Putting Text in Context: How to Estimate Better Left-Right Positions by Scaling Party Manifesto Data using Item Response Theory*

Kenneth Benoit
London School of Economics and Trinity College Dublin
kbenoit@lse.ac.uk

Thomas Däubler
University of Mannheim and MZES
thomas.daeubler@mzes.uni-mannheim.de

May 8, 2014

Abstract

For over three decades, party manifestos have formed the largest source of textual data for estimating party policy positions and emphases, resting on the pillars of two key assumptions: that party policy positions can be measured on known dimensions by counting text units in predefined categories, and that more text in a given category indicates stronger emphasis. Here we revisit the inductive approach to estimating policy positions from party manifesto data, demonstrating that there is no single definition of left-right policy that fits well in all contexts, even though meaningful comparisons can be made by locating parties on a single dimension in each context. To estimate party positions, we apply a Bayesian, multi-level, Poisson-IRT measurement model to category counts from coded party manifestos. By treating the categories as "items" and policy positions as a latent variable, we are able to recover not only left-right estimates but also direct estimates of how each policy category relates to this dimension, without having to decide these relationships in advance based on political theory, exploratory analysis, or guesswork. Finally, the flexibility of our framework permits numerous extensions, designed to incorporate models of manifesto authorship, coding effects, and additional explanatory variables (including time and country effects) to improve estimates.

Key Words: Party manifestos, IRT, Bayesian estimation, Comparative Manifestos Project, policy positions, measurement.

*Prepared for the "Mapping Policy Preferences from Texts" Conference, May 15–16, 2014, Berlin. A much earlier version was presented at the Annual Meeting of the American Political Science Association, August 28–September 1, 2013, Chicago. This research was supported by the European Research Council grant ERC-2011-StG 283794-QUANTESS. We thank Ben Lauderdale and Jouni Kuha for comments and suggestions, and Korinna Veller for research assistance.

By far, measures of *left-right* policy positioning outstrip all other measures of policy distance when comparing political parties across space and time. With roots in early spatial descriptions of the seating in the Constituent Assembly following the French Revolution (see Carlyle, 1888, 92 in Benoit and Laver, 2006, 12–13), this orientational metaphor has proven one of the most resilient of all conceptual frameworks for distinguishing political actors by their policy differences on a single dimension. To make the empirical measurements of distance necessary to test spatial models of political competition, valid and reliable measures of left-right policy positions across countries and times has become the "holy grail" of measurement in comparative political science. While a large body of recent scholarship has sought to define and to locate parties on more specific policy dimensions (e.g. Benoit and Laver, 2006; Bakker et al., 2012), widespread disagreement exists as to how to conceptualise and measure a common left-right dimension, as witnessed in numerous debates over the relative merits of expert surveys (Benoit and Laver, 2007), indexes constructed from the content analysis of manifestos (Budge and Meyer, 2013), debates over how to best construct such indexes (Franzmann, 2013; Jahn, 2014), the validity of scaling roll-call votes (Proksch and Slapin, 2010), and a growth industry using automated and statistical approaches to scale positions from political text (e.g. Laver, Benoit and Garry, 2003; Slapin and Proksch, 2008).

On what points do researchers agree when it comes to defining and measuring left-right policy? First, is is widely believed that it is possible, and more or less valid in most contexts, to differentiate political parties along a single dimension. Early proponents of the spatial model of party competition argued that party and voter positions can be ordered from left to right on a "manner agreed upon by all" (Downs, 1957, 142). While many configurations of positioning on specific policy dimensions are possible, in practice these tend to bundle into a "super issue" (e.g. Gabel and Huber, 2000; Laver and Budge, 1992) that it is meaningful, and certainly *useful*, that is generally called the "left-right" policy dimension. Analysis of expert placements of parties on a "left-right" dimension—without specifying in advance what this should mean—have shown clearly that it is possible to predict party placements on this dimension from policy locations on more specific policy measures (Benoit and Laver, 2006, 141).

Second, however, there is no consensus as to what are the common components of this super-issue called "left-right". In the most comprehensive examinations of this issue to date, Benoit and Laver (2006) found in no uncertain terms that the substantive content of a "left-right" dimension varies significantly across different contexts, to such an extent that "it may be impossible for any single scale to measure this dimension in a manner that can be used for reliable or meaningful cross-national comparison" (143). By implication, claims that indexes with fixed components can apply universally (e.g. Budge and Meyer, 2013, 88) have been shown to be exaggerated or false when stretched to contexts outside of where they were developed (Mölder, 2013). While the key policy differences between classical left and right parties can be found in much classical philosophical thought—including the writings of Marx and Engels on the left on Disraeli and Spencer on the right (Budge and Meyer, 2013)—the intellectual longevity and brilliance of these thinkers of makes the measurement components of a fixed-scale approach to left-right based on their writings no more universally applicable than any other basis for choice. If these components vary across time and space, in other words, there may not be any single "correct" definition of the components of the left-right super-issue.

In what follows, we go straight to the issue of what the left-right dimension means in specific contexts, and whether it is useful or even possible to define it according to a fixed set of components. Drawing on the single largest dataset of evidence on cross-national party positions over time, the Comparative Manifesto Project's dataset of manifesto content analysis of over 3,200 manifestos in 55 countries, we propose a method of constructing scales and comparing the components of these scales across different contexts. Using Item-Response Theory, we extend the "vanilla" method of Gabel and Huber (2000) using a Bayesian framework that allows us to estimate parameters directly on the "items": here, the policy content categories that contribute to the measurement of the left-right super-issue. Contrasting this to other approaches based on static indexes such as the CMP's "Rile" index and validating our measures using external expert placements, we both highlight flaws in existing approaches and offer a constructive method for measuring left-right policy using our IRT approach. In addition, we show how this model can be extended to specific dimensions and multiple dimensions by specifying appropriate prior distributions and extending the model to include multiple latent dimensions.

1 Inductive v. Deductive Approaches to Defining Left-Right

While left-right policy remains the most widely measured dimension of difference in the study of comparative political competition, the method of defining and measuring this dimension is hardly a matter of consensus. Two broad approaches to defining the "left-right" political dimension exist, with different implications for measurement.

A first perspective, termed the *a priori* approach (Benoit and Laver, 2006), specifies the substantive content of the left-right dimension as known, and then seeks to locate the policy positions of political actors on this dimension. Surveys use this method when asking experts or citizens, for instance, about parties' locations on very specific dimensions of policy, for example their preferences for state involvement in the economy or the role of religion in public life (e.g. Benoit and Laver, 2006; Bakker et al., 2012). For parties to be located on a more general or *lower-level* dimension such as left-right policy in the same manner, the components of this dimension must also be specified. Because contemporary scholars disagree over this content, many attempting to define left-right content make recourse to authoritative sources long dead, usually with roots in political theory. For example, Jahn (2011, 750-751) draws on classic distinctions between "left" and "right" attitudes toward equality and the welfare state, found in the thought of Rousseau and Nietzsche. The authoritativeness of the definition is then taken as conveying construct validity to the measurement of left-right according to the pre-defined content. This is the claim made by Budge and Meyer (2013, 89), for instance, who state that the "justification for the Manifesto left-right scale and the basis of its construction [are] that highly influential early modern theorists put them together in their political analyses." Later modern theorists such as Inglehart (1984) or Bobbio (1996) have updated these constructs but take essentially the same deductive approach to identifying the issues that constitute the essential distinctions between left and right ideology in contemporary settings. The difficulty for this approach lies in selecting the components of the left-right dimension and specifying their relative weights in a manner that is generically valid, across different party systems and times.

A second approach reverses the logic of inference about the left-right dimension, identifying the content of the left-right dimension as the simply the sum of whatever parts it comprises, often identified by scaling observable party behaviour.

> This *a posteriori* and quintessentially inductive approach sets its essential empirical task as finding the best-fitting empirical representation of the policy space under investigation, using techniques of dimensional analysis to infer latent policy dimensions and then interpreting the substantive meaning of these dimensions in terms of relative locations of key political agents on these. The approach thus assumes that we know more about the positions of key political actors, relative to each other, than we know about the substantive meaning of key policy dimensions. (Benoit and Laver, 2006, 59)

In the inductive approach, the substantive content of the left-right dimension something to be discovered empirically, through determining which specific dimensions of difference reduce to a single over-arching continuum of difference, in a manner that may well differ depending on the national setting and the time period. According to this perspective, there is no basis for establishing *a priori* the substantive meaning of left-right ideological differences; rather, "the left-right dimension is defined inductively and empirically as the 'super-issue' that most constrains parties' positions across a broad range of policies" (Gabel and Huber, 2000). This approach faces the challenge of interpreting the inferred left-right dimension in terms of substantive content.

The *inductive* approach has been applied in a number of different specifications. The first spatial analyses with the CMP data belong to this group; they were based on two-stage factor analyses to remedy the small-N problem (Budge, Robertson and Hearl, 1987). Klingemann (1995) employed factor analysis to a subset of the CMP categories, some of which were pre-grouped. The most-far reaching formulation of the inductive approach is due to Gabel and Huber (2000), who introduced a "vanilla" approach that conducts a principal component analysis of all category shares. Franzmann and Kaiser (2006) and Franzmann (2013) also proceed primarily in an inductive manner. They use regressions of category shares on party indicator variables to determine which categories differentiate between parties and thus constitute position rather than valence issues. More recently, König, Marbach and Osnabrügge (2013) suggested an approach that is best characterized as mixed. At its heart lies a dynamic factor analysis, but the input data consists of selected and pre-scaled CMP categories, and the inference also includes prior information on party positions from expert surveys. A similar use of expert surveys is made by Bakker (2009). His approach is a form of item-response-theory model, which is applied to a subset of CMP categories that are are pre-grouped into left and right items.

The most commonly used approach towards measuring left-right from the CMP data is the “Rile” index, which comes canned as part of the dataset. As argued by Budge and Meyer (2013, 88), ‘the Manifesto scale is different in having been created *before* its application to the data by characterising the original coding categories as left, right, or neutral on the basis of theoretical writings”. This argument coins “Rile” as the prime example for the *a priori* approach. However, the original description of the procedure, outlined very clearly in Laver and Budge (1992, 26-27), reads differently.[1] “Rile” was constructed in the 1980s applying inductive methods (principal components factor analysis) to CMP data for ten Western European countries from 1945 until the mid-1980s, and using these results to select the components for the final index. It is therefore more appropriate to speak of a mixed approach, and one with strong inductive elements. In a recent contribution, Jahn (2011) has made a renewed point for a primarily deductive approach, although he recognises the necessity to allow for context-specific elements. His measure therefore integrates inductive techniques to weigh the *a priori* selected categories and to include additional components for specific contexts and time periods.

The debate over which approach is superior remains unresolved, as illustrated by the fact that one and the same dataset continues to be used to support both kinds of measures. The main problem of the “deductive” or *a priori* approach is succinctly put by Gabel and Huber (2000, 95): “To our knowledge, however, no rigorous theory based on this first conceptualisation of ideology is sufficiently precise to specify how to use MRG [Manifesto Research Group] data to measure left-right party positions.” This difficulty is also reflected in the fact that even the two approaches from above that rest (in the case of “Rile” allegedly) on ‘deductive” reasoning do include also inductive elements.[2] Proponents of the *a priori* approach argue that the grounding of its definition in a known, and fixed, frame of reference, facilitates comparison of like with like across different contexts. A scale with fixed components, so the argument, broadens

[1] The discrepancy has been noted already by other authors (see Jahn, 2014, 2).

[2] An example where a more specific policy scale, namely an economic left-right index, is constructed without resorting to inductive techniques is Tavits and Letki (2009).

its applicability to not just more, but also *every possible*, context. The Rile index's "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.... They apply universally without having to be adjusted for particular contexts, and thus provide a promise of invariant and reliable measurement across limitations of time and space (Budge and Meyer, 2013, 90)."

At first, this argument may appear intuitive, but a closer look casts strong doubts upon it. Compare the fixed left-right scale to an analogous measure in economics: the Consumer Price Index (CPI). Designed to capture the typical cost of a basket of goods and services consumed by households, the CPI consists of an index constructed from the prices paid by a designated consumer segment for a "market basket" of the goods and services purchased by a household. The CPI has to be time-invariant because its use is explicitly comparative: to track *changes* in inflation across different years. To achieve this comparability, the CPI must be adjusted in three ways. First, the sample of representative goods which the basket comprises must be updated to match changes in consumer consumption, technology, etc. It would be an invalid measure in 2013 to use a basket containing spurs or wax candles, for instance, because consumers no longer ride horses to work nor do scholars write their papers under the shine of sooty candles. Similarly, tablet computers, which were added to the (UK) market basket in 2012 for the first time, would have been unimaginable components 50 or even 20 years ago (Gooding, 2011, 7). Second, the weight of different items must continually be adjusted, to make it relevant to the current period. Finally, the index is only meaningful relative to a certain base, because the very nature of what is being measured (prices) is constantly changing. Using a fixed basket of goods would be problematic, and so is using a fixed set of policy categories to measure left-right positions of political parties.

While the purely inductive method can easily take into account context, it has been criticised on two main grounds, first for being atheoretical, and second for being sensitive to sample composition in terms of countries or time periods included in the analysis (e.g. Jahn, 2011, 748). The first point is correct in a sense that the inductive approach does not specify an *a priori* theory for the *content* of left-right. However, this should not be equated with a complete lack of theoretical foundation. Left-right is a spatial metaphor that reflects the fundamental line of division (in a given context), and the underlying rationale of inductive approaches is to infer this conflict structure. As Fuchs and Klingemann (1990) have argued, the left-right schema should not be interpreted as an ideology, but as an expression of "basic structures of conflict" in the sense of Lipset and Rokkan (1967). The substantive meaning of left-right therefore depends on the nature of political conflict at the time when left-right symbolism became institutionalised, but is generally open to re-specification (Fuchs and Klingemann, 1990, 232-233). Left-right is then a super-issue (Inglehart, 1984; Gabel and Huber, 2000; Franzmann, 2013), also because parties deliberately seek to bundle specific issues and communicate them in a simplified form, by linking them to the overarching policy divide. Thus, there is a theoretical explanation for how the left-right dimension is brought about in the inductive approach, and the specific method for estimating policy positions should be based on such a theory. What the inductive approach does not and need not do is specifying a theory for the *content* of left-right. The latter is context-specific and therefore cannot be derived from a universally applicable ideological framework.[3] As for the second criticism, dependence of results on the specific sample analyzed, the question becomes if this is in fact a weakness or rather a strength of the approach. Suppose one uses the same inductive method on different, but partially overlapping samples and obtains different results. In this case the researcher is alerted to the fact that the assumption of a uniform left-right dimension across all cases does not hold up. The issue can then be further examined, and scholars can take appropriate measures, such as splitting samples appropriately. When using a supposedly invariant *a priori* approach, such heterogeneities among the observations do not even stand a chance of being detected. We empirically demonstrate this later, showing how a fixed components approach can measure the wrong content in many contexts.

Inductive approaches seem to gone out of fashion, however. Of course, one reason for the popularity of the "Rile" index is that it is much easier to use than conducting one's own inductive position estimations first. In addition, existing inductive approaches are not without

[3]This does not imply that it is impossible to formulate theories that explain the *context-specific* content of left-right, although this is difficult and therefore usually not done in the literature related to the estimation of policy positions.

problems. First, the standard statistical technique used within the inductive approach is factor analysis. One reason why end users are skeptical about policy positions estimated on this basis may have to do with the fact that factor analysis is a fairly complex technique, whose rationale is not that easy to grasp. Factor analysis tends to appear as a black box technique, especially if factor loadings, the correlations of the unobserved factors with the observed variables, are not reported or discussed (as in Gabel and Huber, 2000). Second, it is questionable if factor analysis is the appropriate technique for inferring left-right positions from manifesto data. Factor analysis assumes linear relationships between variables, an assumption that is problematic for counts or relative shares of issue statements (Van der Brug, 2001; Gemenis, 2013, 120-121). Van der Brug (2001, 120-121) gives the example of the "Military: Positive" and "Military: Negative" categories. They are meant to reflect opposing poles of one dimension, but are only weakly negatively correlated since centrist parties tend not to mix both kinds of references, but rather ignore issues related to the Military at all. Third, factor analysis is not built on an explicit model of the process that creates the basic data of interest: the number of statements referring to a certain policy issue. Factor analytic methods are models of correlations, which discard very interesting information about the means and the variances of the input variables and (Jackman, 2001, 230). Factor analysis is essentially a data reduction technique (Reckase, 1997). And while there may be an analogy between that and the interpretation of left-right as a super-dimension expressed in issue bundles, factor analysis does not represent a model of how parties reduce the issue space.

The problems associated with a purely "deductive" approach towards measuring left-right are so fundamental that a satisfactory remedy will be hard to find. Concerning inductive approaches, indeed we "need explicit criteria of how categories can be transferred to a left-right scale" (Franzmann and Kaiser, 2006, 166). Even better is an explicit model. We introduce one in form of an item-response-theory model, which offers a number of advantages over existing inductive measurement approaches. Our IRT model constitutes a representation of the actual data generating process, i.e. manifesto writing, with an intuitive interpretation of all the parameters, and model-based uncertainty measures. Using our IRT approach, moreover, we are able to estimate not only the latent party positions, but also at the same time estimate the degree to which each policy categories contributes to the content of the left-right dimension. Last but not least, the IRT model also provides a bridge between the deductive and inductive approaches. In its Bayesian version, the model can easily incorporate different kinds of *a priori* information, in an explicit and formal way. Thus, our model is able to combine many of the best features of the inductive and deductive approaches.

2 How to scale policy dimensions using IRT

2.1 Data: Text category counts (from CMP)

Researchers interested in analyzing the development of party positions over time can hardly avoid using election manifestos as data source. An election manifesto is a text that "can be singled out as a uniquely representative and authoritative characterisation of party policy at a given point in time" (Budge, Robertson and Hearl, 1987, 18). Because manifestos are drawn up for purposes of shaping the frames of its election campaign and setting out its policy positions, parties typically place great care in drafting these texts. While there are other, non-manifesto-based approaches for estimating the policy positions of political actors (see Benoit and Laver, 2006, 56-77), the regular publication and the "official" status of manifestos makes them the first choice to measure time-varying party positions. For these reasons, manifestos have formed the main source of textual data for both manually coded content analysis research such as the long-standing Manifesto Project (e.g. Budge, Robertson and Hearl, 1987; Budge et al., 2001; Klingemann et al., 2006; Volkens et al., 2013) and derivative research, as well as numerous attempts to extract policy positions automatically using supervised (Laver, Benoit and Garry, 2003) or unsupervised (Slapin and Proksch, 2008; Monroe and Maeda, 2004) learning methods.

Of course, manifestos are only useful for quantitative analysis once their features have been abstracted and quantified. By drawing on the rich dataset of coded policy statements from party manifestos, we are drawing on the same dataset used to estimate left-right ideology—whether taking inductive or deductive approaches—by numerous other researchers (e.g. Gabel and Huber, 2000; Franzmann and Kaiser, 2006; Laver and Budge, 1992; Jahn, 2011; Mölder, 2013). The CMP's coding method relies on qualitative content analysis using trained expert coders to classify the sentences of each text into a predefined set of policy categories spanning seven domains.[4] The full information set available from the project therefore consists of a matrix of documents by coding categories, where each cell represents the percentage of sentences in a given document assigned to a given content category. In addition, the dataset provides information on the overall length of a document measured in number of unitized quasi-sentences.

As a first fundamental decision, we convert the data into category counts, by multiplying the shares with the total length. Counts are the natural unit of measurement for political statements in manifestos. Manifestos are written argument by argument, and so they are coded. Therefore, the data generating process should also be modelled as a count process. This procedure is also required since we desire good estimates of the uncertainty associated with the quantities of interest to be inferred. In line with the general insight that more data provides more confident estimates than less data, the amount of information available from an election manifesto should be taken into account. Some have modelled this process explicitly (for instance Benoit, Laver and Mikhaylov, 2009), to estimate the variance of manifesto text as a function of its length. Other models, such as Laver, Benoit and Garry (2003)'s "wordscores" or the Poisson scaling model of Slapin and Proksch (2008) incorporate this as a feature of their estimation method. Whether words, sentences, or "quasi-sentences", models of textual data built on party manifestos share the feature of modelling observed counts of text units and using these counts to estimate features of the party's policy stances. In reality, however, very little is known about the data generating process giving rise to counts of textual units in party platforms.[5] What we do know, is that there is considerable variation.

When we inspect the empirical distribution of CMP category counts, it becomes clear that most of these counts tend to be zero for a given manifesto, because some categories do not apply to a given context (such as EU policy statements in New Zealand) or because some coders simply never use them. Figure 1 shows which categories tended to get the most and the least usage, on a log10 scale. The most frequently used category was "per504: Welfare State Expansion: Positive" to which nearly 10% ($10^{.91}$) of all coded sentences were assigned.

[4] This consists of 56 core policy categories, plus an additional 51 extended categories added to cover policy in countries added since the 1980s. For details see https://manifesto-project.wzb.eu/coding_schemes/1.

[5] Based on interviews with Irish manifesto authors, Däubler (2012) reports that parties' opinions about document length and their motivations for providing longer or shorter documents differ significantly.

At the other end of the scale was “per507: Education Limitation: Positive”, assigned in just $10^{-1.18} = .05\%$ or just half of one percent of all coded sentences. Not only are some categories rarely used, but also most coded manifestos draw on only about half of the available policy categories. The median manifesto used just 27 of the 56 possible “core” coding categories, with the bottom 25% of all manifestos using 19 or fewer categories. The conclusion is that election manifestos and the textual units covering different issues contained in them vary widely in length, across parties, countries, and time. This variation should be used and modeled when inferring party positions.

2.2 The Model

Our fundamental aim is to infer the “left-right” position of a party. All we observe, however, is a set of category codings for the manifesto text. Following Benoit, Laver and Mikhaylov (2009), we start from the notion that the party intends to communicate a certain position, called θ_i in the manifesto i.[6] This position is fundamentally unobservable and uncertain, but will be communicated through the text. As writing proceeds, the party makes various policy statements referring to different issues. The count of statements referring to a certain category is therefore the observed data. Parties vary in how much they write. Partly, this will be due to idiosyncratic features such as the writing style of the authors, but it may also be related to the context of the specific election campaign. Next, some issues are covered briefly, others at greater length. One reason is that some topics (say: economic matters) feature more prominently on the agenda of national parties (and governments) than others (say: culture). And finally, some issues are well suitable for differentiating the party from its competitors and are positional in nature, while others are less so and have the character of valence issues (Stokes, 1963; Franzmann and Kaiser, 2006; Dolezal et al., 2013).

Our model takes into account these considerations by modelling the latent variable as well as the left-right policy components directly using a model based in item-response theory (IRT). Here, observed sentences coded for the CMP content categories are observed data, the categories are the “test items”, and the parties correspond to the test “persons”. Each item’s contri-

[6]For simplicity, we speak of the “party” as a single collective author here. In practice, there are typically multiple authors and these may have of course also have different policy positions.

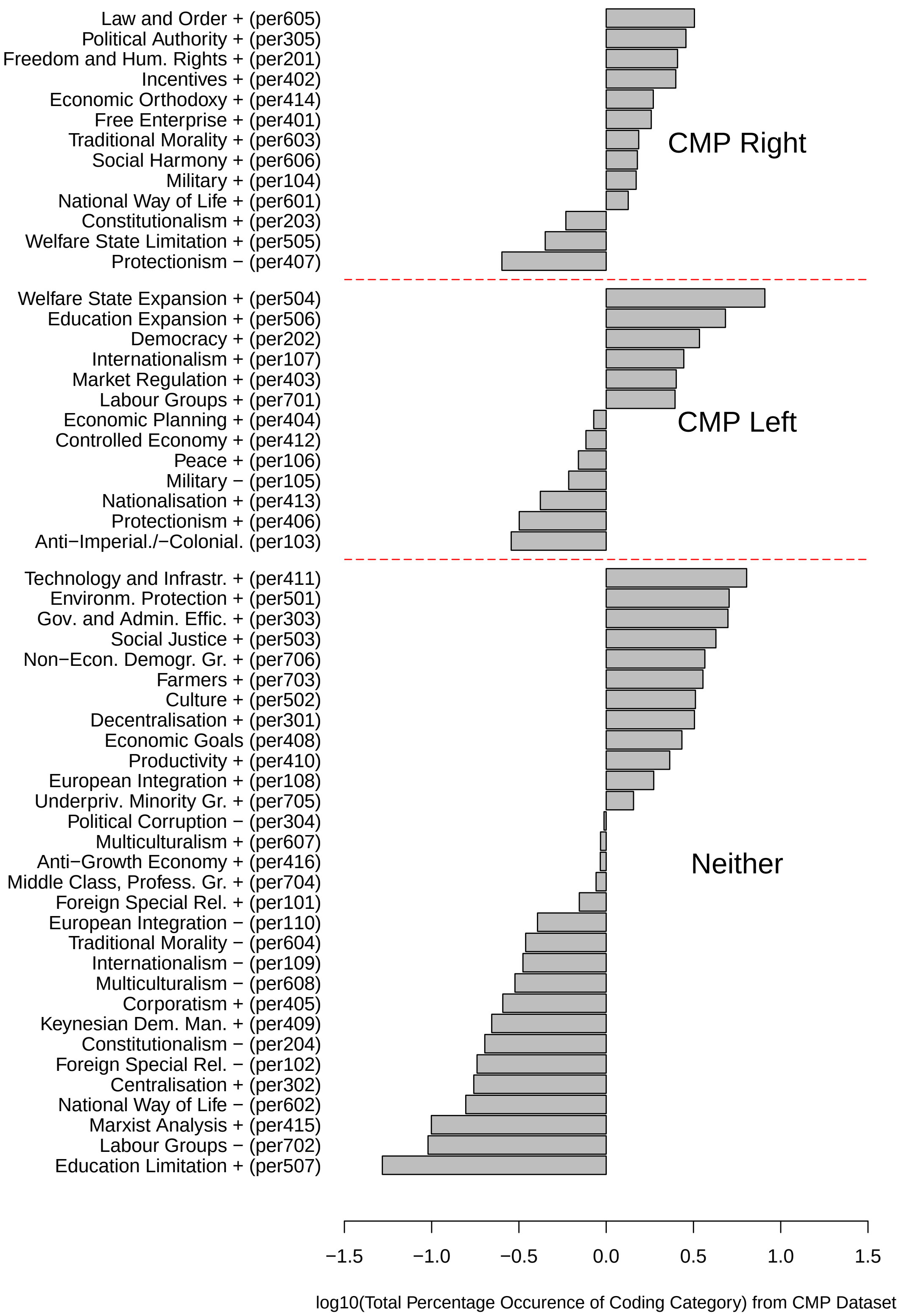


Figure 1: Frequencies of total percentage use of CMP coding categories, log10 scale.

bution to the observed outcomes is mapped via a series of item parameters, given the position of the latent variable θ_i, which must be estimated.

The model is as follows. Letting y_{ij} represent the count in manifesto i of coded sentences in category j, we can model the rate of occurrence of category count y_{ij} in a hierarchical IRT model as:

$$Pr(y|\mu_{ij}) = \frac{e^{-\mu_{ij}}\mu_{ij}^{y_{ij}}}{y_{ij}!} \tag{1}$$

$$\mathrm{E}(y_{ij}) = \mu_{ij}$$

$$\log(\mu_{ij}) = \log(\lambda_i) + \alpha_j + \beta_j\theta_i \tag{2}$$

Observed counts occur according to a Poisson process, with a probability defined as in Eq. 1. This can be decomposed into a log-linear function as per Eq. 2, as a combination of document-level (i) and item-level (j) parameters.

The parameter λ_i functions as an offset parameter denoting the rate of exposure of y_{ij} caused by the fact that manifesto documents vary widely in length, such that $\log(\mathrm{E}(y_{ij})/\lambda_i)) \sim$ Poisson(μ_{ij}) using the usual exponential link function. Here we model the exposure as a random variable because (as we show below) manifesto length is itself a random variable, rather than considering manifesto length to be fixed. When conditioned on λ_i, this Poisson model has an equivalent multinomial formulation. This can be demonstrated as:

$$\eta_{ij} = \log\frac{\pi_{ij}}{\pi_{iJ}} = \alpha_j + \beta_j\theta_i \tag{3}$$

where α_i and β_j function as regression coefficients for contrasts of (with α_j as a constant) each $J-1$ categories relative to an omitted base case. This can be reformulated as

$$\log(\mu_{ij}) = \eta + \lambda_i^* + \alpha_j^* + \beta_j^*\theta_i \tag{4}$$

in an equivalent log-linear model, where the $Y_{ij} \sim$ Poisson(μ_{ij}) as independent Poisson random variables, as is the the total observed sentences in a manifesto $\sum_{j=1}^{J} n_i$. The η provides conditioning on the grand total of counts, while the separate parameter λ_i provides an adjustment for the observed marginals for each document i, assuring exact reproduction of the multinomial denominators n_i (see Agresti 2002, 9 for a derivation, and Lowe and Benoit 2011, 4 for application to this model). In our formulation, we estimate (from Eq. 2) $\log(\lambda_i) = \eta + \lambda_i^*$ from 2, where the $\log(\lambda_i)$ represents the observed length of a document in terms of its total sentence count and functions both as an exposure rate and a means of conditioning on the document marginals across content categories. Unlike in the multinomial model, however, λ_i is a random variable rather than fixed. As we argue below, total manifesto length is indeed a random variable whose systematic features can be explained by political factors.

For the analysis of the content of left-right policy, the item parameters form our focus. The α_j and β_j parameters are a set of "item" parameters for each j category that map the "response" of each manifesto in terms of its observable characteristics conditional on this latent position θ_i. α_j corresponds to the "difficulty" of an item, indicating how extensively parties cover this issue, independent of the position. The β_j constitute discrimination parameters, which indicate how a particular policy category's use varies in response to changes in the latent dimension θ_i. Put differently, the absolute size of β_j reflects the degree to which a category is positional, and its sign shows whether the category is a "left" or a "right" item. One advantage of our approach over the factor analytic method is that the relationship between parties and items is explicitly modelled (Reckase, 1997, 29). Note also an interesting feature of this model: a certain position can be expressed in many different ways. For example, in order to communicate a markedly leftist position, a party may use a one very leftist category a few times, it may use one moderately left category many times, or it may refer to various moderately left categories a few times each.

Using a Bayesian approach to inference, furthermore, we are able to supply informative priors to the item parameters. This allows the incorporation of prior information about the left- or right-leaning values of specific policy categories, or the modelling of category or manifesto parameters hierarchically using additional information. For estimating multiple k dimensions of policy θ_{ik}, furthermore, this allows us to incorporate through informative priors information about which dimensions should be estimated from which policy categories. Finally, using MCMC simulation methods and the Gibbs sampler to estimate the parameters, we are able

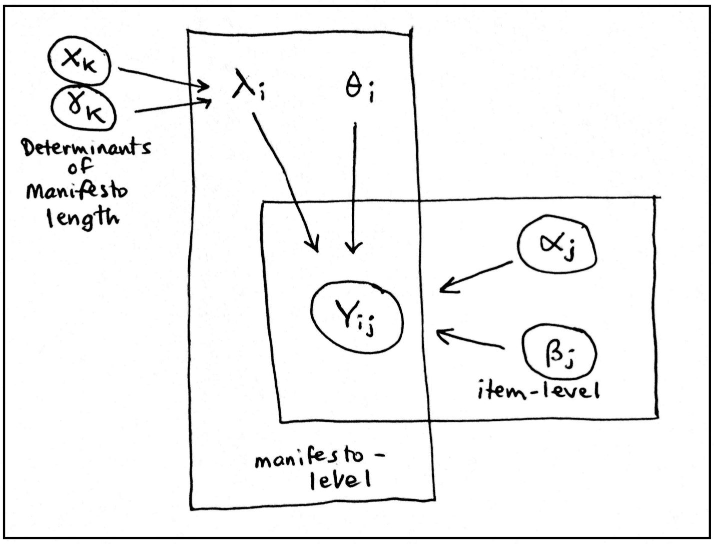


Figure 2: Directed acyclic graph of the one-dimensional Poisson IRT for document and item parameters to category counts Y_{ij}

directly to estimate uncertainty for the parameters of interest from the sampled posteriors.

Eq. 2 is similar in structure to the standard "two-parameter" IRT model widely used in psychometrics and adapted in political science for the analysis of ideal points from binary outcomes (e.g. Clinton, Jackman and Rivers, 2004), but with important differences. First, the outcome y_{ij} is a more informative outcome based on counts rather than the binary outcome common to the logistic or "ogive" (probit) used in nearly all common IRT models. Counts—or more precisely, the logarithm of counts—have been shown to be far more informative than binary models when the features of interest are word or word-like quantities (Lowe et al., 2011). Second, because our outcome y_{ij} is not constrained or conditioned on a rate of exposure known to vary widely (manifesto length), our model has a third parameter λ_i which is not an item-level parameter such as the "guessing" parameter as in the "three-parameter" IRT model (where it represents a subject's probability of correcting answering the item as $\theta_i \rightarrow -\infty$). Our model is very similar in structure to the "wordfish" scaling model for word counts first presented by Slapin and Proksch (2008), although theirs was not explicitly presented as an IRT model, instead using conditional maximum likelihood to estimate θ_i by conditioning on fixed effects for the item and exposure parameters.

An alternative approach to estimate is to reformulate a model fit to the complete set of category counts as a multinomial, by conditioning on the count total. Our approach relies on incorporating information about the document length through a hierarchical model of λ_i^*, something easily possible in the Bayesian framework. Hence we model $\lambda_i^* \sim N(\mu_{\lambda i^*}, \sigma_{\lambda^*})$ where $\mu_{\lambda i^*} = \mathbf{z}_i$ where $\mathbf{z}_i$ is a set of covariates modelling the exposure rate. Below, we demonstrate that this can be used both to incorporate a model predicting the manifesto length, as well as improving overall model identification simply by including the log of the document length as a covariate while still allowing it an error variance σ_{λ^*}.

As expressed, the model parameters in Eq. 2 is unidentified, because any linear transformation of θ_i can be offset by corresponding transformations in the item parameters. To fix the scale and location of θ_i, we impose a mean zero, unit variance restriction on θ_i, a condition that is sufficient for local identification (see Jackman, 2009, 441). For global identification and to prevent the property that any scaling is invariant to reflection, we also fix one of the $\theta_i < \theta_{i'}$ based on prior knowledge about the relative positions of the i and i' manifestos.

To fit the model, we have chosen the following prior distributions for the hyperparameters. As priors we set:

$$\begin{aligned}
\lambda_i &\sim \text{unif}(-20, 20) \\
\alpha_j &\sim N(0, \sigma_\alpha) \\
\beta_j &\sim N(0, \sigma_\beta) \\
\theta_i &\sim N(0, 1) \\
1/\sigma_\alpha^2 &\sim \text{Gamma}(.01, .01) \\
1/\sigma_\beta^2 &\sim \text{Gamma}(.01, .01)
\end{aligned}$$

We estimate all models using MCMC and producing parameter estimates by sampling from the posteriors following a suitable burn-in.[7]

[7]Code to produce these model estimates can be found in the `MCMCirtPoisson1d` code at http://github.com/kbenoit/quanteda

3 Estimating Left-Right as a Latent Variable

In this section we fit the basic model to the core 56 CMP categories as "items" to estimate the single-dimensional latent variable θ_i, and compare this measure to other solutions.

3.1 Party locations on a single dimension

Fitting the core model (Eq. 2) to the core 56 CMP policy counts, we are able to obtain estimates of the policy positions θ_i for each party i on a single dimension of policy. Table 1 presents these results for two models, the first fit to all documents with a free λ_i parameter (following Eq. 2), and a second on a subsample for which we had covariates to model manifesto length in a hierarchical model for λ_i. (This second model is detailed in the next section; here our focus in on the estimates of policy positions.) For the set of all manifestos, we obtained estimates and confidence intervals ("Bayesian credible regions") for each manifesto. The top part of Table 1 presents a set of estimates for selected parties from Germany, the UK, and the United States, ordered from left to right. In each context, the location of the parties has high validity, ordered in a manner which would accord with any informed observer's understanding of party politics in each context. With the move to the centre of Blair's New Labour in 1997, furthermore, the measure tracks Labour's move relative to the more traditional leftist position of Labour in 1987.

The bottom of Table 1 shows correlations with external measures, including expert surveys, Rile, and the "vanilla" method of Gabel and Huber (2000). For the expert surveys, we have also divided the sample into three broad subsets, Western Europe, Eastern Europe, and the Pacific and North America. (No expert surveys were available for other regions, so these are not reported.) The correlation with all expert surveys was 0.70 for the whole sample, and 0.76 for the covariate subset. The correspondence was highest in the Pacific and North America, at 0.86 and 0.92 respectively, and next highest for Western Europe. Correspondence with left-right in Eastern Europe was lowest, indicating that the same patterns that fit overall did not fit particularly well in Eastern Europe. This finding is consistent with earlier research finding that the content of left-right policy is different in post-communist settings (Benoit and Laver, 2006; Mölder, 2013). Correlations with Rile were about 0.80, and 0.64 for the "vanilla" method based

Model	All Documents		Lambda_i Covariates	
	Mean	*95% BCR*	*Mean*	*95% BCR*
Illustrative Estimates of Theta_i				
Germany 2010				
SPD	-0.84	[-0.95, -0.75]	-0.78	[-0.89, -0.66]
FDP	0.03	[-0.06, 0.1]	0.60	[0.51, 0.7]
CDU/CSU	0.08	[0.01, 0.17]	0.53	[0.42, 0.63]
Left	-1.56	[-1.65, -1.46]	-1.98	[-2.12, -1.84]
Greens	-1.20	[-1.28, -1.12]	-1.23	[-1.33, -1.13]
UK 1987				
Labour	-0.55	[-0.72, -0.39]	-0.44	[-0.65, -0.24]
Liberal Democrats	-0.36	[-0.49, -0.24]	-0.11	[-0.26, 0.04]
Conservatives	0.78	[0.69, 0.87]	1.22	[1.11, 1.34]
UK 1997				
Labour	0.16	[0.04, 0.3]	0.20	[0.04, 0.37]
Liberal Democrats	-0.37	[-0.53, -0.23]	0.02	[-0.15, 0.18]
Conservatives	0.66	[0.56, 0.77]	0.78	[0.65, 0.9]
United States 2012				
Democrats	-0.17	[-0.28, -0.03]	-0.05	[-0.19, 0.09]
Republicans	1.15	[1.07, 1.23]	1.47	[1.37, 1.57]
SD(lambda)	1.98	[1.84, 2.13]	1.13	[1.09, 1.16]
SD(beta)	1.79	[1.47, 2.29]	0.76	[0.63, 0.93]
SD(alpha)	2.22	[1.88, 2.74]	1.51	[1.26, 1.85]
Chains		3		3
Burn-in		1,000		1,000
Samples		500		500
N manifestos		3,287		2,400
Gelman-Rubin R^2		-		-
Geweke Diagnostic (% theta_i failing test)		2.0%		0.5%
Correlations of Theta_i				
With All Expert Surveys	0.70	(n=523)	0.76	(n=505)
W Europe subset	0.76	(n=352)	0.81	(n=351)
E Europe subset	0.54	(n=152)	0.63	(n=136)
Pacific & N America subset	0.86	(n=20)	0.92	(n=18)
With CMP Rile	0.80		0.77	
With "Vanilla" Method	0.64		0.95	

Table 1: Basic model based on core 56 CMP category counts estimating $\hat{\theta}_i$, with comparisons to external measures.

on factor analysis.

Figure 3 plots the correlations against expert surveys, indicating a good linear fit, also illustrating the difference in correlations for the subsets by region. The slope of the patterns

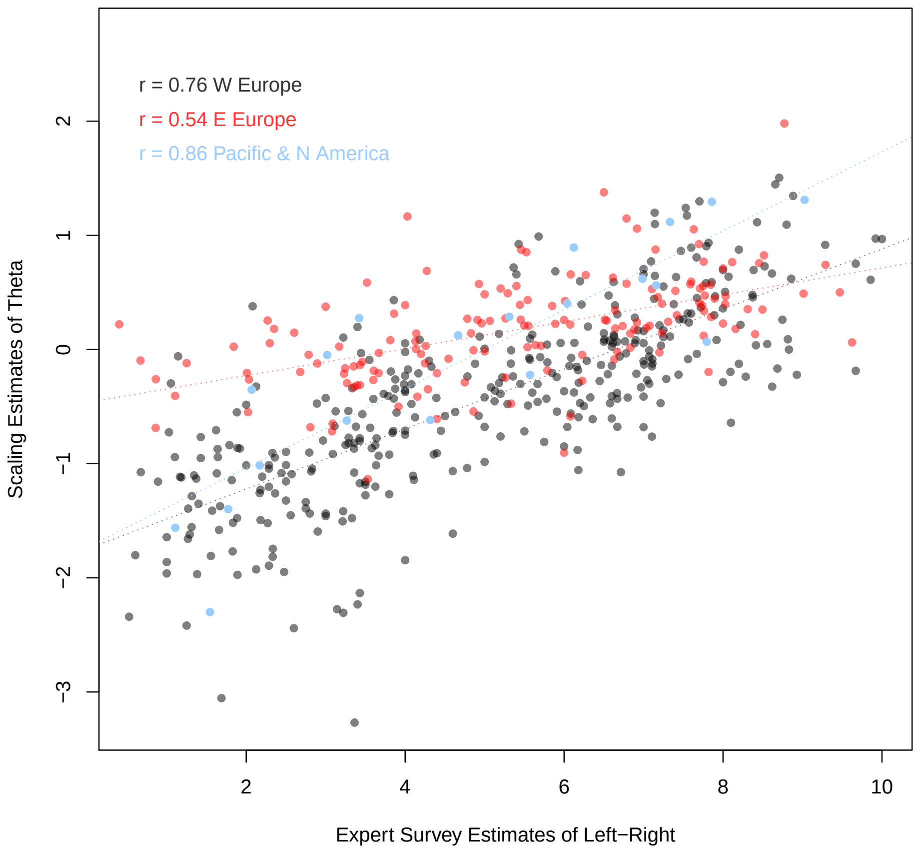


Figure 3: Comparison of $\hat{\theta}_i$ to expert survey estimates for the regional partition model, core 56 categories.

differs with subset, but the patterns indicate a clear linear relationship.

The method of IRT scaling that uses *all* of the policy categories provides a good fit to the data overall, indicating its validity as a measure of party locations on a single axis of policy differences corresponding to what experts judged to be the left-right dimension. The fit varies with context however, indicating that even with inductive approaches, one size does not fit all.

3.2 Modelling manifesto length: A hierarchical model for λ_i

Specifically with regard to CMP data, Benoit, Laver and Mikhaylov (2009) describe the link between unobserved policy position and observed text as governed by a stochastic process where the variance is directly related to the observed text units. Correspondingly, if the observed text is longer, the researcher's inferences about the underlying message become less uncertain.

Their model of stochastic text generation, however, treats the determination of textual length as fixed and unexplained; substantively, a political black box: "The total number of text units found in a manifesto appears to be, absent systematic information or prior expectation on this matter, unrelated to any political variable of interest" (Benoit, Laver and Mikhaylov, 2009, 501). In this section, by contrast, we model the number of textual units in the overall manifesto explicitly. This allows us to test whether manifesto length is essentially idiosyncratic or whether it arises from systematic political factors.

The specific context of manifesto preparation, the resources available and devoted to the task, and the intended uses are shaped by a complex set of factors that are hard to model (Däubler, 2012, 62). However, we can consider how some key variables should impact on the overall length of manifestos. First, larger parties (measured in *% Vote*) should present longer manifestos, for a number of reasons. They usually have more resources at their disposal (more donations, state funding). Second, larger parties represent a larger and thus more heterogenous group of voters. Therefore they need to offer a much broader range of policies in order to appeal to their electorate. Small parties are more likely to be built around a single issue or to cater to the special interests of a small segment of society. Some small parties may not even base a great extent of their appeal on proposing detailed policies, but rather on leader's charisma or on an "anti-establishment" notion. A third reason why larger parties can be expected to have longer manifestos stems from their higher propensity to get involved into government negotiations.

How *outgoing government status* affects overall manifesto length is ambiguous. On the one hand, government parties should have a larger stock of existing policies to work from, since actual policy-making is their everyday business. They also enjoy special access to resources and staff, also from the ministerial bureaucracy, that can assist with general policy development and more specifically with manifesto preparation. On the other hand, opposition parties may be less constrained and face a stronger need to present their case and explain their future policy plans (Garry and Mansergh, 1999). In addition to these party characteristics, there are other factors that are expected to affect manifesto length. Preparing manifestos requires time, which is scarce in case of early elections. This means that less work can be spent on developing the policy material to publication-ready standard. Also, in early elections there may be less need

Variable	IRT	OLS
% Vote	0.0195	0.0178
	(0.002)	(0.002)
Outgoing government status (0/1)	0.0821	0.0281
	(0.056)	(0.050)
Proportion of legisl. period passed	1.0049	0.9854
	(0.107)	(0.103)
Young Democracy (0/1)	-0.1970	-0.2353
	(0.069)	(0.064)
Year	0.0310	0.0299
	(0.001)	(0.001)
Intercept	-61.4047	-55.0023
	(2.799)	(2.626)
Residual Std Error	1.27	1.11

Table 2: Log of document length regressed on covariates. The IRT model is a hierarchical model of $\log(\lambda_i)$.

to introduce new policy ideas since some from the previous election remain up-to-date. As measure, we use the *proportion of legislative period passed*.

Turning to the country level, we expect manifestos to be shorter in *young democracies* (in the first ten years after transition), since it takes time for programmatic party competition to develop. Finally, as anyone can immediately see from a simple inspection of the CMP data, manifestos have strongly increased in size over the *years*. Some authors (e.g. Topf, 1994; Kavanagh, 1996) suggested that this development reflects the trend that governments have taken over a larger number and variety of responsibilities. Another mechanism could be that technological development has lowered the marginal costs of the material production and dissemination of longer documents.[8]

Table 2 shows the results if we add these covariates to predict the λ_i parameter in a linear model with a Gaussian error term (left column). There is clear evidence that overall document length varies systematically with the included explanatory factors. Larger parties publish longer manifestos. For instance, an increase in vote share of ten percentage points (i.e. by 10 on the scale of the variable) results in an increase of 22% in terms of overall manifesto length

[8]In addition to the CMP's information on vote share (that we completed in a number of cases), data sources used for measuring these variables comprise Döring and Manow (2012); Andersson, Bergman and Ersson (2012); Teorell et al. (2013). Details about the operationalizations are available from the authors on request.

(since $exp(.195) \approx 1.22$). Above, it was argued that it is theoretically ambiguous if outgoing government parties publish longer or shorter manifestos than opposition parties. Indeed, while the respective coefficient suggests that government status is associated with an 8.6% increase in length), we cannot say with high certainty that the coefficient is positive. Very clearly, manifestos issued for early elections are shorter than their counterparts from "regular" elections. We would expect that a manifesto presented at an election at the end of the legislative period is longer by 65% ($\approx exp(.5) \approx 1.65$) compared to an election taking place right in the middle of the term. Also, in younger democracies manifestos are considerably shorter The coefficient suggests that during the first ten years after transition, manifestos have a relative length of only 82% of their counterparts in later years. Considerable temporal variation remains despite including the other variables. An additional year later in terms of the election date is associated with a 3% increase in length. For comparison purposes, an OLS model for the log of document length is shown in the left column. The results are very similar, but it is interesting to note that in the IRT model the effect of government status seems to be more pronounced. This is because the IRT model can control for the effect of position on length, and position is also associated with government status. Taken together, the findings for length further support the argument that the IRT approach provides a useful model of the data generating process, allowing for interesting extensions to analyze additional aspects of the data.

3.3 A better estimate of uncertainty: Modelling policy shifts

Because the document-level parameters are stochastic and estimated by sampling from the posterior, it is possible to estimate uncertainty over the left-right positional parameters $\hat{\theta}_i$ through simulating draws from the posterior distribution using the Gibbs sampler used to obtain parameter estimates. For each $\hat{\theta}_i$ representing a manifesto's left-right policy position, in other words, we can estimate the variances directly from the posterior draws once the sampling distribution of the posterior simulations has reached convergence. This is contrasts with the non-parametric simulation approach applied directly to the textual data by Benoit, Laver and Mikhaylov (2009), who assumed that the category frequencies were drawn according to a multinomial distribution and bootstrapped the category counts on this basis to compute a standard error for each Y_{ij} in

	BLM (2009)		Model $\Delta\hat{\theta}_i$	
Statistically Significant Change?	**Elections**	**% of Total**	**Elections**	**% of Total**
No	1,308	62.3%	1,246	51.9%
Yes	791	37.7%	1,154	48.1%
Non-adjacent	778	–	887	–
Total	2,877	100.0%	3,287	100.0%

Table 3: Comparative over-time mapping of policy movement on Left-Right measure, taking into account the statistical significance of shifts – comparing change in IRT estimates to Benoit, Laver and Mikhaylov (2009) estimates from non-parametric bootstrapping.

addition to compound categories such as the additive Rile index.

Using the results of our model estimated on the full sample, we can contrast our results to those of Benoit, Laver and Mikhaylov (2009, Table 1), who reported that using their non-parametric bootstrapping procedure, only 38% of parties' observed left-right "movements" could be declared real rather than the result of stochastic features generating the text from underlying policy positions. Using our much more complete, generative model of policy positions, we find in Table 3 that this rate of change is actually considerably higher, at 48.1%, suggesting that the policy shifts that can be considered real indications of changes in party ideology rather the product of measurement noise is much nearer to half than to previous estimates of closer to a third. This measure is a far more informed estimate of the real policy movement, based on a more complete model that includes the noisiness of the stochastic text described in Benoit, Laver and Mikhaylov (2009) but also incorporating the full information as to how the use of policy statements reflects the dimension of left-right politics based on all of the patterns found in the dataset.

4 Why *a priori* low-dimensional measurement is flawed

For believers in the "deductive" approach to measuring political spaces, the authoritative, *ex ante* definition of the content of a political dimension has distinct advantages for measurement. By specifying that a dimension of environmental policy should consist of two contrasting extremes of "supporting protection of the environment, even at the cost of economic growth" on one extreme, versus "supporting economic growth, even at the cost of damage to the environment" on the other, for example, anchors the dimension in a way that makes it clear to human experts who rate and interpret party placements on this aspect of policy (see Benoit and Laver, 2006). For aggregate or *lower* dimensional constructs, most famously that of "left-right", however, this is much more difficult if not impossible to specify in the same way. Expert surveys typically take the approach of either defining a broad range of policies that are typically associated with left-right (e.g. the Chapel Hill expert surveys) or instruct experts to take all aspects into account without specifying what these should be (e.g. Benoit and Laver, 2006).

The widely used "Rile" index constructed from 26 policy categories of the CMP takes a more fixed and prescriptive approach. Regardless of how these categories were chosen—by exploratory factor analysis or by Karl Marx—the validity of this measure depends on its correspondence with what informed observers, according to "standard accounts" (Budge and Meyer, 2013, 91), would consider to be the left-right positions of the political parties it measures. Substantive invariance is of little benefit for an index, in other words, if it fails to measure the high-level dimension for which it was designed. In this section we examine the association between the CMP policy categories and left-right measures, showing that there is no universally applicable set of policy bundles that measure left-right policy in all settings. Instead, we advocate the inductive approach using the IRT model to link categories appropriately to the higher-level dimensions such as left-right, an approach that also allows us to inspect the item discrimination parameters.

4.1 Back to the future: Laver and Budge (1992) revisited

The original selection of the Rile index's components is described in detail by Laver and Budge (1992, 25–30). With the goal of locating parties in a one-dimensional space, Laver and Budge fit the sample of 10 Western European countries[9] from 1945–1985 using exploratory principal components factor analysis fit on a country-by-country basis. Inspecting each set of results for face validity and making some decisions to combine or drop some categories based on their loadings, the result was the first version of the Rile scale now the most widely used quantity

[9]These were: Austria, Belgium, Great Britain, Denmark, France, Ireland, Italy, Luxembourg, Netherlands, and Sweden.

in the CMP dataset. In fitting the manifesto data to the single dimension of difference that appeared to meaningfully differentiate parties, they emphasize that this process was "based solely on the intrinsic plausibility and coherence of the sets of issues that define the underlying policy dimension" (25). Presumably, this is why categories such as "Political authority: Positive (305)" are considered "right-wing": because in the sample examined, this was the pattern of their association.

Applying our one-dimensional IRT model to the same data, pooled across countries, we also observed a good fit for the Rile index to the CMP policy categories. Figure 4 plots the item discrimination parameters $\hat{\beta}_j$ for each policy category, fit to the original sample of ten countries reported in from Laver and Budge (1992) for the same time period. As can be seen by the positioning of the parameter estimates relative to the dividing line, the Rile left categories were indeed associated with left positions, and the right categories with right positions. Some were far less informative than others, however: Freedom and Human Rights: Positive (201), for instance, could not be statistically distinguished from ideologically neutral categories; neither could the "left" categories of Protectionism: Positive (406) or Anti-Imperialism/Anti-Colonialism (103).

Among the 30 policy categories excluded from the Rile index, furthermore, we see several that are very strongly associated with left-right policy positions: Anti-growth Economy: Positive (416) and Marxist Analysis: Positive (415) on the left, and Internationalism: Negative (109) on the right.[10] *For this sample*, we see from Figure 4 that there are numerous categories not used to estimate left-right context that could have contributed productively to the measurement of party positions along this single dimension. By modelling category counts directly as a function of the responsiveness of the party's manifesto content to their underlying latent position θ_i, the IRT approach uses all available information. Instead of requiring a list of "in" and "out" categories, the IRT approach uses them all and estimates their relative contributions from the data.

[10] We note here the irony of ascribing the construction of Rile to Marx given that he apparently failed to include his own namesake category in the left side of the index.

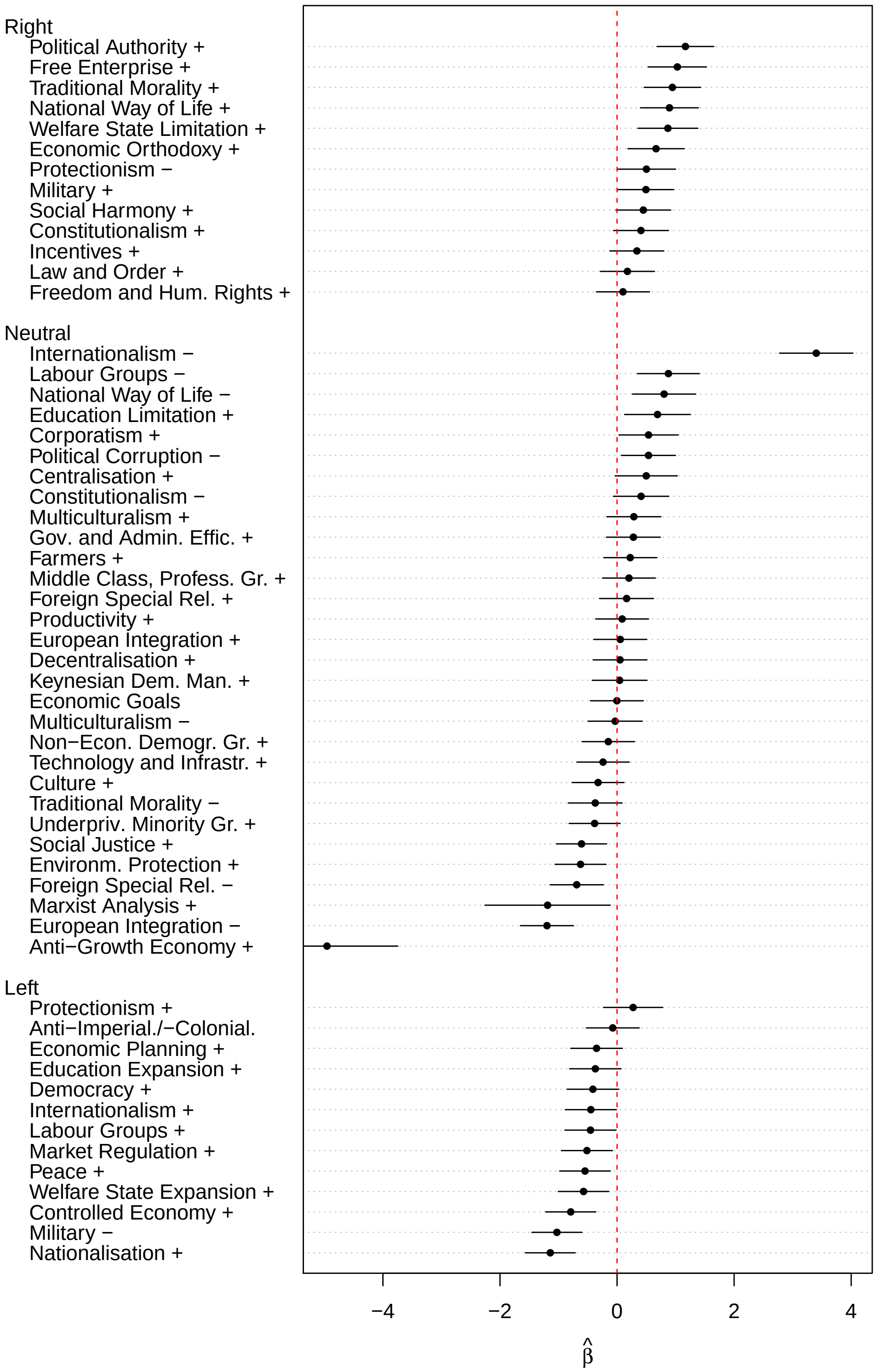


Figure 4: Item discrimination parameter estimates ($\hat{\beta}_j$) for the original 10 countries analyzed in Laver and Budge (1992, Table 2.4).

4.2 Measuring left-right in heterogeneous contexts

The perils of fixed indexes is easily shown when extending the model to other time periods and country contexts. Now covering 55 countries including Eastern Europe, Armenia, Korea, Mexico, and Turkey, it seems appropriate to re-examine the fit of its components in these contexts. Using inductive methods similar to factor analysis, Mölder (2013) found that "Rile" fits very poorly to the post-communist set of countries. Here we extend that analysis to all countries, based on a version of Eq. 2 that decomposes the item parameters into six separate regions. The results are presented in Figure 5, comparing each set of components on the y-axis to the fit from the Western European set on the x-axis. The dashed lines partition the plots into four regions, and the colour present the left categories in blue and the right categories in red. If the Rile items fit a region well, then the blue categories will lie in the lower left quadrant, and the red categories in the upper right, as for the Pacific region (consisting of New Zealand and Australia). If it fits poorly, as for the Far East, then many "left" categories may appear as right (in the upper left quadrant) and many "right" categories may be associated with left-wing positions (in the lower left quadrant).

Figure 5 also indicates which specific policy categories are misfit in each context. In Eastern Europe and the Far East, for instance, category Anti-Imperialism (103) was associated with right-leaning positions, while Free Enterprise: Positive (401) and Economic Incentives: Positive (402) were associated with left-leaning positions. In all but the Pacific region, furthermore, the "right" category of Protectionism: Negative (407) was associated with leftist ideology. In further results below aimed at the content of the economic dimension, we explore the nature of left-right economic content more specifically.

From these very mixed results, we conclude that there is no "holy grail" of pre-specified left-right for a purely "deductive" approach measuring a single dimension of left-right ideological positioning. While comparisons along a single dimension are meaningful, measuring this dimension the same way in every context is not, because the content of left-right varies across countries and times. Just as we must adjust the CPI basket and the weightings of its content to produce a valid measure of inflation, so must we adjust the contents and their weightings to produce valid measures of left-right. Far from undermining its usefulness, this *correct* fit-

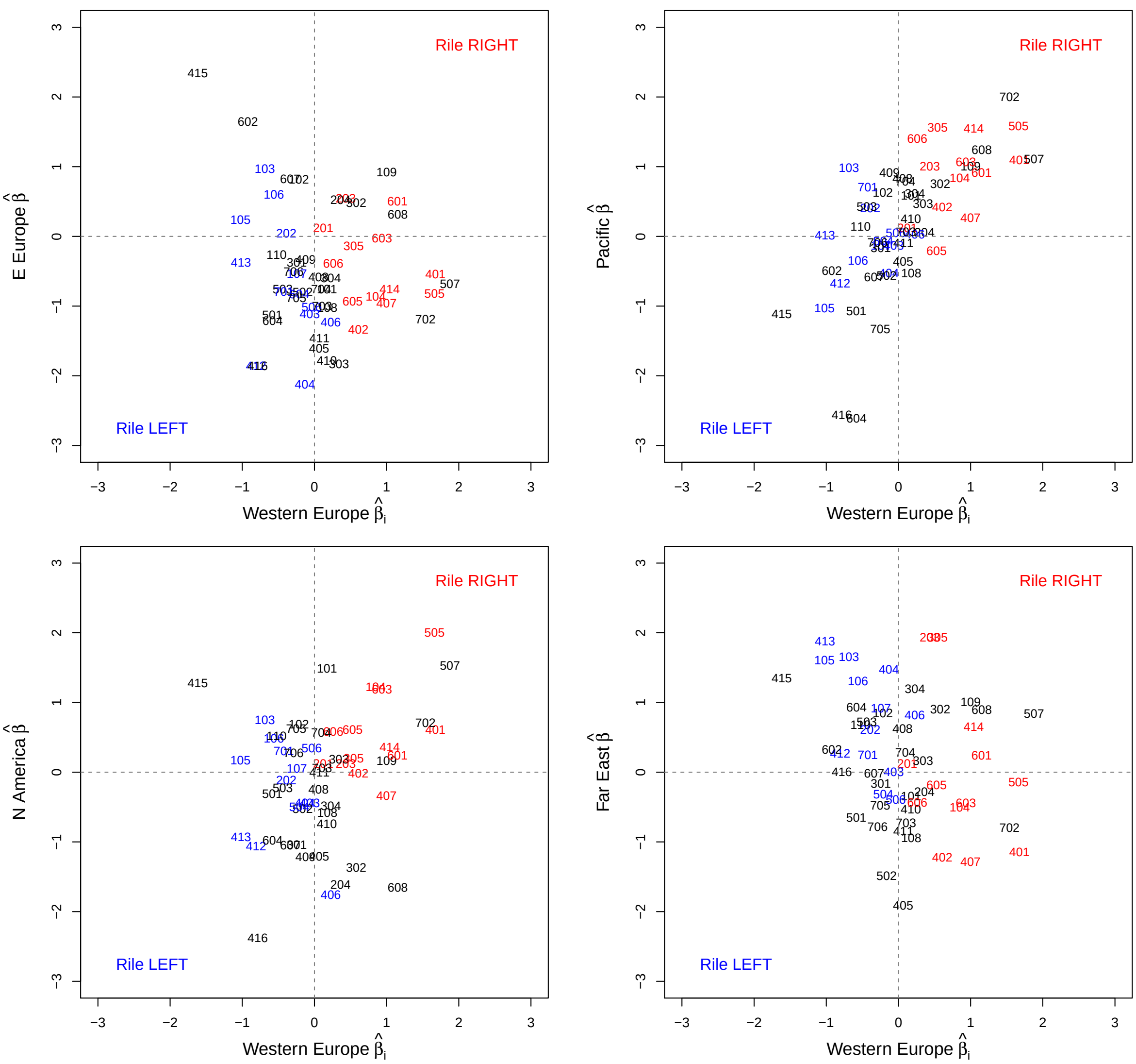


Figure 5: Item discrimination parameter estimates from Western Europe compared to subsets estimates in other regions.

ting to the meaningful single dimension, *in context*, is precisely what makes locations on this dimension comparable across different settings.

5 Using IRT to Estimate Multiple Dimensions of Policy

For many applications, researchers are interested in measuring policy positions in more than one dimension. For instance, many party systems can be adequately characterized by competition along two dimensions, an economic and a "social" one related to moral questions such as abortion and homosexuality (Laver and Hunt, 1992; Benoit and Laver, 2006). Extending the model to two dimensions d is conceptually straightforward. We now model the mean of the counts as

$$\log(\mu_{ij}) = \log(\lambda_i) + \alpha_j + \beta_{1j}\theta_{1i} + \beta_{2j}\theta_{2i} \tag{5}$$

which implies that we estimate two positions θ_{di} and two sets of discrimination parameters β_{di}. The more difficult task, as with any multi-dimensional IRT model, is to impose appropriate constraints in order to reach statistical identification of all the parameters (Jackman, 2001). As in the one-dimensional case, the direction of the scale is not fixed. This applies now to two dimensions, and in addition the dimensions may be swapped, which in total gives 8 possible rotations that are observationally equivalent (Jackman, 2001, 231). In analogy to the one-dimensional case, we restrict the θ_{di} to have a mean of zero and a variance of one. To fix the direction of the scale in each of the dimensions, we use half-normal priors for one β_{dj} value in each dimension. Like this, $\beta_{1;20}$ and $\beta_{2;45}$ (the discrimination parameters of "Free enterprise: positive" on dimension one and "Traditional morality: positive" on dimension two) are constrained to be non-negative, thus associating higher values of θ_{di} with rightist positions.

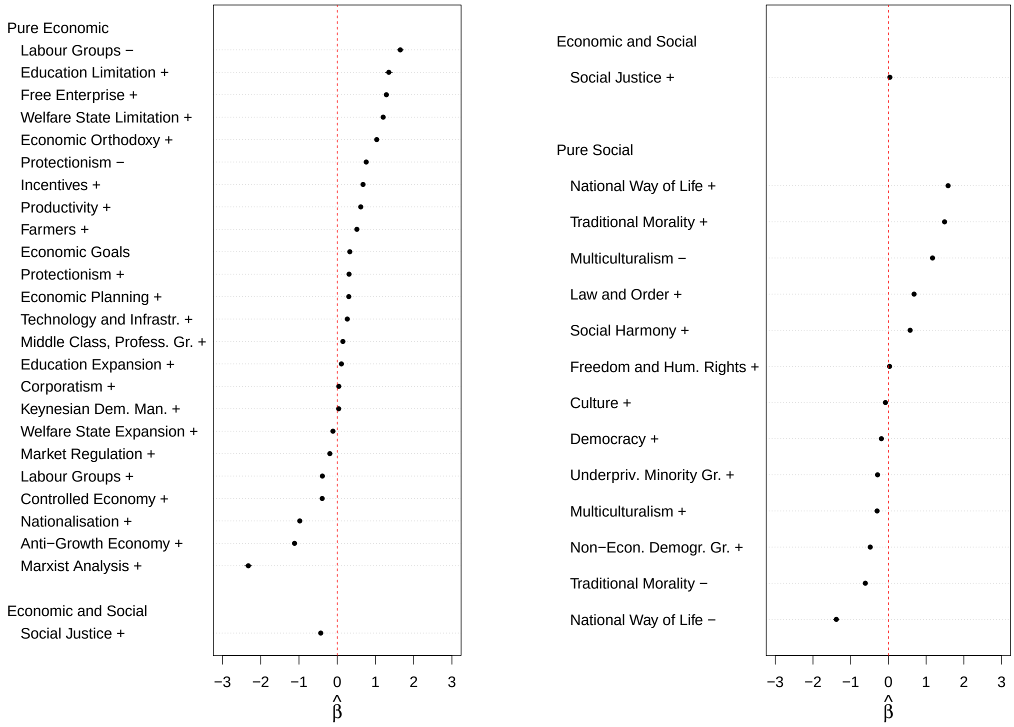
In order to separate the two dimensions and associate them with economic respectively social questions, the β_{dj} are set to zero for "non-economic" items on the economic dimension and similarly for "non-social" items on the social dimension. This implies that we have also a number of categories which do not discriminate on either of the two dimensions, but which are retained in the data and for which we infer an α_j parameter. In addition, in our specification, we have one category that we leave free to be associated with both dimensions, since the item description suggests this: "Social Justice: positive" (503). Note that these choices, however, form just one out of many possible options. A big advantage of the Bayesian IRT approach lies in allowing for the incorporation of prior information about any of the parameters in a very flexible manner.

Our specification of the priors for the two-dimensional model is:

$$
\begin{aligned}
\lambda_i &\sim N(\mu_\lambda, \sigma_\lambda) \\
\mu_\lambda &\sim N(0, 1/\sqrt{1/10000}) \\
\alpha_j &\sim N(0, \sigma_\alpha) \\
\beta_{dj} &\sim N(0, \sigma_\beta) \\
\theta_{di} &\sim N(0, 1) \\
1/\sigma_\lambda^2 &\sim Gamma(.01, .01) \\
1/\sigma_\alpha^2 &\sim Gamma(.01, .01) \\
1/\sigma_\beta^2 &\sim Gamma(.01, .01)
\end{aligned}
$$

with some of the β_{dj} receiving priors of 0. Figure 6 shows the β_{dj} parameters of all items that were selected to contribute to the respective dimension. Again, we can see considerable variation in the extent to which the categories discriminate, this time on an economic left-right dimension. The most rightist items are "Labour Groups: negative", "Education Limitation" and "Free Enterprise: Positive". On the other hand, "Marxist Analysis: positive", "Anti-growth Economy: positive" and "Nationalisation: positive" are associated with strongly left positions. A number of items have clear valence character, as indicated by β_{dj} parameters close to zero. Interestingly, among them is "Welfare State Expansion: positive", the most frequently used of all categories and one that was clearly on the left in the one-dimensional analysis. On a purely economic dimension, it contributes much less to separating left from right parties.

Also on the social dimension, the findings for the β_{dj} parameters make perfect sense. Most clearly discriminating between socially left and right parties are the "National Way of Life" and "Traditional Morality" categories, with the contrasting pairs as expected to be found at the opposing ends of the scale. Categories such as "Freedom and Human Rights: positive", "Culture: positive" and "Democracy: Positive" tend to be of valence character. We allowed "Social Justice: positive" to be *a priori* associated with both dimensions, and it turns out that this category does practically only differentiate parties on the economic dimension, with a negative/left β parameter. Finally it is noteworthy that our uncertainty concerning the β_{dj}



(a) Economic Dimension (b) Social Dimension

Figure 6: Item discrimination parameter estimates $\hat{\beta}_{di}$ from the two-dimensional model fit with subsets of categories selected for possible economic and social policy content.

values is much smaller in the two-dimensional model. Including prior information puts more constraints on the substantive meaning of these dimensions and it becomes easier to infer which categories contribute to the two underlying latent dimensions.

Next, we compare the CMP-based party positions on both dimensions to those from expert surveys as above (Figure 7. On the economic dimension, there is a quite strong positive association, reflected in a respectable $r = .69$. With regard to the social dimension, there is also a positive correspondence, although the observations are much more scattered around the regression line. When assessing the latter figure, it needs to be kept in mind that manifestos do not necessarily cover "social" issues as extensively as they do economic ones, which makes it harder to correctly place the parties on that dimension.

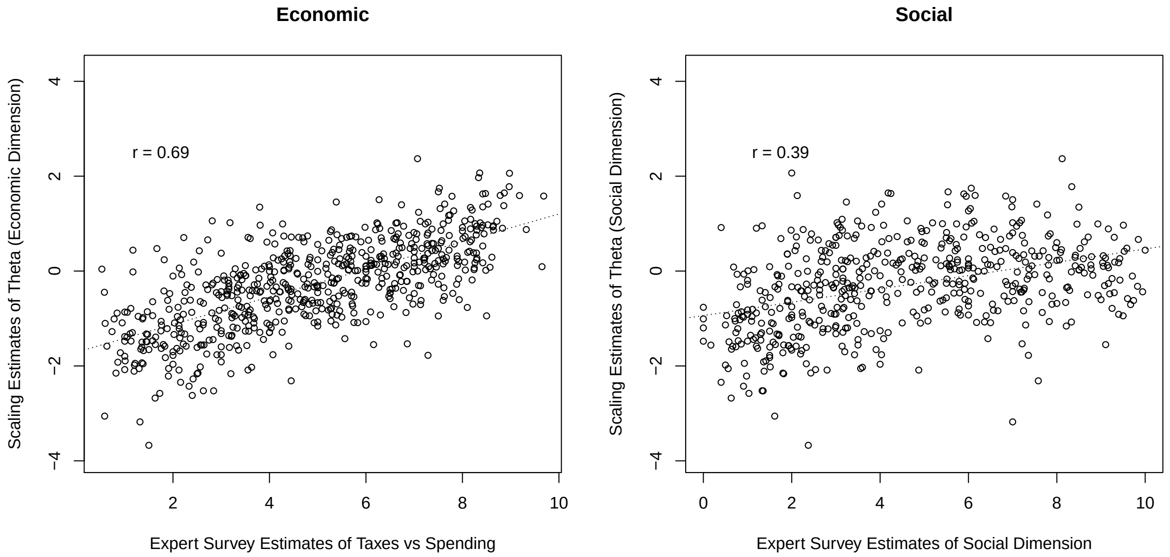


Figure 7: Correlations of two-dimensional model estimates for $\hat{\theta}_{\text{econ } i}$ and $\hat{\theta}_{\text{social } i}$ with expert survey estimates.

6 Using alternative items: The Comparative Agendas Project

To illustrate the flexibility of our model, we also apply it to data from the Belgian part of the Comparative Policy Agendas Project (CAP) (Walgrave and De Swert, 2007; Baumgartner, Green-Pedersen and Jones, 2008). The CAP aims to identify the topic focus of policy documents, media coverage and political events (e.g. cabinet meetings) (Baumgartner, Green-Pedersen and Jones, 2008). Walgrave, Varone and Dumont (2006, 1025) describe their approach used in the Belgian sub-project as follows: “These agendas were encoded in their entirety in order to compute relative issue attention (saliency) in percentage of all issues appearing on these agendas.” In the Belgian case, the coded documents also include party manifestos.[11]

The general coding approach used for the Belgian manifestos resembles that of the CMP, as “(semi)sentences” (Walgrave and De Swert, 2007, 42) were hand-coded into one of 137 (in some cases 143) categories. An important difference between the CAP and the CMP, however, is that the CAP categories are exclusively based on content and thus not intended to be positional. The category scheme spans a wide array of very detailed topics ranging from issues of political organisation (e.g. “State reform, political power and intercommunity conflicts”, code

[11] Unfortunately most country teams that are part of the CAP do not code manifestos. The only other case for which CAP manifesto data are available is Denmark. We are planning to incorporate analyses of the Danish data in future versions of this paper.

012), economic issues (e.g. "trade policy", code 148), social questions (e.g. "migration and integration of immigrants, code 173) to environmental topics (e.g. "water", code 294). Not neglected are important matters such as "conception and contraception" (code 172) and "fishing" (code 318).

These data provide a difficult task for any approach towards measuring party positions, since they are not designed as positional items. It would not be easy at all to come up with a well justified deductive method for estimating policy positions from that data, since it is almost impossible to tell on theoretical grounds which of these policy content categories actually convey information in terms of positions. We infer party positions from the CAP data applying our scaling model to the 39 Belgian manifestos from 1991-2003 that were coded by the Belgian CAP team. The obtained positions are shown in Figure 8, which readers familiar with Belgian politics will recognize as a spatial representation with high face validity. Across elections, the radical right-wing Vlaams Belang is correctly placed, and so are the Green parties at the left end of the political spectrum. The Socialist parties PS and SP are also located towards the left, and the Christian-democratic parties (CVP/CD&V and PSC) can be found in the political center.

To complement these findings, Figure 9 compares the positions inferred from the CAP data to those from other approaches. The graph is based on the set of 13 cases that could be matched with temporally close expert survey results.[12] The results provide further evidence that the IRT model produces valid results also when applied to the CAP data. The retrieved positions correlate at $r = .84$ with the expert surveys, which is slightly higher than for CMP Rile and almost as high as for the IRT model applied to the CMP data. One advantage of the IRT model is of course that we can also easily infer which items distinguish most between the parties. First, and somewhat surprisingly, we actually find that a large number of the categories does provide positional information. For 77% of the 137 items we can say with more than 90% certainty that they discriminate on the left-right dimension. The five most leftist items are Environmental problems with energy (code 301), Biohazard (303), Forestry (317), Specific Industries (354) and Radiation (302), and the five most rightist items are Youth Disaffection

[12] These cases are: Agalev 1999 and 2003, CD&V 2003, CVP 1999, Ecolo 1999, N-VA 2003, PRL-FDF 1999, PS 1999, PSC 1999, SP 1999, VLD 1999 and 2003, VU 1999.

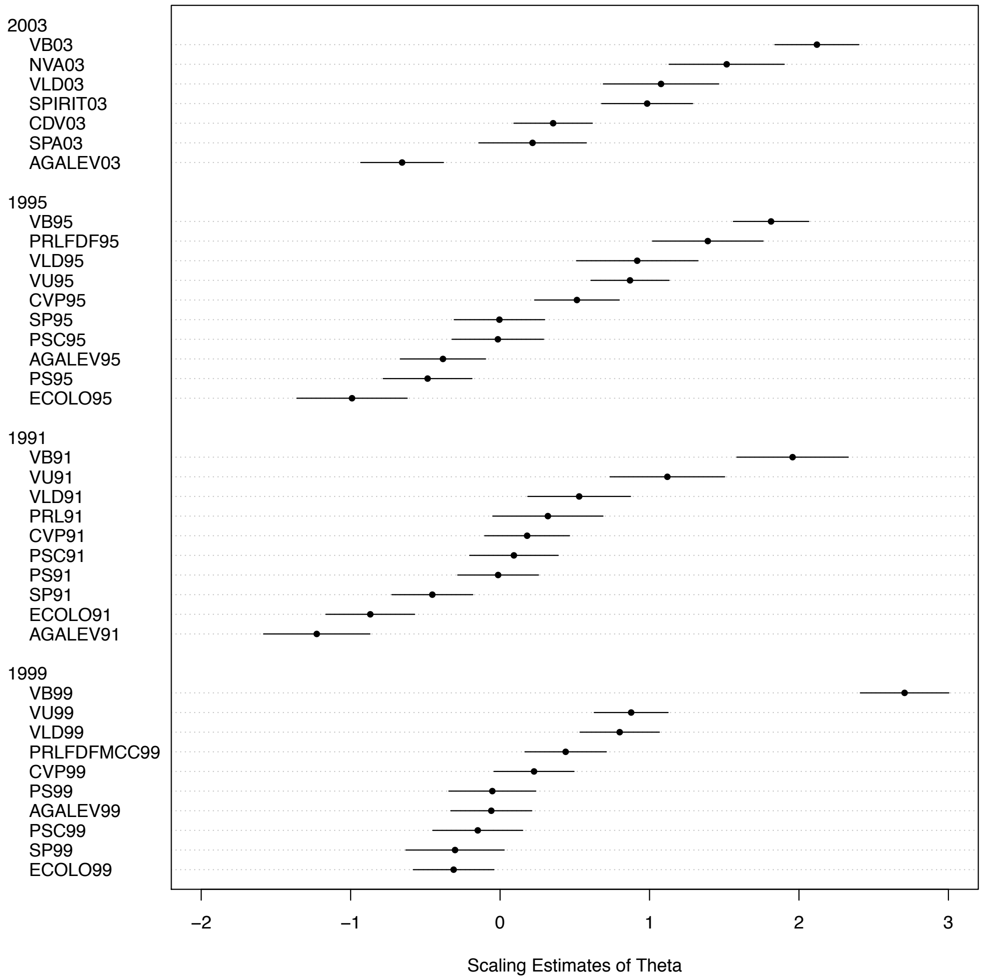


Figure 8: Left-right estimates for Belgian Parties from the Comparative Agendas Project dataset, 1999–2003.

(193), State Reform (12), Private International Law (135) Interior Politics of Other Countries (402) and Family Law (133).

Taken together, using the IRT model on the CAP data provides strong evidence that the important thing about scaling positions is not so much the input in terms of particular items. As long as these contain *some* information that is indicative of policy differences, an appropriate scaling procedure can recover the positions on the latent variable even if a large part of the input data differentiates very little between parties.

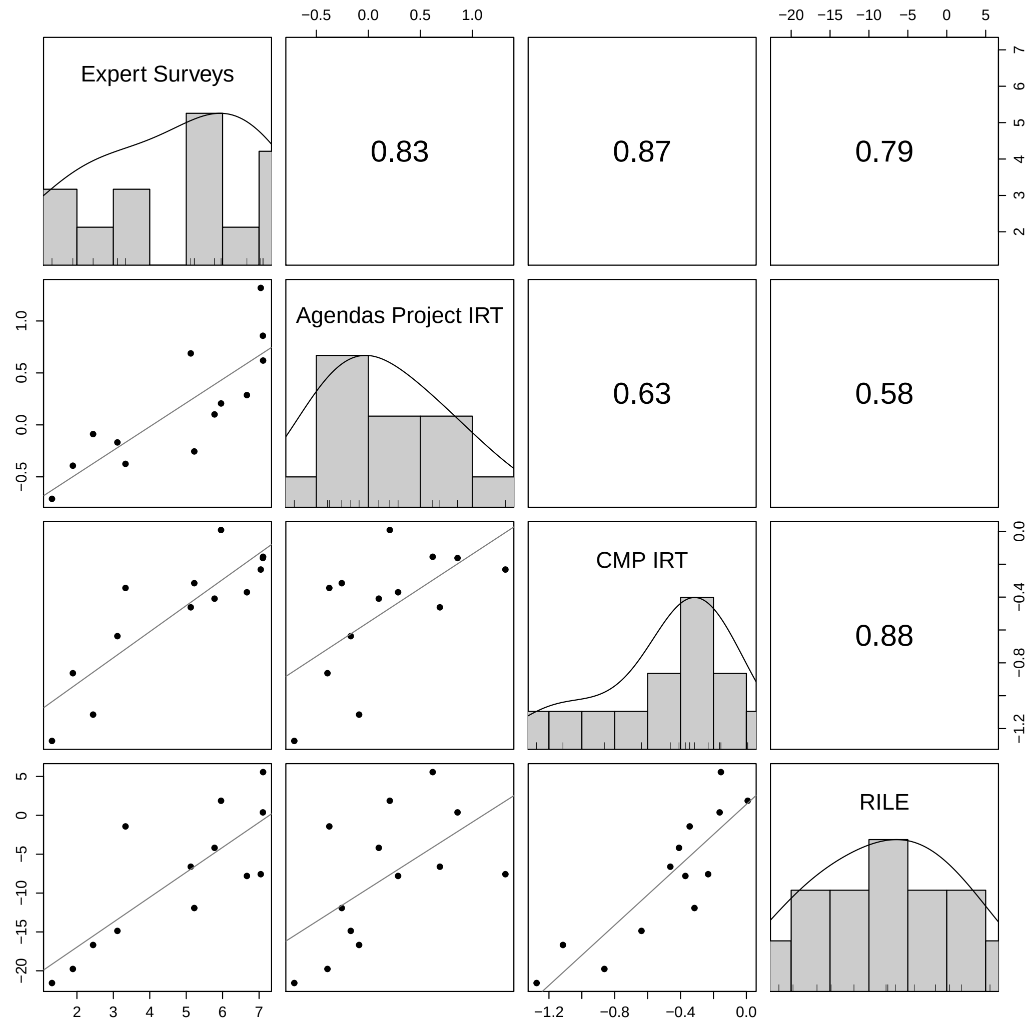


Figure 9: Model estimates of left-right policy positions from the Comparative Policy Agendas Project and the CMP, compared to expert survey results and Rile for 13 Belgian parties 1999–2003.

7 Conclusions

While classic left-right ideology is conceived in terms of broadly similar positioning on bundles of more specific issues, its observation and measurement in practice is not stable or universal with respect to these specific issues. What defines the left-right dimension in Western Europe for instance, is very different from what issues define this dimension in Eastern Europe or settings further east. The implications are that any fixed definition cannot fit all contexts. If a measure built on predefined components fits poorly in a given context, then it fails to provide a valid measure by the most basic definition of validity: that a measure faithfully represents the underlying concept that it purports to measure.

A better alternative to measurement of party locations on a single dimension is to take an inductive approach, letting the data determine the bundling of issue content, using methods that allow direct estimation — and by implication, comparison — of the relative contributions of various policy components to the single dimension in each context. Here we have proposed a measurement approach based on item response theory, permitting the estimation of ideology as a latent "ability" variable, and for the contribution of each element of measured policy to act as "items" for which additional mentions are generated depending on their relationship to the underlying ideology variable. Not only does this approach allow a better estimation of uncertainty about these parameters than other approaches, but also its flexibility permits a more complete model of the political process, including the stochastic generation of overall manifesto length, and the incorporation of additional information as variables to improve the model fit in specific contexts. The remedy for poor fit thus becomes the same remedy as for any omitted variable problem: conditioning on additional information until the fit is restored.

What conclusions can be drawn for the validity of the CMP's Rile index, given our findings? Namely, Rile fails to provide a valid measure of left-right policy in many contexts, echoing earlier findings (e.g. Benoit and Laver, 2007; Mölder, 2013). While Rile continues to fit reasonably well in the context where it was first fit—Western Europe—it travels increasingly poorly to the parties and countries recently added to the CMP's growing collection. This is because *no* single measure constructed in this fashion can have universally good fit, because the meaning of left-right is not universal. Even we accept that most systems do differentiate parties meaningfully along a single heuristic axis, the nature of this axis remains locally determined. There is nothing magical or universal about Rile—or any other constructed index of policy—and because it is a *calculated* quantity rather than an intrinsic part of the manifesto project's research design, there is no reason to cling to it when other approaches have been shown to provide more valid measures of party locations on a single dimension of policy.

The conclusions for the approach to estimating policy positions as latent variables using textual mentions as *items*, by contrast, further demonstrates the great value of coding manifesto content. While it is important the items reflect real policy emphases, our approach means that in measuring latent variables, no fixed decisions about the selection of these items needs to be made at the design stage. Our replication of the left-right policy positions for Belgium using a completely different set of items, from the Comparative Agenda Project's coding of the same party documents, drives this essential point home. Just as in the classical testing framework from which IRT was developed, it is not the exact questions which are of interest, but rather the manner in which patterns of response inform us about the latent quantities of interest. These underlying quantities remain the same for individuals, while tests and their questions differ. Using our inductive approach, we focus directly on the essential quantity—latent ideological positions—while making the most from the items without becoming too obsessed with a debate over their individual contributions to our measure.

Here we have presented the basic model, a model where the conditioning (length) parameter λ_i is modelled as a set of political variables affecting manifesto length, a partition of the items into regions, and a first attempt at estimating multiple (but still low) dimensions of policy using a multidimensional IRT model. In future work we plan to incorporate additional variables, including the item partitioning, more directly into the model to improve fit further, and to explore the relationship of the Poisson scaling model to the classic two-parameter IRT logistic model.

References

Agresti, Alan. 2002. *Categorical Data Analysis*. John Wiley & Sons.

Andersson, Staffan, Torbjörn Bergman and Svante Ersson. 2012. *The European Representative Democracy Data Archive*. Riksbankens Jubileumsfond (In2007-0149:1-E).

Bakker, Ryan. 2009. "Re-measuring left–right: A comparison of SEM and Bayesian measurement models for extracting left–right party placements." *Electoral Studies* 28(3):413–421.

Bakker, Ryan, Catherine De Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Anna Vachudova. 2012. "Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999–2010." *Party Politics* p. 1354068812462931.

Baumgartner, F. R., C. Green-Pedersen and B. D. Jones. 2008. *Comparative Studies of Policy Agendas*. Routledge.

Benoit, Kenneth and Michael Laver. 2006. *Party Policy in Modern Democracies*. London: Routledge.

Benoit, Kenneth and Michael Laver. 2007. "Estimating Party Policy Positions: Comparing Expert Surveys and Hand Coded Content Analysis." *Electoral Studies* 26(1):90–107.

Benoit, Kenneth, Michael Laver and Slava Mikhaylov. 2009. "Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions." *American Journal of Political Science* 53(2, April):495–513.

Bobbio, Norberto. 1996. *Left and right: The significance of a political distinction*. University of Chicago Press.

Budge, Ian, David Robertson and Derek Hearl. 1987. *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*. Cambridge: Cambridge University Press.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford: Oxford University Press.

Budge, Ian and Thomas Meyer. 2013. Understanding and Validating the Left-Right Scale (RILE). In *Mapping Policy Preferences from Texts: Statistical Solutions for Manifesto Analysts*, ed. Andrea Volkens, Judith Bara, Ian Budge, Michael D. McDonald and Hans-Dieter Klingemann. Oxford University Press pp. 85–106.

Carlyle, Thomas. 1888. *The French revolution: a history*. Chapman & Hall, Ld.

Clinton, J., S. Jackman and D. Rivers. 2004. "The statistical analysis of roll call voting: A unified approach." *American Journal of Political Science* 98(2):355–370.

Däubler, Thomas. 2012. "The preparation and use of election manifestos: Learning from the Irish case." *Irish Political Studies* 27(1):51–70.

Dolezal, Martin, Laurenz Ennser-Jedenastik, Wolfgang C Müller and Anna Katharina Winkler. 2013. "How parties compete for votes: A test of saliency theory." *European Journal of Political Research* .

Döring, Holger and Philip Manow. 2012. "Parliament and government composition database (ParlGov): An infrastructure for empirical information on parties, elections and governments in modern democracies. Version 12/10 – 15 October 2012.".

Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2):135–150.

Franzmann, Simon and André Kaiser. 2006. "Locating Political Parties in Policy Space: A Reanalysis of Party Manifesto Data." *Party Politics* 12(2):163–188.

Franzmann, Simon T. 2013. "Towards a real comparison of left-right indices A comment on Jahn." *Party Politics* p. 1354068813499865.

Fuchs, Dieter and Hans-Dieter Klingemann. 1990. The Left-Right Schema. In *Continuities in Political Action: A Longitudinal Study of Political Orientations in Three Western Democracies*, ed. M. Kent Jennings and Jan W. van Deth. Berlin: Walter de Gruyter pp. 203–234.

Gabel, M. J. and J. D. Huber. 2000. "Putting parties in their place: Inferring party left-right ideological positions from party manifestos data." *American Journal of Political Science* 44(1):94–103.

Garry, John and Lucy E. Mansergh. 1999. Irish party manifestos. In *How Ireland voted 1997*, ed. Michael Marsh and Paul Mitchell. Boulder: Westview pp. 82–106.

Gemenis, Kostas. 2013. "What to do (and not to do) with the Comparative Manifestos Project data." *Political Studies* 61(S1):23–43.

Gooding, Philip. 2011. "Consumer Prices Index and Retail Prices Index: the 2011 basket of goods and services." Office for National Statistics.

Inglehart, Ronald. 1984. The Changing Structure of Political Cleavages in Western Society. In *Electoral Change, Realignment and Dealignment in Advanced Industrial Democracies*, ed. Russell Dalton et. al. Princeton: Princeton University Press pp. 25–69.

Jackman, S. 2001. "Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference and model checking." *Political Analysis* 9(3):227–241.

Jackman, Simon. 2009. *Bayesian analysis for the social sciences*. Vol. 846 John Wiley & Sons.

Jahn, Detlef. 2011. "Conceptualizing Left and Right in comparative politics Towards a deductive approach." *Party Politics* 17(6):745–765.

Jahn, Detlef. 2014. "What is left and right in comparative politics? A response to Simon Franzmann." *Party Politics* 20(2):29–301.

Kavanagh, Dennis. 1996. *British politics: continuities and change*. Oxford: Oxford University Press.

Klingemann, Hans-Dieter. 1995. Party positions and voter orientations. In *Citizens and the state*, ed. Hans-Dieter Klingemann and Dieter Fuchs. Oxford: Oxford University Press pp. 183–206.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara and Ian Budge. 2006. *Mapping Policy Preferences II. Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and OECD 1990-2003*. Oxford: Oxford University Press.

König, Thomas, Moritz Marbach and Moritz Osnabrügge. 2013. "Estimating Party Positions across Countries and Time—A Dynamic Latent Variable Model for Manifesto Data." *Political Analysis* 21(4):468–491.

Laver, Michael and Ben W. Hunt. 1992. *Policy and Party Competition*. New York: Routledge.

Laver, Michael and Ian Budge. 1992. Measuring policy distances and modeling coalition formation. In *Party policy and government coalitions*. New York: St. Martin's Press.

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Estimating the policy positions of political actors using words as data." *American Political Science Review* 97(2):311–331.

Lipset, Seymour Martin and Stein Rokkan. 1967. Cleavage Structures, Party Systems and Voter alignments. In *Party Systems and Voter Alignments. Cross-national Perspectives*, ed. Seymour Martin Lipset and Stein Rokkan. Free Press pp. 1–67.

Lowe, W. and K. R. Benoit. 2011. "Practical Issues in Text Scaling Models: Estimating Legislator Ideal Points in Multi-party Systems Using Speeches." Paper presented at MPSA 2011.

Lowe, William, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. "Scaling Policy Preferences From Coded Political Texts." *Legislative Studies Quarterly* 26(1, Feb):123–155.

Mölder, Martin. Forthcoming. "The validity of the RILE left–right index as a measure of party policy." *Party Politics* .

Monroe, Burt and Ko Maeda. 2004. "Talk's cheap: Text-based estimation of rhetorical idealpoints." Working paper: Michigan State University.

Proksch, S.-O. and J. B. Slapin. 2010. "Position taking in the European Parliament speeches." *British Journal of Political Science* 40(3):587–611. forthcoming.

Reckase, Mark D. 1997. "The past and future of multidimensional item response theory." *Applied Psychological Measurement* 21(1):25–36.

Slapin, Jonathan B. and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.

Stokes, Donald E. 1963. "Spatial models of party competition." *The American Political Science Review* 57(2):368–377.

Tavits, Margit and Natalia Letki. 2009. "When left is right: party ideology and policy in post-communist Europe." *American Political Science Review* 103(4):555–569.

Teorell, Jan, Nicholas Charron, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Petrus R Sundin and Richard Svensson. 2013. "The Quality of Government Dataset, version 20Dec13.".

Topf, Richard. 1994. Party manifestos. In *Labour's last chance? The 1992 election and beyond*, ed. Anthony Heath, Roger Jowell, John Curtice and Bridget with Taylor. Aldershot: Dartmouth pp. 149–171.

Van der Brug, Wouter. 2001. Analysing party dynamics by taking partially overlapping snapshots. In *Estimating the Policy Positions of Political Actors*, ed. Michael Laver. Routledge pp. 115–132.

Volkens, Andrea, Judith Bara, Ian Budge and Michael D McDonald. 2013. *Mapping Policy Preferences from Texts: Statistical Solutions for Manifesto Analysts*. Oxford University Press.

Walgrave, Stefaan, Frédéric Varone and Patrick Dumont. 2006. "Policy with or without parties? A comparative analysis of policy priorities and policy change in Belgium, 1991 to 2000." *Journal of European Public Policy* 13(7):1021–1038.

Walgrave, Stefaan and Knut De Swert. 2007. "Where Does Issue Ownership Come From? From the Party or from the Media? Issue-party Identifications in Belgium, 1991-2005." *The Harvard International Journal of Press/Politics* 12(1):37–67.