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

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, 2015

Abstract

Hand-coded party manifestos have formed the largest source of comparative, over-time data for estimating party policy positions and emphases, based on the fundamental assumption that left-right ideological positions can be measured by comparing the relative emphasis of predefined policy categories. We critically challenge this approach by showing that left-right ideology can be better measured from specific policy emphasis using an inductive approach, and by demonstrating that there is no single a priori definition of left-right policy that outperforms the inductive approach across contexts. To estimate party positions, we apply a Bayesian measurement model to category counts from coded party manifestos, treating treating the categories as "items" and policy positions as a latent variable. This approach also produces direct estimates of how each policy category relates to left-right ideology, without having to decide these relationships in advance based on political theory, exploratory analysis, or guesswork. We also demonstrate that the IRT approach can work even when the items are not specifically designed to measure ideological positions. A big advantage of our framework lies in its flexibility: here, we specifically show how two infer policy positions in two dimensions , but there are numerous extensions for future research, such as examining coder effects or adding covariates to predict the model parameters.

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

*An earlier version of this paper was presented at the "Mapping Policy Preferences from Texts" Conference, May 15–16, 2014, Berlin. Predecessors were presented at the Annual Meeting of the European Political Science Association, June 20–22, 2013, Barcelona, and 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, Will Lowe 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 have 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*b*), 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? It is widely viewed as possible, and 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 is meaningful, and certainly *useful*, to label 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).

There is no consensus, however, 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).

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 generalization of a two-parameter IRT model for unordered polytmous responses. Known as the Nominal response model (Bock, 1972), this method has direct equivalencies to existing unsupervised scaling approaches to text analysis (e.g. Slapin and Proksch, 2008). This IRT framework 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. In contrast to a priori approaches based on fixed definitions of left-right ideological content, our inductive method permits all relevant information to be used to scale the left-right super issue, producing measurements from constituent policy statements that better match expert placements. We show how the IRT approach to measuring ideology can be applied to obtain estimates tailored to heterogenous contexts, and extend the model to the two-dimensional case. Finally, using alternative policy statements from a different manifesto coding project,we demonstrate that our measurement approach is by no means confined to an application to a certain dataset.

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 the subject of much debate. Two broad approaches exist to defining the "left-right" political dimension, 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, such as 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*b*). 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 these components, many attempting to identify the ingredients of the left-right content dimension make recourse to authoritative sources long deceased, 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 justify the construction of the Manifesto Project's left-right scale being that highly influential early modern theorists—including Marx and Engels on the left on Disraeli and Spencer on the right—associated certain policy content with their respective positions. Based on these associations, the Manifesto Project's Rile scale selects 13 of its policy categories as pre-defined "left" policy categories, another 13 as "right", and treats its remaining 30 policy categories as unrelated to left-right. 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 remains 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. The best-known general application of the inductive approach is from Gabel and Huber (2000), whose "vanilla" method applied principal components analysis to the manifesto category percentages and scored each manifesto on the first principal component as a measurement of left-right position. From 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, 96). Franzmann and Kaiser (2006) and Franzmann (2013) also proceed primarily in an inductive manner, using regressions of policy category shares on party indicator variables to determine which categories differentiate between parties and thus provide information about positions.

Mixtures of *a priori* and inductive approaches are also possible. König, Marbach and Osnabrügge (2013), for instance, applies dynamic factor analysis on input data consisting of preselected and pre-scaled CMP categories, as well as relying on prior information on party positions from expert surveys for inference. A similar use of expert surveys is made by Bakker (2009), who applied a logistic two-parameter IRT model to a subset of CMP categories that are pre-grouped into left and right items. Even the the original "Rile" scale, despite being held up as a model of the *a priori* approach, was originally based on inductive fitting (see Laver and Budge, 1992, 26-27), by applying 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.[1] 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 the same underlying dataset—the Manifesto Project's database of coded manifesto sentences—continues to form the basis for both *a priori* and inductive measures of left-right policy positions. 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 also include 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 goes the argument, broadens its applicability to not just more, but also to *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.

[1]The discrepancy between claims that Rile was selected on purely a priori grounds (e.g. Budge and Meyer, 2013, 88) and the description in Laver and Budge (1992, 26-27) has been noted elsewhere (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).

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 to specify 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 have 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 estimating one's own set of inductive positions. In addition, existing inductive approaches are not without 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 method which for many appears to be a black box technique. especially if factor loadings, the correlations of the unobserved factors with the observed variables, are not reported or discussed (compare 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

[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.

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 tend to ignore military issues completely.[4]. 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 cannot offer 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 category 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 enables the researcher to decide herself just how much "deductive" information she wants to add to the analysis.

[4]A related point is that linear factor analysis can produce a spurious second factor when applied to data characterized by a bipolar latent dimension (Van Schuur and Kiers, 1994; Maraun and Rossi, 2001)

2 Scaling policy dimensions using IRT

2.1 Data: Policy category counts (from CMP)

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. 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). [5]

We focus not on the proportions of coded categories most commonly used as CMP data, but rather the *counts* of category codes. Counts are the natural unit of measurement for political statements in manifestos. Manifestos are written, and coded, statement by statement. 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

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

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.

2.2 IRT Model for Unordered Categorical Outcomes

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, generating observable data in the form of counts of statements in different policy categories. The configurations of statements made in different party manifestos provide a basis on which we can measure their policy positions, because some policy issues are explicitly positional, or represent valence issues for which differences in emphasis represent differences in position (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). Policy categories form the "test items", parties correspond to the subjects, and the estimate of latent "ability" θ_i represents a party's left-right policy position. Each item's contribution to the observed outcomes is mapped via a series of item parameters that measure their association with the latent ideological dimension.[7]

Applied to the current problem, consider for a single text i that it generates a series of

[6] For simplicity, we speak of the "party" as a single collective author here. In practice, there are typically multiple authors (Däubler, 2012), and these may of course also have different policy positions.

[7] A similar model is introduced in Elff (2013). That article focuses on estimating positions on separate dimensions, for which items are pre-selected, though. Albright (2008) applies a Bayesian binomial model to data for all the CMP categories, but does not consider results for the item parameters at all.

statements x_k where $k = 1, \ldots, v_i$. Each statement that a text i records represents an "item", from which there are a fixed set of J possible unordered categories (statement types). Were there only two possible statement types with $J = 2$, then we could express for category $j = 1$:

$$P(x_k = 1) = \frac{e^L}{1 + e^L} \tag{1}$$

and for category $j = 2$,

$$P(x_k = 2) = 1 - P(x_k = 1) \tag{2}$$

and where the logit transformed quantity L is expressed in the familiar two-parameter logistic item-response formulation as $L = a_j(\theta_i - b_j)$.

In this formulation with just two response categories, θ_i is a latent measure of subject i's "ability" to answer the item with a response $j = 1$ (representing a "correct" answer) versus with a response of $j = 2$ (an incorrect answer). The parameters a_j and b_j represent the discrimination parameters and difficulty parameters of item j, respectively. It is only necessary to speak of a single response category for each item, since the only other response category can be expressed in terms of the probability of this item. For a text coding research design, this would be analogous to having a two category coding scheme, where a text unit might belong only to one category or the other.

Now consider the case where $J > 2$. In this situation, we replace the binomial logistic formulation with the multinomial generalization. Here,

$$Pr(x_k = j) = \frac{e^{L_k^j}}{1 + \sum_{j=1}^{J-1} e^{L_k^j}} \tag{3}$$

where L_k^j represents a "multivariate logit" (Bock, 1972), and there is a vector of J such logits for each item. This formed the basis for Bock (1972)'s *nominal response model* (NRM), generalising the two-parameter logistic IRT model from the binomial to the multinomial case of a multiple, unordered categorical response structure. In the multivariate formulation (for a given text i),

$$L_k^j = \zeta_k^j + \lambda_k^j \theta_i \tag{4}$$

where the quantities ζ_k^j and λ_k^j represent the item parameters for the jth category of response for the "item" k. (Note: k is a statement choice that is made, where this statement has to be assigned to one statement type category.) We note here that the general function in Eq. 4 is not identified, because it is invariant with respect to the translation of L_k^j. We discuss how to constrain and estimate this model below.

For a fixed and equal number of items $n_1 = n_2 = \ldots = n_I$, we could then observe counts of each category j across a set of i individuals, in the same way that we could tabulate response categories across test takers, with the important proviso that in this setup, each item would have identical response categories. This would give us $Y_{ij} = \sum_k x_{ik}$, corresponding to a matrix of counts for each response category j for each text i. In text analysis and many other settings, however, the number of items differ across cases, for example when texts differ in length. Put differently, the number of items $Y_{i.} = \sum_j Y_{ij}$ varies across i. To accommodate this, we can reformulate the NRM as a log-linear model for the expected value μ_{ij} of the counts Y_{ij} :

$$\log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_j \theta_i, \tag{5}$$

where the α_i is a parameter that represents variable text length.

As probability mass function we use the Negative Binomial (cp. Cameron and Trivedi, 1986, :32-33)

$$Pr(Y_{ij} = y_{ij} | \mu_{ij}, \phi_j) = \frac{\Gamma(y_{ij} + \phi_j)}{\Gamma(y_{ij}+1)\Gamma(\phi_j)} \left(\frac{\phi_j}{\mu_{ij}+\phi_j}\right)^{\phi_j} \left(\frac{\mu_{ij}}{\mu_{ij}+\phi_j}\right)^{y_{ij}}$$

The expected value of the counts is given by $E(Y_{ij}) = \mu_{ij}$, and the variance by $Var(Y_{ij}) = \mu_{ij}(1+\phi^{-1})$. The parameter ϕ^{-1} represents the extra variance in the data relative to the special case of the Poisson ($\phi^{-1} = 0$), for which the variance is equal to the mean. One reason why we observe overdispersion may be unobserved heterogeneity at the level of the counts Y_{ij}. Put differently, there are random effects that influence the counts in the cells of the table representing the frequency of category use for each document. To allow for the possibility that the variance systematically differs across categories, we infer a different ϕ_j parameter for each category.

In the case of $\phi^{-1} = 0$, our model is algebraically equivalent to the "wordfish" scaling model for word counts first presented by Slapin and Proksch (2008) (see Appendix for details). Their model, however, 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.[8]

One advantage of the IRT approach over the factor analytic method is that the relationship between parties and items is explicitly modelled (Reckase, 1997, 29). From the inferred item parameters we can learn much about the content of the latent dimension. To start with, after a bit of algebraic manipulation we can see the relation of the item parameters to their counterparts in the standard 2PL-IRT model wherein $L = a_j(\theta_i - b_j)$:

$$a_j = \lambda_j \tag{6}$$

$$b_j = \frac{\zeta_j}{\lambda_j} \tag{7}$$

The λ_j therefore form the "discrimination" parameters a_j, indicating indicate how a particular policy category j's use varies in response to changes in the latent dimension θ_i. Put differently, the absolute size of λ_j reflects the degree to which a category is positional, and its sign shows whether the category is a "left" or a "right" item. Note, however, that the "difficulty" parameter b_j in the standard model is a combination of the values of the two item-level parameters in the NRM. This equivalency was also noted by Baker (1992), who cautioned that the values of the NRM parameters do not have a simple formulation in terms of the standard (e.g. 2PL) IRT model, because they describe the discrimination and location of specific item category response functions, whose shapes and locations depend on the way the parameter values from all the categories combine (Ostini and Nering, 2009, 18).

[8]More recently, Lo, Proksch and Slapin (2014) extended their model introducing an overdispersion parameter at the document level.

This complication is shown in Figure 1, which refers to a hypothetical example with five item categories that differ with regard to their λ_j value.[9] The α parameter was set to 1, and so were all five ζ_j parameters, implying that the baseline frequency of the five categories (i.e. the part unrelated to the latent position) is the same. The left panel shows how the expected number of items falling in the respective category varies with the latent position θ_i, depending on λ_j. The expected number of statements increases (decreases) monotonically over the range of the latent position if λ_j is positive (negative), and does so more strongly the more extreme λ_j. When $\lambda = 0$, the category is not responsive to the latent position at all, and the curve is flat at the level of the baseline frequency determined by the combination of α_i and ζ_j. Also note that in the case of the example, all curves cross in the same point at $\theta = 0$, due to the baseline frequency being equal as all ζ_j were chosen the same. The right-hand panel now illustrates how the parameter values *jointly* form the expected probability that an item falls in a certain category (or equivalently, the expected proportion of items in a category). Here, due to the interdependence, only the item response functions for the two categories with the lowest/highest λ_j are monotonically decreasing/increasing, approaching zero and one in the limit, as the latent position ranges from plus to minus infinity. In other words, an infinitely rightist/leftist document would only consist of words/statements falling into the rightmost/leftmost category. The curves for the remaining $J-2$ categories, in contrast, follow a unimodal shape. As in the left panel, the equality of the ζ_j implies the curves intersect at $\theta = 0$, with the associated y-coordinate in the right graph equal to 1/J=0.2. Also note that the plot in the right panel holds for any α, i.e. regardless of the total number of items.

In this context, note another interesting implication of the IRT 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 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.[10]

[9] Elff (2013, :221) briefly discusses a plot similar to the right panel of Figure 1.

[10] Implicit in our model is a quadratic spatial utility function (compare Lowe, 2014, :Appendix A): document parameter θ_i and item parameter λ_j are multiplied, which can be shown to result from the difference between a document location and a category location being squared (see also Elff, 2013). We do not try to uncover these locations, since we see little analytical value added by projecting both parties and items into the same space (which lacks a clear-cut interpretation, since there is no status quo and an alternative such as in the context of roll-call voting).

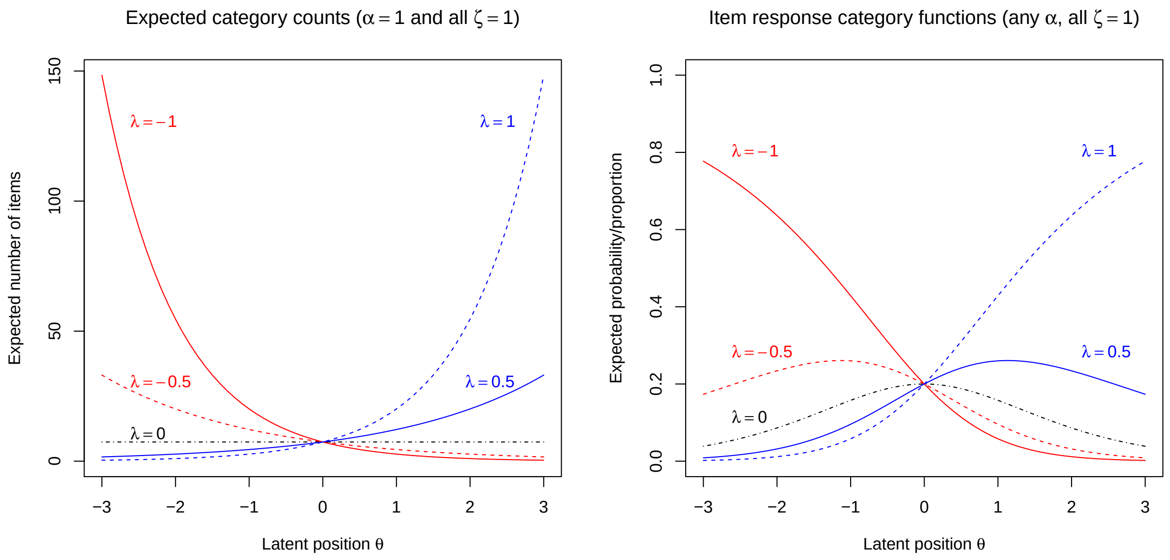


Figure 1: Item category characteristic curves for a hypothetical example with five categories.

The model in Eq 5 requires additional constraints for identification, as there are five fundamental indeterminacies. First, shifts in the mean of the ζ_j can be compensated by shifts in the mean of the α_i – put differently, it is impossible to infer whether all categories are jointly more (less) frequent or whether the manifestos are longer (shorter) overall. So the location of the ζ_j needs to be fixed by a mean or corner constraint. Second, changes in the mean of the λ_j can be set off by changes in α_i – when all the categories are jointly more (less) responsive to position, the resulting addition (loss) of text can be offset by higher (lower) alpha values (with the specific amount for text i depending on θ_i). This issue can be addressed e.g. by imposing that the mean of λ_j equals zero. Third, the mean of the positions θ_i is not fixed. Increases (decreases) can be set off by shifting the mean of the ζ_j, i.e. making the baseline frequency of all categories smaller (larger). We can resolve this indeterminacy by constraining the mean of the positions. Fourth, only the ratio of the variance of λ_j and the variance of θ_i is fixed. We can infer the variation in positions relative to the variation of the discrimination parameters, but not their absolute levels. This implies that either of the two variances needs to be constrained (for instance to be one), while the other is allowed to vary. Fifth, the polarity of the positions θ_i is not determined, since larger values can represent either more rightist or more leftist positions (we can multiply both λ_j and θ_i by negative one and get the same result). This reflection invariance can be prevented by constraining the order of either two positions $\theta_i < \theta_{i'}$

or discrimination parameters $\lambda_j < \lambda_{j'}$.

In a Bayesian framework, these restrictions can in practice be implemented as either hard constraints or as soft constraints through the choice of priors. We use a hard constraint (choosing a reference category j for which $\zeta_j = 0$) to address the first indeterminacy, and soft constraints through priors for the second to fourth:

$$
\begin{aligned}
\alpha_i &\sim N(\mu_\alpha, \sigma_\alpha) \\
\zeta_j &\sim N(\mu_\zeta, \sigma_\zeta) \\
\lambda_j &\sim N(0, \sigma_\lambda) \\
\theta_i &\sim N(0, 1) \\
\mu_\alpha &\sim N(0, 5) \\
\mu_\zeta &\sim N(0, 5) \\
\sigma_\alpha &\sim \text{Half-Cauchy}(0, 5) \\
\sigma_\zeta &\sim \text{Half-Cauchy}(0, 5) \\
\sigma_\lambda &\sim \text{Half-Cauchy}(0, 5) \\
\phi^{-1} &\sim \text{Uniform}(0, 200)
\end{aligned}
$$

We leave resolving reflection invariance to the post-processing stage, when we invert the scale (if necessary) so that increasing θ_i and λ_j represent more rightist positions (as judged on the basis of prior knowledge). We also "harden" the soft constraints post-hoc by mean-deviating λ_j and standardising θ_i in each draw, and mean-center the ζ_j (which makes the specific choice of the reference category irrelevant). We simulate all models using Hamiltonian Monte Carlo (Hoffman and Gelman, 2014), as implemented in the software package Stan (Stan Development Team, 2015), by sampling from the posteriors following a suitable warm-up period.

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. 5) 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.[11] We restricted our sample to manifestos issued in democratic countries after the first oil crisis which arguably changed politics considerably.[12] Table 1 presents these results for the IRT model, the first with the Poisson variance and the second with a separate variance parameter ϕ_j estimated at the policy category level. 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 plus North America. (No expert surveys were available for other regions, so these are not reported.) The correlation with all expert survey postions was 0.75 for the negative binomial model, and 0.73 for the Poisson. The correspondence was highest in the Pacific and North America, at 0.89 and 0.82 respectively (negative binomial). 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, and that manifestos may also be a less reliable

[11] We aggregated the extended four-digit category codes into their respective three-digit "parent" categories.

[12] To be precise, starting from the 2014b edition of the data (Volkens et al., 2014) we use post-1972 manifestos from countries with a Polity-IV rating of at least seven, or a Freedom House rating of at least nine (when no Polity-IV rating was available). We dropped duplicate data entries (manifestos associated with several parties) and cases based on estimates and those with missing document length information.

Model		Poisson		Negative Binomial
	Mean	*95% BCR*	*Mean*	*95% BCR*
Illustrative Estimates of θ_i				
Germany 2010				
Left	-1.91	[-2,05, -1,78]	-2.23	[-3,04, -1,41]
Greens	-1.36	[-1,45, -1,28]	-1.30	[-2,08, -0,52]
SPD	-0.90	[-1, -0,8]	-0.89	[-1,68, -0,1]
CDU/CSU	0.37	[0,27, 0,46]	0.41	[-0,34, 1,15]
FDP	0.37	[0,28, 0,46]	0.37	[-0,24, 0,99]
UK 1987				
Labour	-0.57	[-0,78, -0,36]	-0.57	[-1,28, 0,14]
Liberal Democrats	-0.28	[-0,44, -0,12]	-0.07	[-0,73, 0,6]
Conservatives	1.11	[1, 1,23]	1.28	[0,53, 2,02]
UK 1997				
Labour	0.26	[0,11, 0,41]	-0.10	[-0,85, 0,66]
Liberal Democrats	-0.17	[-0,34, -0,01]	0.06	[-0,67, 0,79]
Conservatives	0.91	[0,8, 1,02]	0.87	[0,06, 1,69]
United States 2012				
Democrats	-0.11	[-0,23, 0,01]	-0.14	[-0,73, 0,45]
Republicans	1.56	[1,47, 1,65]	1.00	[0,45, 1,56]
Chains		2		2
Burn-in		300		1,000
Samples per chain (after thinning by)		250 (2)		500 (4)
N manifestos		2,250		2,250
Correlations of θ_i				
With All Expert Surveys	0.73	(*n*=525)	0.75	(*n*=525)
W Europe subset	0.78	(*n*=362)	0.82	(*n*=362)
E Europe subset	0.57	(*n*=143)	0.55	(*n*=143)
Pacific & N America subset	0.86	(*n*=20)	0.89	(*n*=20)
With CMP Rile	0.78		0.69	
With "Vanilla" Method	0.84		0.94	

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

data source in that region. This finding is consistent with earlier research findings that the content of left-right policy is different in post-communist settings (Benoit and Laver, 2006; Mölder, 2013). Correlations with Rile were 0.69 and 0.78 for the negative binomial and Poisson models respectively, and 0.94 and 0.84 for the “vanilla” method based on principal components analysis.[13]

[13]This confirms what Lowe (2013) shows theoretically: the “vanilla” approach implicitly approximates the type of model we describe.

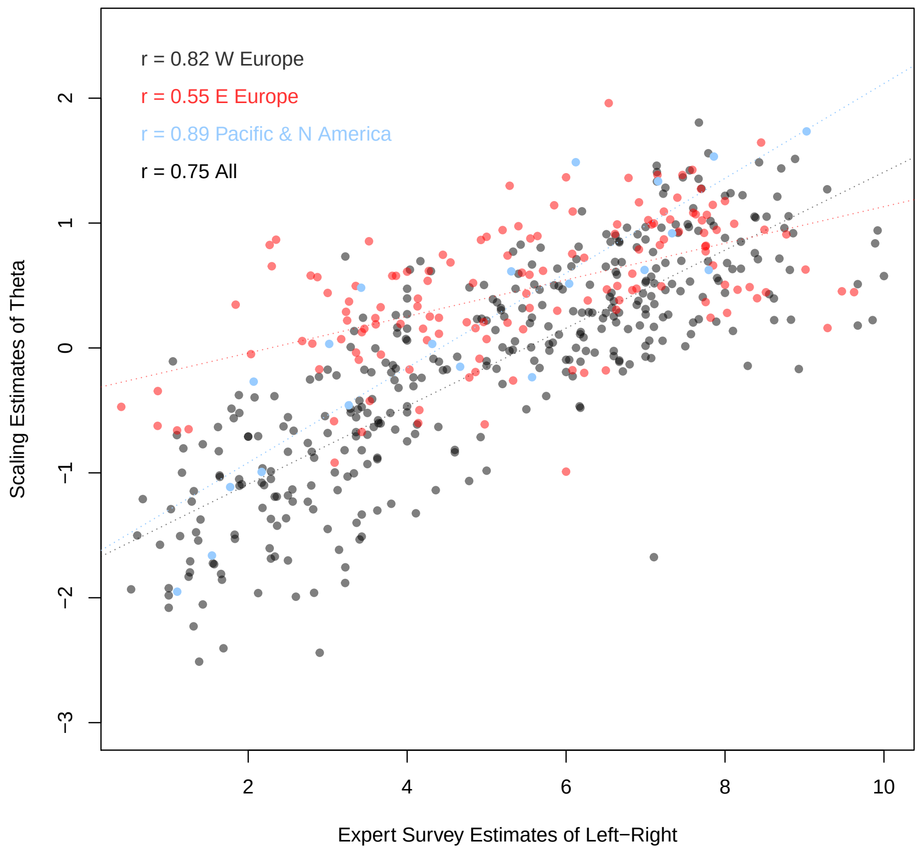


Figure 2: Comparison of $\hat{\theta}_i$ to expert survey estimates for the post-1972 sample (from Table 1).

Figure 2 plots the correlations against expert surveys, indicating a good linear fit, also illustrating the difference in correlations for the subsets by region.[14] The slopes of the patterns differ according to the 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.

[14]Expert data are from Benoit and Laver (2006); Steenbergen and Marks (2007); Hooghe et al. (2010); Bakker et al. (2012*a*). We match the expert placement to the temporally closest manifesto, if a document is available within three years before or after the survey date.

Statistically Significant Change?	BLM (2009) Elections	%	Poisson $\Delta\hat{\theta}_i$ Elections	%	Neg Binomial $\Delta\hat{\theta}_i$ Elections	%
No	1,308	62.3%	836	52.7%	1,532	96.7%
Yes	791	37.7%	749	47.3%	53	3.3%
Non-adjacent	778	–	703	–	703	–
Total	2,877	100.0%	2,288	100.0%	2,288	100.0%

Table 2: 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.

3.2 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 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 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 2 that this rate of change is actually far lower, at just 3.3% of changes, suggesting that the policy shifts by parties from one election to the next on the left-right "super-issue" tend in fact to almost never occur. 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. The contrast of the negative binomial results with those from the Poisson model, furthermore, illustrates one of the significant consequences of using the restrictive and unrealistic variance assumption of the Poisson model, which leads to significantly underestimated parameter uncertainty in $\hat{\theta}_i$ (Lowe and Benoit, 2013).

4 Left-right as a super-issue in different contexts

4.1 The policy components of the left-right dimension

For proponents of 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, especially that of a single left-right ideological dimension, 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 Manifesto Project's widely used "Rile" index constructed from 26 policy categories of the CMP takes a more fixed and prescriptive approach. However these categories were chosen, 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.

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[15] 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 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 3 plots the item discrimination parameters $\hat{\lambda}_j$ for each policy category, fit to post-1972 coded manifestos. As can be seen by the positioning of the parameter estimates relative to the dividing line, most of the the Rile left categories were indeed associated with left positions, and most of the right categories with right positions. Some were far less informative than others, however, and were not estimated as corresponding to left or right, including Political Authority: Positive (305), Social Harmony: Positive (606), Constitutionalism: Positive (203), Freedom and Human Rights: Positive (201), Education Expansion: Positive (506), and Protectionism: Positive (406). Some of these categories, such as Political Authority: Positive, are known to cause problems in indexing left-right positions by biasing leftist parties to the right, such as the Italian Communist Party, a far-left party erroneously scored as far right in the 2000s because of its high proportion of

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

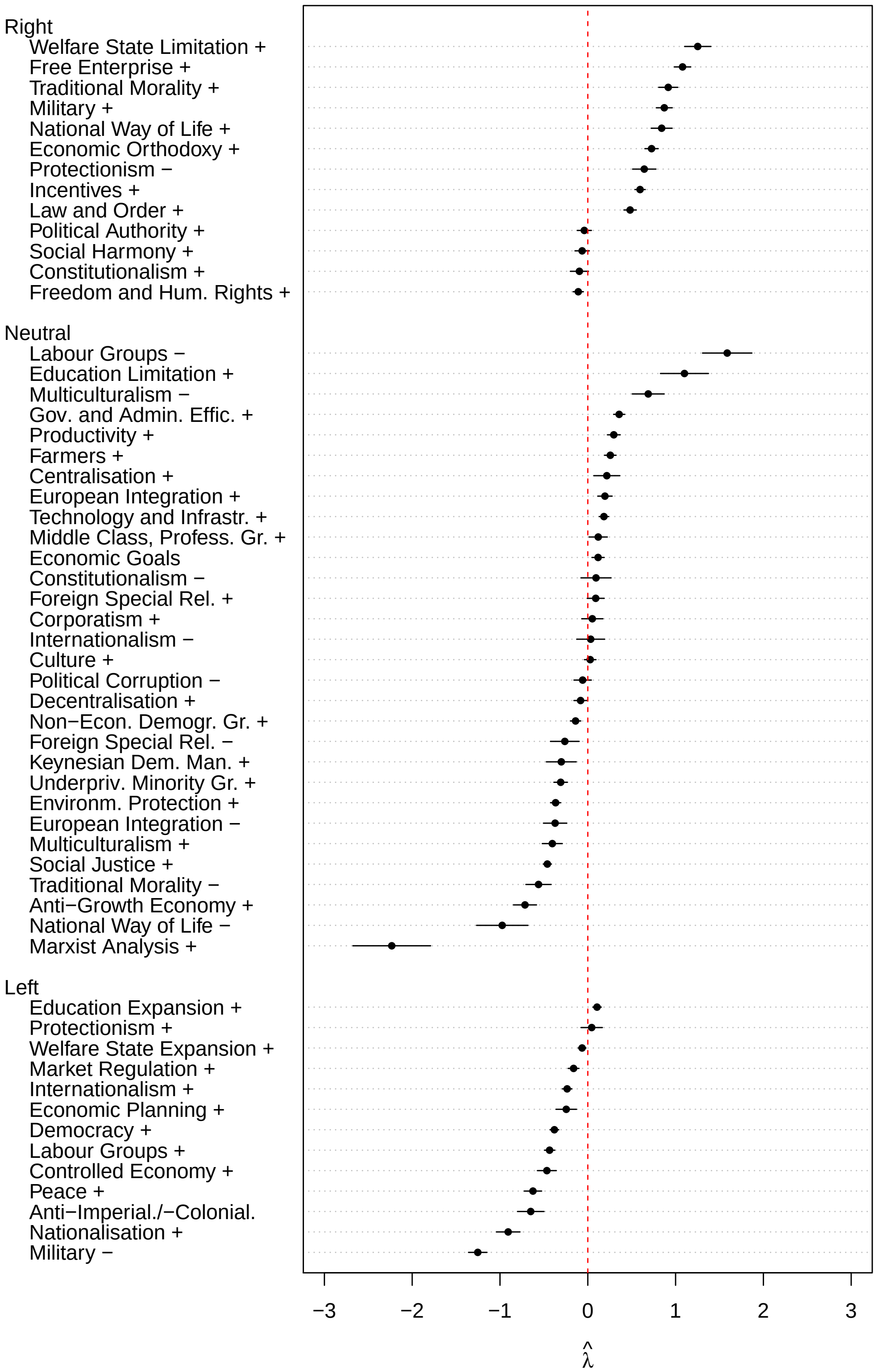


Figure 3: Item discrimination parameter estimates ($\hat{\lambda}_j$) by Rile category, for the post-1972 sample.

statements coded Political Authority: Positive (see Benoit and Laver, 2007, 97–98).

Among the 30 policy categories excluded from the Rile index, furthermore, we see several that are very strongly associated with left-right policy positions: Marxist Analysis: Positive (415) and National Way of Life: Negative (602) on the left, for instance, and Labour Groups: Negative (702) and Multiculturalism: Negative (608) on the right. *For this sample*, we see from Figure 3 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.

4.2 Left-right in heterogeneous 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 fitting the IRT model separately to five separate regions. The results of the estimates of λ_j are plotted in Figure 4, which compares the item discrimination parameter estimates from each region 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 4 also indicates which specific policy categories are misfit in each context. In Eastern Europe, for instance, the Rile right categories of Protectionism: Negative (407) and Constitutionalism: Positive (203) were strongly associated with left-leaning positions. In North America, Rile left categories such as Anti-Imperialism (103) and Military: Negative (105) were neutral categories, as was Rile right category National Way of Life: Positive (601). In the Far

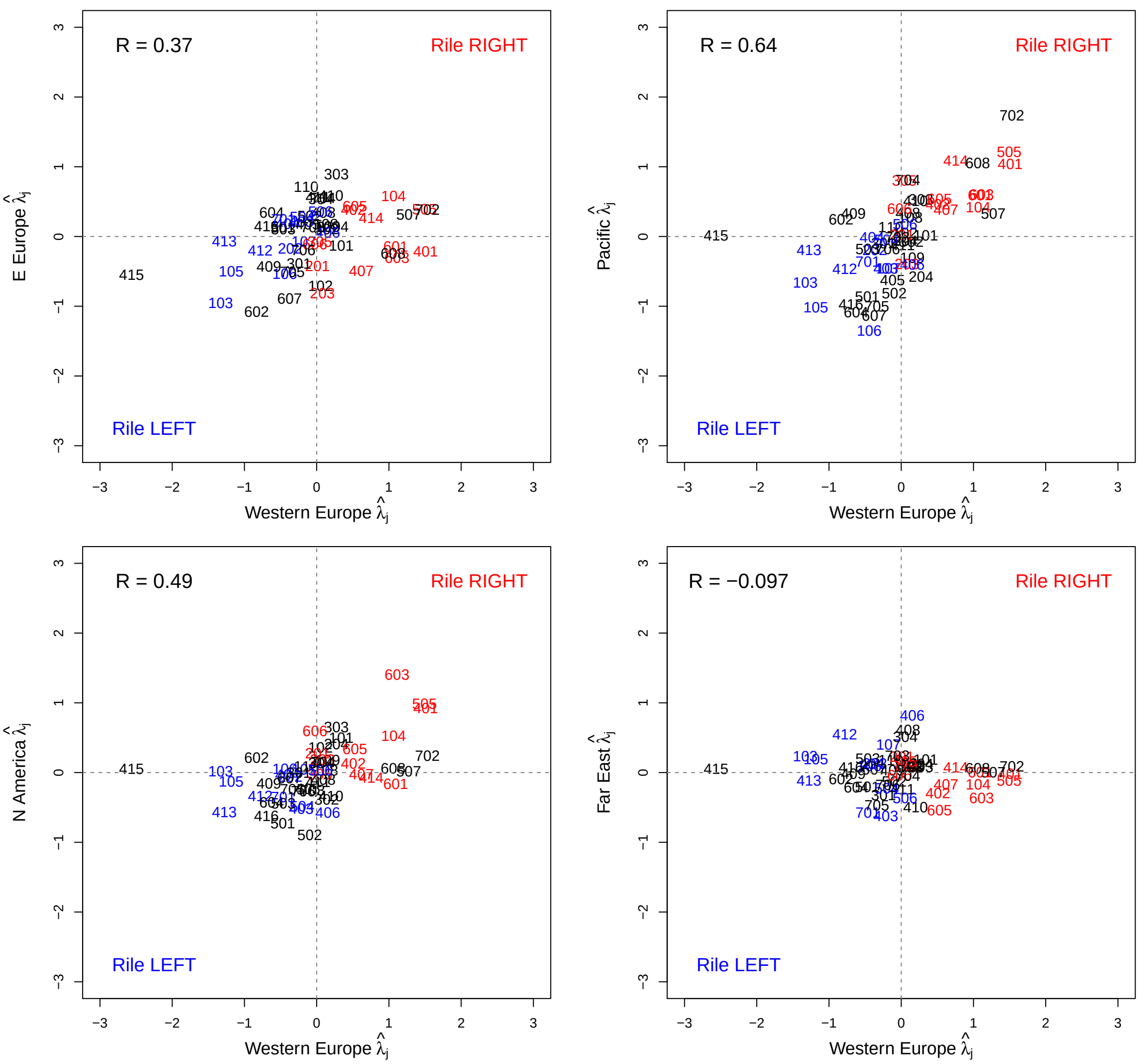


Figure 4: Item discrimination parameter estimates from Western Europe compared to subsets estimates in other regions.

East, Rile right categories Economic Incentives: Positive (402) and Protectionism: Negative (407) were neutral categories. Even the seemingly clearly leftist category of Marxist Analysis: Positive (415) was associated with neutral positions in non-European contexts.

In general, the further the context was from the Western European context where the Manifesto Project's fixed left-right index was developed, the worse the fit. In the Far East, the association of the items to those estimated from the Western European sample was nearly zero ($R = -0.097$). From these results, we conclude that there is no single set of fixed ingredients that can reliably and consistently measure the left-right ideological divide. 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* fitting 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). One approach for obtaining policy positions in multiple dimensions is to assign categories *a priori* to the different policy fields/dimensions, and conduct the scaling on a dimension-by-dimension basis (Elff, 2013). Proceeding this way is not recommendable, however, when we have reasons to believe that there are categories which are linked to more than one of the latent dimensions.

For this purpose, we can extend the model to two dimensions d, which is conceptually straightforward. We now model the mean of the counts as

$$\log(\mu_{ij}) = \alpha_i + \zeta_j + \lambda_{1j}\theta_{1i} + \lambda_{2j}\theta_{2i} \tag{8}$$

which implies that we estimate two positions θ_{di} and two sets of discrimination parameters λ_{dj}. 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; Rivers, 2003).

To fix the variances and the covariance of the discrimination parameters, we use a category j for which $\lambda_{1j} = 0$ and $\lambda_{2j} = 1$, and a category j' for which $\lambda_{1j'} = 1$ and $\lambda_{2j'} = 0$. The means of the positions in each dimension are set to zero, too. These six restrictions constitute the first set of constraints. For further identification, the λ_{dj} are set to zero on the economic dimension for items that we believe are certainly "non-economic", and to zero on the second dimension for items that we judge to be clearly unrelated to "social" questions as understood here. This implies that we also have 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 a ζ_j parameter. Of particular importance, however, is that we leave some items free to be associated with both dimensions.[16] The constraints are completed by setting one of the ζ_j to zero, as in the one-dimensional case.

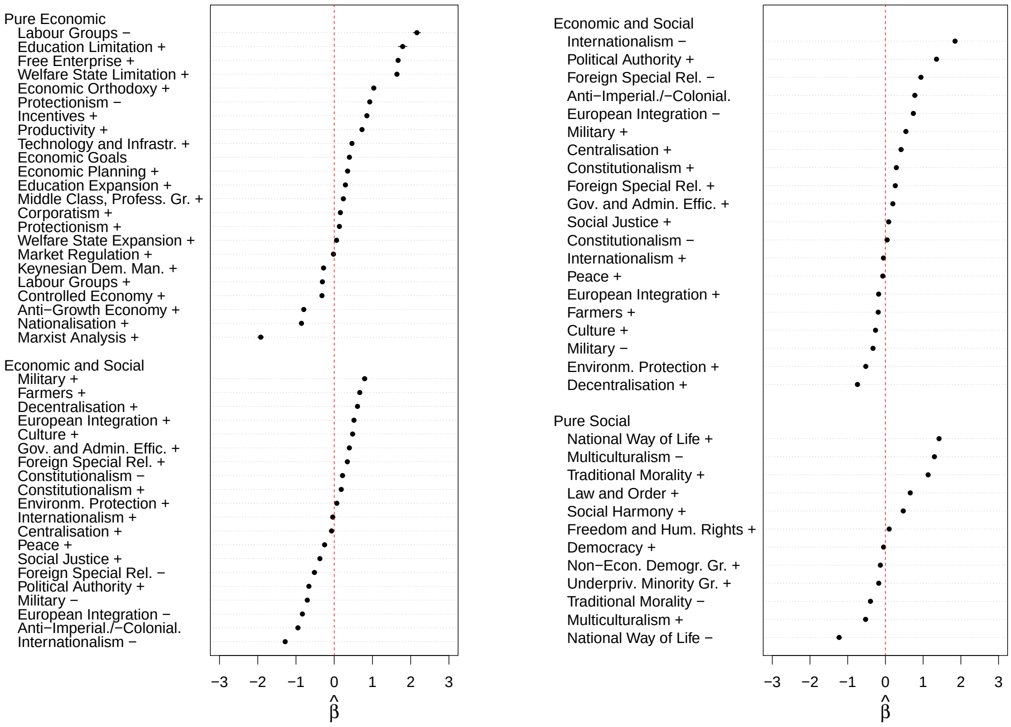
Priors for the two-dimensional model are chosen as follows:

$$\begin{aligned}
\alpha_i &\sim N(\mu_\alpha, \sigma_\alpha) \\
\zeta_j &\sim N(\mu_\zeta, \sigma_\zeta) \\
\lambda_{dj} &\sim N(\mu_{\lambda_d}, \sigma_{\lambda_d}) \\
\theta_{di} &\sim N(0, \sigma_{\theta_d}) \\
\mu_\alpha &\sim N(0,5) \\
\mu_\zeta &\sim N(0,5) \\
\mu_{\lambda_d} &\sim N(0,5) \\
\sigma_\alpha &\sim \text{Half-Cauchy}(0,5) \\
\sigma_\zeta &\sim \text{Half-Cauchy}(0,5) \\
\sigma_{\lambda_d} &\sim \text{Half-Cauchy}(0,5) \\
\sigma_{\theta_d} &\sim \text{Half-Cauchy}(0,5)
\end{aligned}$$

with (hard) constraints applied to some of the λ_{dj}, as just described. For the two-dimensional model, we set $\phi^{-1} = 0$, i.e. choosing a Poisson likelihood *a priori*.[17]

[16] While we do not attempt to prove this, our experience suggests that the model is just identified when applying one constraint on a λ_{1j}, one constraint on a λ_{2j}, one double constraint on λ_1 and λ_2 for the same category, and two further constraints on any λ_{dj} (in addition to the first set described at the beginning of the paragraph). So what we present below is an over-identified model. Note that these constraints also resolve reflection invariance.

[17] Relaxing this assumption led to convergence problems. The way the constraints are set requires that both the variances of the policy positions and the variances of the discrimination parameters are inferred from the data. It seems that this already tricky inference problem is exacerbated when adding an error variance in form of the overdispersion parameter. We post-process the HMC draws so that the variances of the positions (in each dimension) equal one in each draw.



(a) Economic Dimension

(b) Social Dimension

Figure 5: Item discrimination parameter estimates $\hat{\lambda}_{dj}$ from the two-dimensional model fit with subsets of categories selected for possible economic and social policy content.

Figure 5 shows the λ_{dj} parameters of all items that were selected to contribute to the respective dimension (economic in the left panel of the graph, and "social" in the right one).[18] Again, we can see considerable variation in the extent to which the categories discriminate. For the purely economic items, "Labour Groups: negative" and "Education Limitation" are most rightist, whereas "Marxist Analysis: positive" "Nationalisation: positive" are the most leftist ones. Considering the "social dimension", the results also correspond to the expectations, with the contrasting pairs related to "National Way of Life" and "Traditional Morality" to be found at the opposing ends of the scale.

Particularly interesting insights can be gained from the results for the items that were allowed to contribute to both dimensions. We may expect that most of these items are either predominantly associated with only one dimension, or that economically rightist (leftist) dis-

[18]The categories whose λ_{dj} was set to one were "Free enterprise: positive" for the first and "Traditional morality: positive" for the second dimension. As item for which $\lambda_1 = \lambda_2 = 0$ we choose "Political corruption: negative".

crimination parameters tend to go along with socially conservative (liberal) ones. Indeed, we find such categories. For example, positive references to the military represent positions on the right side on both dimensions. And "Environmental protection: positive" is one category whose usage in manifestos is mostly explained by a party's position in one of the dimensions, in this case notably the second, "social" rather than the first, economic dimension.

In addition, there are a number of items which follow a more complex pattern. The prime example is the "Political Authority: positive" category (which is one of the rightist items in the CMP's fixed "rile" scale). In the second dimension, its λ value is indeed positive, implying "socially" conservative positions. In the economic realm, however, the category is associated with the left political spectrum. Thinking about this for a moment, this result makes perfect sense, as political authority in the economic context corresponds to a more active role of the state in the economy. The extra analytical leverage we receive from the two-dimensional solution is also shown by the two categories referring to European integration. Pro-integration statements reflect positions that are economically rightist, and "socially" somewhat left. Negative remarks about European integration can result from economically leftist views, or from "socially" conservative ones.[19]

Next, we compare the CMP-based party positions on both dimensions to those from expert surveys as above (Figure 6, using the "taxes vs. spending" and "social policy/social lifestyle" party ratings. On the economic dimension, there is a quite strong positive association, reflected in a respectable $r = .73$. With regard to the social dimension, there is also a positive correspondence, although the observations are much more scattered around the regression line. Note, however, that we let a broad range of categories (rather than just a few related explicitly to morality issues) contribute to the second dimension, and that manifestos in general do not necessarily cover "social" issues as extensively as they do economic ones. This will make it harder to correctly place the parties on the second dimension.

[19]The uncertainty concerning the λ_{dj} values is smaller in the two-dimensional model. One reason is that we use a Poisson rather than a negative binomial likelihood here. In addition, there are also more constraints on the discrimination parameters λ_{dj}.

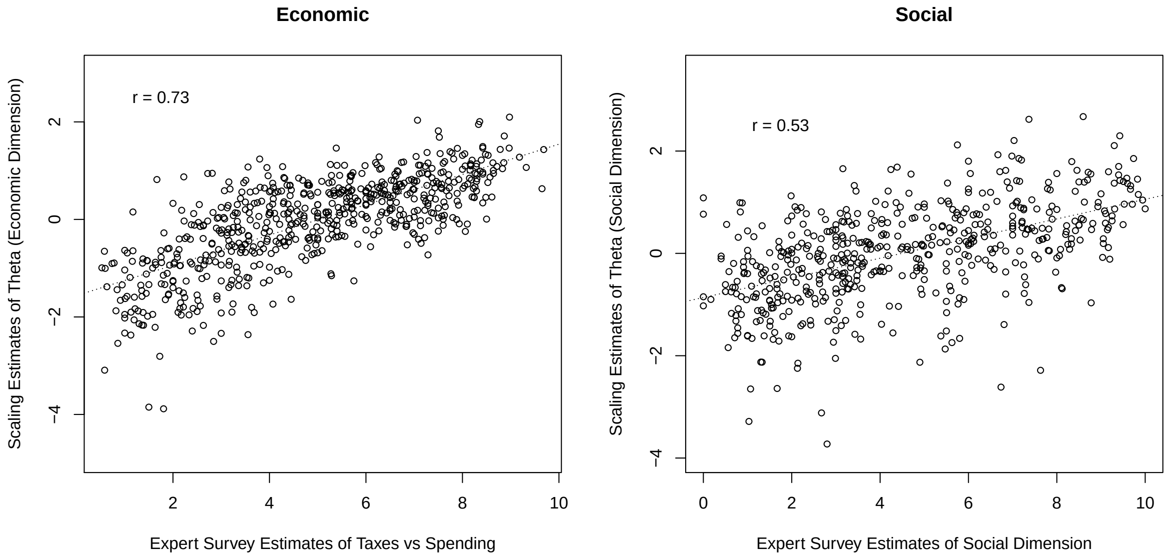


Figure 6: Correlations of two-dimensional model estimates for $\hat{\theta}_{\mathrm{econ}\ i}$ and $\hat{\theta}_{\mathrm{social}\ i}$ with expert survey estimates.

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.

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 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 the negative binomial scaling model to the 39 Belgian manifestos from 1991-2003 that were coded by the Belgian CAP team (2003 data include Flemish-speaking parties only). The obtained positions are shown in Figure 7, which readers familiar with Belgian politics will recognize as a spatial representation with high face validity. In three out of four elections, the radical right-wing Vlaams Belang is inferred to be the most rightist party. Only in 1995, the two liberal parties PRL and VLD are placed to the right of VB, which should be due to their economic rather than “social” positions. For all four elections, the Green parties (AGALEV and ECOLO) and the Socialist parties (PS and SP) are consistently located on the left side of the political spectrum, and the Christian-democratic parties (CVP/CD&V and PSC) can be found in the political center.

To complement these findings, Figure 8 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.[20] The results provide further evidence that the IRT model produces valid results also when applied to the CAP data. The retrieved positions correlate at $r = .83$ with the expert surveys, which is slightly higher than for CMP Rile and basically the same as for the IRT model applied to the CMP data. One advantage of the IRT model is of course that we can learn something about the items and thus about the content of political competition. First, it is interesting to note that even with the non-positional CAP data and allowing for item-level overdispersion, we find that 31% of the 137 items discriminate on the left-right dimension (judged on the basis of whether 90% of the posterior distribution of a λ_j are to the left or right of zero). The three most leftist items are “Environmental problems

[20]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.

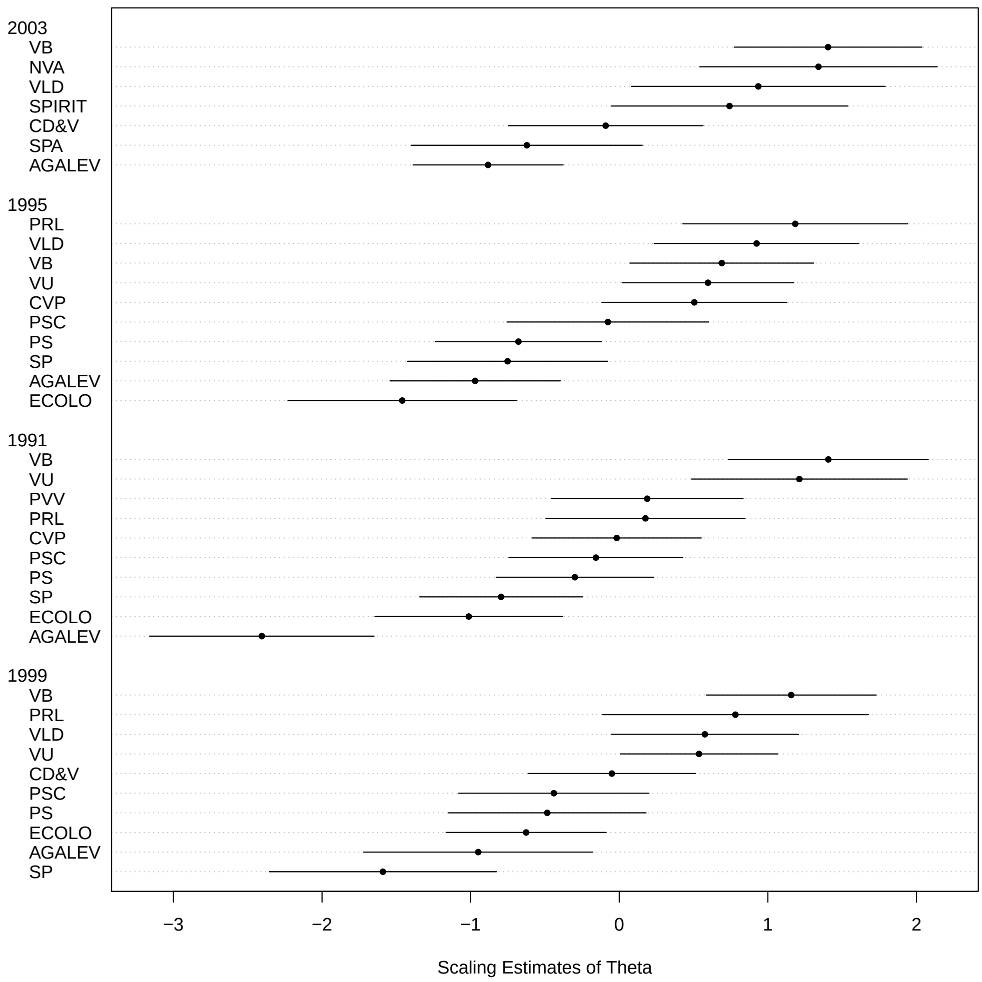


Figure 7: Left-right estimates for Belgian Parties from the Comparative Agendas Project dataset, 1999–2003.

with energy" ,"Forestry" and "Biohazard", and the three most rightist categories are "State Reform", "Migration" and "Asylum". This suggests that the category codings in the Belgian CAP represent issues that practically work as valence issues, where differences in emphasis are linked to parties' positions, and the scaling model picks up this information.

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.

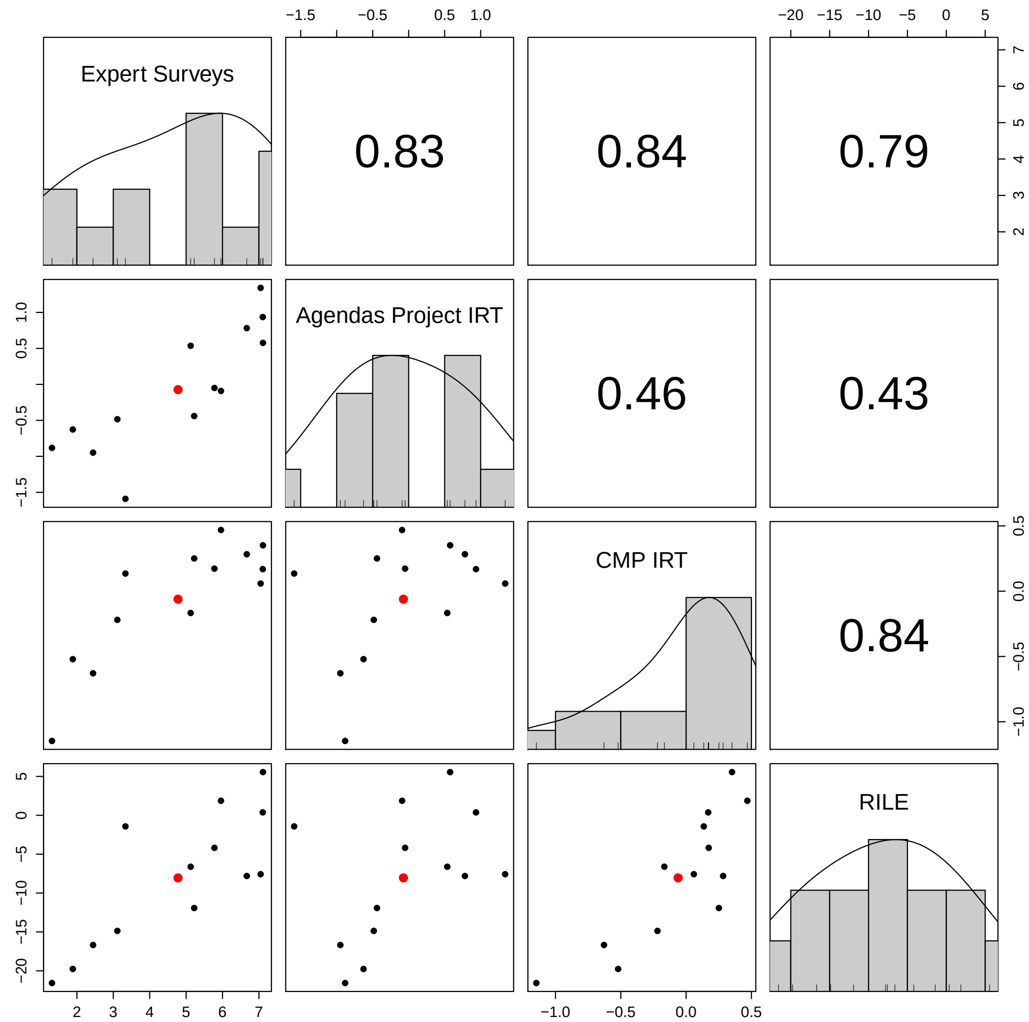


Figure 8: Model estimates of left-right policy positions from the Comparative Policy Agendas Project and the CMP, compared to expert survey results and Rile for 13 Belgian parties 1999–2003.

7 Conclusion

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 it permits more realistic modelling of the stochastic process that generates the observed counts of specific categories of statements as a nominal response framework. Our approach was to model counts as negative binomial, estimating an additional variance component for each category, providing a better fit and more realistic error estimates for the resulting political quantities. In addition, the flexibility of the Bayesian IRT approach permits a more complete model of the political process, including potentially the stochastic generation of overall manifesto length (represented by the α_i parameter), 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 a basic one-dimensional model applied to the full sample, a one-dimensional model fitted to sub-samples and have extended the model to the two-dimensional case. In future work we plan to extend the model, allowing for the possibility that item parameters vary across context, but are still inferred within one and the same model. This would be analogous to a multi-level IRT model, with random effects partitioning the items (see Fox, 2010, section 6.3) or random effects with a model of the covariance structures (see Curtis, 2010, Table 9). Using the flexibility of IRT scaling of these quantities by treating coded categories as items to which the manifesto statements respond, extensions of the model permit the direct modeling of a variety of interesting political questions without having to decide *a priori* how the manifesto content relates to the quantities being estimated.

References

Albright, Jeremy J. 2008. "Bayesian Estimates of Party Left-Right Scores.". Unpublished paper, Indiana University.

Baker, F. B. 1992. *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.

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 E. Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Vachudova. 2012*a*. "Measuring Party Positions in Europe: The Chapel Hill Expert Survey Trend File, 1999-2010." *Party Politics* .

Bakker, Ryan, Catherine De Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Anna Vachudova. 2012*b*. "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):495–513.

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

Bock, R Darrell. 1972. "Estimating item parameters and latent ability when responses are scored in two or more nominal categories." *Psychometrika* 37(1):29–51.

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.

Cameron, A Colin and Pravin K Trivedi. 1986. "Econometric models based on count data. Comparisons and applications of some estimators and tests." *Journal of Applied Econometrics* 1(1):29–53.

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

Curtis, S McKay. 2010. "BUGS Code for Item Response Theory." *Journal of Statistical Software* 36.

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* .

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

Elff, Martin. 2013. "A dynamic state-space model of coded political texts." *Political Analysis* 21(2):217–232.

Fox, J P. 2010. "Bayesian item response modeling: Theory and applications.".

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.

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.

Hoffman, Matthew D. and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15(Apr):1593–1623.

Hooghe, Liesbet, Ryan Bakker, Anna Brigevich, Catherine De Vries, Erica Edwards, Gary Marks, Jan Rovny, Marco Steenbergen and Milada Vachudova. 2010. "Reliability and validity of the 2002 and 2006 Chapel Hill expert surveys on party positioning." *European Journal of Political Research* 49(5):687–703.

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.

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.

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.

Lo, James, Sven-Oliver Proksch and Jonathan B Slapin. 2014. "Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos." *British Journal of Political Science* pp. 1–20.

Lowe, Will. 2013. "There's (basically) only one way to do it." Unpublished paper, University of Mannheim.

Lowe, Will. 2014. "Putting it all on the line. Measurement models for event data." Unpublished paper, University of Mannheim.

Lowe, William and Kenneth Benoit. 2013. "Validating Estimates of Latent Traits From Textual Data Using Human Judgment as a Benchmark." *Political Analysis* 21(3):298–313.

Maraun, Michael D and Natasha T Rossi. 2001. "The extra-factor phenomenon revisited: Unidimensional unfolding as quadratic factor analysis." *Applied Psychological Measurement* 25(1):77–87.

Mölder, Martin. 2013. "The validity of the RILE left-right index as a measure of party policy." *Party Politics* pp. 1–12.

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

Ostini, Remo and Michael L Nering. 2009. "Polytomous Item Response Theory Models." pp. 1–121.

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.

Rivers, Douglas. 2003. "Identification of multidimensional spatial voting models.". Unpublished manuscript, Stanford University.

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.

Stan Development Team. 2015. "Stan: A C++ Library for Probability and Sampling, Version 2.6.0.".
URL: *http://mc-stan.org/*

Steenbergen, Marco R and Gary Marks. 2007. "Evaluating expert judgments." *European Journal of Political Research* 46(3):347–366.

Stokes, Donald E. 1963. "Spatial Models of Party Competition." *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.

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.

Van Schuur, Wijbrandt H and Henk A Kiers. 1994. "Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what model to use instead." *Applied Psychological Measurement* 18(2):97–110.

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

Volkens, Andrea, Pola Lehmann, Nicolas Merz, Sven Regel and Annika Werner. 2014. "The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2014b.".

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.

A Equivalency to the Poisson regression model

Put in terms of Poisson regression, we can re-express α_i in Eq. 5 as $\log(t_i)$ to represent a variable exposure rate based on document length. Setting $t_i = e^{\alpha_i}$, Eq. 5 is equivalent to

$$\log(\mu_{ij}) = \log(t_i) + \zeta_j + \lambda_j \theta_i \tag{9}$$

$$\log(\mu_{ij}) - \log(t_i) = \zeta_j + \lambda_j \theta_i \tag{10}$$

$$\log\left(\frac{\mu_{ij}}{t_i}\right) = \zeta_j + \lambda_j \theta_i \tag{11}$$

B Equivalency to Slapin and Proksch's "wordfish" model.

Eq. 9 is equivalent to Slapin and Proksch (2008)'s unidimensional Poisson scaling model of document positions θ_i for a document-term matrix, expressed as:

$$\log(\mu_{ij}) = \alpha_i + \psi_j + \lambda_j \theta_i \tag{12}$$

These equivalencies are mapped in Table 3.

	Our IRT Model		Slapin and Proksch (2008)
Qty.	Interpretation	Qty.	Interpretation
θ_i	latent "ability"	θ_i	Ideological position
α_i	denominator for multinomial equivalence	α_i	Fixed document length effect
ζ_j	Conditional "difficulty" parameter	ψ_j	Fixed word effect
λ_j	Item discrimination parameter	β_j	Word sensitivity to position θ_i

Table 3: Equivalencies between IRT Model and "Wordfish"