior means
- Prediction uncertainty: (simulated) posterior variances
- Interval predictions: (simulated) quantiles of posterior distribution
- Simulation using acceptance-rejection sampling

Applications

1 The Problem

2 The Model

3 Estimation and Prediction

4 Applications

5 Conclusion

The Comparative Manifestos Project

The Comparative Manifestos Project

- Also known as: Manifesto Research Group, Manifesto Project

The Comparative Manifestos Project

- Also known as: Manifesto Research Group, Manifesto Project
- Party manifestos of major democracies, 1945–2008

The Comparative Manifestos Project

- Also known as: Manifesto Research Group, Manifesto Project
- Party manifestos of major democracies, 1945–2008
- 56 policy objectives

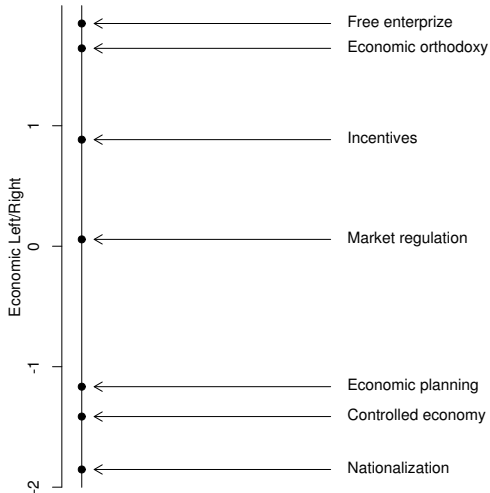
The Comparative Manifestos Project

- Also known as: Manifesto Research Group, Manifesto Project
- Party manifestos of major democracies, 1945–2008
- 56 policy objectives
- Segmentation of manifestos into ‘quasi-sentences’

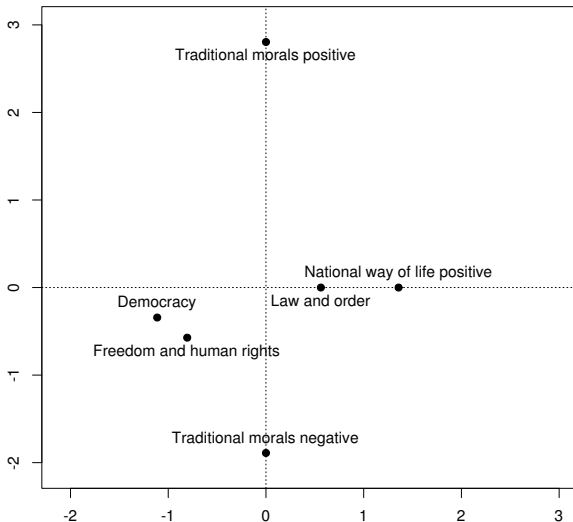
The Comparative Manifestos Project

- Also known as: Manifesto Research Group, Manifesto Project
- Party manifestos of major democracies, 1945–2008
- 56 policy objectives
- Segmentation of manifestos into ‘quasi-sentences’
- Data set contains percentages, but counts can be recovered (taking into account the total number of coded quasi-sentences)

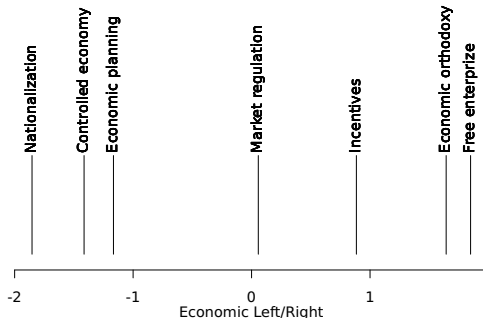
Locations of policy objectives in economic policy space



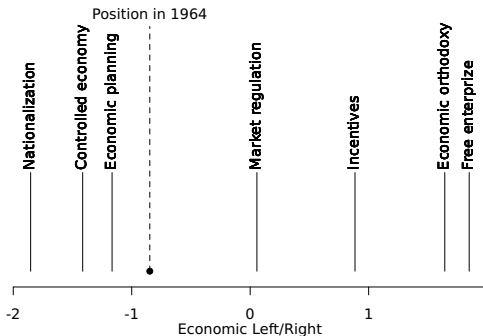
Locations of policy objectives in socio/cultural policy space



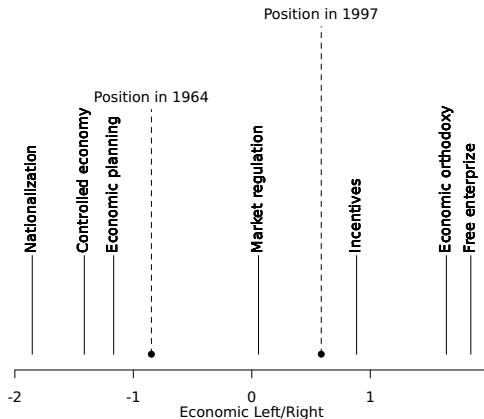
Political positions of the UK Labour Party on the economic left/right dimension



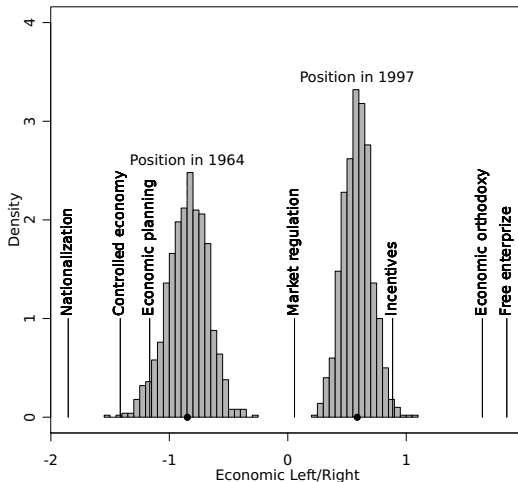
Political positions of the UK Labour Party on the economic left/right dimension



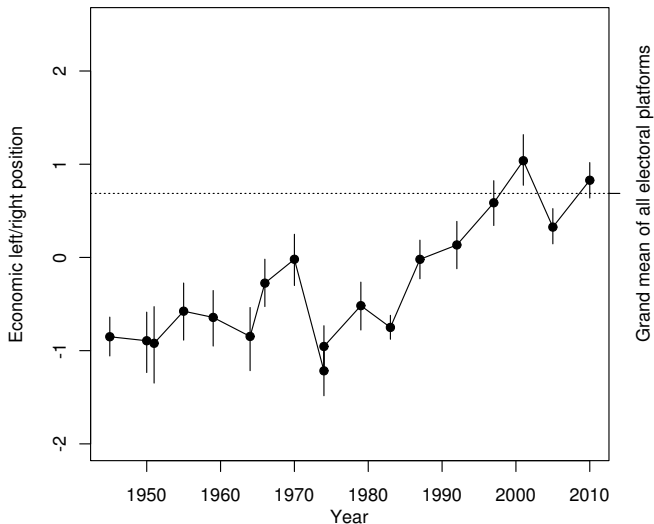
Political positions of the UK Labour Party on the economic left/right dimension



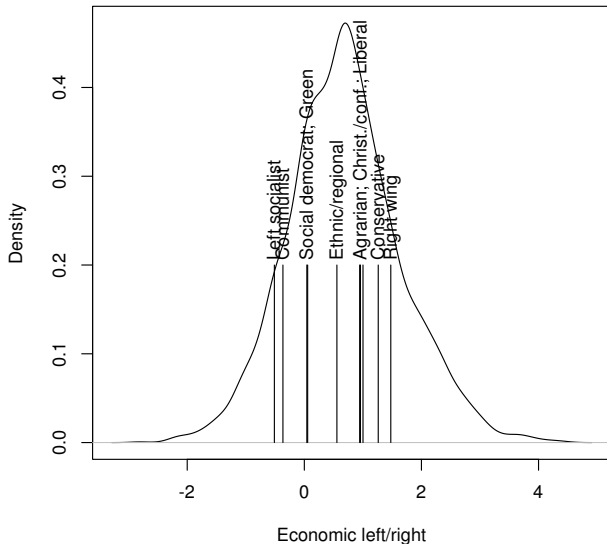
Political positions of the UK Labour Party on the economic left/right dimension



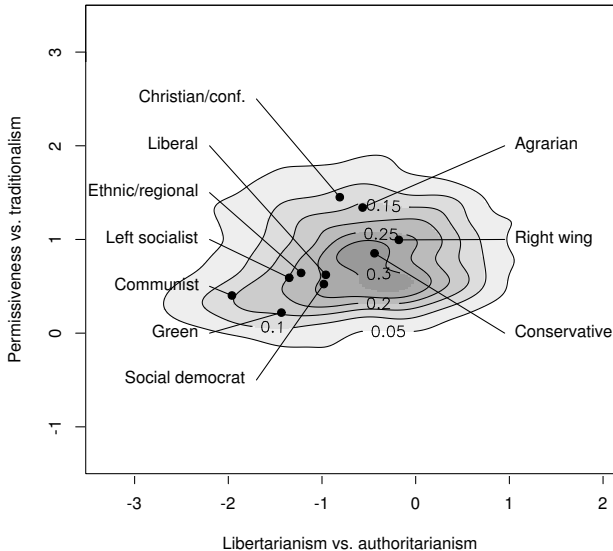
Political positions of the UK Labour Party on the economic left/right dimension



Average political positions of party families in the economic policy space



Average political positions of party families in the socio-cultural policy space



- 1 The Problem
- 2 The Model
- 3 Estimation and Prediction
- 4 Applications
- 5 Conclusion**

Summary and further remarks

Summary and further remarks

- Separation of salience and position

Summary and further remarks

- Separation of salience and position
- Explicit modelling of dynamics

Summary and further remarks

- Separation of salience and position
- Explicit modelling of dynamics
- Avoids incidental parameter problem and enables out-of-sample predictions

Summary and further remarks

- Separation of salience and position
- Explicit modelling of dynamics
- Avoids incidental parameter problem and enables out-of-sample predictions
- Measures of uncertainty are provided

Summary and further remarks

- Separation of salience and position
- Explicit modelling of dynamics
- Avoids incidental parameter problem and enables out-of-sample predictions
- Measures of uncertainty are provided
- Using ML framework allows to test hypotheses (e.g. LM tests)

Summary and further remarks

- Separation of salience and position
- Explicit modelling of dynamics
- Avoids incidental parameter problem and enables out-of-sample predictions
- Measures of uncertainty are provided
- Using ML framework allows to test hypotheses (e.g. LM tests)
- Applied to CMP data, but can in principle applied to other coded corpora of political text

Discussion

- Identifying restrictions necessary

- Identifying restrictions necessary
- Model search difficult

- Identifying restrictions necessary
- Model search difficult
- Computationally intensive

- Identifying restrictions necessary
- Model search difficult
- Computationally intensive
- Implemented in *R* and *C++* (using *R* packages *Rcpp* and *RcppArmadillo*)