Party Positions from Wikipedia Classifications of Party Ideology*

Michael Herrmann
University of Konstanz

Holger Döring
University of Bremen

Abstract

We develop a new measure of party position based on a scaling of ideology tags supplied in infoboxes on political parties' Wikipedia pages. Assuming a simple model of tag assignment, we estimate the locations of parties and ideologies in a common space. We find that the recovered scale can be interpreted in familiar terms of 'left vs. right'. Estimated party positions correlate well with ratings of parties' positions from extant large-scale expert surveys, most strongly with ratings of general left-right ideology. Party position estimates also show high stability in a test-retest scenario. Our results demonstrate that a Wikipedia-based approach yields valid and reliable left-right scores comparable to scores obtained via conventional expert coding methods. It thus provides a measure with potentially unlimited party coverage. Our measurement strategy is also applicable beyond Wikipedia.

*Contact: michael.herrmann@uni-konstanz.de and doering@uni-bremen.de. We thank panelists at the 2019 Annual MPSA Conference, and participants in research colloquia at the Universities of Bremen, Konstanz, and Mannheim for their feedback, especially Christopher Hare, Christopher Wratil, Samuel Baltz, Markus Tepe, Dominic Nyhuis, Sascha Göbel, Thomas Malang, Benjamin Guinaudeau, Paul Bederke, Susumu Shikano, Konstantin Käppner, Christian Breunig, Thomas Gschwend, and Hans-Dieter Klingemann.

1 Introduction

The idea that parties occupy different positions on an ideological continuum is fundamental to theories of the political process. The need for scoring parties’ on such a continuum is evidenced by the sheer diversity of approaches and the creativity political scientists devote to obtaining such scores, which range from expert codings of party manifestos (Volkens et al., 2013), to left-right placements of parties by expert judges (Kitschelt, 2014) or voters (Lo, Proksch and Gschwend, 2014), to supervised scaling of word frequencies in party manifestos (Laver, Benoit and Garry, 2003), to unsupervised scaling approaches that extract positions from expert-coded manifesto content (Däubler and Benoit, 2017) or raw word frequencies (Slapin and Proksch, 2008) in political manifestos, to analyses of roll call votes (Bräuninger, Müller and Stecker, 2016) and parliamentary speech (Lauderdale and Herzog, 2016; Peterson and Spirling, 2018). Together, these approaches enable the measurement of party positions across time, space, levels of government, and policy areas.¹

We propose a new measure of party left-right position, one that offers greater party and country coverage than any existing measure. The measure is derived from semi-standardized information about party ideology available on the English Wikipedia.² In particular, we draw on ideological keywords (e.g., socialism) that Wikipedia editors use to tag parties and to link them with Wikipedia articles on those ideologies. We develop an ideal point model of how these tags are assigned and use it to scale over 2,000 parties and their associated ideologies on a latent dimension. Our scaling approach is based on the idea that co-occurrences of ideological keywords across parties reveal information about the closeness of parties and ideologies. Keywords that often occur in the same parties (e.g., socialism vs. social democracy) should be closer in space than keywords that rarely occur together (e.g., socialism vs. conservatism). Likewise, parties sharing the same ideological keywords should occupy more similar positions in political space than parties sharing few keywords. To capture these dependencies, we formulate a model in which parties are more likely to get tagged with keywords that are close to them. In a second step, we extend the model to address the fact that some keywords on Wikipedia are inherently ordered.

¹ Similar efforts have been made in order to obtain left-right scores for actors other than parties, for example voters, political candidates, and industries (e.g., Bonica, 2014; Temporão et al., 2018).

² Wikipedia is the largest knowledge database in the world. The English Wikipedia alone has over 5.8 million articles (<https://en.wikipedia.org/wiki/wikipedia> accessed March 2019). Wikipedia is also the most used knowledge database in the world with over 500 million unique visitors each month, ranking among the top 10 most visited websites for years (<http://alexa.com/topsites>, accessed March 2019).

Our analysis demonstrates that keyword summaries provided by Wikipedia editors enable valid and reliable inferences about party left-right position. Based on our model of keyword assignment, which allows for misclassification and differences in keyword informativeness, we recover a scale from Wikipedia classifications that conforms with common intuitions of left vs. right. We show that estimates of party position on this scale correlate with ratings of party position from the largest available expert surveys, and most strongly with ratings of general left-right position. We further demonstrate the reliability of our estimates over repeated measurements with Wikipedia data collected months apart. Together, our results indicate that Wikipedia classifications allow for extracting left-right scores comparable to scores obtained via conventional expert coding methods.

Our findings are in line with studies showing that political information on Wikipedia is often factually correct³ (Brown, 2011; Göbel and Munzert, 2019; Poschmann and Goldenstein, 2019), and that Wikipedia can be used for extending political science measurement to a much larger universe of cases (Munzert, 2018). Our results also tie in with studies demonstrating the validity of crowd-sourcing information in political research (e.g., Sumner, Farris and Holman, 2020; Winter, Hughes and Sanders, 2020). In particular, a number of studies find that assessments by large groups of non-expert coders (i.e., crowd-coding) of the content of party manifestos or parliamentary debates can give estimates of political position as good as assessments by expert coders (cf. Benoit et al., 2016; Haselmayer and Jenny, 2017; Horn, 2019; Lehmann and Zobel, 2018).

Compared to existing approaches, a Wikipedia-based approach offers potentially unlimited coverage of parties. To get some sense of the scope afforded by our measure, note that the largest available databases in political science together include over 4,000 unique parties (Döring and Regel, 2019). Extant measures of party left-right position cover only a minority of these. For example, the largest data source in terms of countries covered, the Democratic Accountability and Linkages Project (DALP), provides expert ratings of left-right position for 506 parties from 88 electoral democracies (see Table 1). The data source with the largest party coverage, the Manifesto Project, includes 1,170 parties from 60 countries. The use of Wikipedia allows us to expand that coverage considerably: based on the largest available compilation of parties, drawn from major political science datasets (Party Facts: Döring and Regel, 2019), we identify about 3,900 parties (third column) that possess a Wikipedia article. Of these, about 2,100 (fourth column) contain scalable information

³A number of studies examines the accuracy of Wikipedia vis à vis traditional expert sources (for a review, see Mesgari et al., 2014). Unlike the majority of these studies, which employ a “small-n, every-detail approach” (Brown, 2011, 340) in which one or several field experts scrutinize the entire content of a small and often selective sample of Wikipedia articles, our analysis is more akin to Brown’s “large-n, specific-facts approach” (2011, 340).

Table 1: Coverage of the largest data source on parties (Party Facts), the largest data sources on party left-right positions (MAP: Manifesto Project, DALP: Democratic Accountability and Linkages Project, CHES: Chapel Hill Expert Survey), and our measure

Continent	Party Facts	Wikipedia	Scalable	MAP	DALP	CHES
Africa	672	601	261	6	68	0
Americas	854	696	381	90	103	0
Asia	871	732	375	216	109	10
Europe	1916	1712	1079	836	214	467
Oceania	170	144	51	22	12	0
TOTAL	4,483	3,885	2,147	1,170	506	477

on party ideology. With the steady expansion of Wikipedia, this number is likely to grow.

While our focus in this paper is on Wikipedia data, the measurement approach we pursue is applicable more generally. For example, selected experts or hired crowd-coders could be tasked with the indexing of parties with keywords. The keywords could be predetermined by the researcher or chosen by the coders. Aside from parties, researchers could also scale other actors, be it politicians, judges, or interest groups. Our approach should be applicable whenever political entities are being indexed with keywords that relate them to some unobserved latent dimension.

2 Party ideology classifications on Wikipedia

We focus on information from ‘party infoboxes’ (see Figure 1 for an example). Infoboxes are placed at the top of a Wikipedia article and give a quick summary of facts on a topic. According to Wikipedia guidelines, they

“contain important facts and statistics of a type which are common to related articles. For instance, all animals have a scientific classification (species, family and so on), as well as a conservation status. Adding an [infobox] to articles on animals therefore makes it easier to quickly find such information and to compare it with that of other articles.”⁴

Unlike the textual format of a basic Wikipedia page, infoboxes restrict editors to submitting information for a defined set of categories; for parties, these are, for instance, the party’s name, its

⁴<https://en.wikipedia.org/wiki/Help:Infobox> (accessed March 2019)

La République En Marche!

From Wikipedia, the free encyclopedia

La République En Marche !^[a] (frequently abbreviated **REM**, **LRM** or **LRM**, officially **LaREM**; possible translation: "The Republic on the move!"), sometimes called **En Marche !** (French: [ɑ̃ maʁʃ]; English translation: "Forward!",^{[7][8]} "Onward!",^[9] "Working!" or "On The Move!"),^[10] is a centrist^[11] liberal^[12] and social-liberal^[13] political party in France. It was founded on 6 April 2016 by Emmanuel Macron, a former Minister of Economy, Industry and Digital Affairs, who was later elected President of France in the 2017 presidential election with 66.1% of the second round vote. It considers La République En Marche! a progressive movement, uniting both the left and the right.^[14]

The party ran candidates in the 2017 legislative elections^[15] including dissidents from the Socialist Party, The Republicans and minor parties. It won an absolute majority in

La République En Marche !

LA RÉPUBLIQUE En Marche !

Executive Officer	Vacant
General Secretary	Stanislas Guerini
President in the National Assembly	Gilles Le Gendre
President in the Senate	François Patriat
Founder	Emmanuel Macron
Founded	6 April 2016; 2 years ago
Headquarters	63, Rue Sainte-Anne 75002 Paris, France
Youth wing	Les Jeunes avec Macron
Membership (2019)	▲ 411,892 claimed adherents ^[1]

Ideology	Liberalism ^{[2][3]} Social liberalism ^[4] Pro-Europeanism ^[5]
Political position	Centre ^[2]
European affiliation	None

Figure 1: Infobox in the Wikipedia article for La République En Marche! and associated tags

founding year, or the name of its leader. The categories of a party infobox are fixed and cannot be altered by editors.⁵ Their content should be “comparable”, “concise”, “relevant to the subject”, and “already cited elsewhere in the article”.⁴ While the categories of a party infobox are pre-defined, their use is optional. Hence, some categories, or even the entire infobox, may be missing from a party’s Wikipedia article.⁴

Of interest to us are the infobox entries (henceforth tags) provided in the categories “ideology” (henceforth ideology) and “political position” (henceforth lr-position). As can be seen from Figure 1, these categories enable Wikipedia editors to index parties with political philosophies and inclinations, and to establish links between their respective Wikipedia pages. For example, the French party La République En Marche!, formed by Emmanuel Macron in the run-up to his bid for presidency, is tagged with the ideologies liberalism, social liberalism, and pro-Europeanism, with links given for all tags. In terms of lr-position, the party is tagged as centre, including a link to an associated Wikipedia page.

There are no restrictions (that we know of) on which or how many ideology tags a party can have. Our data suggests, however, that editors use tags sparingly, and that they often draw on

⁵The party infobox template is protected, see: https://en.wikipedia.org/wiki/Template:Infobox_political_party (accessed March 2019).

existing tags rather than generating new ones. The large majority of parties in our data only receive between one and four ideology tags. Furthermore, some tags are assigned much more often than others (see Figures 2 and 3).

In the political position category, editors mostly draw upon a set of seven tags for classifying a party as far-left, left-wing, centre-left, centre, centre-right, right-wing, or far-right (see Online Appendix E for details). Compared to ideology tags, lr-position tags tend to get used somewhat less often: among all parties that exhibit a tag, only 4 out of 5 have an lr-position tag, while nearly all (97 percent) have an ideology tag. Roughly 3 out of 5 parties receive one lr-position tag and roughly 1 out of 5 parties receive two such tags (see Figure 2). In the latter case, the assigned tags nearly always represent adjacent positions on the political spectrum. Only very few parties receive more than two lr-position tags.

3 A model of tag assignment

We assume that tag assignment is driven by similarity: a party gets tagged if its platform is perceived to be in agreement with the tagged ideology’s basic tenets. We further assume that agreement between party platforms and ideologies can be represented as distances on a latent dimension. Let $i = 1, 2, \dots, N$ be an index of parties, let $j = 1, 2, \dots, J$ be an index of ideology tags, and let y be an N by J matrix with binary entries y_{ij} indicating the occurrence of tag j in party i ’s Wikipedia article. We represent these occurrences by an ideal point model similar to the one suggested in Lowe (2008),

$$\Pr(y_{ij} = 1) = F(\alpha_j - \beta_j(o_j - x_i)^2), \quad (1)$$

where F is the inverse logit transformation, x_i and o_j are the positions of party i and ideology j on the latent dimension, β_j is a tag-specific discrimination parameter, and α_j is a tag-specific constant.

As can be seen, the probability of observing tag j in party i is maximized when their positions on the latent dimension coincide (i.e., when $x_i = o_j$). Thus o_j is the modal value of the response function and the maximum probability at o_j is $F(\alpha_j)$. The discrimination parameter β_j measures how strongly the probability of observing tag j in party i depends on their closeness on the latent dimension, i.e., how rapidly the probability of observing j decreases as we move away from o_j . A high value of β_j implies a peaked response curve in which a party’s probability of exhibiting tag

j decreases quickly with increasing distance from j ; a low value implies a wide response curve in which the probability that a party exhibits tag j does not depend much on its position on the latent dimension. Tags with low β are thus less informative about party position. If $\beta_j = 0$, tag j occurs with probability $F(\alpha_j)$, regardless of party position. Parameters β and α thus allow tags to differ in their informativeness and prevalence.

The above ideal point model is closely connected to extant approaches to scaling word frequencies in political manifestos.⁶ Lowe (2008, 365f.) discusses how the Wordscores and Wordfish approaches can be subsumed under an ideal point model with the same parametric structure as Eq. (1). Our application of the model is different in that we only observe binary word presences and absences.⁷ The main substantive difference to the commonly-used Wordfish approach is that Eq. (1) allows words to have distinct locations. As explained in Lowe (2008), Wordfish implements a reduced version of the ideal point model that allows word usage only to rise or fall along the latent dimension. By contrast, we assume that keywords (i.e., ideologies) occupy positions on the underlying dimension such that their occurrence may rise and fall, i.e., parties may be too far to the right as well as too far to the left for being tagged with a particular ideology. Word frequency scaling approaches typically also include constant terms for the manifestos to correct for the obvious fact that longer manifestos contain more words (Slapin and Proksch, 2008). We do not apply such a correction as parties do not vary much in their propensity of being tagged: inspection of the marginal distribution shows that the number of tags per party varies little (see Figure 2) and is small compared to the number of available tags (see Figure 3 for the most prominent ones).

As explained in the previous section, Wikipedia offers two kinds of tags: ideology and lr-position. Unlike ideology tags, lr-position tags directly encode regions (i.e., far-left, left-wing, centre-left, etc.) on the underlying scale. We shall treat these tags as coarse, graded indicators of party position in one of seven consecutive intervals along the latent dimension and assume that a party gets tagged if its position is perceived to fall within the interval implied by the tag. To achieve this, we represent the assignment of lr-position tags by an ordered logit model with unknown x (Caughey and Warshaw, 2015; Treier and Jackman, 2008). Numbering tags from left to right, and letting $z_i = k$ denote the presence of tag $k = 1, 2, \dots, 7$ in party i s Wikipedia article,

⁶Similar approaches can also be found in the literatures on item response theory (i.e., unidimensional unfolding; see, e.g., Andrich, 1988) and community ecology (i.e., Gaussian ordination; see, e.g., ter Braak, 1985; ter Braak and Šmilauer, 2015).

⁷In Online Appendix A we elaborate on this connection by showing that Eq. (1) and the model studied in Lowe (2008) are special cases of the same generalized linear model.

$$\Pr(z_i = k) = F(\tau_k - \gamma x_i) - F(\tau_{k-1} - \gamma x_i), \quad (2)$$

where γ is a discrimination parameter and τ are tag-specific cut points with $\tau_0 = -\infty$, $\tau_7 = \infty$ and $\tau_{k-1} < \tau_k$ for all k .

Our ordered logit model in Eq. (2) is essentially a one-item version of the graded response model from item response theory (Samejima, 1969). In this model, the probabilities of tag assignment are single-peaked, except for the left- and rightmost tags. The model parametrizes these outcome probabilities via $\Pr(z_i \leq k) = F(\tau_k - \gamma x_i)$, the cumulative probability of observing tag k or lower (i.e., the probability that party i is tagged as k or further to the left). These cumulative probabilities define a set of binary logit models, each with its own intercept τ_k and a common slope parameter γ . The slope parameter γ measures how rapidly the probability of tag assignment changes in response to party position on the latent dimension. Values of γ further from zero imply steeper cumulative response curves and more peaked outcome response curves.

The intercepts τ_k define a set of adjacent intervals on the latent dimension, corresponding to the outcome values. The boundaries of these intervals can be interpreted as the points at which the cumulative probabilities are tied, and they are given by τ_k/γ , for $k = 1, 2, \dots, 6$. For example, a party located at $x = \tau_2/\gamma$ has $F(\tau_2 - \gamma x) = F(0) = 0.5$, and thus a 50 : 50 chance of being tagged as left-wing or something further to the left, as opposed to being tagged as centre-left or something further to the right. By construction, the interval boundaries are defined with respect to the cumulative responses, not the observed responses. However, the boundaries are also related to the observed responses in that the probability of outcome k peaks in the middle of the interval associated with it.

As explained in the previous section, for some parties we observe more than one lr-position tag. In these instances we treat each outcome value as an independent realization of z , conditional on x , and model their joint probability. Formally, this means that instead of z_i in Eq. (2), we model the outcome variable z_{il} , where $l = 1, 2, \dots, 7$ indexes a party's first, second, etc. observed lr-position tag.

In what follows, we pursue two approaches to recovering party positions from Wikipedia articles:

1. Estimating Eq. (1) with data on ideology tag assignment.
2. Jointly estimating Eqs. (1) and (2) with data on ideology and lr-position tag assignment.

The first approach recovers the positions of parties and tags solely from their co-occurrences without any prior assumptions about their locations on the latent dimension. This approach has applicability beyond Wikipedia. It can be used whenever political entities are indexed with a set of keywords, following a logic of ‘pick any(-thing that applies)’ (Levine, 1979).

When applied to Wikipedia tags, a downside of the approach is that it discards the additional information contained in lr-position tags. Since we know the ordering of these tags, their relative locations on the latent dimension do not need to be estimated. This allows us to constrain the estimation problem and recover the locations of parties and tags with greater precision. The fact that we know the ordering of lr-position tags also helps with the identification of the latent dimension. To see why, note that a scaling of ideology tags alone can reveal such a dimension only if neighboring tags occur jointly in some parties. The inclusion of lr-position tags makes this precondition unnecessary: even if ideology tags occurred in completely separate (i.e., non-overlapping) groups of parties, their ordering could still be inferred from the lr-position tags with which they co-occur.⁸ The key assumption behind combining ideology and lr-position tags in this way is that both sets of tags can be placed on the same latent dimension. As a simple check of this assumption, we compare the placement of ideology tags under both estimation approaches. Strong differences in results would be an indication that the two sets of tags might not form a common scale.

4 Estimation

We employ Bayesian Markov Chain Monte Carlo (MCMC) methods to obtain a posterior distribution for all model parameters (Albert and Chib, 1993). MCMC simulations are performed with JAGS, Version 4.3.0 (Plummer, 2017). JAGS code used to estimate both models is provided in Online Appendix C.

4.1 Identification and parametrization

Like all latent variable models, Eqs. (1) and (2) are not identified without some restrictions on the parameters. Three constraints are necessary to identify all parameters in a unidimensional model (Rivers, 2003). To resolve invariance to addition, we center the underlying scale at 0; to resolve invariance to multiplication, we standardize the scale to units of $SD(x)$; and to resolve invariance

⁸We thank an anonymous reviewer for pointing out this particular fact to us.

to reflection, we impose the constraint $\bar{x} < o_m$, where m is the index value for the tag conservatism. The identifying restrictions are imposed on each posterior draw, with offsetting transformations on the other parameters to keep outcome probabilities unchanged (see Online Appendix B for further details).

To facilitate convergence to the target distribution, we estimate a re-parametrized version of Eq. (1). As shown in Lowe (2008, 366), Eq. (1) can be equivalently stated as

$$\Pr(y_{ij} = 1) = F(\delta_j + \lambda_j x_i - \beta_j x_i^2), \quad (3)$$

where $\delta_j = \alpha_j - \beta_j o_j^2$ and $\lambda_j = 2\beta_j o_j$, and where the negative sign on β follows from the assumption of concavity (i.e., the response function must be single peaked, not single dipped). Eq. (3) has the same number of parameters as Eq. (1) but is quadratic only in x , while Eq. (1) is quadratic in x and o . In practice, we find that parametrizing the model as in Eq. (3) yields faster convergence to the target distribution both in terms of iterations and runtime.⁹ We therefore use Eq. (3) as our estimation equation. To obtain estimates of the parameters in Eq. (1), we transform posterior draws of λ , δ , and β into posterior draws of α , β , and o , using the re-parametrization relations stated above, and subsequently apply the identifying restrictions.

4.2 Priors

We assign standard normal priors to x (Albert and Johnson, 1999), and normal priors with mean zero and variance 5 to δ and λ . To enforce the concavity constraint in Eq. (3), we assign log-normal priors with log-mean zero and log-variance parameter 2 to β , which implies a prior variance on β of about 47. To see what these priors mean, consider the prior variation of o relative to x (cf. Clinton and Jackman, 2009, 601-2): Monte Carlo simulation shows that the prior variances on λ and β imply a prior 95% credibility interval for o of about $[-9.6, 9.6]$. Party positions are thus a priori interior to ideologies with the prior variance of ideologies being large relative to that of party positions. Since the locations of ideologies and parties are only identified relative to each other, because the scale of the latent dimension is unknown, the wide range of ideologies relative to party positions suggests that the priors are permissive enough to allow the data to inform the estimation result. Using wider priors on ideologies leaves results unchanged (but slows down convergence).

⁹Following Bafumi et al. (2005, 176), we use potential scale reduction factors to compare convergence under both parametrizations for a given number of iterations.

Likewise, using somewhat tighter priors also yields similar results.

For γ in the ordered logit model, we use a normal prior with mean zero and variance 25. For the category cut points we choose normal priors with mean zero and variance 25, subject to the order constraint $\tau_1 < \tau_2 < \dots < \tau_6$.

4.3 Starting values

We employ correspondence analysis (CA) to generate starting values for party positions. CA is a deterministic dimension-reduction technique, similar to principal components analysis (Greenacre, 2010). Our use of it is motivated by a result in ter Braak (1985) proving that first-dimension coordinates from a CA of the binary data matrix yield approximate maximum likelihood (ML) estimates of x and o if the data-generating process adheres to Eq. (1).¹⁰ This property of CA has made it particularly popular in scaling applications involving sparse data matrices (i.e., matrices with many more zeroes than ones), which are common in other fields (see, e.g., Smith and Neiman, 2007; ter Braak and Šmilauer, 2015) and which we also encounter in our application. ML estimators tend to be numerically unstable in these situations, while CA is guaranteed to yield a solution regardless of how large or sparse the data matrix is (ter Braak and Šmilauer, 2015). Moreover, CA tends to approximate ML more closely when the data are sparse (ter Braak, 1985, 863).¹¹ In our MCMC estimation approach, numerical instability is not an issue; however, convergence to the target distribution can be slow with sparse data. To facilitate convergence, we follow the proposal in ter Braak (1985) and use CA estimates of x and o as starting values. Online Appendix G compares our initial CA estimates to the final Bayesian estimates.

To generate starting values for the remaining parameters, we first estimate logistic regression models—one for each ideology—of the form given in Eq. (1) using CA estimates of o and x as inputs. This gives us estimates of α and β . We then transform those estimates, using the re-parametrization relations given in section 4.1, to obtain starting values for δ and λ . To obtain starting values for γ and τ , we estimate the ordered logistic regression model in Eq. (2) using CA

¹⁰The result of ter Braak (1985) also applies if the observed data are frequency counts rather than presences or absences. Lowe (2008) introduces correspondence analysis to the literature on political text scaling by showing its similarity to the Wordscores method, as well as discussing its properties as an estimator of word and document scores.

¹¹That said, CA estimates can never be equal to ML estimates and they show some well-known biases. In one dimension, the CA solution compresses the ends of the scale, pulling estimates of x and o that lie at the boundaries toward the middle of the scale (ter Braak, 1985, 864). This defect does not take away the usefulness of CA in identifying the underlying dimension, but it highlights the added value from estimating the full parametric model in Eq. (1) (on this point, see Lowe, 2008, 369).

estimates of x as inputs.

4.4 Convergence

We set up four parallel chains and run each model for 22,000 iterations, discarding the first 2,000 draws as burn-in. We thin the result, keeping every 10th draw, to obtain 8,000 samples from the posterior distribution. Assessment of posterior draws via traceplots and potential scale reduction factors suggest convergence of all parameters to their target distribution (see Online Appendix D for details).

5 Selection of parties and tags

Collecting information from Wikipedia requires a list of parties that defines our target population. To maximize coverage, we draw on the largest list of parties that is currently available in political science: the Party Facts database (version 2019a; Döring and Regel, 2019; Bederke, Döring and Regel, 2019).

Party Facts is a collaborative project that aims to solve the problem of delineating the universe of political parties. It offers an authoritative reference list of relevant parties in the world, based on the parties included in the CLEA, ParlGov, Manifesto Project, PolCon and a number of smaller datasets. Party Facts includes nearly all parties that won at least 5% seat share in a national election, as well as parties with at least 1% vote share for some countries. Of all political science datasets it currently provides the largest coverage of political parties worldwide, with over 4,500 parties and urls to the Wikipedia pages of about 3,900 of these (see Table 1).

We collect all ideology and lr-position tags for parties that have a Wikipedia url in Party Facts and an infobox on Wikipedia. Data collection took place on April 27, 2019. We consider only tags that are linked to an associated Wikipedia article and code a party as tagged with the ideology or lr-position to which the link refers (see Online Appendix E for further details).

To get some insight into the variety of tags and their usage, Figures 2 and 3 give a breakdown of the raw numbers. The first thing to note is that tags are used sparingly. As Figure 2 shows, Wikipedia editors typically use one to four tags to describe party ideology; only a small fraction of parties show more than seven ideology tags. For lr-position tags, the modal frequency is one but a considerable number of parties also receive two such tags. A closer inspection of these latter cases

Figure 2: Number of tags per party

reveals that editors almost always assign adjacent lr-position tags to parties (e.g., centre-right and right-wing).

In addition to being used sparingly, tags are also used unequally: the 20 most often-used ideology tags, which make up less than 5 percent of all observed tags, account for over 50 percent of total tag usage. Among this core of widely applied tags, we find well-known, major ideologies such as liberalism, conservatism, and social democracy (see Figure 3).

To reduce the potential for bias and inaccuracy in our data, we focus on tags that are widely used on Wikipedia, as such tags should get more exposure and, as a result, be better known and more easily scrutinized than tags referring to rare or obscure ideologies. To achieve this, we implement a threshold of 50 tag occurrences. Tags observed in fewer parties are omitted from the analysis. This restriction allows us to rule out narrow ideologies that depend on national context (e.g., Basque nationalism), as well as peculiar ones (e.g., eurocommunism) about which only a small number of editors probably know enough to be able to apply them adequately, and spot and correct mistakes.¹² For parties, we choose an inclusive threshold of at least two tags. In sum, this gives us scalable information for more than 2,100 parties (see Table 1).

6 The resulting scale

We begin by inspecting the scale that we obtain for its face validity. Figures 4 and 5 summarize the estimation result via the estimated response curves and party positions for Models 1 and 2. Each response curve indicates the probability that a particular tag is assigned to a party with a

¹²We employ the 50-tag threshold to be conservative. Our results remain substantively and quantitatively the same if we include tags that are used fewer than 50 times.

Figure 3: Frequency of usage for tags that are used at least 20 times

given position on the underlying dimension. For example, a party at position 0 in Figure 5 has an estimated probability of about 60 percent of being tagged with liberalism, about 30 percent probability of being tagged with social liberalism, about 20 percent probability of being tagged with pro Europeanism, etc.¹³ The use of politically informative keywords thus lends a substantive interpretation to areas of the scale based on the ideologies that are particularly prominent there. In turn, this is reflected in party positions.

Figures 4 and 5 generally support the idea of a global left-right dimension underlying Wikipedia classifications of party ideology. Ideologies that are traditionally associated with the labels ‘left’ and ‘right’ (e.g., communism, socialism, social democracy, liberalism, conservatism, national conservatism) line up in the familiar order. These ideologies’ response curves are also steep in the sense that their probability of being assigned to a party rises and falls quickly along the underlying spectrum. The ideologies thus help discriminate parties on the underlying dimension. By contrast, ideologies such as agrarianism or regionalism have flat response curves and thus contribute less information on where to place parties on the underlying dimension. The fact that major ideologies line up in the familiar order suggests that party position estimates can be interpreted in terms of left vs. right. In other words, the scale has face validity.

We emphasize that we obtain substantively the same result, regardless of whether we include lr-position tags or not (compare Figures 4 and 5). Ideology tags alone are thus sufficient to obtain a meaningful left-right scale. However, the fact that the assignment of ideology tags follows a familiar left-right reasoning suggests that Wikipedia’s lr-position tags can be used to tighten our inferences about the underlying scale (besides allowing us to include even more parties in the estimation) by construing them as ordered indicators of adjacent intervals on the underlying dimension—the assumption that gives rise to Model 2.

As Figure 5 shows, the inclusion of lr-position tags leaves the ordering of ideologies largely unchanged but helps distinguish some ideologies more clearly from one another. For example, social democracy and democratic socialism are now separated more clearly, with the former falling firmly into the centre-left bracket and the latter into the left-wing bracket. Likewise, Christian democracy now clearly peaks left of conservatism and within the centre-right bracket. Lastly, the inclusion of lr-position tags reduces the probability of tag assignment for most ideology tags, especially for some on the far-right end (i.e., right-wing populism, national conservatism, etc.).

¹³Point estimates for the locations of all ideologies (i.e., the locations where the probability curves peak) along with credible intervals are given in Online Appendix F.

Figure 4: Response curves and estimated party positions (indicated by tick marks) from a scaling of ideology tags only; 1,367 parties and 27 tags (some response curves are unlabeled to avoid clutter)

Figure 5: Response curves for ideology tags, estimated intervals for lr-position tags, and estimated party positions (indicated by tick marks) from a joint scaling of ideology and lr-position tags; 2,147 parties and 35 tags (some response curves are unlabeled to avoid clutter)

This is due to the higher frequency with which lr-position tags are observed compared to most ideology tags, as well as their patterns of co-occurrence. In particular, the rightmost ideology tags often occur in conjunction with the far-right tag as well as the right-wing tag. This leads to more equal probabilities of far-right and right-wing parties being tagged with one of these ideologies.

7 Validation with expert surveys

Having established our estimates' face validity, we now consider their convergent and discriminant validity (cf. Adcock and Collier, 2001). Accordingly, a valid measure should correlate highly with other measures of the same construct, while showing lower correlations with measures of related but different constructs.

To test this, we draw on expert ratings of party left-right position as well as party position on other dimensions from the two largest existing expert surveys, the Democratic Accountability and Linkages Project (Kitschelt, 2014) and the Chapel Hill Expert Survey (CHES) (Polk et al., 2017). The DALP covers 506 parties from 88 electoral democracies worldwide. Ratings were collected between 2008 and 2009 and may not be entirely accurate with respect to parties that have recently altered their position significantly. Nevertheless, no other expert survey provides broader country coverage. From the CHES we use the most recent set of ratings, which were collected in 2014 and cover 268 parties from 31 European democracies.

Regarding convergent validity, Figure 6 shows that our estimates correlate well with expert ratings of party left-right position. Estimates from Model 2 generally fit expert ratings more closely than those of Model 1, suggesting that the inclusion of lr-position tags adds useful information above and beyond ideology tags. Table 2 further shows how the correlation between our scores and expert ratings varies over countries. Since expert surveys ask for placements of parties only within a given polity (i.e., each expert rates parties from his or her country), pooling of ratings across countries may introduce measurement error due to differential item functioning (Hare et al., 2014; Struthers, Hare and Bakker, 2019).¹⁴ Comparing estimates on a country-by-country basis, the median correlation is above 0.9 and the 25th percentile is above 0.75 in all four comparisons. Thus, for three-quarters of the countries covered by either the DALP or the CHES, our estimates of party positions accord well—and for half the countries covered, they accord very well—with the

¹⁴For the CHES, Bakker et al. (2014) show that differential item functioning in experts' ratings of parties' economic left-right position is small.

Figure 6: Comparison of party position estimates to expert ratings from the Democratic Accountability and Linkages Project (top row) and from the 2014 Chapel Hill Expert Survey (bottom row); Model 1: $N = 321$ and $N = 203$; Model 2: $N = 435$ and $N = 247$

Table 2: Country-wise correlations with expert ratings (means and percentiles)

		Mean	10%	25%	50%	75%	90%
DALP	Model 1	0.72	0.26	0.77	0.91	0.97	1.00
	Model 2	0.74	0.42	0.76	0.90	0.97	0.99
CHES	Model 1	0.83	0.66	0.79	0.90	0.96	0.98
	Model 2	0.90	0.75	0.90	0.94	0.97	0.99

judgment of country experts.

Regarding discriminant validity, we compare our estimates to expert ratings of party position on economic policy, redistribution, GAL-TAN, and identity politics. As Table 3 (first column) shows, experts’ perceptions party left-right position are substantially correlated with their perceptions of party stances on all other measured dimensions. This suggests that each of the more specific dimensions reflects some aspect of left vs. right. Reassuringly, our estimates also correlate with every other dimension measured in expert surveys. However, the correlations remain weaker than those with left-right, suggesting that our estimates most likely represent general left-right position, as opposed to more specific policy stances.

Closer inspection of Table 3 reveals further nuances. While experts’ judgments of party left-right position relate most strongly to their perceptions of parties’ stances on economic or redistributive policy, Wikipedia-based estimates are more strongly associated with expert ratings of parties’ positions on GAL-TAN or identity politics than on economic policy. With the inclusion of lr-position tags, correlations increase somewhat between our estimates and ratings of parties’ economic and redistributive stances (compare Model 1 and Model 2) but, overall, our scores remain slightly more associated with social and identity politics. Although small, these differences suggests that Wikipedia editors’ understanding of left vs. right is driven more by value considerations than by economic considerations, whereas experts tend to view left vs. right somewhat more in terms of economic policy and redistribution than social policy.

8 Reliability

As content on Wikipedia is open to instant updating and revision by anyone, the raw data is constantly evolving. Any day, new tags might get added to a party’s infobox, others deleted. This raises the question of how robust our estimation is to Wikipedia’s openness: Can we rely on estimates of

Table 3: Correlations with expert ratings on other dimensions

		Expert L-R	Model 1	Model 2
DALP	Left-Right	1.00	0.75	0.79
	Economy	0.72	0.47	0.53
	Redistribution	0.67	0.44	0.49
	GAL-TAN	0.54	0.53	0.54
	Identity	0.65	0.65	0.65
CHES	Left-Right	1.00	0.84	0.90
	Economy	0.81	0.53	0.63
	Spend-Tax	0.77	0.51	0.61
	GAL-TAN	0.70	0.77	0.75

party positions from a single point in time, when the raw data is subject to constant revision (i.e., additions, refinements, mistakes, and corrections of tags)?

To gauge this uncertainty, we estimate party positions using ideology and lr-position tags observed at different points in time. Specifically, we compare our current scores to scores obtained in the same way using data collected some 15 weeks earlier (on January 11, 2019). The assumption is that during this short period of time the underlying party positions remain the same. Changes in tags will be due to editors adding new classifications or modifying what they see as incorrect classifications, either through deletion or replacement. Given Wikipedia’s 500 million monthly visitors, there should be ample opportunity for observing such changes in the given time period. By comparing the positions of parties that experience a change in their tags we gain some insight on the sensitivity of our measure to ongoing editing activity. A reliable measure should yield similar position estimates despite overt variation in tags.

Across the 15-week period considered, we find that 111 and 279 of the parties included in the estimation of Models 1 and 2 see a change in their tags. Comparing those parties’ estimated scores, we find a close correspondence between them (Figure 7). The average absolute difference between scores is only 0.20 and 0.17, respectively. If we take a difference of 0.5 as the minimum for a substantial change—a difference of 0.5 roughly corresponds to a move from one left-right bracket to the next (see Figure 5)—we find that 10 and 13 parties, respectively, experience such a change in their positions (see Appendix H for a detailed inspection of these cases). The strong correlation between both sets of scores indicates a high test-retest reliability.

The fact that changes in tags lead to small variation in party positions suggests that the observed

Figure 7: Reliability of party position estimates. The x-axis shows party position estimates obtained from Wikipedia classifications collected 15 weeks prior to those on which our current estimates are based, which are shown on the y-axis. A 90-degree line is superimposed. Model 1: $N = 111$; Model 2: $N = 279$

edits are mostly refinements rather than sweeping revisions of parties' infobox classifications.¹⁵ Perhaps this result should not be too surprising. After all, if there was little stability in parties' infobox tags, we would not expect a scaling of tags collected at an arbitrary point in time to yield a strong and coherent left-right dimension. Nonetheless the result is reassuring. Monitoring the evolution of party scores in the long run will, of course, be necessary for maintaining confidence in the measure.

9 Generalizing beyond expert surveys

Parties differ greatly in how much attention they receive by the public, with some parties being much better known than others. This could affect how much attention and scrutiny they receive on Wikipedia. Prominent parties are likely to attract greater interest from Wikipedia editors, and with more resources going into the creation and monitoring of those parties' articles, their tags are

¹⁵For example, adding the tag liberal conservatism to a party tagged with liberalism and conservatism would be a refinement that does not alter the party's estimated position to any significant degree.

likely to be more accurate than those of less-well known parties.¹⁶ Our above comparisons with expert ratings might thus represent a best-case scenario, as expert surveys tend to include the most important parties in a polity. For parties not included in expert surveys, the accuracy of our scores might be lower.

To assess this possibility, we formulate two hypotheses. Hypothesis one is that greater scrutiny of a party's Wikipedia article improves the accuracy of its estimated left-right score. Hypothesis two is that articles of parties included in expert surveys are scrutinized more than those of other parties. If both hypotheses hold true, we can expect our scores to be less accurate for parties not included in expert surveys, and more so the stronger the observed relationships are. We use two indicators for an article's level of scrutiny (see Online Appendix I for details): its number of editors; and the (average) experience of its 50 most active editors, where experience is defined as an editor's total number of live edits on Wikipedia. For comparison, we also include two indicators of party prominence in the analysis: article pageviews and vote share as recorded in Party Facts.

To assess the first hypothesis, we study the closeness of fit between our scores and the expert ratings shown in Figure 6. If more editors and greater editor experience improve the accuracy of information on a party's Wikipedia page, one would expect the party's estimated position to correspond more closely to the experts' views and hence be closer to the regression line than that of other parties. We employ a linear regression model with variance heterogeneity (King, 1998; Verbyla, 1993) to examine this conjecture (see Online Appendix J for a description of the model). This approach allows us to model the residual variation around the regression line as a function of additional covariates, thus providing a natural way to study correlates of inaccuracy in party scores.

We find that more editors and greater editor experience are associated with a tighter fit between our scores and both the DALP and the CHES ratings (see Table 4). By contrast, pageviews and vote share are not consistently associated with greater accuracy of party scores (the correlation between pageviews and the number of editors is 0.79). Similar results hold if we test each predictor in isolation. In both models editor experience shows the strongest effect: the standard error of the regression decreases by 8 percent (i.e., by a factor of $\exp(-0.084)$) and by 10 percent, respectively, for a one-unit increase in (the log of) editor experience. Given the observed ranges and variances on all the variables (see Table 5), editor experience is thus associated with the biggest reduction in regression error.

¹⁶Another possibility for tags of less well-known parties to be less accurate is that less information is available about their ideology. This is a generic problem that arises with any measure of party position, including expert surveys.

Table 4: Predictors of party position estimates (DALP: N = 430; CHES: N = 244)

		DALP		CHES	
		Coef	SE	Coef	SE
Mean	constant	−1.815	0.076	−1.843	0.065
	left-right expert	0.330	0.013	0.360	0.011
Variance	constant	0.063	0.313	−0.053	0.372
	vote share	0.165	0.167	−0.419	0.383
	pageviews (log)	−0.001	0.031	0.021	0.040
	number of editors (log)	−0.023	0.035	−0.041	0.038
	editor experience (log)	−0.084	0.031	−0.107	0.034

To assess the second hypothesis, Table 5 compares parties covered either by the CHES or the DALP surveys to all other parties that have a Wikipedia page. Consistent with the fact that expert surveys deliberately select the most important parties in a polity, we find that the parties included in expert surveys tend to have a higher vote share and receive more pageviews, on average, than other parties on Wikipedia. Consistent with the notion of greater interest, these parties also attract more editors, on average, than other parties on Wikipedia. For editor experience, we find the opposite pattern with parties not included in expert surveys having more seasoned article editors, on average. Apart from these differences, we find that the two samples are surprisingly similar in terms of variation and spread. In fact, the samples overlap to a large extent on each of the indicators.

The results suggests that both groups of parties have something speaking for them. Parties included in expert surveys tend to attract more editors than other parties on Wikipedia. Yet, their editors tend to be less experienced. Since editor experience contributes to the accuracy of tag assignment, Wikipedia-based scores need not be less accurate for parties not covered by expert surveys. For example, we would expect in-survey parties' scores to be 5 percent more accurate, on average, based on their higher number of editors, but we would expect out-of-survey parties' scores to be 9 percent more accurate, on average, based on their editors' greater experience.¹⁷ The observed correlation between a party's number of editors and the average experience of its editors is about -0.66 , so a party with fewer editors tends to have more experienced editors and vice versa. This as well as the large overlap between the two samples on each of the relevant indicators gives us reason to expect a Wikipedia-based measure to not perform worse (than expert surveys would)

¹⁷This calculation uses the coefficient estimates from the CHES model in Table 4 and the differences in sample means from Table 5.

Table 5: Comparing parties included in expert surveys (N = 610) to other parties on Wikipedia (N = 3,275)

	In survey	Mean	SD	Min	25%	75%	Max
vote share	Yes	0.21	0.19	0.00	0.06	0.31	1.00
vote share	No	0.17	0.22	0.00	0.03	0.21	1.00
pageviews (log)	Yes	7.03	2.20	0.00	5.84	8.46	12.64
pageviews (log)	No	5.49	1.83	0.00	4.30	6.69	11.91
editors (log)	Yes	4.27	2.05	0.00	3.50	5.63	8.51
editors (log)	No	3.09	1.59	0.00	2.20	4.14	7.98
editor experience (log)	Yes	7.25	1.55	3.50	6.21	7.97	12.43
editor experience (log)	No	8.10	1.63	3.09	6.92	9.16	12.63

in estimating positions for parties for which we currently lack such estimates.

10 Conclusion and outlook

Party positions are central to explanations of politics. In this paper we entertained the hypothesis that semi-standardized classifications of party ideology provided on Wikipedia carry valid and reliable information about party position. We stipulated an ideal point model that allows us to exploit variation in ideological keywords across parties and obtained scores of party position on a latent dimension whose substantive interpretation fits common intuitions of left vs. right. In line with our hypothesis, we found party scores to correlate well with independent expert judgments of party left-right position and we found them to be reliable in a test-retest scenario. This demonstrates that our approach yields a novel measure of party left-right position.

As a proof of concept, our results hold great promise for future research. Party ideology and the notion of ideological differences between parties are key causal factors in explanations of the political process. They feature prominently in theories of political change and stability, legislative decision making, coalition politics, public policy, economic growth, inequality and redistribution, as well as in accounts of democratization, state building, and violent conflict. Conversely, parties' ideological positions are the object of study in many research areas, from electoral competition to the quality of representation to political fragmentation and polarization. With its potentially unlimited coverage, our measure of party position opens up new possibilities for researchers to study these and other topics on a much larger universe of cases. This includes many parties for

which we currently lack ideological placements, as well as new countries not covered by extant data sources.

Our measure is not limited to parties that currently exist. Unlike many expert surveys, our scaling includes parties of the past, provided they have a Wikipedia article, and provided Wikipedia editors assign to these parties some of the same tags that they assign to other parties. In addition, our Wikipedia-based approach may provide estimates for parties founded just recently, which expert surveys may not (yet) cover. One database that does include past parties and also regularly updates its sample is the Manifesto Project. Compared to manifesto-based measures of party position, our measure does not depend on the existence of an electoral manifesto. This allows for the inclusion of parties that do not produce such documents, including nascent parties, non-democratic parties, or parties in regimes with restricted electoral competition.

References

- Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3).
- Albert, James H and Siddharta Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(422):669–679.
- Albert, James H. and Valen E. Johnson. 1999. *Ordinal Data Modeling*. New York: Springer.
- Andrich, David. 1988. "The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude." *Applied Psychological Measurement* 12(1):33–51.
- Bafumi, Joseph, Andrew Gelman, David K. Park and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13:171–187.
- Bakker, Ryan, Seth Jolly, Jonathan Polk and Keith Poole. 2014. "The European Common Space: Extending the Use of Anchoring Vignettes." *Journal of Politics* 76(4):1089–1101.
- Bederke, Paul, Holger Döring and Sven Regel. 2019. "Party Facts – Version 2019a." *Harvard Dataverse* doi: 10.7910/DVN/FTQAYT.
- Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver and Slava Mikhaylov. 2016. "Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110(2):278–295.
- Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2):367–386.
- Bräuninger, Thomas, Jochen Müller and Christian Stecker. 2016. "Modeling Preferences Using Roll Call Votes in Parliamentary Systems." *Political Analysis* 24(2):189–210.
- Brown, A. R. 2011. "Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage." *PS: Political Science and Politics* 44(2):339–343.
- Caughey, Devin and Christopher Warshaw. 2015. "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model." *Political Analysis* 23(2):197–211.
- Clinton, Joshua D. and Simon Jackman. 2009. "To Simulate or NOMINATE?" *Legislative Studies Quarterly* 34(4):593–621.
- Däubler, Thomas and Kenneth Benoit. 2017. *Estimating Better Left-Right Positions Through Statistical Scaling of Manual Content Analysis*. Working paper: University of Mannheim.
- Döring, Holger and Sven Regel. 2019. "Party Facts: A Database of Political Parties Worldwide." *Party Politics* 25(2):97–109.

- Göbel, Sascha and Simon Munzert. 2019. *The Comparative Legislators Database*. Working paper: University of Konstanz and Hertie School of Governance.
- Greenacre, Michael J. 2010. "Correspondence Analysis." *Wiley Interdisciplinary Reviews: Computational Statistics* 2(5):613–619.
- Hare, Christopher, David A. Armstrong II, Ryan Bakker, Royce Carroll and Keith T. Poole. 2014. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59(3):759–774.
- Haselmayer, Martin and Marcelo Jenny. 2017. "Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding." *Quality & Quantity* 51(6):2623–2646.
- Horn, Alexander. 2019. "Can the Online Crowd Match Real Expert Judgments? How Task Complexity and Coder Location Affect the Validity of Crowd-Coded Data." *European Journal of Political Research* 58(1):236–247.
- King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: Michigan University Press.
- Kitschelt, Herbert. 2014. *Dataset of the Democratic Accountability and Linkages Project (DALP)*. Duke University.
- Lauderdale, Benjamin E. and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24:374–394.
- Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(2):311–331.
- Lehmann, Pola and Malisa Zobel. 2018. "Positions and Saliency of Immigration in Party Manifestos: A Novel Dataset Using Crowd Coding." *European Journal of Political Research* 57(4):1056–1083.
- Levine, Joel H. 1979. "Joint-Space Analysis of "Pick-Any" Data: Analysis of Choices from an Unconstrained Set of Alternatives." *Psychometrika* 44(1):85–92.
- Lo, James, Sven Oliver Proksch and Thomas Gschwend. 2014. "A Common Left-Right Scale for Voters and Parties in Europe." *Political Analysis* 22:205–223.
- Lowe, Will. 2008. "Understanding Wordscores." *Political Analysis* 16:356–371.
- Mesgari, Mostafa, Chitu Okoli, Mohamad Mehdi, Finn Arup Nielsen and Arto Lanamäki. 2014. "'The Sum of All Human Knowledge': A Systematic Review of Scholarly Research on the Content of Wikipedia." *Journal of the Association for Information Science and Technology* 66(2):219–245.
- Munzert, Simon. 2018. "Measuring the Importance of Political Elites." *SocArXiv* doi: 10.31235/osf.io/t8gs5.

- Peterson, Andrew and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26:120–128.
- Plummer, Martyn. 2017. *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. Version 4.3.0.*
- Polk, Jonathan, Jan Rovny, Ryan Bakker, Erica Edwards, Liesbet Hooghe, Seth Jolly, Kostelka Filip, Gary Marks, Gijs Schumacher, Marco Seenbergen, Milada Vachudova and Marko Zilovic. 2017. "Explaining the Salience of Anti-Elitism and Reducing Political Corruption for Political Parties in Europe with the 2014 Chapel Hill Expert Survey Data." *Research and Politics* January-March:1–9.
- Poschmann, Philipp and Jan Goldenstein. 2019. "Disambiguating and Specifying Social Actors in Big Data: Using Wikipedia as a Data Source for Demographic Information." *Sociological Methods and Research* doi: 10.1177/0049124119882481.
- Rivers, Douglas. 2003. *Identification of Multidimensional Spatial Voting Models.* Typescript: Stanford University.
- Samejima, Fumiko. 1969. "Estimation of Latent Ability Using a Response Pattern of Graded Scores." *Psychometrika Monograph Supplements* 17.
- Slapin, Jonathan B. and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.
- Smith, Karen Y. and Fraser D. Neiman. 2007. "Frequency Seriation, Correspondence Analysis, and Woodland Period Ceramic Assemblage Variation in the Deep South." *Southeastern Archeology* 26(1):47–72.
- Struthers, Cory L., Christopher Hare and Ryan Bakker. 2019. "Bridging the Pond: Measuring Policy Positions in the United States and Europe." *Political Science Research and Methods* doi:10.1017/psrm.2019.22.
- Sumner, Jane Lawrence, Emily M. Farris and Mirya R. Holman. 2020. "Crowdsourcing Reliable Local Data." *Political Analysis* 28(2):244–262.
- Temporão, Mickael, Corentin Vande Kerckhove, Clifton van der Linden and Julien M Hendrickx. 2018. "Ideological Scaling of Social Media Users: A Dynamic Lexicon Approach." *Political Analysis* 26:457–473.
- ter Braak, Cajo J. F. 1985. "Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model." *Biometrics* 41(4):859–873.
- ter Braak, Cajo J. F. and Petr Šmilauer. 2015. "Topics in Constrained and Unconstrained Ordination." *Plant Ecol* 216:683–696.
- Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.

- Verbyla, A. P. 1993. “Modelling Variance Heterogeneity: Residual Maximum Likelihood and Diagnostics.” *Journal of the Royal Statistical Society. Series B (Methodological)* 55(2):493–508.
- Volkens, Andrea, Judith Bara, Ian Budge, Michael McDonald and Hans-Dieter Klingemann. 2013. *Mapping Policy Preferences from Texts III: Statistical Solutions for Manifesto Analysts*. Oxford: Oxford University Press.
- Winter, Nicholas J. G., Adam G. Hughes and Lynn M. Sanders. 2020. “Online Coders, Open Codebooks: New Opportunities for Content Analysis of Political Communication.” *Political Science Research and Methods* 8(4):731–746.

Online Appendix

Party Positions from Wikipedia Classifications of Party Ideology

Michael Herrmann

University of Konstanz

Holger Döring

University of Bremen

A Connections to other scaling models

Our model of ideology tag assignment is related to extant models for word counts in party manifestos through the following generalized linear model

$$\text{link}(E(y_{ij})) = a_j + b_j x_i + c_j x_i^2, \quad (4)$$

where x is an unobserved latent variable, a , b , c are unknown parameters, and y is a matrix of either

1. word (j) presences and absences in party infoboxes (i), or
2. word (j) counts in party manifestos (i).

Case 1 yields our model of tag assignment: Since, for binary data, $E(y_{ij}) = \Pr(y_{ij} = 1)$, our model as stated in Eq. (3) is equivalent to Eq. (4) with a logit link function and the constraint $c_j < 0$.

Case 2 yields the model in Lowe (2008) and the Wordfish approach: Applying a log link function and constraining $c_j < 0$, Eq. (4) becomes Eq. (12) in Lowe (2008, 366). As shown there, the Wordfish model results from replacing $c_j < 0$ with $c_j = 0$.

B Identification

The identifying restrictions stated in section 4.1 imply the following offsetting transformations on the other parameters. Let the unidentified parameters be denoted with a star and the identified parameters without. Consider first Eq. (1). For identification to leave outcome probabilities unchanged, we require that

$$\alpha_j - \beta_j(o_j - x_i)^2 = \alpha_j^* - \beta_j^*(o_j^* - x_i^*)^2.$$

Substituting the identifying restrictions $x_i = (x_i^* - \bar{x}^*)/\text{SD}(x^*)$ and $o_i = (o_i^* - \bar{x}^*)/\text{SD}(x^*)$, the left-hand side of the above relation becomes

$$\alpha_j - \beta_j(o_j^* - x_i^*)^2/\text{SD}(x^*)^2,$$

from which we deduce that $\beta_j = \beta_j^*\text{SD}(x^*)^2$ and $\alpha_j = \alpha_j^*$ are the offsetting transformations required for identification.

Consider now Eq. (2). For outcome probabilities to remain unchanged, we require that

$$\tau_k - \gamma x_i = \tau_k^* - \gamma^* x_i^*,$$

which, after imposing the identifying restriction $x_i = (x_i^* - \bar{x}^*)/\text{SD}(x^*)$, implies the offsetting transformations $\gamma = \gamma^*\text{SD}(x^*)$ and $\tau_k = \tau_k^* + \gamma^*\bar{x}^*$.

C JAGS code

The input data for Model 1 are y , a binary matrix with parties i in rows and ideology tags j in columns, D , a vector containing the dimensions of y , and con , the column number of the tag conservatism.

```
1  model {
2
3    # likelihood
4    for (i in 1:D[1]) {
5      for (j in 1:D[2]) {
6        logit(pi[i, j]) <- delta[j] + lambda[j] * xstar[i] - beta[j] * xstar[i] ^ 2
7        y[i, j] ~ dbern(pi[i, j])
8      }
9    }
10
11   # priors
12   for (i in 1:D[1]) {
13     xstar[i] ~ dnorm(0, 1)
14   }
15   for (j in 1:D[2]) {
16     delta[j] ~ dnorm(0, 0.2)
17     lambda[j] ~ dnorm(0, 0.2)
18     beta[j] ~ dlnorm(0, 0.5)
19   }
20
21   # transformation and identification of posterior draws
22   xbar <- mean(xstar)
23   sdx <- sd(xstar)
24   ocon <- lambda[con] / (2 * beta[con])
25   polarity <- ifelse(xbar < ocon, 1, -1)
26   for (i in 1:D[1]) {
27     x[i] <- polarity * (xstar[i] - xbar) / sdx
28   }
29   for (j in 1:D[2]) {
30     a[j] <- delta[j] + beta[j] * (lambda[j] / (2 * beta[j])) ^ 2
31     b[j] <- beta[j] * sdx ^ 2
32     o[j] <- polarity * ((lambda[j] / (2 * beta[j])) - xbar) / sdx
33   }
34 }
```

The primary input data for Model 2 are y_ideo , a binary data matrix with parties i in rows and ideology tags j in columns, and y_1r , a matrix with two columns. Column $y_1r[, 2]$ is a stacked

vector of lr-position tags (coded 1 through 7) that are observed for each party. The vector is stacked to allow for multiple observations of lr-position tags within parties. A corresponding index vector `y_lr[, 1]` identifies each party. Auxilliary inputs `D` and `con` are as stated in Model 1.

```

1  model {
2
3    # likelihood
4    for (i in 1:D[1]) {
5      for (j in 1:D[2]) {
6        logit(pi[i, j]) <- delta[j] + lambda[j] * xstar[i] - beta[j] * xstar[i] ^ 2
7        y_ideo[i, j] ~ dbern(pi[i, j])
8      }
9    }
10   for (n in 1:N) {
11     mu[n] <- gamma * xstar[y_lr[n, 1]]
12     y_lr[n, 2] ~ dordered.logit(mu[n], tstar[1:6])
13   }
14
15   # priors
16   for (i in 1:D[1]) {
17     xstar[i] ~ dnorm(0, 1)
18   }
19   for (j in 1:D[2]) {
20     delta[j] ~ dnorm(0, 0.2)
21     lambda[j] ~ dnorm(0, 0.2)
22     beta[j] ~ dlnorm(0, 0.5)
23   }
24   for (k in 1:6) {
25     tau[k] ~ dnorm(0, 0.04)
26   }
27   tstar[1:6] <- sort(tau)
28   gamma ~ dnorm(0, 0.04)
29
30   # transformation and identification of posterior draws
31   xbar <- mean(xstar)
32   sdx <- sd(xstar)
33   ocon <- lambda[con] / (2 * beta[con])
34   polarity <- ifelse(xbar < ocon, 1, -1)
35   for (i in 1:D[1]) {
36     x[i] <- polarity * (xstar[i] - xbar) / sdx
37   }
38   for (j in 1:D[2]) {
39     a[j] <- delta[j] + beta[j] * (lambda[j] / (2 * beta[j])) ^ 2

```

```

40     b[j] <- beta[j] * sdx ^ 2
41     o[j] <- polarity * ((lambda[j] / (2 * beta[j])) - xbar) / sdx
42   }
43   for (k in 1:6) {
44     t[k] <- tau[k] - gamma * xbar
45   }
46   g <- polarity * gamma * sdx
47 }

```

D Assessing parameter convergence

We ran four chains for 20,000 iterations each, after a burn-in of 2,000 iterations. Keeping every 10th posterior draw from each chain, we obtain 8,000 posterior samples for each parameter. Given the large number of parameters, we assess potential non-convergence via scale reduction factors (see below), with additional inspection of traceplots for tag parameters (not shown).

Tables D.1 and D.2 give a summary of the scale reduction factors $\hat{R}$ for all parameters in Models 1 and 2. By inspection, there are no signs of non-convergence. Scale reduction factors are generally close to one and no value exceeds 1.1. Tables D.3 and D.4 give an indication of the efficiency of the MCMC samplers for each model. Effective samples sizes suggest that 8,000 posterior samples yield sufficient information to quantify parameter uncertainty at conventional levels, in particular for parameters x and o .

Table D.1: Scale reduction factors $\hat{R}$ for parameters of Model 1

	Min	Mean	Max	50%	75%	90%	95%	99%	Npar
α	1.00	1.00	1.01	1.00	1.01	1.01	1.01	1.01	27
β	1.00	1.01	1.02	1.00	1.01	1.01	1.02	1.02	27
o	1.00	1.00	1.01	1.00	1.00	1.01	1.01	1.01	27
x	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1,367

Table D.2: Scale reduction factors $\hat{R}$ for parameters of Model 2

	Min	Mean	Max	50%	75%	90%	95%	99%	Npar
γ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1
τ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	6
α	1.00	1.01	1.08	1.00	1.01	1.03	1.04	1.07	28
β	1.00	1.00	1.03	1.00	1.00	1.02	1.02	1.02	28
o	1.00	1.01	1.08	1.00	1.02	1.03	1.04	1.07	28
x	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2,147

Table D.3: Effective sample sizes for parameters of Model 1

	Min	Mean	Max	1%	5%	10%	25%	50%	Npar
α	333	3,583	8,000	339	359	443	978	2,440	27
β	196	2,120	8,000	206	243	267	501	1,638	27
o	795	4,524	8,000	822	923	1,141	2,164	3,729	27
x	890	6,366	8,000	1,546	2,242	2,903	4,668	8,000	1,367

Table D.4: Effective sample sizes for parameters of Model 2

	Min	Mean	Max	1%	5%	10%	25%	50%	Npar
γ	5,321	5,321	5,321	5,321	5,321	5,321	5,321	5,321	1
τ	2,856	5,350	8,000	2,875	2,950	3,044	3,670	5,007	6
α	288	5,159	8,000	293	532	963	3,093	5,298	28
β	101	4,319	8,000	117	164	175	1,181	3,734	28
o	414	5,691	8,000	447	608	1,168	3,757	7,454	28
x	1,016	6,675	8,000	1,914	2,779	3,472	5,355	8,000	2,147

E Data collection and coding of tags

As stated in the main text (section 5), we consider only tags that include a link to an associated Wikipedia article and we code a party as tagged with the ideology or lr-position to which the link refers.

Technically, a tag offers two pieces of information: a link and an associated label that is presented in the infobox. Our coding of tags focuses on the targets of links rather than their labels. On their own, labels are merely words that editors can choose freely in order to describe a party's ideology. This lack of constraint in labels can make it difficult or even impossible for other users to verify their meaning and accuracy. Links resolve this difficulty by requiring editors to commit to already existing content on Wikipedia. Their use ensures that the veracity of tags in a party's infobox can be assessed and verified by other Wikipedia users. To reduce measurement error, we therefore use link targets instead of labels.

Links sometimes refer to Wikipedia urls that redirect to another Wikipedia article. For example, the urls for 'liberalist' and 'liberal politics' both redirect to the Wikipedia article Liberalism. In such cases, the final destination article is what we code as a party's tag.

We use Python and R scripts to extract the desired information from Wikipedia and to create a dataset for analysis. Our data collection proceeds in five steps:

1. We retrieve the entire infobox content (all categories) from parties' Wikipedia articles.
2. We create a dataset that stores the downloaded information and the raw Wikitext markup code for all infoboxes.
3. We select all entries for the infobox categories ideology and position and extract their links.
4. We download the titles of the Wikipedia articles (including redirect targets) to which the links refer.
5. We create a stacked dataset with the following variables: party, category (ideology vs. position), link, tag (i.e., article title) from which we construct all input data used in the estimation of Models 1 and 2.

To distinguish ideology tags from lr-position tags, we employ the following definition: tags referring to the Wikipedia articles Far-left politics, Left-wing politics, Centre-left politics, Centrism,

Table E.5: Tags that are observed in the infobox category on political position

	position	ideology
centre-right politics	612	0
centre-left politics	578	1
centrism	455	66
right-wing politics	422	0
left-wing politics	337	1
far-right politics	166	1
far-left politics	131	0
big tent	28	49
syncretic politics	6	5
third position	5	3
radical centrism	3	5
left of center (turkey)	1	1
pro-europeanism	1	217
third way	1	25

Centre-right politics, Right-wing politics, or Far-right politics are lr-position tags. All other tags are ideology tags. For ease of exposition, we refer to lr-position tags as far-left, left-wing, centre-left, centre, etc. (i.e., we refer to their labels), rather than using their full article titles in Figures 3 and 5.

Our above definition is consistent with the modeling assumption that lr-position tags encode successive intervals on an underlying continuum (see section 3). It is also consistent with where we find tags in infoboxes. Table E.5 lists all the tags that we observe in the infobox category on political position. It shows their respective frequency in that category, as well as their frequency in the ideology category. We find that lr-position tags, as defined above, are used almost exclusively in the position category of an infobox (with the exception of centrism). Conversely, ideology tags, as defined above, are used almost exclusively in the ideology category, as the table lists only seven such tags.

An alternative way of defining ideology and lr-position tags would be solely with reference to their infobox category. As Table E.5 shows, such a definition would lead to inconsistencies. Tags used across both categories would have to be treated in the estimation as ideology as well as lr-position tags, depending on their category. The tag centrism in particular is used in both the ideology and position category in a total of 36 parties. Furthermore, tags in the lower half of Table E.5 do not refer to successive intervals on the underlying dimension, which makes them incompatible with our assumptions about the ordering inherent in lr-position tags (see section 3).

Treating these tags as ideology tags allows us to estimate their locations on the underlying scale freely.

F Point estimates for ideologies

Figure F.1: Point estimates and 95% credibility intervals for the locations (i.e., optima) of ideology and position tags. Panel A shows tags whose 95% credibility intervals lie within the range of party positions (-2.5 to 2.5), as indicated by the dashed lines; panel B shows tags whose estimated 95% credibility intervals lie beyond the range of party positions.

G Comparison to CA estimates

Figure G.2 compares our Bayesian estimates to estimates obtained through correspondence analysis of the data matrix. For parties from left to center both approaches yield very similar results, and CA apparently gets better at distinguishing center-right parties from more right-wing parties, once lr-position tags are included (notice the longer linear slope in the right-hand panel). Yet, CA does not satisfactorily distinguish right-wing and far-right parties, ‘squeezing’ them together relative to left-of-center parties. Figure G.3 shows that CA nevertheless yields useful estimates of party positions that successfully distinguish left from right. The overall correlation between CA estimates and expert ratings is not much worse than for our Bayesian estimates, especially after including lr-position tags. Yet, the correlation does not hold up for both sides of the political spectrum. In particular for right-of-center parties with an expert score of 6 or higher, the correlation disappears.

Figure G.2: Comparison of party position estimates to their starting values, i.e., estimates obtained via correspondence analysis

Figure G.3: Comparison of CA estimates to expert ratings of party positions: left-hand panels show results based on ideology tags only (i.e., the data used to estimate Model 1); right-hand panels show results based on ideology and lr-position tags (i.e., the data used to estimate Model 2)

H Inspection of changes in tags

In the main text (section 7) we conclude that edits to tags are mostly refinements rather than sweeping revisions of a party's ideology classification. Here we examine whether the larger deviations that we saw in the test-retest might reflect the opposite, namely controversy and divergent views among editors over a party's ideology. We consider the 13 parties mentioned in section 7 whose position changed by more than 0.5 units on the latent dimension (i.e., a move from one left-right bracket to the next).

First, we consider the identity of the tags that changed. If there is controversy over tags, we might expect most disagreement over the assignment of extreme tags such as far right, far left, right-wing populism, or socialism. Among the 13 parties with the strongest change in position, the tags that are being added or removed most often are centre right (6 times), right wing (5), Christian democracy (3), liberalism (3), and pro-Europeanism (2). All other tags only change in one party (see below).

Second, we examined for each party which of its tags did and did not change. To assess whether editors diverge in their views on a party's position, we compare the ideological range of the party's tags at both points in time. The ideological range is the difference between a party's rightmost and leftmost tags. We take an increase in the ideological range over time as evidence for potential divergence in editors' views. To obtain the ideological range, we treat the point at which a tag's response curve peaks as its location (see Figure 5 or Figure F.1). For lr-position tags, this point is the midpoint of each left-right interval. For tags whose estimated location falls outside the range of party positions, we substitute the values -2.5 and 2.5 . Our assessment of each party's tags is given below. For reference, we also give each party's estimated positions and expert rating if available. Overall, we find more indication of convergence than divergence of tags over time.

Convergence of tags (ideological range decreases):

- The Slovenian National Party exhibits the largest change in estimated position (from 0.10 to 1.81). It loses the tag left wing, while gaining Euroscepticism, right-wing populism, social conservatism, and right wing. Its other tag is far right. The range of tags decreases. The party's DALP expert survey score is 7.9 on a scale from 1 to 10.
- The Alliance for Change (of Guyana) (positions: -0.91 and -0.30) loses the tag socialism. Its other tags are nationalism and progressivism. The range of tags decreases.

- The Association for the Rebirth of Madagascar (positions: -1.08 and -0.11) gains the tag liberalism. Its other tags are nationalism and socialism. The range of tags decreases.
- The German Social Union (of the former German Democratic Republic), a historical party (positions: 0.96 and 1.46), loses the tags Christian democracy and centre right. Its other tags are national conservatism, social conservatism, and right wing. The range of tags decreases.
- The New Liberal Alliance (of Denmark) (positions: 0.18 and 0.72) loses the tag liberalism, while gaining right wing. Its other tags are Euroscepticism and centre right. The range of tags decreases. The party's DALP expert survey score is 7.8 on a scale from 1 to 10 ; its CHES score is 7.9 on a scale from 0 to 10 .
- The People's Front for Democracy and Justice (of Eritrea) (positions: -1.38 and -0.83) loses the tag Marxism-Leninism. Its other tags are left wing nationalism, socialism, secularism and big tent. The range of tags decreases.
- The Popular Action (of Peru), a historical party (positions: 0.34 and -0.22), loses the tags centre right and nationalism, while gaining the tags big tent, liberalism, progressivism, and social democracy. Its other tag is centre. The range of tags decreases.
- The Revolutionary Nationalist Movement (of Bolivia) (positions: 0.32 and 0.87) loses the tag left-wing nationalism. Its other tags are centre right and social conservatism. The range of tags decreases. The party's DALP expert survey score is 6.5 on a scale from 1 to 10 .
- The Rastakhiz Party (of Iran), a historical party (positions: 0.89 and -0.10), loses the tags anti-communism and right wing. Its other tags are populism and secularism. The range of tags decreases.
- Volya (of Bulgaria) (positions: 1.07 and 1.75) loses the tag centre right, while gaining the tags right wing and far right. Its other tags are Euroscepticism and right-wing populism. The range of tags decreases.

Divergence of tags (ideological range increases):

- The League of Polish Families exhibits the second largest change in estimated position (2.02 to 0.88). It gains the tags Christian democracy, conservatism, pro Europeanism, and centre

right. Its other tags are Euroscepticism, social conservatism, and far right. The range of tags increases. The party's DALP expert survey score is 9.3 on a scale from 1 to 10. Inspection of the party's Wikipedia article shows that its tags are divided into "current" and "historical". According to the article, the party moved from a far right toward a more centrist position in 2010 (after the DALP survey was administered). Our approach currently ignores this additional information, lumping all tags together. The "change" in estimated position is thus due to measurement error rather than disagreement between editors.

- Change 90 (of Peru), a historical party (positions: 1.54 and 0.92), loses the tag right wing, while gaining the tags centre right and Christian democracy. Its other tags are right-wing populism and social conservatism. The range of tags increases.

Change of editor perception:

- The Liberal People's Party (of Norway) (positions: 0.56 and -0.16) loses the tags conservative liberalism, economic liberalism, and centre right, while gaining the tags pro-Europeanism, social liberalism, and centre. No tags remain unchanged. In both cases the range of tags is small. The changes are not indicative of starkly opposing views of the party's position (i.e., its position shifts from about center right to the center of the scale).

I Collection of metadata on Wikipedia articles

The metadata on Wikipedia articles that we present in the main text (section 9) was retrieved from Wikipedia using the XTools interface. R scripts were used to collect the data. We considered all parties for which the Party Facts database records a Wikipedia url. For each of these parties' articles, we requested

1. its number of pageviews,
2. its number of editors from the earliest available date until the day of tag data collection, excluding bots and spiders, and
3. the names of its top 50 editors from the earliest available date until the day of tag data collection, excluding bots and spiders.

For each of the ($> 100k$) editors identified in step 3, we then retrieved the total number of live edits made on Wikipedia up until the date of tag data collection. To compute our measures, we take the log of each count, adding 1 to pageviews to avoid $\log(0)$. For our measure of editor experience, we compute the average of the log number of live edits over each party's top 50 editors.

J Modeling the accuracy of Wikipedia-based estimates

As stated in the main text (section 9), we assess the accuracy of our scores by modeling how closely they approximate the expert ratings. For simplicity, we assume that error in expert ratings is negligible so that all measurement error rests with our scores. If both measures tap into the same basic dimension, they should linearly map into each other. Deviations from the linear mapping then represent measurement error in our scores. We formalize these assumptions with the following model

$$\hat{x}_i \sim \text{Normal}(\mu_i, \sigma_i), \quad (5)$$

$$\mu_i = \mathbf{v}e_i, \quad (6)$$

$$\sigma_i = \exp(\eta z_i), \quad (7)$$

where $\hat{x}_i$ is a party's Wikipedia-based position estimate (by construction, these estimates are normally distributed), $\mathbf{v}e_i$ is a scalar product that linearly maps a party's expert rating into its Wikipedia-based position estimate with $\mathbf{v}$ being a two-element parameter vector and e_i a vector containing a party's expert rating (taken either from the DALP or the CHES survey) and a leading 1; likewise ηz_i is a scalar product of further parameters η and additional covariates z_i including a leading 1.

The proposed model is a linear regression with variance heterogeneity (King 1998, Verbyla 1993). Of particular interest to us is Eq. (7), which conditions the standard error of the regression on additional covariates. This feature allows us to study to what extent measurement accuracy in our scores varies with attributes of the parties' Wikipedia articles such as their number of editors or editor experience. We obtain estimates of $\mathbf{v}$ and η via maximum likelihood.

K Software statement

Python 3.7

- <https://github.com/siznax/wptools>
- <https://github.com/5j9/wikitextparser>
- <https://pandas.pydata.org/>

R 3.5

- `crch`
- `curl`
- `data.table`
- `ggrepel`
- `glue`
- `jagsUI`
- `jsonlite`
- `MASS`
- `plyr`
- `rio`
- `stargazer`
- `tidyverse`
- `xtable`

JAGS 4.3.0

- <http://mcmc-jags.sourceforge.net/>