Challenges in Tracking Parties' Positions in Multi-dimensional Policy Spaces by Manifesto Project data: A Dynamic Idealpoint Model and its Extensions.

Martin Elff

Zeppelin University, Friedrichshafen

`martin.elff@zu.de`

The compilation of coded manifestos by the Manifesto Project is an invaluable resource for the reconstruction of parties political positions and their changes over time. Nevertheless, the extraction of political positions from these data is confronted by some considerable challenges. First, the relation between positions and counts or percentages – as present in these data – is essentially non-linear. Second, Manifesto data may reflect both political positions and the salience of policy areas. Third, the policy space wherein parties take positions is not necessarily uni-dimensional. Fourth, positions taken by parties can be expected to be autocorrelated, since parties do not invent their positions in each election *ex nihilo*. Fifth, it is unlikely that even the Manifesto data is free of measurement error and and that the policy topic categories used may vary in terms of their discriminatory power for policy positions. The paper points out how these problems are addressed by recently developed models of the generating process of the Manifesto data. It further proposes some extensions to these models to address the remaining challenges.

1 Introduction

Political positions of parties play a central role in many areas of research in comparative politics. In contemporary research on the formation and duration of coalition governments, political positions of potential coalition member parties play a vital role (e.g. Laver and Hunt 1992). Also, the full potential of spatial models of voting can be realized only if measures of candidates' and parties' political positions are used that are independent of the voters' perceptions. Without such independent measures, spatial analyses of voting may fall victim to projection effects in voters' assessment of the positions of candidates and parties (Brody and Page 1972; Wilson and Gronke 2000; Merrill et al. 2001).

For the purpose of an independent reconstruction of parties' political positions various methods have been discussed in recent years. The reconstruction of political positions from political texts usually starts with the identification of semantic or grammatical units, such as words (Laver and Garry 2000; Laver et al. 2003; Slapin and Proksch 2008), sentences or "quasi-sentences" (Budge et al. 1987, 2001a; Volkens et al. 2010), and proceeds with classifying these units into politically relevant categories and concludes with counting the occurrence of these political categories. Word count approaches have gained some popularity during the last couple of years, owing to the recent availability of easy-to-use software for automatically generating such word counts. While some argue that word-count based reconstruction of political positions are at least competitive to sentence-classification based reconstructions (Laver and Garry 2000; Laver et al. 2003), the word-count approaches have not yet gained the same acceptance in the political science community as sentence-based approaches. One reason may be the wide availability of the data produced by the *Manifesto Project*, which are based on the categorization of quasi-sentences.

The contribution of the Manifesto Project (Budge et al. 1987, 2001a; Volkens et al. 2010, 2013) to the reconstruction of political positions from political texts, in particular party manifestos – documents published by or on behalf of parties on occasion of elections – can hardly be overstated. There is no source on Earth that rivals it in terms of temporal and geographical breadth when it comes to collecting, processing, and publishing such data. And a simple review of all the published literature that makes use of it may fill a whole volume. Nevertheless, the Manifesto Project data have not yet been brought to their full potential. They are most widely used in the condensed form of the RiLe index, which summarises the many policy topics uncovered, recorded, and coded by the Manifesto Project and its precursors. This condensation of the rich Manifesto Project data is motivated by the notion of an all-encompassing ideological left-right dimension, but there is more to find in manifestos and more to find in Manifesto Project data than that. Parties touch upon in their manifestos on a multitude of political topics that are relevant for a multitude of policy domains and it might well be that the positions parties take in their manifestos cannot be reduced to a single dimension, at least not without losing potentially valuable information.

The present paper discusses a recent attempt at bringing the data of the Manifesto Project to its full potential, the dynamic idealpoint model of manifestos (Elff 2013). The main purpose of this paper is to illustrate how this model can be used to reconstruct the positions of parties in specific policy domains, which may be multidimensional, how to describe the evolution of such positions, and how these positions can be used to distinguish party families. In addition, it indicates a couple of challenges that confront this dynamic idealpoint model and points to some ways to address these challenges in its further development. The paper is composed as follows: The next section sets out the basic ideas behind the development of the dynamic idealpoint model. It gives a theoretical justification for the use of manifestos as a base for the reconstruction of political position, resting on a notion of selective emphasis, although this notion may be at odds with some of the original ideas that guided some of the principle investigators of the Manifesto Project. This section also discusses the fundamental challenges in using the Manifesto Data, which have lead to the dynamic idealpoint model around which this paper is organised. The third section discusses some formal and technical aspects of the model. The fourth section illustrates the application of the dynamic idealpoint model to the reconstruction of parties position in two different policy domains, one unidimensional, the other two-dimensional, and to identifying the typical average positions of various party families. This section also discusses the use of such reconstructed political positions as independent variables in the analysis of patterns of voting behaviour. The fifth section discusses the challenges that confront the dynamic idealpoint model and possibilities of its further development. A sixth section discusses the problems that arise if the complexities that the idealpoint model addresses are disregarded. As usual, the paper is concluded by a brief summary.

2 The Ideas Behind the Dynamic Idealpoint Model

2.1 Rhetoric, Position and Selective Emphasis

When politicians make public speeches they usually claim what good, they have done for country and people or what good they plan to do, whether this means increasing prosperity, defending public security or upholding morality. Also, it has often been claimed that many voters do not so much care about where parties or candidate stand on issues, but rather how parties or candidates have performed in office. These observations or assumptions have given rise to the concept of *valence issues* (Stokes 1963) and to the "valency and saliency" theory of party competition (Robertson 1976; Budge and Farlie 1983a). This theory also is said to have guided the coding of electoral platform by the Manifesto Research Group and the Comparative Manifestos Projects. The view that this theory purports has been aptly summarized by Ian Budge (in Budge et al. 2001b: 82):

> *1. Party strategists see electors as overwhelmingly favoring one course of action*

on most issues. Hence all party programmes endorse the same position, with only minor exceptions.

2. *Party strategists also think that electors see one party as more likely that the others to carry through the favored course of action.*

3. *Hence each party has a set of issues that 'belong' to it, in the sense that the centrality of these issues in an election will increase its vote.*

4. *A party therefore emphasizes its 'own' issues in its election programme, in an attempt to increase the salience of these for voters. It emphasizes 'rival' issues less or not at all.*

5. *Policy differences between parties thus consist of contrasting emphases placed on different policy areas.*

This view also seems to have been corroborated by the fact that the electoral platforms coded by the Manifesto Research Group (MRG) and the Comparative Manifesto Project (CMP) mostly contain positive references to policy goals and hardly ever negetive references. Prima facie, assumptions such as these are necessary to justify the relatively low proportion of coding categories employed by the MRG/CMP that explicitly oppose a specific objective or state of affairs. Also, in Budge's view the "valency and saliency theory" is well corroborated both by results of analyses of the manifesto data and by results of other scholars (Budge et al. 2001b: 82–83). However, some critical reflections may lead to the conclusion that, on the one hand, these assumptions *contradict* rather than justify the use of MRG/CMP data to determine parties' genuine political or ideological positions — quite in contrast to the common use of these data to assign left-right positions to political parties. On the other hand, these assumptions are neither completely plausible nor are they necessary to justify the use of the MRG/CMP coding procedures for manifesto texts.

The first assumption explicitly states that all party platforms endorse the same position while the other assumptions state that differences between platforms of different parties come about by different nuances of this common position. Further, these nuances are mainly instrumental, so as to gain the most votes by emphasizing one's own strengths by virtue of "issue ownership." If these assumptions were true, then electoral platforms and party programmes might be used to explain and predict election results (Budge and Farlie 1983b), but using them to reconstruct genuine political/ideological positions of various political parties would be a hopeless endeavour, as there would not be any differences in the positions to begin with.

Unless one stretches the meaning of the concept of valence issue to a degree that its delimitation to the concept of position issue vanishes, one can hardly state that the most issues are valence issues without being challenged by numerous and prominent counterexamples. For example, both supporters and opponents of legalized abortion may frame their arguments as the positive affirmation of certain values, culminating in the phrases "pro-life" and "pro-choice", but rhetoric devices aside, the contrasting positions on issues like this are more than

just different emphases of different items from the same bundle of consensual values. Even if one grants that most parties, extremist parties aside, mention in their platforms subsets of the same set of basic values, this does not mean that they all endorse essentially the same position. Avoiding explicitly confrontative statements does not imply that different priorities among the same set of values does not preclude that these priorities express opposing ideological positions.

Manifesto Project data have not only be used to describe parties' appeals to groups of voters or to make predictions about their government policies in case they gain office, but also to describe the *ideological outlook*. The fact that the Manifesto Project and it precursors construct a scale that assigns parties' manifestos a general left-right scores – i.e. scores that are intended to represent their position on an overarching "ideological" left-right dimension – suggests that it is actually one of Manifest Project's intentions to reconstruct parties' ideologies. It is perhaps one of the many overlooked contributions of Anthony Downs to political science to have given a quite brief and nevertheless useful clarification of the concept of ideology. According to Downs, an ideology is "a verbal image of a good society and the chief means of constructing such a society" (Downs 1957: 96). This can be paraphrased as an ideology being a specification of the *fundamental aims* of politics and the *principles of policy* to reach these aims. Looking at the way that party manifestos are coded by the Manifesto Project, it is remarkable how well it fits with an interpretation of manifestos as an expression of ideology: The Manifesto Project codes manifestos according to the degree they emphasise a variety of policy goals or policy objectives, such as *Free Enterprise*[1] (per401) or *Social Justice* (per503), but also general policy principles, such as *Nationalization* (per413) or economic *Incentives* (per402). The ideology of a party will then be reflected in the selective emphasis of those policy objectives and policy principles that are coherent with its ideological position and the avoidance of policy objectives and policy principles that are incoherent with if not opposing its ideological position. For example, a socialist party will emphasise *Social Justice* and *Nationalization* rather than *Free Enterprise* or *Economic Orthodoxy* (per414), which are rather typical for the ideological positions of economically conservative parties. Of course, this is already reflected in the work of the Manifesto Project, in so far as in the construction of their RiLe index emphases of *Social Justice* and *Nationalization* are treated as indicators of "leftist" positions, while emphases of *Freedom of Enterprise* and *Economic Orthodoxy* are treated as indicators of "rightist" positions. The assumption of selective emphasis can be restated, for the purpose of formal or statistical modelling, as follows:

Assumption 1 *Other things being equal, the emphasis of a policy objective or policy principle i used to express the political position of party j in the manifesto published on occasion of an election at time t decreases monotonically with the distance $\Delta_{ijt} = ||\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt}||$ between the location $\boldsymbol{\alpha}_i$ of the policy objective/principle and the position $\boldsymbol{b}_{jt}$ of the party in the political space.*

---

[1]The labels used in Budge et al. (2001b) and Klingemann et al. (2006) are used here. The codebook that is currently available online uses other designations for the same variables.

The assumption explains why the model discussed in the paper is indeed an *idealpoint* model: The emphasis of policy objectives/principles in manifestos declines with the distance from the position of the parties, that is, the emphasis will be maximal if the objective/principle coincides with the position. This is analogous to idealpoint models of choice (e.g. voting) where an actor prefers choice options that are closer to her position in a political space to choice options that are farther away, and would prefer an option that coincides with her position (if such a choice were available) to all other options.

That political or ideological positions are reflected in the selective emphasis of certain policy objectives and policy principles does not mean that *all* of those objectives are positional in nature. The concept of "valence issues" or rather *valence objectives* is not without plausibility and again such valence objectives can be found in the Manifesto Project coding scheme. Among such valence objectives may be general *Economic Goals* (per408), *Productivity* (per410), *Freedom and Human Rights* (per201) and *Democracy* (per201) even though the latter two seem to be used more often by left-libertarian parties (see below). Whether such valence objectives can be used in the construction of scales of political positions is an open question. If such valence objectives can be "owned" in the sense of issue-ownership and if issue-ownership derives from past successes in government, there is little reason to expect that the emphasis of these objectives has any systematic relation with the political/ideological position of parties. E.g after the Great Depression the US Democrats could perhaps claim "ownership" of the "valence issue" economic growth, while in Germany it is often believed that the CDU/CSU has a claim to ownership of this issue. Yet one can even doubt whether the ownership of a "valence issue" by another party will lead a party to avoid it. E.g. if the British Conservative might claim ownership of the economic growth issue, Labour might not really afford to ignore it in its manifestos but may as well try to make up for the lack of ownership by emphasising that they care for economic growth as least as much as the Conservatives. On the other hand, if the valence objectives are truly consensual, than they should be emphasised more or less by all parties independent from their position. What then could drive variation in their emphasis might be the urgency of social, economic or political problems connected to this goal. In other words, the emphasis of valence objective would indeed reflect the *salience* of the policy area which they belong to: Economic growth then is likely to be emphasised in times of recession. The fact that some of the policy objectives mentioned in manifestos can be considered as valence objectives poses some challenges for the model developed here and is not yet reflected in its present form. These challenges are discussed later in this paper.

2.2 Positions within and Salience of Policy Domains

While the argumentation of the previous section indicates that different political and ideological positions will be expressed in different emphases of different political objectives, the emphasis of political objectives may also vary other grounds. The *salience* of more general

areas of policy may vary over time (Laver 2001): For example, while in times of prosperity economic objectives may be important for the political discussion and have less appeal to voters, conflicts in the field of economic policy may heighten in times of recession, massive state deficits or inflation. The distinction between the emphasis of a policy objective as a means to espouse a political-ideological position and its emphasis because of the salience of a general range of policy problems suggests, that the policy objectives considered by the Manifesto Project data can be grouped into *policy domains*. This idea can be formalised by the following assumption:

Assumption 2 *Other things being equal, the emphasis by party j at time t of a policy objective or policy principle i that belongs to the policy domain $\mathcal{D}$ (i.e. $i \in \mathcal{D}$) increases monotonically with the salience of the policy domain at time t.*

The relation between political positions, the salience of areas of policy, and the emphases of policy goals in parties' platforms can best be explained by way of a simplified example. Consider the following four political goals: (1) taking the economy under government control (*Controlled Economy*, `per412`), (2) protecting the freedom of business enterprises (*Free Enterprize*, `per401`), (3) protecting the freedoms, citizen rights and human rights of individuals from infringements by agencies of state authority, such as the police, (*Freedom and Human Rights*, `per201`) and (4) upholding "law and order", that is the authority of state institutions (*Law and Order*, `per605`). On the one hand, all these three policy goals are positively formulated, that is, they are in favour of something. On the other hand, these policy goals seem to reflect contrasting political or ideological positions. The first goal, taking control of the economy, would usually be perceived as a "socialist", hence "leftist" goal. The second goal, referring to the freedom of enterprise, would however be usually viewed as a "pro-market", hence "rightist" goal. Likewise, the protection of individuals' rights and freedoms and upholding law and order are usually viewed as "liberal" (in the American sense) or "leftist" and "conservative" or "rightist" goals, respectively. While each of these four goals can be seen as being "leftist" or "rightist" in some respect, it should be stressed here that they refer to different aspects of society that are the object of policy, to different *policy domains*. The first to refer to economics while the second to refer to aspects of domestic policy. The crucial point of this example is that there are no cogent reasons *a priori* to believe that "leftist" positions with respect to economic policy and "liberal" positions with respect to domestic policy coincide. In fact, classical Liberalism combines a strong "pro-market" attitude with a strong emphasis on the protection of individuals from intrusion into their liberty by state authority either in the name of law and order or in the name of public morality. Also one could argue that East European communism combines a socialist or interventionist views with an emphasis of law-and-order guaranteed by a strong, if not authoritarian, state.

Now if disagreements between parties on economic policy become more salient, e.g. because of economic crises, one could expect that they will give more emphasis on the first two

of the four goals in the example. "Centrist" parties might try to strike a balance between the first two, emphasising both of them. "Leftist" parties might give more emphasis to the first goal, while "rightist" might give more emphasis to the second goal. As a consequence, "leftist" parties and "rightist" parties will seem to be more polarised, while in fact they are just giving more attention to economic policy, without necessarily having changed their position. Supposed that there is a trade-off between the emphases of different policy topics because there are limits to the attention parties can give to various policy domains and to the size of manifestos, then more emphasis is given to the economic policy topics will lead to less emphasis being given to the domestic policy topics. As a consequence, while the polarisation about economic policy *seems* to increase it also seems to decrease about domestic policy. As long as "leftist" positions coincide in the two policy areas and "rightist" positions coincide in the two policy areas, any general assignment of "left-right" scores will not be much affected by this. But if positions in the first policy area can be opposite to positions in the second policy area, then left-right scores will be affected by changes in the salience of economic policy relative to domestic policy and vice versa.

The possibility that policy topics may belong two different policy domains raises a couple of questions: First, how can it be established whether a policy topic belongs to the same or a different policy domain than another policy topic? Second, how can it be established whether a policy topic is opposite to another policy topic within a particular policy domain? Third, how can it be established whether parties' positions within two policy domains tend to coincide or vary independent from another? Principle Components Analysis or related multivariate analysis may seem to give an answer to these questions. Indeed such methods have been used by various authors to answer not exactly these but somewhat related questions. Gabel and Huber (2000) use Principal Factors Analysis (a method closely related to PCA) in the attempt to show that, however diverse the policy domains that policy topics belong to, parties positions can be aligned mainly along a single dimension. Bartolini and Mair (1990) have used PCA to reconstruct parties' positions with respect to the field of economic policy. However, PCA does not seem to work well for this purpose. For example, if one conducts a PCA of variables from which the RiLe scores in a recent edition of the Manifesto Project data set (Volkens et al. 2010) are constructed, one will find that as many as 20 principal components are "extracted" by the usual Kaiser criterion of an eigenvalue of at least unit size, with the first principal component capturing just 5.07 per cent of the total variance of the principal components. There three problems that undermine the usefulness of PCA for finding an answer of those three questions.

The first problem is directly related to the nature of the questions at hand: PCA will group together policy topics that have a positive correlation with one another and will assign policy topics to opposite poles of a principle component that have a negative correlation with one another. Now if two policy topics belong to the same policy domain and the salience of this policy domain varies considerably across countries and time points then this will shift the

correlation between these topics into a positive direction, other things being equal, while the correlation between two topics that belong to different policy field will be shifted towards a negative correlation. As a consequence, the loadings of policy topics not only reflect their location within particular policy domains but also how the saliency of these policy domains varies.

The second problem is that a policy domain does not need to be unidimensional. That formal spatial models of of political competition often rest on the assumption of uni-dimensionality is that multidimensional spaces of political competition do not lead to the equilibria format theorists seek to establish (e.g. McKelvey 1976; Plott 1967). Yet the theoretical inconvenience of multidimensionality is not an argument against its empirical existence. Rather it should be regarded as an empirical question whether a particular policy domain is unidimensional or multidimensional. But if one admits policy domains to be multidimensional, PCA or similar models may not be of much help in deciding whether two policy topics belong to the same policy domains or two different ones. Consider the case that a PCA of a set of policy topics supports a five-dimensional solution. Then this leads to several interpretations: One interpretation would be of five different policy domains, each being unidimensional and no systematic relation between the saliences of the individual policy domains. Another interpretation would be that of two two-dimensional policy domains with a strong trade-off in terms of the saliences of the two policy domains. And yet another interpretation would be that of two policy domains, one uni-dimensional, the other three-dimensional, and again a strong trade-off in terms of salience. It seems hard to decide between these interpretation without additional information. Nevertheless, this additional information may exist in so far there may be substantial (but not statistical) grounds on which one can assign policy topics to policy domains. In the example given above, it has been quite obvious that *Controlled Economy* and *Free Enterprise* belong to the field of economic policy, while *Freedom and Human Rights* and *Law and Order* belong to the field of domestic policy.

The third problem is of a more technical nature: Emphases of policy topics in manifestos are strictly non-negative and, if the emphasis is conceived of the proportion of a manifesto given to it, are bounded from above. E.g. in the Manifesto Project data set, the emphasis of policy topics is recorded as the percentage of quasi-sentences that refer to a particular policy topic relative to the total number of quasi-sentences in the manifesto. As a consequence, unless the variation of the emphases is low – which is not what one would hope for if one uses e.g. Manifesto Project data – and is clustered near 50 per cent – which is unlikely if the number of policy topics is sufficiently large – the statistical association between emphases of policy topics will be *non-linear* and therefore is not well represented by correlation coefficients.

That this is not a mere technicality or methodological idiosyncrasy can demonstrated by an example with real data from the Manifesto Project. Consider the two Manifesto Project variables *Controlled Economy* (per412) and *Free Enterprise* (per401). Since these two variables represent policy objectives that directly contradict one another, if a linear correlation would

Figure 1: Percentages of sentences emphasizing the policy objectives "Controlled economy" and "Freedom of enterprise" in electoral platforms of parties from 55 countries, 1945-2008.
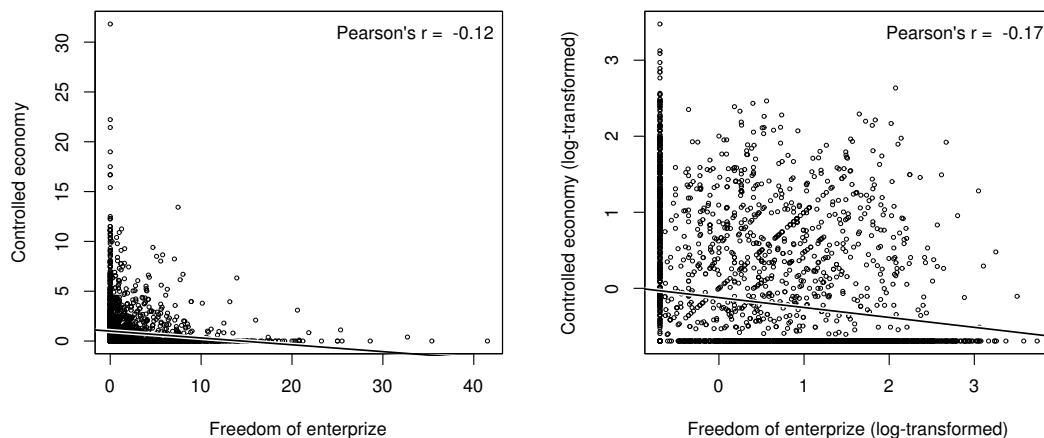
Data source: The Manifesto Project (Volkens et al. 2010). In the right-hand panel, the percentages p are transformed into $\ln(p + 0.5)$.

be an appropriate measure of the relation of the emphases of these topics it could be strongly negative. But in fact, the correlation between these two variables is very weak only -0.12. Judging from the size of the correlation alone, one could be led to the conclusion that the two policy topics are only weakly opposed to one another, despite the obvious contradiction among them. However, as can be seen in the left-hand panel of Figure 1, which contains a scatterplot of the emphasis percentages of these two policy topics, there does exist a strong negative relation between the emphasis of the two policy topics: Whenever the emphasis percentage of one of the policy topics tends to be large the emphasis percentage of the other policy topic is close to zero. What attenuates the correlation coefficient between the two is that the relation is intrinsically non-linear and therefore not well presented by a linear relation, which is demonstrated by the best-fitting regression line included in the scatter plot. Given that the relation among the Manifest Project variables are intrinsically non-linear it is hardly surprising when Principle Components Analysis does give little useful guidance in the construction of scales for political positions.

In view of this non-linearity, Elff (2009) and Lowe et al. (2011) propose to base analyses on log-transformed percentages. However, such a transformation of the data does not help much to expose the contrast between policy objectives in terms of substantive policy content. This is illustrated by the right-hand panel of Figure 1, which shows a scatterplot of log-transformed percentages of references to the policy topics *Controlled Economy* (`per412`) and *Free Enterprise* (`per401`). While the distribution of the dots looks less skewed than in the left-hand panel, the correlation between the log-transformed percentages is only very slightly

larger in absolute size ($-.17$) than the correlation between the raw percentages ($-.12$). Also, a principal components analysis of such log-transformed percentages results in the extraction of 17 principal components (instead of 20 with the untransformed percentages), with the first principal component capturing 7.48 of the total variance.

That the log-transformation does not solve the problem of non-linearity to a degree that is fully satisfactory may have two explanations: Firstly, the log-transformation of percentages that are computed from raw counts does not really eliminate the consequences of the percentages being bounded from below. In the Manifesto Project data, the variables contain many zeros, which will lead to negatively infinite values if the logarithm is applied directly. Such infinite values can be avoided by adding a small quantity to the percentages before taking logarithms, but that will mean that the logarithms of the percentages are bounded from below by the logarithm of this added quantity. Secondly, the relations between the emphases of policy topics may still be non-linear after a monotonic transformation such as taking logarithms. Consider for example a policy topic that expresses "centrists" positions between the "leftist" and the "rightist" positions in a policy space. Such a centrist topic will be increasingly be emphasised as the position of a party is moved from the "left" end of the policy space to its "centre", but it will be less and less be emphasised as the position is moved further from the "centre" to the "right".

The problem of zeros in the emphasis counts and percentages can be addressed by moving from an ad-hoc adjustment to a genuine probability model, in which the *expected value* (in the stochastic sense) of an emphasis is a non-linear function of the (squared) distance between the location of a policy topic and a party's position in a policy space. Such an expected value may vary smoothly between zero and arbitrary large bounds, without ever reaching zero exactly. Nevertheless, the smaller the expected value of the observed emphasis count or observed emphasis percentage the higher the probability that its observed values is zero. The formal specification of such a probability model as well as ways of estimating its parameters are discussed in the next section.

3 Technical Aspects of the Dynamic Idealpoint Model: Formal Specification and Estimation

3.1 A Probabilistic Link between Position and Relative Emphasis

The previous section spelled out the substantial ideas, assumptions, and problems that led to the dynamic idealpoint model of party manifestos. In this section its more technical aspects will be discussed. The first step in this is the specification of the probability model that links parties' positions with the emphasis of policy topics in their manifestos. It is given by the following assumption:

Assumption 3 *Let m_{ijt} denote the number of times policy topic i is emphasized in the political text that political actor j has published on occasion t, then m_{ijt} is the realization of a random variable M_{ijt} with Poisson distribution that has a mean parameter μ_{ijt} given by*

$$\mathrm{E}(M_{ijt}) = \mu_{ijt} = \mathrm{e}^{v_{jt}}\,\mathrm{e}^{-\frac{1}{2}\Delta_{ijt}^2} \Leftrightarrow \ln\mu_{ijt} = v_{jt} - \frac{1}{2}\Delta_{ijt}^2 = v_{jt} - \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt})'(\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt}) \qquad (1)$$

where $\mathrm{e}^{v_{jt}}$ reflects the saliency of the policy domain in the manifesto and $\mathrm{e}^{-\frac{1}{2}\Delta_{ijt}^2}$, reflects the relative emphasis given to policy topic i in the manifesto.

In the previous section, a reason were already given for why such a probability model might be preferable to a simple transformation of emphasis counts or emphasis percentages as they can be found in Manifest Project data. With the above formulation this reason can be stated more precisely: If the number of times a policy topic is emphasised follows a Poisson distribution then the probability that a topic is not emphasised at all, that is, that $M_{ijt} = 0$, is

$$\Pr(M_{ijt} = 0) = \frac{\mu_{ijt}^0\,\mathrm{e}^{-\mu_{ijt}}}{0!} = \exp\left(-\mathrm{e}^{v_{jt}}\,\mathrm{e}^{-\frac{1}{2}\Delta_{ijt}^2}\right)$$

so that $\Pr(M_{ijt} = 0) \to 1$ as $\Delta_{ijt} \to \infty$, but still for finite distances $\Delta_{ijt} > 0$, $\Pr(M_{ijt}) > 0$. This means one does not need to conclude from $m_{ijt} = 0$ that $\Delta_{ijt} = \infty$, which would have been the case with the simple logarithmic transformation of the observed data $\hat{\Delta}_{ijt} = \ln(m_{ijt})$. Another advantage of a probability model is that it allows to account for classification error (Benoit et al. 2009), since it does not require that the observed counts are identical to the expected emphasis count implied by the position of a party and the location of the policy topic.

Under the assumption of a Poisson distribution, the positions of the parties taken in their manifestos and the locations of the policy topics in the policy domains can be reconstructed independent from the salience of a policy domain, if one conditions on the number of all emphases of policy topics that belong to a particular policy domain: Let $M_{1jt}, ldots, M_{Ijt}$ be all emphasis counts of the policy topics that belong to the same policy domain then the sum $M_{1jt} + \cdots + M_{Ijt} = M_{+jt}$ has a Poisson distribution with mean parameter $\mu_{+jt} = \mu_{1jt} + \cdots + \mu_{Ijt}$ and consequently the joint distribution of $M_{1jt}, ldots, M_{Ijt}$ has, conditional on $M_{jt} = m_{+jt}$ a multinomial distribution (Agresti 2002: 8f):

$$\Pr(M_{1jt} = m_{1jt}, \ldots, M_{Ijt} = m_{Ijt} | N_{jt} = n_{jt})$$

$$= \frac{m_{+jt}!}{\mu_{+jt}^{m_{+jt}}\,\mathrm{e}^{-\mu_{+jt}}} \prod_i \frac{\mu_{ijt}^{m_{ijt}}\,\mathrm{e}^{-\mu_{ijt}}}{m_{ijt}!} = \frac{m_{+jt}!}{\prod_i m_{ijt}!} \prod_i \left(\frac{\mu_{ijt}}{\mu_{+jt}}\right)^{m_{ijt}} \qquad (2)$$

$$= \frac{m_{+jt}!}{m_{1jt}! \cdots m_{Ijt}!}\pi_{1jt}^{m_{1jt}} \cdots \pi_{Ijt}^{m_{Ijt}},$$

Figure 2: The Relation between the positions of political actors, the location of policy objectives, and the emphasis of policy objectives in political texts, if the salience of the corresponding policy domain is held constant.
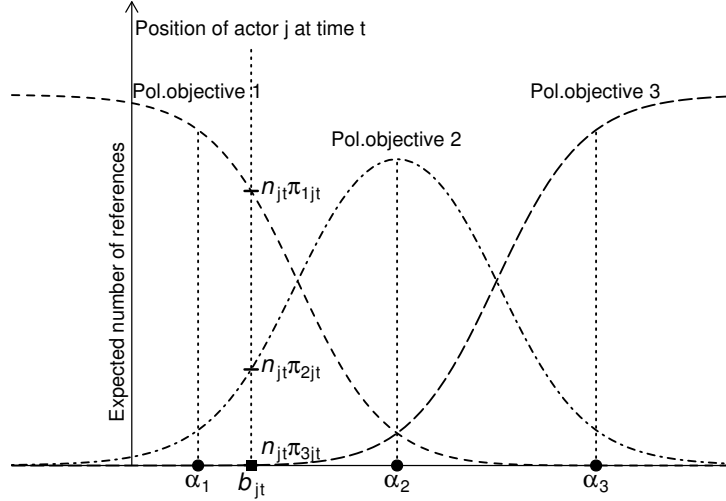
where

$$
\pi_{ijt} = \frac{\mu_{ijt}}{\mu_{+jt}} = \frac{\mu_{ijt}}{\sum_h \mu_{hjt}} = \frac{\exp\left(v_{jt} - \frac{1}{2}\Delta_{ijt}^2\right)}{\sum_h \exp\left(v_{jt} - \frac{1}{2}\Delta_{hjt}^2\right)} = \frac{\exp\left(-\frac{1}{2}\Delta_{ijt}^2\right)}{\sum_h \exp\left(-\frac{1}{2}\Delta_{hjt}^2\right)} \tag{3}
$$

Figure 2 illustrates how the expected number of emphases of policy topics is related to the position taken by a political actor in in a one-dimensional space of a policy domain, where the policy field contains three policy objectives. The horizontal axis represents locations in the space, where the three policy topics with locations α_1, α_2, and α_3 are represented by round dots on the axis. The variously dashed curves indicate how the expected value of the emphasis of each of the policy objectives varies with possible positions of a political actor if the salience of the policy space in question is held constant. The figure also contains the political position taken by an example actor j at time point t. This position b_{jt} is indicated on the horizontal axis by a square dot. Above b_{jt} the expected values μ_{1jt}, μ_{2jt} and μ_{3jt} of emphases of the policy topics are shown. Since the position of actor j at t is closer to the location α_1 of the first objective than to the location α_2 of the second topic, the the expected emphasis μ_{1jt} of the first topic is higher than the expected emphasis μ_{2jt} of the second topic. The expected emphasis of the third topic is close to zero, because the actor's position is quite far from the location of the third topic.
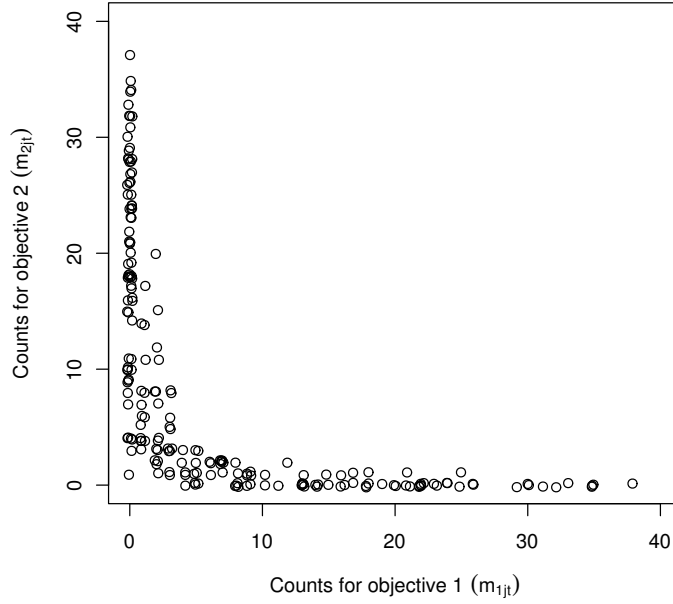
Figure 3: Emphases m_{1jt} and m_{2jt} of two hypothetical policy objectives in poltical texts if political actors' positions $\boldsymbol{b}_{jt}$ in a uni-dimensional political space have a standard normal distribution and if the policy objectives are located at $\boldsymbol{\alpha}_1 = -1.5$ and $\boldsymbol{\alpha}_2 = 1.5$.

As another illustration of the implications of a spatial model of political texts as discussed in this paper, Figure 3 shows how observed counts look like if one has two policy objectives located at points −1.5 and 1.5 in a one-dimensional political space and if policy positions of political actors have a standard normal distribution. Each dot represents a simulated emphasis count of the first and the second objective in a political text of one of the political actors. The horizontal coordinate of a dot represents a Poisson-distributed count with the expected value determined by the distance between an actors's position and the location of the first objective (displaced by some jitter), while the vertical coordinate represents a Poisson-distributed count the expected value of which is determined by the distance actor's position and the location of the second objective (again displaced by some jitter).[2]

The pattern of the bivariate distribution of these simulated counts exhibits a similar non-linear pattern as observed in the left panel of Figure 1. Many of the dots are either close to the horizontal or to the vertical axis. That is, in many cases the emphasis counts of the first objective are zero while the emphasis counts of the second objective are not and vice versa. Of course, the depiction of such simulated data generated from a hypothetical configuration

---

[2]Because counts are always integer, counts with similar expected values are likely to be identical, that is, several actors with positions close to one another may have identical emphasis counts if the salience of the political space is held constant. In order to make such multiply identical counts distinctive in the diagram the coordinates of the counts are displaced by jittering.

conforming to the model presented in this paper cannot prove that the model is correct. It can only make plausible that such a model is applicable to the Manifesto Project data. The goodness of fit of such a model is of course an empirical question.

3.2 Dynamics

A drawback of any method that try to reconstruct the positions that a party j takes at each time point t individually is that it disregards the information contained in the sequence in which parties' positions evolve. The positions of parties typically are not created *ex nihilo* on occasion of each election, but gradually evolve over time. The reason is that parties cannot abandon previously proposed objectives completely, without losing all their credibility.

If the positions of parties are treated as parameters of a probability model, trying to estimate these positions individually will lead to an *incidental parameter* problem: Each new manifesto with new sets of emphases of policy topics will lead to another set of parameters. Estimation in such a situation is inefficient at best and inconsistent (in the sense of asymptotic statistics) at worst. The solution of such an incidental parameter problem – already established in the context of multilevel modelling – is to reformulate it into a model involving latent variables with a probability distribution the parameters of which are to be estimated instead of incidental parameters. Treating the positions of a party j taken in their manifestos published at successive points in time $t = 0, \ldots, T_j$ as a latent variable in addition allows to model the over-time evolution of political positions as well as formulating and testing hypotheses about the distribution of these positions. A simple way in which political positions of a party can be modelled is this: The position of a party at time t, on occasion of an election, is a random step from its position on occasion of the previous election at time point $t - 1$.

Apart from more technical reasons just mentioned, treating parties' positions as latent data evolving in a stochastic manner has the additional advantage of opening up the possibility to "fill in" parties' positions for elections where no manifesto data is available: Suppose that no manifesto data are available at time point t, so that the corresponding emphases of policy topics are to be regarded as missing, the knowledge about the dynamic of the evolution of political positions can be used to predict the position at t from the manifesto data available from the previous time points form 0 to $t - 1$ and the following time points from $t + 1$ to T_j.

Based on the reasons just given, the following assumptions form also form the base of the dynamic idealpoint model discussed here:[3]

Assumption 4 *A political actor's position b_{j0} at time $t = 0$ has a normal distribution with zero mean and variance matrix Σ_0, or*

$$b_{j0} \sim \mathcal{N}(0, \Sigma_0) \tag{4}$$

---

[3]These assumptions differ slightly from Elff (2013), where the positions may have a non-zero mean, while linear constraints are imposed on the location parameters α_i. By lifting the constraint on the location parameters and setting the mean parameter of the positions to zero, both sets of assumption lead to observationally equivalent results.

Assumption 5 Conditional *on the previous time $t-1$ the position $\boldsymbol{b}_{jt}$ of a political actor j at time t has a normal distribution with mean $\boldsymbol{\Gamma b}_{jt-1}$ and variance matrix Σ_1, where $\boldsymbol{\Gamma}$ is a matrix of auto-regression slopes. That is, the positions of the political actors take a vector auto-regression (VAR(1)) form:*

$$\boldsymbol{b}_{jt} = \boldsymbol{\Gamma b}_{jt-1} + \boldsymbol{e}_{tj}, \qquad \boldsymbol{e}_{tj} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_1) \tag{5}$$

3.3 Estimation

If the party positions $\boldsymbol{b}_{jt}$ were known, the parameters $\boldsymbol{\theta}$ (the positions of the objectives $\boldsymbol{\alpha}$, the variance components Σ_0 and Σ_1 and the auto-regression coefficients $\boldsymbol{\Gamma}$ of the positions) of the model could be estimated by maximizing the likelihood function $\mathcal{L}_c = \prod_j \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j)$ or rather the log-likelihood function $\ell_c = \sum_j \ell(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j) = \sum_j \ln \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j)$ for $\boldsymbol{\theta}$, where $\boldsymbol{m}_j$ and $\boldsymbol{b}_j$ are constructed by "stacking" the $\boldsymbol{m}_{jt}$ and $\boldsymbol{b}_{jt}$, e.g. $\boldsymbol{b}_j' = (\boldsymbol{b}_{j1}', \dots, \boldsymbol{b}_{jT_j}')$.

However since the party positions are unobserved, maximum likelihood estimates are obtained by maximizing the *marginal* likelihood, or rather the marginal log-likelihood. Both are obtained by "integrating out" the unobserved data:

$$\mathcal{L} = \prod_j \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j) = \prod_j \int \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j) \partial \boldsymbol{b}_j, \tag{6}$$

$$\ell = \sum_j \ell(\boldsymbol{\theta}; \boldsymbol{m}_j) = \sum_j \ln \int \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j) \partial \boldsymbol{b}_j. \tag{7}$$

The ML estimation equation therefore is

$$0 = \frac{\partial \ell}{\partial \boldsymbol{\theta}} = \sum_j \int \frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j)}{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{m}_j)} \frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j)}{\partial \boldsymbol{\theta}} \partial \boldsymbol{b}_j = \sum_j \int \mathcal{P}(\boldsymbol{b}_j | \boldsymbol{m}_j; \boldsymbol{\theta}) \frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j)}{\partial \boldsymbol{\theta}} \partial \boldsymbol{b}_j$$

$$= \mathrm{E}\left(\frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j)}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{m}_j \right)$$

which thus suggests an EM algorithm for its solution (Dempster et al. 1977).

The problem with the involved integrals is that they have no closed-form solution and therefore need to be computed numerically. Elff (2013) uses a combination of a Monte-Carlo method for the computation of the integrals and an EM algorithm to maximize the marginal likelihood. While the details of this method shall not be discussed here, as they are spelled out in the online appendix to the paper just cited, later in this paper the challenges of this method for its "every day" usage will be discussed.

Based on an MLE $\hat{\boldsymbol{\theta}}$ *predictions* of the unobserved party positions can be obtained from the posterior distribution

$$\mathcal{P}(\boldsymbol{b}_j | \boldsymbol{m}_j; \hat{\boldsymbol{\theta}}) = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}; \boldsymbol{m}_j, \boldsymbol{b}_j)}{\mathcal{L}(\hat{\boldsymbol{\theta}}; \boldsymbol{m}_j)} = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}; \boldsymbol{m}_j, \boldsymbol{b}_j)}{\int \mathcal{L}(\hat{\boldsymbol{\theta}}; \boldsymbol{m}_j, \boldsymbol{b}_j) \partial \boldsymbol{b}_j}.$$

Posterior means as "point predictions" (with credibility intervals) as well as sample values from the posterior distribution for positions of parties in an economic policy space and a non-material policy space (which are discussed further below) can be obtained from the author's website.

# 4 Application of Reconstructed Positions

4.1 An Illustration

According to the dynamic idealpoint model discussed in this paper and introduced by Elff (2013). parties take a position in the policy space by virtue of the location of their electoral platform relative to the policy objectives that characterize the policy space. The closer the position of a party is to a particular policy objective, other things being equal, the higher is the expected value of the number of quasi-sentences that refer that objective relative to the references of the other objectives that belong to the same policy domain. In addition, the model allows the positions of the parties to change from election to election, so that the positions that they take with their electoral platforms form an auto-regressive time series for each party. It should be noted that the model allows for several policy domains covered in an electoral platform, so that parties pick positions in several policy spaces at the same time. The emphasis given to a policy objective then not only reflects the position of a party in the respective policy domain, which is expressed in the emphasis *relative* to the other policy objectives in the same domain, but also the salience of the policy domain as a whole. For example, if the domain of economic policy becomes more salient because economic policies get more pressing, a party will give more room to this domain in its electoral platform whatever its position is in this domain (Laver 2001).

In the following, the results that can be achieved by the method discussed in this paper concern two policy domains, the domain of economic policies and the domain of social/moral policies. The following Manifesto variables are considered as representing the domain of economic policy: *Nationalization* (per413), *Controlled Economy* (per412), *Economic Planning* (per404), *Market Regulation* (per403), *Incentives* (per402), *Economic Orthodoxy* (per414), and *Free Enterprise* (per401). The left-hand diagram in Figure 4 gives a graphical representation of the estimated locations of the policy objectives in this policy space. For the space of social/moral policies the following Manifesto variables are used: *Democracy* (per202), *Freedom and Human Rights* (per201), *Law and Order* (per605), *National Way of Life Positive* (per601), *Traditional Morals Positive* (per603), and *Traditional Morals Negative* (per604). This policy space is assumed to be two-dimensional and its axes are interpreted as a "authoritarian/libertarian" and a "traditionalist/permissive" one. The locations of the policy objectives in this space are illustrated by the right-hand diagram in Figure 4.

As there is no algorithm to automatically sort Manifesto policy topics into policy domains,
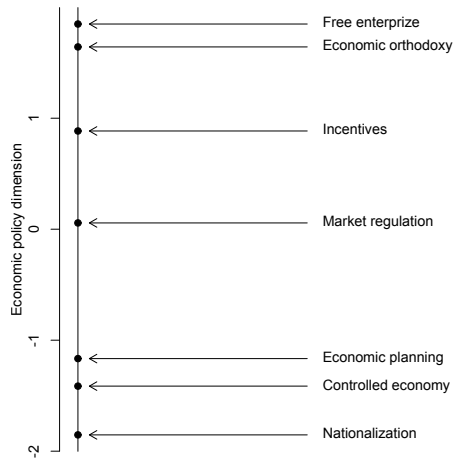
(a) Economic policy space
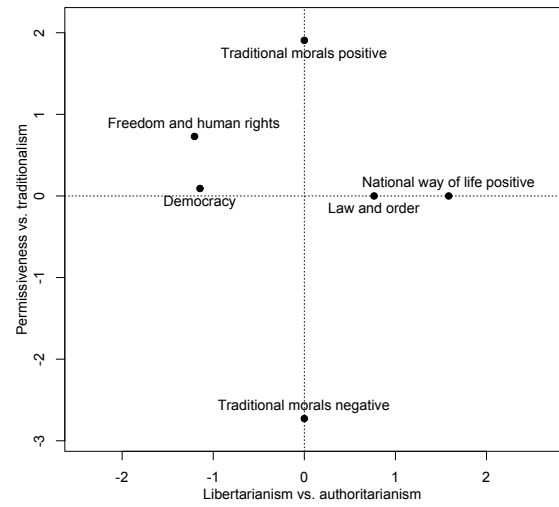
(b) Social/moral policy space

Figure 4: Location of policy objectives in policy spaces. The numerical estimates are available in Elff (2013).

the categories to represent a particular policy domain have to be carefully chosen. In the present case, the selection of policy goals is based on a trial-and-error process where various choices of policy goals were screened for giving interpretable results. This screening used a simplified method that was already used in Elff (2009) with some success for the explanation of the role of parties positions for the strength of the relation between class, religiosity and voting. Similarly it has to suffice as a justification of the choice of the policy objectives, whether or not they lead to meaningful results in terms of allowing to distinguish between party families.

Once the parameters of the latent state-space model are estimated one can obtain predictions about the positions that parties take with their electoral platforms in the form of empirical Bayes posterior distributions. The role of point estimators of these positions can then be played by the means of these posteriors. As an illustration, Figure 5 shows the posterior distributions of the positions of the Labour Party and the Conservative Party of the UK in 1964.
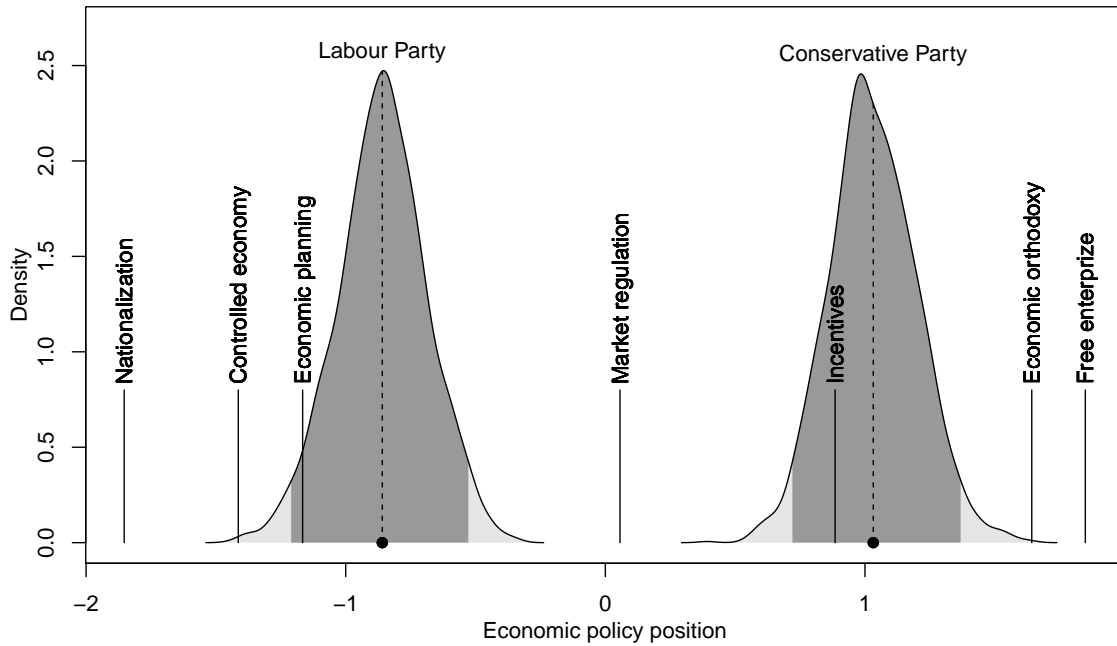
Figure 5: Positions of the Labour Party and the Conservative Party of the UK in 1964: Empirical Bayes posterior distributions. The lines and light gray areas represent the posterior densities, the dark gray areas represent 95 per cent credibility intervals and the dots represent posterior means. The means and posterior densities are estimated by kernel methods based on 2,000 simulations from the posterior distribution of the two parties' positions.

4.2 Positions of Party Families

The ideological currents of Liberalism, Conservatism, Socialism, Communism, and Nationalism have been shaping the politics of Europe since the emergence of representative democracy in the 19th century. Many of the older European parties can be traced to one of these ideological currents, an observation that has led to the notion of party families (???). It is now demonstrated how the typical positions of party families can be recovered using the dynamic idealpoint model discussed here. In accordance with the party systems literature, the party families considered here are (1) liberal parties, (2) conservative parties, (3) socialist and social democratic parties, (4) communist parties (5) conservative parties, (6) Christian, Christian democratic and confessional parties, (7) right-wing parties, (8) ethnic/linguistic/regional parties, (9) agrarian/rural parties, (10) left socialist and new left parties, and (11) environmentalist or "green" parties. The appendix shows a list of parties considered in this section and the party families which they were sorted into.

The "profile" of the party families is illustrated by figure 6. For each of the party families, the diagram shows its average position on the one axis of the economic policy space and on

the two axes of the space of social/moral policies, that is, the authoritarian/libertarian and the traditionalist/permissive axis. To facilitate the comparison of parties' positions on these three axes, the coordinate values on the axis are all standardized to an average of zero and unit standard deviation. The dots in the diagrams represent the average position of electoral platforms of the parties from the respective families, the grey areas connect quartiles of the distribution of these positions and the solid horizontal lines connect the 2.5 and 97.5 percent quantiles, so that they cover 95 per cent of the distribution of the positions of the platforms.

With regards to the "ideological" party families of the communist, left-socialist, social democratic, liberal, conservative, and right-wing populist parties it appears that they are more polarized on the single dimension of the economic policy space than on any of the social/moral dimension. Further, for five of these party families one can state that whenever it their positions are concentrated on the state-intervention side of the economic policy axis they also tend to be on the libertarian and the permissive side of the two axes of the social/moral policy space, and whenever they are positioned on the market-liberal side of the economic policy axis they also tend to be positioned on the authoritarian and on the traditionalist sides of the respective axes of the social/moral policy space. While this pattern seems to support the idea of an overarching ideological left-right dimension, it is already broken by one of the "ideological" party families, the family of the liberal parties: Consistent with the ideas of classical liberalism, they tend to support more market-liberal positions, but at the same time more libertarian and, to a somewhat lesser degree, permissive positions in the space of social/moral policies.

Figure 7 summarizes the average positions of the party families in terms of their positions in the domestic and social policy space. It superimposes a contour plot of a two-dimensional kernel density estimate (Venables and Ripley 2002) of the empirical distribution of the parties' positions with dots that indicate the average positions of the party families in this policy space. These average positions suggest that it is indeed reasonable to distinguish between these two dimensions of the domestic and social policy space: It turns out that the most distinctive feature of families of Christian parties and of agrarian parties is their support for traditional moral norms rather than a support for state authority. What make green parties distinctive as a party family also becomes obvious: They are the party family with the most permissive positions in terms of social policy and they are second in terms of libertarian positions on the libertarian/authoritarian dimension only to the family of the communist parties. The libertarian position of the communist party family, however, is quite at odds with the fact that communist parties constitute a dogmatic and in some respect authoritarian wing of the labor movement. This anomalous result may be reflect the fact that communist parties, presenting themselves as a fundamental opposition to the existing political and social order will strongly distance themselves from policy objectives like "Law and order" or support for the "National way of life". Another speculative explanation is that this position of the communist party family is the result of an emphasis of democracy in the guise of "workplace democracy"
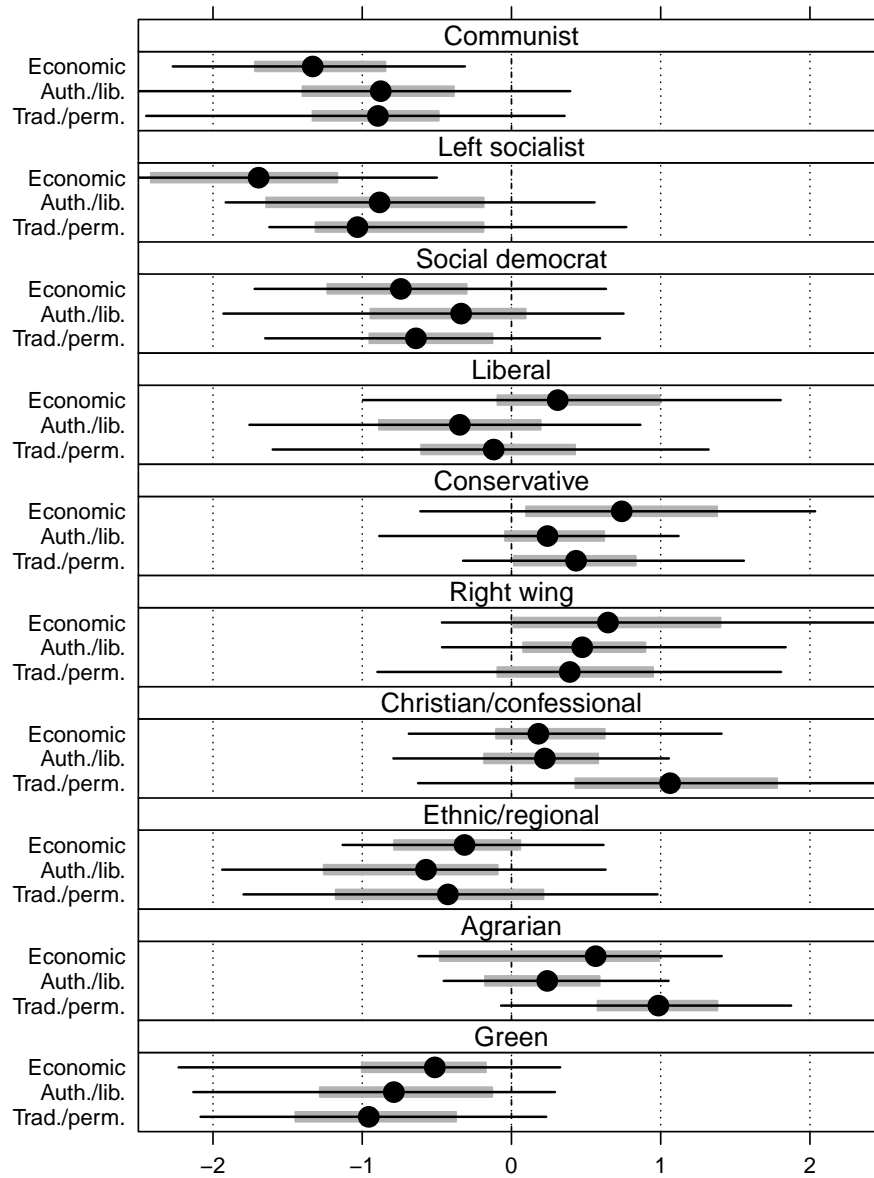
Figure 6: Positions space of electoral platforms of West European parties in different party families on the axis of the economic policy and the two axes of the social/moral policy space (the authoritarian/libertarian and the traditionalist/permissive axis). Means, 2.5, 25, 75, and 97.5-quantiles of parties posterior mean positions.

Figure 7: Positions of party families in the space of domestic and social policy — (empirical Bayes) posterior expectations and averages broken down by party family.
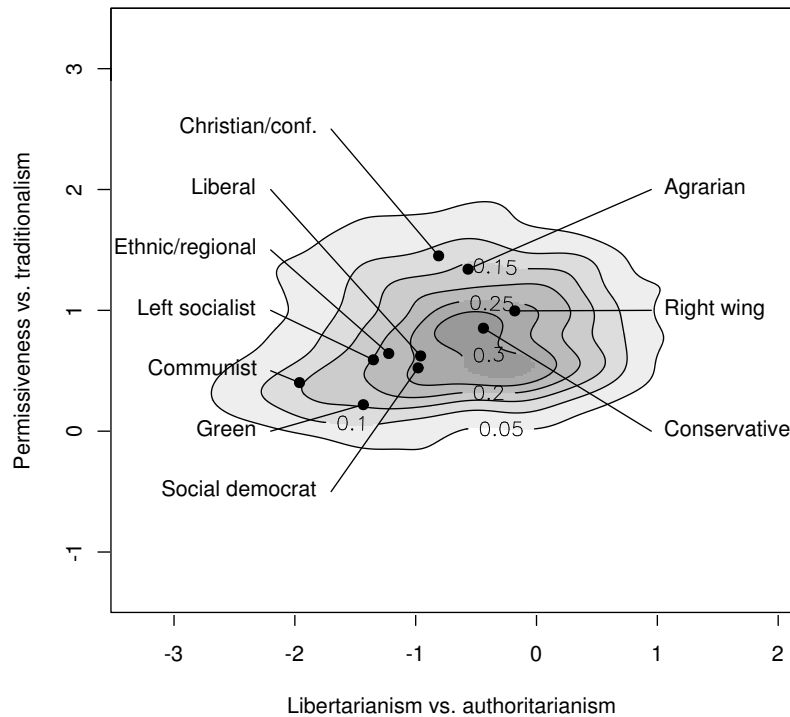
Note: The figure shows a two-dimensional kernel estimate of the distribution of parties mean positions and the group means of party families based on accept-reject samples from the posterior distribution of political positions taken in electoral platforms.
Data source: The Manifesto Project (Budge et al. 2001b; Klingemann et al. 2006; Volkens et al. 2010).

or "soviet democracy". Whether this is the case or an emphasis of libertarian positions is only a common rhetorical device used by communist parties, cannot be decided at the current stage of analysis. Answering this question would perhaps need a more detailed analysis of communist electoral platforms beyond the coding schema of the Manifesto Project.

4.3 Party Positions and Voting Behaviour – Remarks on Using Reconstructed Party Positions as an Independent Variable

Uncovering the influence of parties' political positions on the patterns of voting behaviour, in particular on the relation between social divisions and voting, was one of the motivations that originally led the author of this paper to consider the reconstruction such positions from Manifesto Project data. However, there are not yet any such applications of this kind using the political positions as reconstructed from the dynamic idealpoint model of this paper. Elff (2009) uses reconstructed political positions, that are reconstructed based on ideas de-

veloped in this paper, but the application is more simplistic, resting on a combination of log-transforming the percentages of policy topic emphases obtained from an earlier version of the Manifesto Data set with metric multidimensional unfolding. As a consequence no empirical example can be given. Instead only the formal principles can be given here, which makes the following passages more technical than the previous parts of the present section.

Suppose $\boldsymbol{y}$ is a vector of observations of voting decisions, X a matrix of voter characteristics and $\boldsymbol{b}$ a vector of positions of voting alternatives. Suppose the relation between voting decisions, characteristics, and positions is expressed by a parameter vector ξ and the positions are observed, then ξ can be estimated by maximizing the log-likelihood $\ell_{\mathrm{c}}(\xi; \boldsymbol{y}|X, \boldsymbol{b})$. However, the positions of the voting alternatives – the parties – are unobserved, but the estimated the joint distribution

$$\hat{f}(\boldsymbol{m}_j, \boldsymbol{b}_j) = \mathcal{L}(\hat{\theta}; \boldsymbol{m}_j, \boldsymbol{b}_j) = \mathcal{P}(\boldsymbol{b}_j|\boldsymbol{m}_j; \hat{\theta})\mathcal{L}(\hat{\theta}; \boldsymbol{m}_j)$$

with the emphasis counts obtained with the help of the dynamic idealpoint model are available. Under the assumption that $\boldsymbol{y}$ and $\boldsymbol{m}_j$ are conditionally independent given $\boldsymbol{b}_j$ the appropriate marginal likelihood (with $\boldsymbol{y}$ and X appropriately segmented) that can be used to estimate ξ is

$$\mathcal{L}(\xi; \boldsymbol{y}|X, \boldsymbol{m}) = \prod_j \int \mathcal{L}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j) f(\boldsymbol{m}_j, \boldsymbol{b}_j) \partial \boldsymbol{b}_j$$

$$= \prod_j \mathcal{L}(\hat{\theta}; \boldsymbol{m}_j) \int \mathcal{L}_{\mathrm{c}}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j) \mathcal{P}(\boldsymbol{b}_j|\boldsymbol{m}_j; \hat{\theta}) \partial \boldsymbol{b}_j$$

with corresponding log-marginal likelihood:

$$\ell(\xi; \boldsymbol{y}|X, \boldsymbol{m}) = \sum_j \ln \int \mathcal{L}_{\mathrm{c}}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j) \mathcal{P}(\boldsymbol{b}_j|\boldsymbol{m}_j; \hat{\theta}) \partial \boldsymbol{b}_j + \sum_j \ell(\hat{\theta}; \boldsymbol{m}_j)$$

Fortunately as a side-product of estimating the dynamic idealpoint model, samples from the posterior distribution $\mathcal{P}(\boldsymbol{b}_j|\boldsymbol{m}_j)$ are available. Suppose we have R samples for each j i.e. $\boldsymbol{b}_j^{(r)}, r = 1, \ldots, R$, then we can approximate

$$\ell(\xi; \boldsymbol{y}|X, \boldsymbol{m}) \approx \sum_j \ln \left[\frac{1}{R} \sum_r \mathcal{L}_{\mathrm{c}}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j^{(r)}) \right] + \sum_j \ell(\hat{\theta}; \boldsymbol{m}_j) = \ell_R^*(\xi; \boldsymbol{y}|X, \boldsymbol{m}), \text{ say.}$$

The approximated log likelihood function has the gradient

$$\frac{\partial \ell_R^*(\xi; \boldsymbol{y}|X, \boldsymbol{m})}{\partial \xi} = \sum_j \sum_r \frac{\mathcal{L}_{\mathrm{c}}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j^{(r)})}{\sum_s \mathcal{L}_{\mathrm{c}}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j^{(s)})} \frac{\partial \ell_{\mathrm{c}}(\xi; \boldsymbol{y}_j|X_j, \boldsymbol{b}_j^{(r)})}{\partial \xi}$$

the structure of which again suggests an EM algorithm for finding the MLE of ξ.

It should be noted that this marginal likelihood approach differs from simply "plugging in"

a Monte Carlo sample from $\mathcal{P}(\boldsymbol{b}_j|\boldsymbol{m}_j)$ into the complete-data log-likelihood. This intuitively more simple approach leads to

$$\vec{\ell}_R^{\dagger}(\boldsymbol{\xi};\boldsymbol{y}|X,\boldsymbol{m}) = \sum_j \frac{1}{R} \sum_r \ln \mathcal{L}_c(\boldsymbol{\xi};\boldsymbol{y}_j|X_j,\boldsymbol{b}_j^{(r)}) = \sum_j \frac{1}{R} \sum_r \ell_c(\boldsymbol{\xi};\boldsymbol{y}_j|X_j,\boldsymbol{b}_j^{(r)}),$$

which is an approximation to

$$\ell^{\dagger}(\boldsymbol{\xi};\boldsymbol{y}|X,\boldsymbol{m}) = \sum_j \int \ln \mathcal{L}_c(\boldsymbol{\xi};\boldsymbol{y}_j|X_j,\boldsymbol{b}_j)\mathcal{P}(\boldsymbol{b}_j|\boldsymbol{m}_j;\hat{\boldsymbol{\theta}})\partial\boldsymbol{b}_j,$$

where integration and logarithm are exchanged relative to $\ell(\boldsymbol{\xi};\boldsymbol{y}|X,\boldsymbol{m})$. It is not straight-forward to establish whether estimates based on maximizing $\vec{\ell}_R^{\dagger}(\boldsymbol{\xi};\boldsymbol{y}|X,\boldsymbol{m})$ are consistent and asymptotically equivalent with those based on maximizing $\ell_R^*(\boldsymbol{\xi};\boldsymbol{y}|X,\boldsymbol{m})$. This hinges on whether the approximation

$$\frac{\mathcal{L}_c(\boldsymbol{\xi};\boldsymbol{y}_j|X_j,\boldsymbol{b}_j^{(r)})}{\sum_s \mathcal{L}_c(\boldsymbol{\xi};\boldsymbol{y}_j|X_j,\boldsymbol{b}_j^{(s)})} \approx \frac{1}{R}$$

is "good enough". This requires that $\mathcal{L}_c(\boldsymbol{\xi};\boldsymbol{y}_j|X_j,\boldsymbol{b}_j)$ does not vary too much with $\boldsymbol{b}_j$ – which means that party positions exert little influence on the response – or that the variance of $\boldsymbol{b}_j^{(r)}$ between replications $r = 1,\ldots,R$ is low – which means that we have relative precise information about the party positions. Since an EM algorithm only requires a moderate amount of additional programming relative to the "plug-in" approach, the former seems to be the "safer bet".

5 Challenges

5.1 Different Vagueness of Values and Variant Degrees of Contention

Attempts at reconstructing different policy positions or ideological positions of parties from manifesto data have to assume that the policy objectives more or less emphasised in the manifestos are divergent or contested. If having a "leftist" position means to emphasise different policy objectives than having a "rightist" position, then these policy objectives can hardly seen as consensual. Yet the assumption that *all* policy objectives are contested in this way seems as unrealistic as the assumption that the emphasis of *all* policy objectives follows the "saliency and valency" logic that was originally guiding the early efforts of the Manifesto Research Group. For example, *Nationalization* is clearly a contested policy objective typically emphasised by parties of the (radical) left, *Economic Growth: positive* is unlikely be de-emphasised by any party other than the most ardent supporters of a steady-state economy among green parties. Other examples might be policy objectives such as *Peace* – which one may hardly

expect to be openly opposed by other parties than warmongering fascist parties – or *Freedom and Human Rights* or *Democracy* – which can be expected to be supported by all parties except those at the left or right extreme end of the political spectrum.

Furthermore, an explicit or implicit assumption that forms the base both of approaches resting on difference scores (such as the RiLe-scores created by the Manifesto Project) and resting on a distance model (such as the dynamic idealpoint model so far discussed in the present paper) is that all policy objectives, in so far as they are divergent or contested, discriminate to the same degree between different (e.g. "left" and "right") policy positions. In other words, it is assumed that all policy objective exhibit the same degree of vagueness or acuity. Yet this may also be an assumption that is not really realistic. For example, the policy objective *Freedom and Human Rights* may seem consensual, but perhaps it is not but only vaguely contested and ambiguous: Left-libertarian may rather emphasise the "human rights" aspect as the need to protect citizens from excessive incursions from public authorities, while liberal-conservative parties may emphasise the "freedom" aspect as protecting the existing order from Communism. At the same time an emphasis of "democracy" may refer to an expansion of public participation in the form of referenda by parties of the left, whereas it may refer in particular to conventional representative democracy by parties of the centre-right, again in opposition to Communism and the radical Left. In the field of economic policy, a topic such as *Marxist Analysis* may be highly divisive, used rarely outside the group of radical left-wing and Communist parties.

In the context of the idealpoint model discussed in this paper, the above considerations call for flexible distance functions. Recall that in the dynamic idealpoint model so far discussed, the relative emphasis of the a policy objective is a function of the unweighted (squared) distance between the location of a policy objective and the policy position or the ideological position that a party takes by virtue of its manifesto:

$$\pi_{ijt} = \frac{\exp(\eta_{ijt})}{\sum_h \exp(\eta_{hjt})} \text{ where } \eta_{ijt} = -\frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt})'(\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt}).$$

A more flexible specification of the relative emphasis would have

$$\pi_{ijt} = \frac{\exp(\eta_{ijt})}{\sum_h \exp(\eta_{hjt})} \text{ where } \eta_{ijt} = \tau_i - \frac{\varphi_i}{2}(\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt})'(\boldsymbol{\alpha}_i - \boldsymbol{b}_{jt}). \tag{8}$$

The parameter τ_i allows to include policy topics into the model that are rarely used in manifestos, not or not only because they have an extreme location in the policy space or ideological space, but also because they are more or less expressive for *any* political position however centrist or extreme it may be. The parameter φ_i allows to represent different degrees of vagueness or contestedness of policy objectives or manifesto topics. For those policy objectives or topics that are relatively consensual φ_i may have relatively low values, whereas for highly divisive objectives or topics φ_i may be relatively large.

Figure 8: Relative emphasis of four policy objectives with rigid and flexible distance function

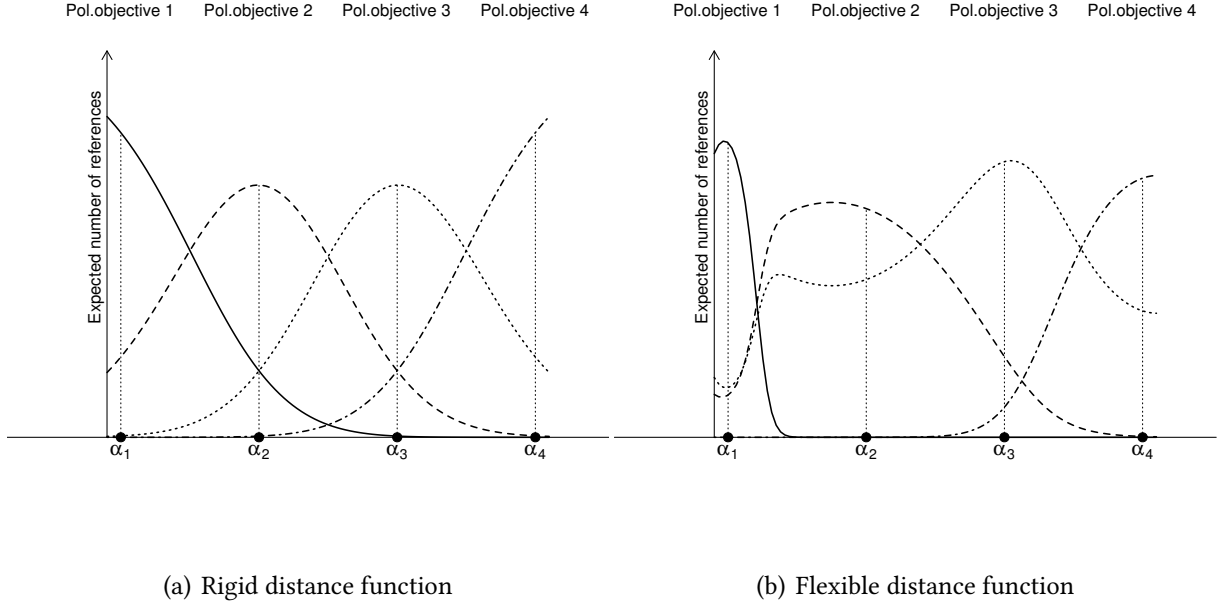(a) Rigid distance function

(b) Flexible distance function

Figure 5.1 demonstrates the difference between rigid and flexible distance functions in a unidimensional policy space. The left-hand panel demonstrates how political positions determine the expected relative emphasis if the distance function is as rigid as in the dynamic idealpoint model introduced by Elff (2013). In the more flexible framework of this subsection, such a rigid model posits that $\tau_i = 0$ and $\varphi_i = 1$ for all policy objectives $i = 1, \ldots, I$ in the policy space in question. The right-hand panel demonstrates how the expected emphasis may be related to political positions if the distance function is more flexible, allowing for $\tau_i \neq 0$ and/or $\varphi_i \neq 1$. Policy objective 1 in this example is not only located in the relative extreme of the policy space, but also highly divisive (such as *Marxist Analysis* might be in the economic policy space). Policy objectives 2 and 4 have a "medium level" of divisiveness, with policy objective 4 being somewhat more divisive than policy objective 2. Policy objective 3 is relatively consensual, being emphasised by all positions except the most extreme that are close to policy objective 3.

Increasing the flexibility of a model by introducing additional parameters sometimes is like opening a Pandora's box of identification problems. The first identification problem arises from the ratio form of the relative emphasis as specified by equation (8): Suppose $\eta_{+jt} = \sum_i \eta_{ijt}$ is unrestricted, then it cannot be estimated because any unrestricted offset that is the same across η_{ijt} for $i = 1, \ldots, I$ cancels out of the ratio. This identification problem can be addressed by requiring that

$$\sum_i \tau_i = 0 \quad \text{and} \quad \sum_i \varphi_i \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_i = 1.$$

and further that the mean of the political positions is zero, i.e. $\boldsymbol{\beta} = 0$.[4] While the restrictions on the τ_i parameters is linear and therefore relatively easy to enforce (e.g. by setting $\tau_I = -\sum_{i<I} \tau_i$) the second set of restrictions requires the use of Lagrange multipliers, thus complicating the estimation procedure somewhat.

The second identification problem arises from the fact that if the weighting parameters φ_i are unrestricted, the scale of the distances becomes indeterminate, unless one imposes restrictions on the variance parameters of the political positions in Σ_0 and Σ_1. One option may be to set e.g. $\varphi_1 = 1$, but a more natural restriction may be to set Σ_0 equal to an identity matrix, which might also contribute to solve some identification problems with respect to rotation.

5.2 Slow Convergence of EM and Computational Intensity

Very often, theory can give little guidance for the reconstruction of latent traits from sets of indicators. Instead, whether an indicator "works" is often a question that can only answered empirically, after the specification of an appropriate measurement model. In addition, it is not always straightforward to established that a particular specification of a measurement model with a given set of indicators is identified. For LISREL models there are some rules of thumb, but they may well break down if some indicators are too much or to little correlated. In the end, the construction of a "working" multiple-indicator model is often a matter of trial-and-error. In the case of the dynamic idealpoint model discussed in this paper, the situation is not much different. Due to its complexity, it offers a wide array of options of identifying restrictions that may or may not be necessary, that may or may not be sufficient, and that may or may not lead to results that have a straightforward or plausible interpretation even if a good fit between model and data can be reached.

The method to estimate the parameters of the dynamic idealpoint model proposed by Elff (2013) is an EM algorithm with Monte Carlo integration, that is, an MCEM algorithm. EM algorithms are known to be numerically stable and to maximize marginal likelihood functions, but also to be slow to converge, especially near the maximum of the log-likelihood function. Monte Carlo integration may allow to compute the log-likelihood function and its derivatives at any required degree of precision, yet it can be highly demanding in terms of the required computation. As a consequence, the MCEM approach does not lend itself easily to the trial-and-error process that often is involved in the search for the appropriate model for a particular policy space.

There are various acceleration schemes available for EM algorithms that can greatly reduce

---

[4]In the original formulation of the dynamic idealpoint model proposed by Elff (2013), identification problems due to the observational invariance of the model with respect to rotation and translation where addressed by setting some elements of the policy objective location vectors $\boldsymbol{\alpha}_i$ to zero or unity, and by requiring that $\sum_i \boldsymbol{\alpha}_i = 0$ (to allow for the mean $\boldsymbol{\beta}$ of the policy positions to be different from zero).A wiser alternative now seems to assume the mean policy position to be zero, thus eliminating the parameter $\boldsymbol{\beta}$ and to drop the linear restriction $\sum_i \boldsymbol{\alpha}_i = 0$.

the number of iterations needed provided that the algorithm has already come close to the maximum of the log-likelihood function (McLachlan and Krishnan 2007). That way, the computational time may in the present case be reduced from a whole day to a few hours. Yet this still is way too long a time to wait for results in the context of a trial-and-error specification search. For the purpose of finding a model that considers the "right" set of policy objectives for a given policy space, a simple approximation of the log-likelihood function leading to at least approximately consistent, if not asymptotically efficient estimates might be sufficient. For this purpose, a *Laplace approximation* of the log-likelihood function can be used. It is widely used for the computation of so-called PQL estimates of mixed-effects logistic regression and other generalised linear mixed models (Breslow and Clayton 1993), yet it is biased in even moderate-sized samples (Breslow and Lin 1995). For the marginal log-likelihood (7) the Laplace approximation is

$$\ell = \sum_j \ln \int \mathcal{L}(\theta; m_j, b_j) \partial b_j \approx \sum_j \ell(\theta; m_j, \tilde{b}_j) - \sum_j \frac{1}{2} \ln \det(K(\theta; m_j, \tilde{b}_j))$$

where $\tilde{b}_j$ is the value of b_j that maximizes the complete-data log-likelihood $\ell(\theta; m_j, b_j)$ for given θ and

$$K(\theta; m_j, b_j) = -\frac{\partial^2 \ell(\theta; m_j, b_j)}{\partial b_j \partial b_j'}$$

is the negative of the matrix of the second derivatives of the complete-data log-likelihood with respect to b_j. For $m_{ijt} \to \infty$ this Laplace approximation converges towards the marginal log-likelihood, but again may lead to biased estimates for small to moderate-sized values of m_{ijt}. On the other hand, Laplace approximation may reduce the computation time from several ours to a few minutes, because it does not need Monte Carlo samples and can be maximized by conventional algorithm such as Newton-Raphson or Fisher-scoring. That is, Laplace approximation may not be the method of choice to obtain final estimates of the dynamic idealpoint model to be used in publications, but it may be useful in exploratory work.

6 Discussion: Why all the Complexity?

Faced with the seeming complexity of the dynamic idealpoint model and all the mathematical technicalities of its construction and estimation, applied researchers may ask whether it is really necessary and not just an exercise in practically irrelevant methodological purism. Part of the attractiveness of e.g. an indicator for party positions based on Manifesto Project data such as the RiLe index is the simplicity of its construction. So the question arises what the real costs are of using such and index. Answering this question is the topic of the following paragraphs.

The fundamental assumption for using Manifesto Project data for the reconstruction of par-

ties' positions in (unidimensional or multidimensional) policy spaces is that these positions are expressed by the selective emphasis of policy objectives or policy topics as they are reflected in the variables of a Manifesto Project data set. An ideal situation for this might seem to be that there are several policy objectives or policy topics that are emphasised and de-emphasised as any party potentially moves its position from one extreme of a unidimensional policy space (the "extreme left") to the other extreme of the policy space (the "extreme right") and that the locations in the policy space where one and only one of the policy objectives or policy topics is emphasised the most are evenly distributed in the policy space as in the top-left panel of Figure 9. Now suppose a RiLe-type index is constructed by subtracting the sum of the emphases of the "leftist" policy objectives 1 through 5 from the sum of the emphases of the "rightist" policy objectives 6 through 10.

The relation between the latent positions of the parties and the corresponding RiLe scores thus constructed is depicted in the top-right panel of Figure 9. This plot reflects an idealised situation in so far as there is neither random noise nor any discontinuities due to the fact that percentages will in reality computed from counts. The plot looks at first glance favourable: RiLe scores do increase monotonously with the latent left-right position of the parties. Yet the plot also shows that the relation is highly linear and that the RiLe scores are very sensitive to variations within the centre of the left-right spectrum and not sensitive at all to variations within the extremes of the spectrum. As a result, RiLe scores might lead to an overstatement of shifts of parties in centre and will be almost oblivious with respect to shifts from the centre to the extreme. This excessive sensitivity to positional differences at the centre of the political space might be thought to be offset by dropping some of the more central policy objectives from the construction of the RiLe-scores. The bottom-left panel shows the relation between the latent positions and the RiLe-scores if the four central policy objectives are removed from the construction of the latter. Now the RiLe scores are a bit more sensitive to differences outside the centre of the political but they are not sensitive at all towards policy differences at the centre.

A comparison between the top-right and the bottom-left panel of Figure 9 seems to suggest that a weighted sum of these two kinds of RiLe-Indices may improve the situation, by combining them in form of a weighted sum. Such a weighted sum could result e.g. from a principle components analysis in the form of factor scores. Indeed, factor scores from principle components analysis have already been used in the literature. The bottom-right panel shows how factor scores generated from a principle components analysis of the selective emphases are related to the underlying political positions. The shape of the curve does not seem to look very comforting, because it is non-monotonic. While the PCA scores seem to be sensitive to differences between position both in the centre and in the moderate "leftist" and "rightist" regions of the political space, its sensitivity goes in the wrong direction in more "radical" regions. As a consequence, if a party moves from a moderately radical position to an extreme position this might then appear as a shift towards the centre!
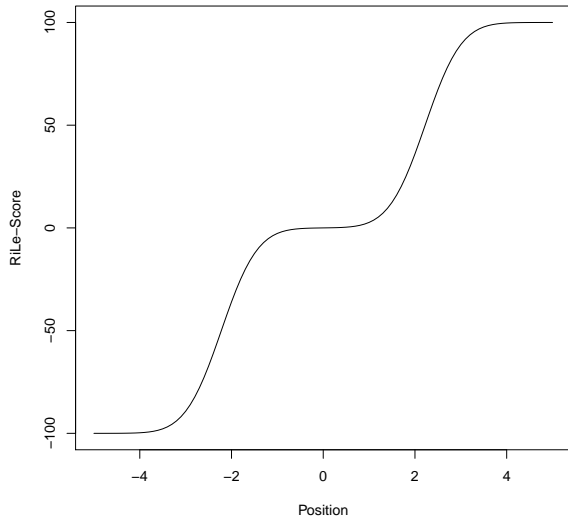
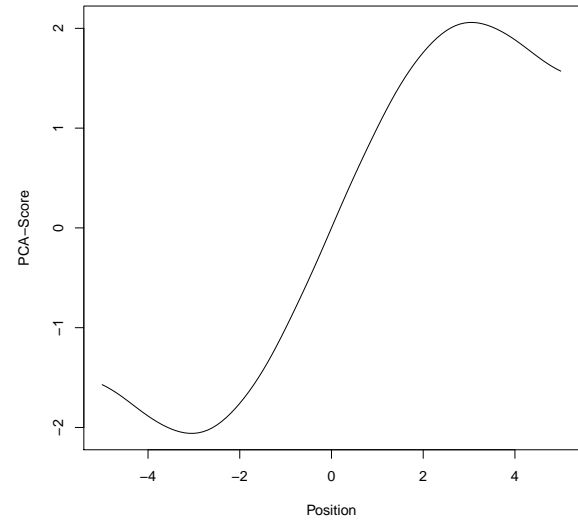Figure 9: Selective Emphasis of Policy Objectives and Variants of an RiLe Index
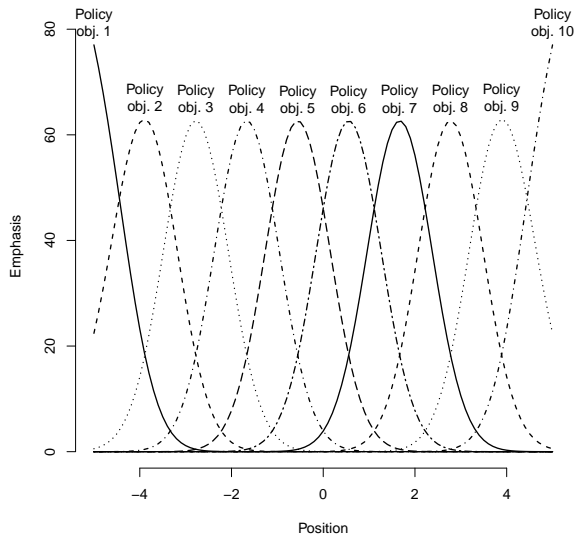
(a) Selective emphasis of policy objectives
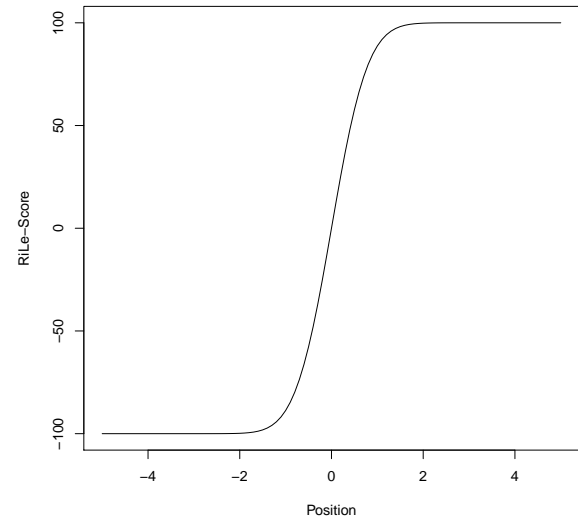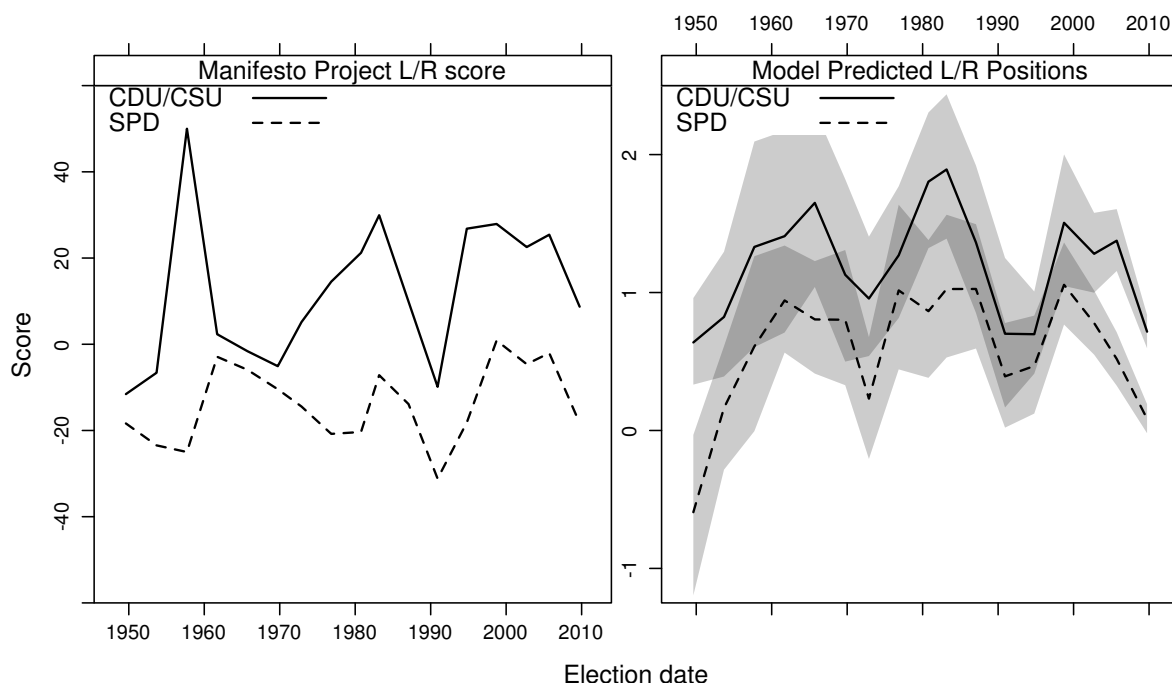
(b) RiLe Index, full set

(c) RiLe Index, four central policy objectives removed

(d) Scores from PCA

Figure 10: A Comparison of Political Positions of the Two Major German Parties Reconstructed with Manifesto RiLe Scores and with the Dynamic Idealpoint Model

It could still be argued that the preceding discussion hinges on an idealised situation and that when confronted with real data, using RiLe scores or a refined method such as that discussed in this paper does not make much of a difference. This is however not the case, as can be seen Figure 10. The left-hand panel in the figure shows the development of the left-right positions of the CDU/CSU and the SPD, the two major parties in Germany, as measured by the RiLe scores published as part of the Manifesto Project data sets. Two features of the development of the positions measured by RiLe scores are particularly remarkable. The first and quite conspicuous feature is the right-wing "spike" of the CDU/CSU in the election of 1957 and the rapid convergence of the two parties right after 1957, when both parties' positions are quite close to one another. Even if the campaign for the election of 1957 was conducted in a relatively polarised manner, such a strong rightward movement does not appear very plausible and might be an artefact of an excessive sensitivity of RiLe scores to movements in the central regions of the political space. In fact it was seeing such fluctuations in the positions of the German parties that led the author of the present paper to look for alternatives for the RiLe scores as measures for parties' political positions. The second feature is that the absence of any traces of the more Marxist positions of the SPD before the Godesberg Programme. The right-hand panel of Figure 10 shows the development of the positions of the two major German parties in the space of *economic* policy. Here the spike in the development of the CDU/CSU does not appear, however the SPD shows clear left-wing positions in the elections before the Godesberg Programme, which it never returns to after 1959. (In contrast

to the leftist position of the SPD in 1990 according to the RiLe scores.) This comparison shows that, on the one hand, the way in which parties' positions are reconstructed does matter in practice and that, on the other hand, looking at positions for parties in specific policy spaces might reveal political movements that may become obscured by a focus on a general left-right dimension.

7 Conclusion

This paper discusses an dynamic idealpoint model of party manifestos, which is essentially an application of the state-space model of coded political text published in *Political Analysis* by the author of this paper. It was of course not the purpose of this paper to restate that work, but to illustrate its motivation stemming from the desire to make the best use of the data of the Manifesto Project, and to demonstrate its application. While the model is of considerable formal complexity, this complexity can be justified from the principles that also can be used for the justification to use coded manifestos at all for the reconstruction of parties' political positions. At was also discussed in the paper, that the dynamic idealpoint model still has its limitations, which should lead to its further development as well as of the associated methods of estimation. Despite these limitation it should also have become clear, that the problems addressed the model cannot be safely ignored.

References

Agresti, A. (2002). *Categorical Data Analysis* (Second ed.). New York: Wiley.

Bartolini, S. and P. Mair (1990). *Identity, Competion, and Electoral Availability: The Stabilisation of European Electorates 1885–1985*. Cambridge: Cambridge University Press.

Benoit, K., S. Mikhaylov, and M. Laver (2009). Treating words as data with error: Uncertainty in text statements of policy positions. *American Journal of Political Science 53*(2), 495–513.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*(421), 9–25.

Breslow, N. E. and X. Lin (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika 82*(1), 81–91.

Brody, R. A. and B. I. Page (1972). Comment: The assessment of policy voting. *The American Political Science Review 66*(2), pp. 450–458.

Budge, I. and D. Farlie (1983a). Party competition selective emphasis or direct confrontation? An alternative view with data. In H. Daalder and P. Mair (Eds.), *Western European Party Systems Continuity and Change*, pp. 265–305. Beverly Hills: Sage.

Budge, I. and D. J. Farlie (1983b). *Explaining and Predicting Elections: Issue Effects and Party Strategies in Twenty-Three Democracies*. London: George Allen & Unwin.

Budge, I., H.-D. Klingemann, A. Volkens, J. Bara, E. Tanenbaum, et al. (2001a). Mapping policy preferences: Estimates for parties, electors, and governments 1945-98.

Budge, I., H.-D. Klingemann, A. Volkens, J. Bara, E. Tanenbaum, et al. (2001b). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945-98*. Oxford: Oxford University Press.

Budge, I., D. Robertson, and D. Hearl (1987). *Ideology, Strategy, and Party Change: Spatial Analysis of Post-War Election Programs in 19 Democracies*. Cambridge: Cambridge Univerity Press.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(1), 1–38.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harber & Row.

Elff, M. (2009). Social divisions, party positions, and electoral behaviour. *Electoral Studies 28*(2), 297–308.

Elff, M. (2013). A dynamic state-space model of coded political texts. *Political Analysis 12*(2), 217–232.

Gabel, M. and J. 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.

Klingemann, H.-D., A. Volkens, J. Bara, I. Budge, and M. McDonald (2006). *Mapping Policy Preferences II. Estimates for Parties, Electors, and Governments in Eastern Europe, the European Union and the OECD, 1990-2003*. Oxford: Oxford University Press.

Laver, M. (2001). Position and salience in the policies of political actors. In M. Laver (Ed.), *Estimating the Policy Positions of Political Actors*, pp. 66–75. London and New York: Routledge.

Laver, M., K. Benoit, and J. Garry (2003). Extracting policy positions from political texts using words as data. *American Political Science Review 97*(2), 311–331.

Laver, M. and W. B. Hunt (1992). *Policy and Party Competition*. New York: Routledge.

Laver, M. J. and J. Garry (2000). Estimating policy positions from political texts. *American Journal of Political Science 44*(3), 619–634.

Lowe, W., K. Benoit, S. Mikhaylov, and M. Laver (2011). Scaling policy preferences from coded political texts. *Legislative Studies Quarterly 34*(1), 123–155.

McKelvey, R. (1976). Intransitivities in multidimensional voting bodies and some implications for agenda control. *Journal of Economic Theory 30*, 283–314.

McLachlan, G. and T. Krishnan (2007). *The EM Algorithm and Extensions.* John Wiley & Sons.

Merrill, S., B. Grofman, and J. Adams (2001, October). Assimilation and contrast effects in voter projections of party locations: Evidence from norway, france, and the USA. *European Journal of Political Research 40*(2), 199–221.

Plott, C. R. (1967). A notion of equilibrium and its possibility under majority rule. *American Economic Review 57*(4), 787–806.

Robertson, D. (1976). *A Theory of Party Competition.* London: Wiley.

Slapin, J. B. and S.-O. Proksch (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science 52*(3), 705–722.

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

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.

Volkens, A., J. Bara, I. Budge, M. D. McDonald, and H.-D. Klinemann (2013). *Mapping Policy Preferences from Texts III: Statistical Solutions for Manifesto Analysts.* Oxford: Oxford University Press.

Volkens, A., O. Lacewell, S. Regel, H. Schultze, and A. Werner (2010). *The Manifesto Data Collection.* Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Wilson, J. M. and P. Gronke (2000). Concordance and projection in citizen perceptions of congressional Roll-Call voting. *Legislative Studies Quarterly 25*(3), 445–467.

Appendix: Classification of Parties into Party Families

Australia: Liberal: LPA Liberal Party of Australia, AD Australian Democrats; Conservative: NPA National Party of Austrialia, DLP Democratic Labour Party; Social democrat: ALP Australian Labour Party

Austria: Liberal: LF Liberal Forum; Social democrat: SPÖ Social Democratic Party of Austria; Christian/confessional: ÖVP Austrian People's Party; Right wing: FPÖ Freedom Party of Austria; Green: GA Green Alternative

Belgium: Liberal: PLP-PVV Liberal Party, PVV (Flemish) Party of Freedom and Progress, PRL (Francophone) Liberal Reformist Party, PLDP (Brussels) Liberal and Pluralist Party, PRL/FDF Francophone Liberal and Democratic Front; Social democrat: PSB-BSP Socialist Party, SP (Flemish) Socialist Party, PS (Francophone) Socialist Party; Christian/confessional: PSC-CVP Christian People's Party, CVP (Flemish) Christian People's Party, PSC (Francophone) Christian Social Party; Right wing: VB Flemish Block; Ethnic/regional: VU Flemish Peoples Union, FDF Francophone Democratic Front, RW Walloon Rally; Green: Ecolo (Francophone Ecologists), Agalev (Flemish Greens)

Canada: Liberal: LP Liberals Party; Conservative: PCP Progressive-Conservative Party; Social democrat: NDP New Democratic Party; Ethnic/regional: BQ Bloc Quebecois

Denmark: Liberal: RV Radical Liberals (Radikale Venstre), V Liberals (Venstre), RF Justice Party, DS Danish Union, LC Liberal Center, CD Center Democrats; Conservative: KF Conservative People's Party, DU Danish Independents; Social democrat: SD Social Democrats; Communist: DKP Communist Party of Denmark; Christian/confessional: KrF Christian Peoples Party; Right wing: FP Progress Party; Left socialist: SF Socialist People's Party, VS Left Socialists, FK Common Course; Green: EL Red-Green Unity List

Finland: Liberal: LKP Liberal People's Party, SK Finnish Center; Conservative: KK National Coalition Party; Social democrat: SSDP Finnish Social Democratic Party; Communist: SKDL Finnish People's Democratic League, DEVA Democratic Alternative; Christian/confessional: SKL Finnish Christian Union; Ethnic/regional: RKP/SFP Swedish People's Party; Agrarian: SMP Finnish Rural Party; Left socialist: TPSL Social Democratic League of Workers and Smallholders, VL Left Alliance; Green: VL Green Union

France: Conservative: Gaullists (UNR/UDT/RPR), Conservatives (PRL/IR/CNIP, etc.), Rally for the Republic; Social democrat: PS Socialist Party; Communist: PCF Communist Party of France; Christian/confessional: MRP Popular Republican Movement, Democratic Center, CDP Center for Democracy and Progress, MR Reformer's Movement, UDF Union for French Democracy; Right wing: Poujadists, FN National Front; Left socialist: RRRS Repubican and Radical Socialist Party; Green: Ecologists/Greens, FRA: Ecology Generation

Germany: Liberal: FDP Free Democratic Party of Germany; Conservative: DP German Party; Social democrat: SPD Social Democratic Party of Germany; Communist: KPD Communist Party of Germany; Christian/confessional: CDU/CSU Christian Democratic Union/ Christian Social Union, DZ German Center Party; Right wing: DKP-DRP German Reich

Party; Ethnic/regional: BP Bavarian Party, SSW South Schleswig Electors' Coalition; Left socialist: PDS Party for Democratic Socialism; Green: Greens, Greens-Alliance 90, Alliance 90-Greens

Great Britain: Liberal: Liberal Party, LDP Liberal Democratic Party; Conservative: Conservative Party; Social democrat: Labour Party, SDP Social Democratic Party

Greece: Liberal: EDIK Union of the Democratic Center; Conservative: ND New Democracy, EP National Alignment, Pola Political Spring; Social democrat: PASOK Panhellenic Socialist Movement, DIKKI Democratic Social Movement; Communist: KKE Communist Party of Greece; Left socialist: SAP Progressive Left Coalition

Ireland: Liberal: PD Progressive Democratic Party; Social democrat: LP Labour Party; Left socialist: WP Workers Party, DLP Democratic Left Party; Green: Greens

Israel: Conservative: Likud; Social democrat: Israeli Labour Party; Ethnic/regional: United Arab List

Italy: Liberal: PRI Republican Party of Italy, PLI Liberal Party of Italy, PR Radical Party, LR La Rete, PI Pact for Italy, AD Democratic Alliance, RI Italian Renewal; Conservative: FI Forza Italia; Social democrat: PSI Socialist Party of Italy, PSDI Social Democratic Party of Italy; Communist: PCI Communist Party of Italy/PDS Democratic Party of the Left, RC Communist Refoundation; Christian/confessional: DC Christian Democrats/PPI Italian People's Party, CCD Christian Democratic Center; Right wing: AN National Alliance; Ethnic/regional: LN Northern League; Left socialist: PSU United Socialist Party, PdUP Proletarian Unity, DP Proletarian Democracy; Green: FdV Green Federation

Japan: Liberal: CGP Clean Government Party, NLC New Liberal Club, JRP Japan Renewal Party, JNP Japan New Party, DPJ Democratic Party of Japan; Conservative: LDP Liberal Democratic Party; Social democrat: DSP Democratic Socialist Party, SDF Social Democratic Federation; Communist: JCP Japan Communist Party, JSP Japan Socialist Party

Luxembourg: Social democrat: POSL LSAP Luxembourg Socialist Workers' Party; Communist: PCL KPL Communist Party of Luxembourg; Christian/confessional: PCS CSV Christian Social Party; Green: GAP Green Alternative Party, GLEI Green List, Ecological Initiative, GLEI/GAP Greens

Netherlands: Liberal: VVD People's Party for Freedom and Democracy, D'66 Democrats 66, PPR Radical Political Party; Social democrat: PvdA Labor Party, DS'70 Democratic Socialists 70; Christian/confessional: KVP Catholic People's Party, ARP Anti-Revolutionary Party, CHU Christian Historical Union, CDA Christian Democratic Appeal; Green: Green Left

New Zealand: Conservative: NP National Party, NZFP New Zealand First Party; Social democrat: LP Labour Party; Left socialist: Alliance

Norway: Liberal: V Liberal Party (Venstre), DLF Liberal People's Party; Conservative: H Conservatives (Høyre); Social democrat: DNA Norwegian Labor Party; Communist: NKP Communist Party of Norway; Christian/confessional: KrF Christian Peoples Party; Right wing: FrP Progress Party; Agrarian: SP Center Party; Left socialist: SV Socialist Left Party

Portugal: Liberal: MDP Portuguese Democratic Movement, PRD Democratic Renewal Party; Conservative: PSD Social Democratic Party, PPM Popular Monarchist Party; Social democrat: PSP Portuguese Socialist Party, ASDI Association of Independent Social Democrats, UEDS Union of the Socialist and Democratic Left; Communist: UDP Popular Democratic Union, PCP Portuguese Communist Party; Christian/confessional: PP Popular Party; Green: PEV Ecologist Party the Greens

Spain: Liberal: PL Liberal Party; Conservative: PCE-IU Communist Party of Spain/Unified Left, AP Popular Alliance/PP Popular Party, PDP Popular Democratic Party; Social democrat: PSOE Socialist Worker's Party of Spain, CDS Democratic and Social Center; Christian/confessional: UCD Union of the Democratic Centre; Ethnic/regional: EE Basque Left, PNV/EAJ Basque National Party, PAR Aragonese Regionalist Party, ERC Catalan Republican Left, PA Andalusian Party, CiU Convergence and Unity, EA Basque Solidarity

Sweden: Liberal: FP Liberal People's Party; Conservative: Hogerpartied/MSP Moderate Coalition Party; Social democrat: SdaP Swedish Social Democratic Labor Party; Communist: Vp Left Party; Christian/confessional: KdS Christian Democratic Community Party; Right wing: NyD New Democracy; Agrarian: Farmers' Party/CP Center Party; Green: Environmentalist Party the Greens

Switzerland: Liberal: FDP-PRD Radical Democratic Party; Conservative: SVP-UDC Swiss People's Party; Social democrat: SPS-PSS Socialist Party of Switzerland; Christian/confessional: CVP-PDC Christian People's Party, EVP-PEP Protestant People's Party; Right wing: SD Swiss Democrats, FPS Freedom Party of Switzerland; Green: Greens

United States: Liberal: Democrats; Conservative: Republicans