Political Science Research and Methods (2026), 1–20
doi:10.1017/psrm.2026.10138

ORIGINAL ARTICLE

Coalition differentiation at scale: measuring intra-coalition rhetorical distinctiveness with supervised classification

Christine Sheldon

Bennett School of Public Policy, University of Cambridge, Cambridge, UK
Email: crs80@cam.ac.uk

(Received 16 August 2025 UTC; revised 27 March 2026 UTC; accepted 29 May 2026 UTC)

Abstract

Coalition parties face a unity-differentiation dilemma: they must uphold compromise to govern effectively, while maintaining distinct identities to maximize electoral success. Existing measures of coalition differentiation almost exclusively rely on static proxies to infer behavior, and the few dynamic measures often only cover a limited sample of governments, limiting comparative analysis. This article introduces a comparative, dynamic measure of rhetorical differentiation based on supervised classification of legislative speeches. Higher classification accuracy indicates greater rhetorical distinctiveness between coalition partners, while misclassification indicates greater rhetorical overlap. The measure provides monthly party-level scores for close to 50 coalition governments, across six European democracies, and thereby enables the dynamic analysis of coalition party differentiation in a public-facing arena, for a sizable comparative sample.

Keywords: coalition differentiation; coalition governance; machine learning; parliamentary speech

1. Introduction

Political parties joined together in government are subject to two competing forces. In order to enjoy incumbency benefits, it is imperative for parties to maintain coalition compromise. Yet simultaneously, once the coalition terminates, planned or otherwise, each party is accountable to the electorate as a distinct entity. This, in turn, is an incentive to publicly distinguish oneself during the incumbency instead. Boston and Bullock (2010) identify these orthogonal incentives as the unity differentiation dilemma. When navigating this dilemma, parties are restricted in the extent to which they can differentiate. Throughout incumbency, coalition parties are bound by collective cabinet responsibility (Fortunato, 2021). This responsibility "stipulates that the government present a united front on all policy matters, *particularly* after the policy has passed the parliament" (Fortunato, 2021, p. 93). As such, the legislative arena provides the main opportunity for coalition parties to differentiate themselves publicly without directly violating collective cabinet responsibility. Even during this time, however, differentiation comes with a price. Effective, and thus united, coalition governance and differentiation are mutually exclusive. Emphasizing differences over coalition compromise can make governing together more challenging and publicly expose tensions between coalition partners (Fortunato, 2021). As such, whether differentiation occurs can provide insight into the dynamics of coalition cooperation and government functioning over time.

© The Author(s), 2026. Published by Cambridge University Press on behalf of EPS Academic Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psrm.2026.10138

By far, the most common practice for capturing how conflictual or differentiated a coalition government is consists of measuring the likelihood that disagreement or divergence will occur during incumbency. This includes, for instance, quantifying when institutional design allows ministers to submit legislative proposals closer to their preferences, rather than the coalition consensus, referred to as ministerial drift (Martin and Vanberg, 2011); the number of parties in a coalition, as an indication of bargaining complexity (Warwick, 1994; Zubek, 2015); or the ideological distance between parties based on their electoral manifestos, signifying the size of the compromise required for cooperation (Carroll and Cox, 2012; Dörrenbächer *et al.*, 2015; Klüver and Spoon, 2017, 2020; Klüver and Bäck, 2019; König *et al.*, 2022; Meyer *et al.*, 2023). These approaches assume that the incentive, and therefore the propensity for differentiation, is determined *a priori* and remains stable throughout the incumbency. Lupia and Strøm (2010), however, emphasize that coalition party behavior during incumbency is dynamic. The utility of behavioral choices is continuously assessed. The choice to differentiate is no exception. A party's decision to act on either of the mutually exclusive incentives of the unity-differentiation dilemma is the result of a careful weigh-off between the costs and benefits of each option. Whether it makes sense to prioritize boosting electoral prospects over generating incumbency benefits depends on different variables, many of which are likely to vary over time (such as polling performance (Fortunato, 2019)). The prevailing option is therefore not static or consistent throughout a government's incumbency and instead is likely to vary throughout. This signifies that there is a clear mismatch between the measurement and the actual process of coalition differentiation. Static measures solely capture differentiation at the start of a coalition government, and therefore fail to unveil exactly what differentiation dynamics are at play during incumbency.

Some studies have set out to dynamically measure the degree of differentiation in a coalition. This includes capturing differences between coalition parties in parliamentary applause (Imre *et al.*, 2022), parliamentary speech (Martin and Vanberg, 2008; Fortunato, 2021), and tabled amendments (Fortunato, 2019, 2021). Yet either limitations in data availability or the intensive nature of coding the measures has restricted these efforts to small samples of governments. Considering many outcomes of interest related to coalition dynamics are rare, such as terminations or cabinet reshuffles, inferences made on such small samples do not generalize well.

This paper presents a novel dynamic measure of coalition differentiation, which is easy and consistently applicable across large samples of coalition governments. Using supervised classification of parliamentary speeches delivered by coalition parties, I generate monthly differentiation scores for six Western European democracies. The classification algorithm is first trained to identify the party label of a speech, based on the words used by members of parliament (MPs) and ministers from each respective coalition party. The trained model is then used to make out-of-sample predictions. The resulting performance indicator of this exercise signifies how well the algorithm is able to distinguish parties based on their word use in parliament. If it does so successfully, parties are differentiating through distinct rhetoric, whereas in instances where the classification performs poorly, coalition parties are less distinguishable in their public presentation in parliament. The classification pipeline presented in this paper generates monthly differentiation scores for all parties comprising a coalition government, for 48 governments from six European countries, over the course of 20 years. This method thus generates a dynamic coalition differentiation metric for a substantive sample of governments and can also easily be extended to other party-labeled coalition government corpora.

A battery of validation tests indicates that this measure corresponds to established static proxies of coalition differentiation and division in the expected directions. Additional analyses of pre-termination trends show that rhetorical differentiation tends to increase in the months leading up to discretionary cabinet termination. This suggests that increases in rhetorical differentiation may be associated with discretionary early termination.

https://doi.org/10.1017/psrm.2026.10138

In sum, this measure opens up the analysis of coalition differentiation by capturing behavior over time and disaggregated by party. This can shed light on how differentiation varies across different party types and strategies and with what consequences. It also allows for the examination of how public rhetorical differentiation relates to cabinet instability, and whether observable differentiation functions as an early warning signal of termination. Below, I will elaborate on past measures of coalition differentiation and their limitations, the selection of supervised classification as a suitable similarity measure for this purpose, and the classification pipeline. This is followed by a presentation of the resulting scores, a number of validation tests, and discussion of the contribution of this measure to the study of coalition governance.

2. Measuring coalition differentiation dynamically

A handful of measures attempt to capture the extent to which coalition parties differentiate during incumbency. One of the earliest examples of measuring coalition differentiation is from Martin and Vanberg (2008). These authors code the length of legislative speeches from German and Dutch coalition governments as an indication of differentiation: the longer one speaks, the more opportunity for differentiation. This naturally is a strong assumption to make, and as such, other measures of differentiation consider more explicit party behavior. Fortunato (2019), for instance, considers legislative amendments made by coalition parties to government bills to be a form of differentiation. To capture this process, he codes all amendments tabled by coalition parties to alter legislative proposals submitted by their coalition partners. Imre *et al.* (2022), in turn, consider legislative applause in their measure of coalition mood. This measure infers the extent to which a coalition is cooperative or competitive, through the frequency with which coalition parties applaud each other's speeches in parliament. And finally, Plescia and Kritzinger (2022) generate a comparative measure of coalition conflict based on reporting on the state of coalition governments in both national and international media.

What all these measures have in common is that they are only produced for relatively small samples. Coding amendments is very time-intensive, and hence Fortunato's differentiation measure is only available for three European countries. Applause, in turn, is rarely consistently codified in debate transcripts, restricting Imre *et al.* (2022) to solely German and Austrian coalitions. And Plescia and Kritzinger (2022) stress how "given the relatively low number of events available for each month in each country, the ICEWS data set did not allow us to take into consideration the dynamic aspect of intra-coalition conflict" (Plescia and Kritzinger, 2022, p. 40).

One way to address these challenges of coding and data limitation is to turn to computational text analysis, which has enabled efficient large-scale measurement of political text data. Sagarzazu and Klüver (2017), for instance, conceptualize coalition differentiation as occurring through variations in issue attention. Coalition parties distinguish themselves by addressing distinct topics from their coalition partners. In measuring this, they use a Bayesian hierarchical topic model, developed by Grimmer (2010), to classify German coalition party press releases into policy areas. They then interpret differences in attention devoted to each of these areas as strategic coalition differentiation behavior. Another example comes from Fortunato (2021), who uses Slapin and Proksch (2008) Wordfish model to generate daily ideological positioning scores from parliamentary speech for the parties of the 2010–2015 United Kingdom (UK) coalition. The distance between these generated scores is then interpreted as the degree of differentiation.

Both topic modeling and Wordfish are forms of unsupervised machine classification. This means these methods are trained to recognize patterns, or groupings, in data without being given prior instructions about the nature of these groupings. By locating and interpreting these clusters, one can make inferences about how these are grouped in this space, and what the distance between these clusters indicates. The one key challenge with unsupervised classification is best summarized by Goet (2019), when he states that: "the scores that these models produce, identify which speakers tend to

https://doi.org/10.1017/psrm.2026.10138

use similar words to one another […]. These estimates may or may not prove to have anything to do with the party, or indeed to have any stable structure over different debates or sessions" (p. 7). This implies that the clusters found by an unsupervised classification algorithm do not necessarily correspond with what the researcher sets out to capture. The risk of inconsistencies between results from different applications means careful, case-by-case *post hoc* validation to assess the nature of the clusters is essential. Thus, although these computational methods are able to process large volumes of data with relative ease, the thorough *post hoc* validation to ensure it is, in fact, differentiation between parties as opposed to some other dimension of difference, requires significant time investment. It is, therefore, unsurprising that both unsupervised classification measurements introduced here are still restricted to a single country or government.

Existing measures of differentiation offer only limited and inconsistent coverage across time and cases. A measure that dynamically captures differentiation and is generated for a sizable sample of coalition governments still remains absent. In the rest of this paper, I will argue that the challenges preventing such robust measurement can be overcome by using supervised classification instead.

3. Supervised classification to capture differences

The application of supervised classification algorithms to predict the party label of political text, most notably parliamentary speech, is an increasingly popular method to measure latent variables in text (Yu *et al.*, 2008; Høyland *et al.*, 2014; Peterson and Spirling, 2018; Goet, 2019; Wäckerle and Silva, 2023; Ishima, 2024). In this method, a supervised learning classifier is trained to assign a text to a political party on the basis of the words used in the speech. The trained classifier is then applied to an out-of-sample set of texts to predict their party label. The resulting performance of this classification exercise can be considered an indication of distinctiveness: the better a classifier is able to recognize party affiliation of a speech, the more distinct one party's rhetoric is from the other parties in the sample. Yet the worse it performs, the more alike speeches are, and thus the more similar parties speak. This method, therefore, overcomes the challenge to the reliability of results of unsupervised methods by imposing the dimension of difference: party label. As a result, "these estimates are guaranteed to be driven by the 'party factor,' regardless of the number of topics that are addressed, or other sources of variation in word use" (Goet, 2019, p. 7); e.g., the differences supervised classification picks up on will always be differences based on party groupings. Such a consistent result allows for interpretation of results across time and cases with relative ease.

Previous applications of this method have predominantly been used to measure ideological polarization in the US Congress (Yu *et al.*, 2008), UK Parliament (Peterson and Spirling, 2018; Goet, 2019), and the European Parliament (Høyland *et al.*, 2014). For the purpose of this paper, I combine this literature on measuring ideological polarization with theories on coalition differentiation. By applying a supervised classification algorithm to a subset of solely coalition party speeches, one can interpret the performance metric of the classifier not as ideological polarization but as rhetorical differentiation. More precisely, interpreting the fine-grained differences in word use between parties as ideological polarization relies on the assumption that all speakers in parliament are transparent about their ideological positions in their parliamentary addresses. Yet theories on coalition differentiation stress how members of coalition parties are not always at liberty to do so and instead are required to publicly defend a coalition consensus. From this perspective, one can consider the differences in speech as the extent to which distinct party characteristics actually shine through. When a supervised classifier is trained only on speeches by government parties, its resulting performance is therefore better understood as capturing the degree of rhetorical differentiation between coalition partners than their overall level of ideological polarization.

The differences between parties, which supervised classification algorithms identify, are those in word use. These can be driven by a number of different factors, such as addressing different topics (Sagarzazu and Klüver, 2017), discussing the same topic with different vocabularies (Goet, 2019), or

https://doi.org/10.1017/psrm.2026.10138

even explicitly choosing to (de)emphasize their party membership over coalition membership. This measure does not distinguish which source of variation is driving the results and instead considers all of the above as meaningful and contributing to coalition rhetorical differentiation.[1] Moreover, a core assumption of this method is that coalition parties behave as unitary actors in the legislative arena. Empirical evidence does suggest that access to the parliamentary floor is regulated by party elites, and thus that there is a central strategy to party speech behavior. Proksch and Slapin (2015) find that party elites "still have a choice over which backbenchers should engage in monitoring through debate and could employ loyalist backbenchers for the task" (p. 44). This is also reflected in the findings of Bäck *et al.* (2019), who find that speechmaking is restricted to fewer, more high-ranked legislators as incumbency draws to a close and communication of a consistent party label becomes more important. Taken together, this suggests that parliamentary speech can reasonably be treated as reflecting party-level patterns of communication rather than purely individual-level variation. That being said, legislative discipline is included as a scope condition in the sample selection to follow, and the validity of this unitary actor assumption is assessed in Appendix 3.

4. Method

Unlike traditional supervised classification approaches, the pipeline generating this measurement is not designed to optimize classifier performance. On the contrary, the metric of interest is grounded in misclassification. The objective is to find consistent error driven by linguistic similarities, as opposed to achieving the highest accuracy. The pipeline is designed to do so, and to be easily extended to other parliamentary arenas not included in this sample, as well as applied to party-labeled coalition corpora from other sources.

4.1. Data and pre-processing

The measure is generated for a sample of six countries: Austria, Denmark, Germany, Ireland, the Netherlands, and Sweden. These countries are not only characterized by consistent coalition governance but also have a practice of parliamentary party discipline, which is a scope condition of the sample selection. The speech data from these countries come predominantly from ParlSpeech V2 (Rauh and Schwalbach, 2020), with the exception of Ireland, for which the speech data is from the Database of Parliamentary Speeches in Ireland, 1919–2013 (Herzog and Mikhaylov, 2017). In total, the sample includes 48 coalition governments observed over the last 20 years. Metadata on coalition membership and duration are taken from the Party Government in Europe Database (PAGED) (Bergman *et al.*, 2021).

The speech data is grouped per cabinet and excludes speeches by the speaker of the house, those of the opposition, and any utterances of fewer than 40 words. The remaining corpus includes speeches by both legislators and members of the executive.[2] The corpus is vectorized using a bag-of-words approach, with removal of stop words during TF-IDF (term frequency-inverse document frequency) vectorization using country-specific stopword dictionaries from the Natural Language Toolkit (Bird *et al.*, 2009). TF-IDF vectorization considers how important a word is to a document relative to the full cabinet corpus. This measure of importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the entire corpus, helping to downweight common words and highlight more distinctive terms. This process improves the

[1] Another driver of differences is the gender composition of parties, with women speaking differently to men (Wäckerle and Silva, 2023), or, similarly, regional dialects. Since the classifiers are trained on data from a full cabinet, as described below, I assume these differences remain relatively stable over time within cabinets and, as a result, should not introduce significant amounts of noise in the measure.

[2] Appendix 3 dissects whether conflating these two types of speakers challenges the unitary actor assumption and finds little evidence for concern.

https://doi.org/10.1017/psrm.2026.10138

representation of meaningful terms for the subsequent classification. Though simple, this bag-of-words method of text vectorization is efficient and, moreover, performs comparably to more advanced techniques like document embeddings as displayed in Appendix 1.

One thing to note is the challenge of vocabulary size influencing classifier performance, as noted by Peterson and Spirling (2018) and Gentzkow *et al.* (2019). Specifically, the size of the vocabulary of unique words found within the text for a classification (in this case, one cabinet) can influence the subsequent performance metric of choice. The columns in a TF-IDF matrix correspond to the unique terms found in the corpus. As a result, vectorization of different corpora will result in matrices of varying dimensions. Yet the number of columns will influence the amount of information, as well as the sparsity to be considered in the classification. Both are likely to influence its resulting performance. These structural differences would introduce noise to the signal and make it difficult to capture meaningful variation across cabinets. To account for this, I generate fixed vocabularies of unique words appearing in all the (cleaned) speech data used per country, as per Peterson and Spirling (2018). Using these vocabularies to subsequently generate the TF-IDF matrices for individual cabinets of the same country means every classification feature space has the same number of columns, enabling cross-cabinet comparison. Differences in vocabulary size and structural factors *between* countries remain, yet this can be mitigated by including country fixed effects in multi-country models.

4.2. Classification

4.2.1. Classifier selection

I employ a logistic regression classifier from the Python module scikit-learn (Pedregosa *et al.*, 2011). This model executes both binary and multi-class classification, implying it can be applied to data from coalition cabinets of all sizes. It also performs reliably with high-dimensional, sparse feature matrices typical of text data. The classifier is implemented using the SAGA solver, which processes large volumes of data with efficiency, with L2 regularization to prevent overfitting and address the correlated features characteristic of text data. I also assess the robustness of the resulting classifier performance by comparing it to two alternative classifiers: a linear support vector classifier and a stochastic gradient descent classifier. Even though these models rely on different optimization strategies, the results their scores as reported in Appendix 2 closely correlate with those generated by the logistic regression classification. Since all three classifiers produce substantively similar results, I conclude that they are capturing the same underlying pattern of rhetorical differentiation in parliamentary speech and construct the final scores using solely a logistic regression classifier.

4.2.2. Classification pipeline

To obtain performance metric scores for each month of incumbency, I employ a cross-validation technique. More precisely, for each cabinet, the data is randomly split into 10 stratified samples. Each of the classifiers mentioned above is then trained and tested 10 times, each time training the data on nine of the sections and applying the trained classifier to the remaining section to make out-of-sample predictions. As it stands, the training data is rarely balanced. This implies that the number of speeches per party will differ. To ensure the classifier is trained on roughly equal amounts of information per party, I apply synthetic minority oversampling technique to each of the folds' training sets. This oversampling technique generates synthetic observations in the minority class, or classes, to match the number of observations in the majority class. By replicating observations based on the k-nearest neighbors of observations in the minority class, the class is expanded without adding new information to the model. As a result, all parties in each cabinet will have the same number of speeches fed to the classifier.[3]

[3]This does not account for structural differences in the length of speeches, i.e., if smaller coalition parties systematically speak more succinctly, they would remain underrepresented in the training data even with the same number of speeches. To account for this, I include speech length as an additional feature in the classification.

https://doi.org/10.1017/psrm.2026.10138

Within each fold, the model is trained to recognize the typical speech profile of each coalition party over the course of incumbency. The trained model is then applied to a stratified out-of-sample set of speeches to assess how clearly those speeches correspond to, or diverge from, the party's characteristic speech profile. By doing this 10 times using a cross-validation split, each speech in the corpus is included in a test set once and will be assigned predicted probabilities of belonging to each of the class labels or coalition parties.

4.2.3. Score calculation

I then calculate monthly party scores by aggregating the predicted probabilities per month, for each of the parties comprising the coalition. Previous studies employing this method have typically considered the overall accuracy of a classifier, thereby capturing distinguishability for parliaments as a whole. To measure individual party distinctiveness, I focus on recall instead. Recall, or sensitivity, reflects the proportion of a party's speeches that are correctly classified as its own (Rainio *et al.*, 2024). High recall indicates that a party's rhetorical profile is distinct enough for the classifier to identify it, while low recall suggests lower distinguishability. In this context, false negatives (speeches from a party that are misclassified as belonging to another) are more meaningful than false positives (speeches from another party that are incorrectly attributed to the party in question), as they directly represent moments when a party's rhetoric blends into that of its coalition partners. Party-level recall therefore best captures coalition party differentiation, by illustrating how each party remains rhetorically distinguishable from its partners over time.

Additionally, existing applications of this method have almost exclusively been applied to two-party systems. These applications, therefore, did not compare results across classifications with different numbers of classes. Yet the number of parties, and thus the number of classes in the classification, will affect the baseline recall, or the level of performance at which a classifier is as good as random. In a binary classification, this is at 50%, whereas in a four-class classification, this is 25%. Comparison between coalitions of different sizes is impossible without correcting for this. I do so by using recall surplus as the final measure of rhetorical differentiation. This is the rate exceeding the baseline score of each month.[4] The formula for this measure is:

$$\text{Recall surplus} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} - \frac{1}{n}.$$

Using recall surplus captures the extent to which the classifier performs above chance and makes scores comparable across cabinets of different sizes. Finally, since some scores are substantially noisy for months with very few speeches (such as summer and winter recesses), I drop the scores from months in the bottom 5% of speech frequency. I employ a relative cutoff point, as each country will differ in the number of speeches its parliamentary speech rules will allow for on average.

4.3. What the scores capture and what they do not

The differentiation scores are generated for each coalition party in each month of incumbency. High recall surplus indicates that a party's parliamentary speech is more rhetorically distinguishable from that of its coalition partners, whereas low recall surplus indicates greater rhetorical overlap. The measure therefore captures the extent to which coalition partners are publicly distinguishable from one another in parliamentary speech over time. Important to note is that this measure is relational. A party-level score does not reflect the distinctiveness of a party's own speech in isolation, but rather captures how distinguishable that speech is relative to those of its coalition partners. The scores should therefore be interpreted as indicators of rhetorical differentiation between coalition partners, and not of individual party choice.

[4]This is akin to the established practice of considering classifier performance superseding a baseline (Megahed *et al.*, 2024).

https://doi.org/10.1017/psrm.2026.10138

Table 1. Dashboard summary of scores and sample by country

Country	Scores	Governments	Unique parties	Year range	Original dataset
AT	447	9	3	1996–2018	ParlSpeech V2
DE	563	8	4	1991–2018	ParlSpeech V2
DK	423	9	6	1997–2018	ParlSpeech V2
IE	491	8	6	1992–2013	DPI 1919–2013
NL	638	10	6	1994–2019	ParlSpeech V2
SE	504	4	6	1991–2018	ParlSpeech V2
Total	**3,066**	**48**	**31**		

The classification pipeline identifies only the observable pattern of rhetorical distinctiveness, not the process that produces it. Similar score patterns may arise from multiple underlying mechanisms, which the measure itself does not disentangle. More precisely, multiple processes may produce high or low levels of rhetorical distinctiveness. Differences may arise from varying ideological positioning, issue emphasis, or rhetorical styles. The classification pipeline does not distinguish which of these mechanisms is driving a given score; instead, it captures the overall observable distinguishability in parliamentary speech. Party-level recall surplus should thus be interpreted as a measure of how rhetorically distinguishable each coalition party is in a given month. Averaging these party-level scores then gives a coalition-level differentiation score, representing the degree of rhetorical differentiation within the coalition as a whole at that point in time.

Finally, what constitutes a relatively high or low score is context-dependent. Absolute score levels are shaped not only by coalition speech patterns but also by linguistic, institutional, and parliamentary conventions that are relatively stable within countries but may vary across them. For example, in parliaments where speakers frequently refer to fellow politicians by name, such references will be mechanically associated with particular parties and therefore more easily identifiable by a classifier. This can inflate absolute score levels without necessarily indicating greater rhetorical differentiation in a substantively meaningful sense. Consequently, the measure is best suited to analyzing within-coalition and within-country variation over time.

5. Results

Table 1 summarizes the dataset generated by the classification pipeline set out above. Across six countries, this dataset includes a total of 3,066 monthly coalition party differentiation scores collected from 48 distinct government cabinets, representing 31 unique parties over time periods ranging from 1991 to 2019, depending on the country.

The distinctive attribute of this measurement is the generation of dynamic, time-varying scores for every party in a cabinet. To illustrate the value of this granularity, Figure 1 plots all scores from two successive Austrian governments (Schüssel I and II), both comprising the same parties. In this comparison, the number of parties, as well as the ideological distance between the parties in office, is thus kept approximately constant from one cabinet to the next. As mentioned previously, static cabinet characteristics, such as the size and ideological distance, are popular measures indicating the state of a coalition government. When these characteristics remain relatively stable across successive governments, existing proxies imply little change, yet the present measure shows distinct patterns in rhetorical differentiation from one cabinet to the next.

The illustrative example of Figure 1 displays the two Schüssel cabinets. The first Schüssel cabinet (2000–2002) was made up of the Austrian People's Party (ÖVP) and the Freedom Party of Austria (FPÖ), with both parties holding 52 parliamentary seats. This marked the first time the populist FPÖ entered a coalition government, although its tenure proved short-lived. Owing to pronounced internal divisions within the FPÖ leadership, Wolfgang Schüssel, the ÖVP chancellor, deemed continued cooperation unsustainable and dissolved the cabinet after 2 years (Müller, 2021). The modest increase

https://doi.org/10.1017/psrm.2026.10138

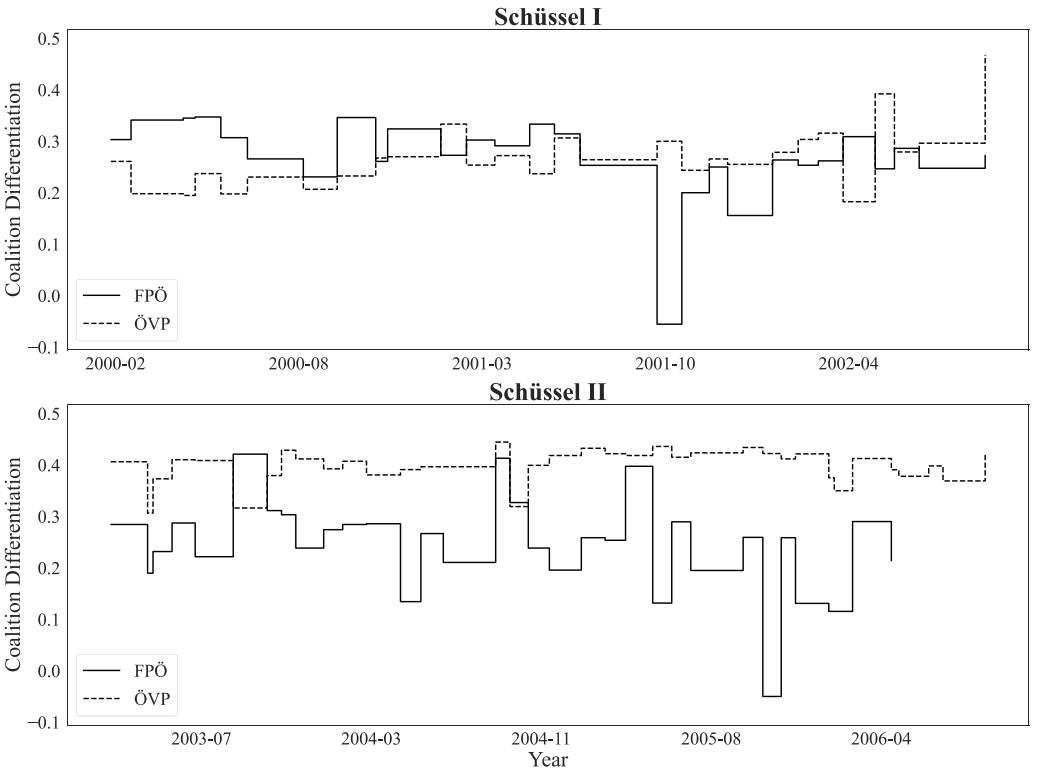


Figure 1. Coalition party rhetorical differentiation for the two consecutive Schüssel governments of Austria.

in differentiation from the ÖVP over the course of the incumbency, shown in the top panel of Figure 1, may indicate the growing discontent of the ÖVP in this cabinet. In the subsequent election, the ÖVP increased its seat share from 52 to 79, while the FPÖ suffered substantial losses, falling to 18 seats. Despite these changes, Schüssel renegotiated a coalition agreement with the FPÖ, and the same parties re-entered government, yet under markedly different conditions (Schüssel II, 2002–2006) (Müller, 2021). The resulting power asymmetry is reflected in the lower panel of Figure 1, where the FPÖ appears systematically less distinguishable than the ÖVP.

This example might give the impression that distinctiveness is merely a function of seat share. Two consecutive Irish governments, again of the same composition, suggest not only that this is false but also illustrate the prevalence of differentiation patterns in stable cabinets. The two cabinets led by Bertie Ahern (1997–2007) both completed their full terms (Mitchell, 2021). Each consisted of Fianna Fáil (FF) and the Progressive Democrats (PD). The first government was a minority coalition (FF 77 seats, PD 4) and relied on the support of independent MPs in parliament. In the subsequent term, the same parties together held a majority (FF 80, PD 8) (Mitchell, 2021). The behavior of the PD, as shown in Figure 2, is particularly notable. During the first Ahern cabinet, the party exhibited a markedly lower level of differentiation. In the second, once the coalition became a minimum winning cabinet, its differentiation began at a higher level than in the previous cabinet and increased steadily until it matched that of the chief executive party, despite the substantial disparity in seat share. This change could well represent the PD's more strategic position in the second Ahern government, where it was now required for a legislative majority. Nonetheless, the causes and implications of these differentiation patterns should be subjected to further investigation.

https://doi.org/10.1017/psrm.2026.10138

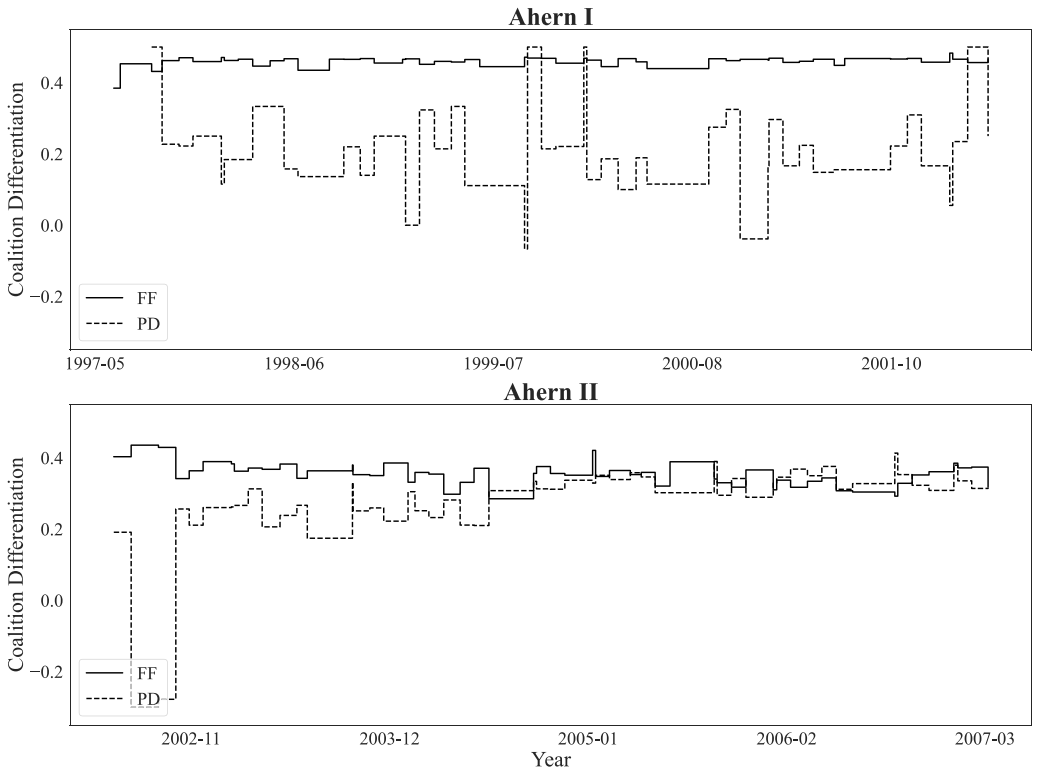


Figure 2. Coalition party rhetorical differentiation for the two consecutive Ahern governments of Ireland.

Both examples highlight that the descriptive trends in party distinctiveness from one cabinet to the next are not wholly captured by existing static measures of coalition differentiation. Instead, the dynamic classification-based measure appears to track variation in ways that are broadly consistent with what is historically known about the dynamics of these cabinets. The following section moves beyond this initial face validity and presents a series of validation tests to assess the extent to which these scores capture meaningful distinctions between coalition parties over time.

6. Validation of differentiation scores

6.1. Convergent validity: static variables

As noted above, most traditional indicators of coalition differentiation are static. These typically assign a single value to each cabinet based on institutional or compositional characteristics thought to influence whether coalition partners become more or less distinguishable. Consequently, the validation of the measure presented in this paper starts with examining whether it corresponds to these established proxies, in theoretically expected directions. To do so, I average the monthly party-level differentiation scores within each cabinet to generate a single cabinet-level measure and compare these scores to a range of commonly used coalition characteristics using metadata from PAGED (Bergman *et al.*, 2021). Because this validation exercise compiles a dynamic party-level measure into a static cabinet-level value, the resulting estimates should solely be interpreted as broad tests of convergent validity. The descriptive relationships shown below are supplemented by simple cabinet-level regression estimates reported in Appendix 4, which indicate that the observed associations are directionally consistent with theoretical expectations, while remaining statistically insignificant.

https://doi.org/10.1017/psrm.2026.10138

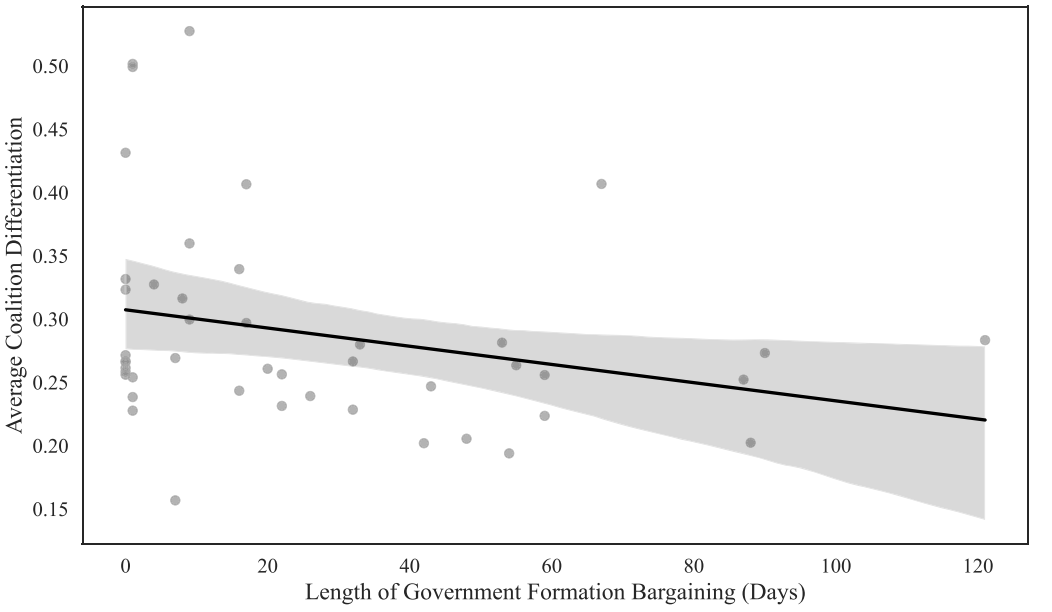


Figure 3. Average coalition differentiation compared to duration of coalition formation bargaining.

I first consider *ex ante* cabinet characteristics commonly associated with a more complex coalition bargaining environment. Figure 3 shows the relationship between averaged differentiation scores and the duration of government formation negotiations, measured in days. The negative association indicates that as formation duration increases, cabinets tend to display less differentiation behavior during incumbency. This is what one would expect theoretically, since a protracted formation process likely implies that a wider range of issues have been discussed and agreed upon prior to taking office. The more policy disagreement is settled in advance, the less opportunity there is to diverge from each other while in government. A similar logic applies to the length of the coalition agreement. Figure 4 plots the negative relationship between differentiation scores and the number of words in the cabinet's coalition agreement. The longer the agreement, the more policy detail is codified before governing begins, leaving less room for public differentiation once in office.

Subsequently, Figures 5 and 6 present average differentiation patterns by cabinet type and cabinet size, respectively. Coalition size is often treated as a proxy for bargaining complexity, and the descriptive pattern in Figure 6 is broadly consistent with that expectation: average rhetorical differentiation tends to increase with the number of coalition parties.[5] Figure 5 likewise suggests consistent patterns across cabinet types. Minority cabinets depend on support from actors outside the cabinet to secure parliamentary majorities, which may reduce the need for coalition partners to present themselves as a unified bloc. Minimal winning coalitions, in contrast, depend on their coalition partners to pass legislation, which, therefore, limits how much parties can express rhetorically distinguishable party labels. Surplus majority cabinets, which contain more parties than are necessary for a legislative majority, often allow for more than one legislative majority configuration. As a result, these cabinets therefore have more opportunities for coalition partners to be publicly distinct. The measure indeed captures variation that corresponds with these expected differences across cabinet types.

[5]This pattern remains visible despite the correction for the number of classes built into the classification pipeline.

https://doi.org/10.1017/psrm.2026.10138

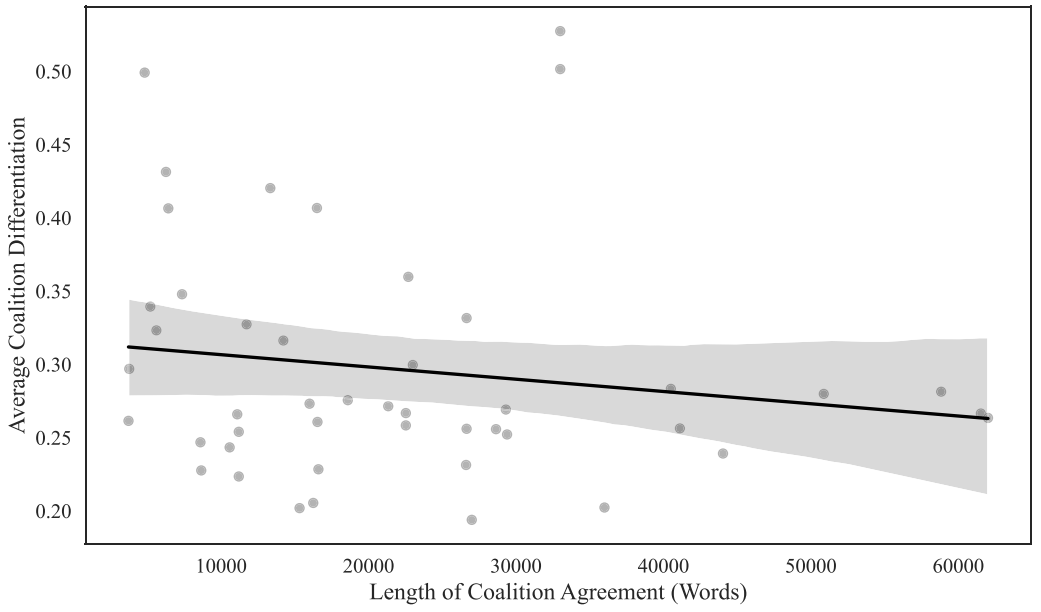


Figure 4. Average coalition differentiation compared to length of coalition agreement document.

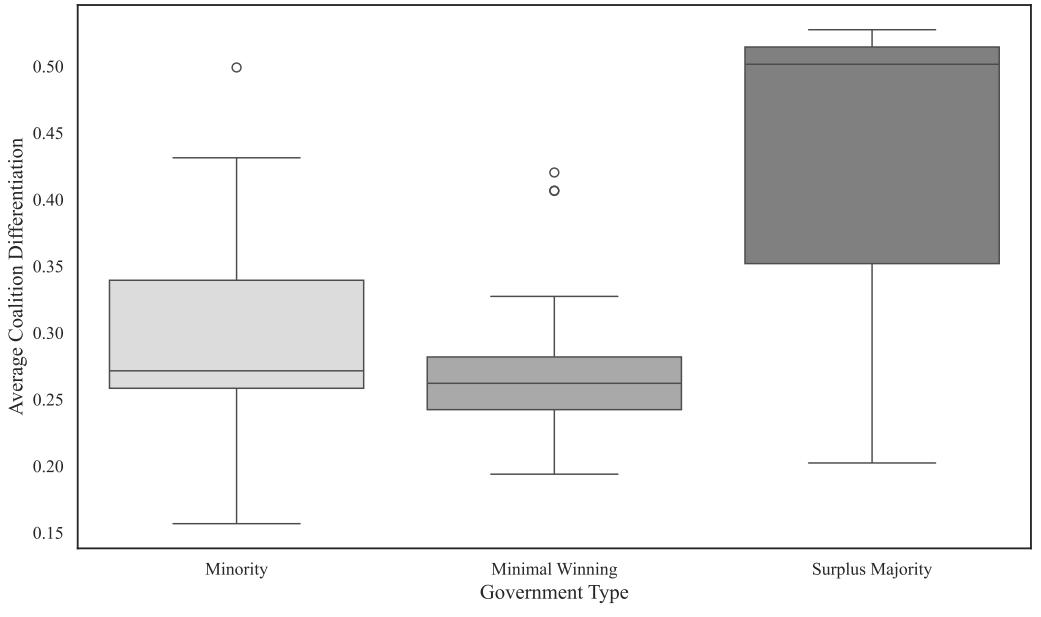


Figure 5. Average coalition differentiation for different cabinet types.

6.2. Convergent validity: dynamic variables

Dynamic measures of coalition differentiation are rare, especially for sizable samples. Yet the few existing dynamic measures, which document month-to-month fluctuations, provide an opportunity to examine whether the classification-based measure captures variation over time. To do so, I compare the monthly classification-based coalition differentiation scores against two such measures. The

https://doi.org/10.1017/psrm.2026.10138

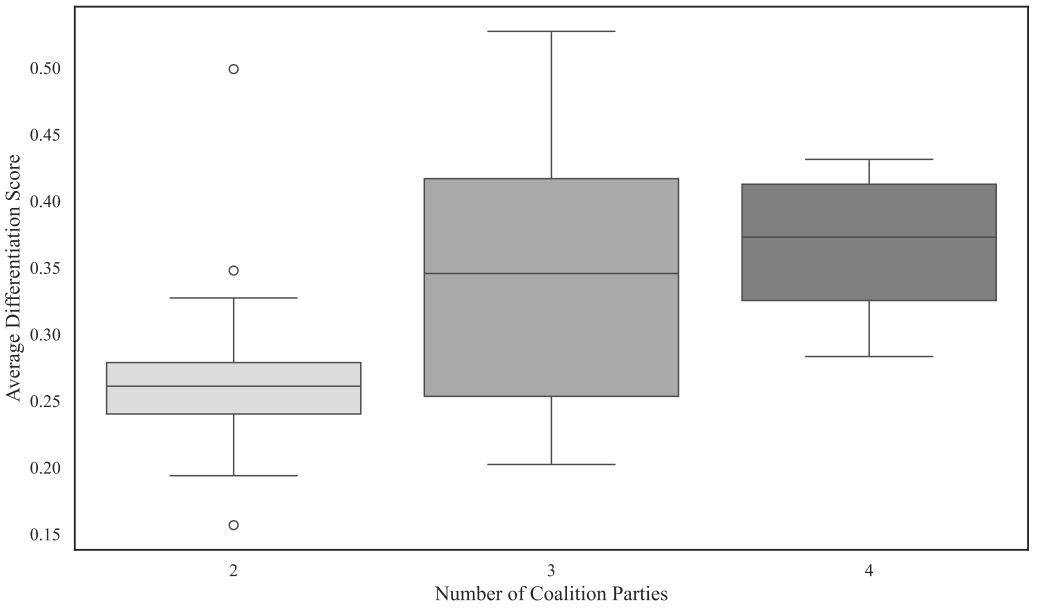


Figure 6. Average coalition differentiation for different coalition sizes.

Table 2. Regression results: dynamic measures predicting coalition differentiation

	Cabinet-month (1)	Cabinet-month (2)
Intercept	0.204***	0.283***
	(0.015)	(0.045)
Coalition mood	0.011***	
	(0.002)	
Issue diversity (logged)		0.010
		(0.012)
Country fixed effects	✓	
Number of countries	2	1
Observations	323	113
R^2	0.110	0.006
Adjusted R^2	0.104	−0.003

Note: $^{***}p < 0.01$.

results of simple linear regression models, estimated separately for each of these comparisons and including country fixed effects where appropriate, are reported in Table 2.

The first comparison with an existing dynamic measure of coalition differentiation is with coalition mood, developed by Imre *et al.* (2022) from parliamentary applause data in Austria and Germany. Here, I aggregate the monthly differentiation scores to the cabinet level by taking their average and regress these cabinet-level averages on the corresponding monthly coalition mood scores. The coefficient in Model 1 of Table 2 is significant and positive, indicating that higher coalition mood is associated with higher differentiation. This does stand in contrast to the expectation that more applause, and hence a cooperative coalition mood, should coincide with reduced differentiation. This finding may reflect that applause captures a broader engagement or salience in parliamentary debate, rather than solely approval or coalition cooperation. It should be noted that Figure 7 shows that the relationship is driven primarily by Austria, suggesting that contextual or institutional factors specific

https://doi.org/10.1017/psrm.2026.10138

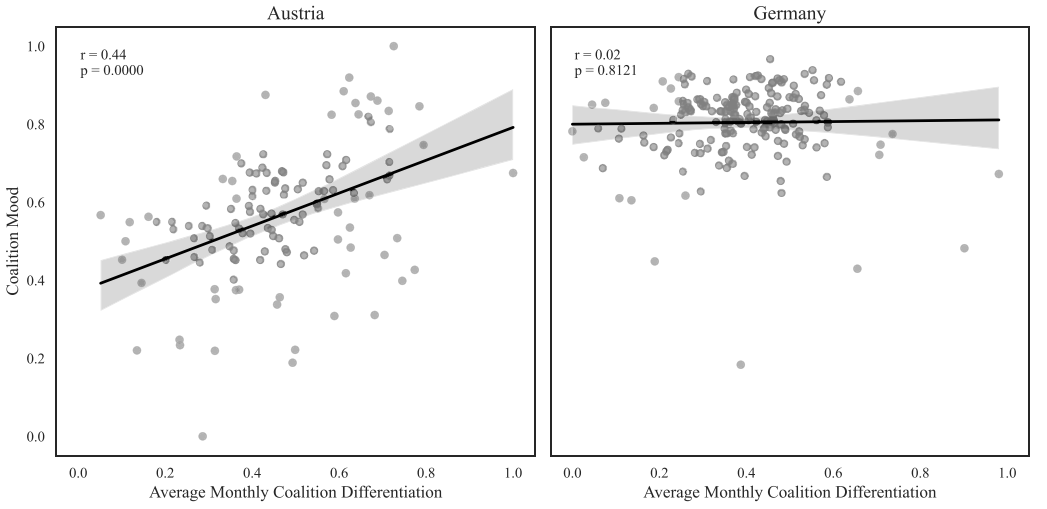


Figure 7. Average monthly coalition differentiation and coalition mood.

to the Austrian parliamentary setting may explain this pattern. Based on these findings, the specific mechanisms underlying the relationship between differentiation and coalition mood should be considered more closely.

Next, I compare my measure to the variance in issue attention across coalition party press releases, introduced by Sagarzazu and Klüver (2017) for a set of German cabinets. This captures the extent to which coalition partners prioritize different topics in their public communication. Model 2 of Table 2 shows a positive coefficient, suggesting that higher differentiation in press releases is associated with higher rhetorical differentiation in parliamentary speech. While this relationship is not statistically significant, the direction of the association is consistent with the expectation that coalition differentiation is likely engaged with similarly across different arenas. This convergence remains suggestive, given that the two measures are based on entirely separate corpora (press releases versus parliamentary debate) and capture related dimensions of coalition differentiation.

Overall, these results provide cautious evidence of convergent validity for this measure of coalition differentiation. Across both measures, the variations are positively associated, even if some coefficients remain small. The alignment with two independently constructed dynamic measures does suggest that the monthly differentiation scores are sensitive to temporal variation in coalition behavior, and thus that this measure captures shifts in party positioning beyond static characteristics. Although the positive relationship with coalition mood in Austria differs from theoretical expectations, the overall pattern across measures supports the conclusion that this classification-based approach captures temporal variation in coalition differentiation.

6.3. *Pre-termination differentiation dynamics*

Finally, I examine whether differentiation follows different temporal patterns depending on cabinet termination type. I do so by estimating an event-time model centered on cabinet termination. I first restructure the data to a cabinet-month format by taking the maximum differentiation score within the cabinet in that respective month.[6] Then I construct a backward-looking time variable, τ, which measures the number of months remaining until termination. The analysis window is restricted to the

[6]This, therefore, remains agnostic about which party could be driving differentiation.

https://doi.org/10.1017/psrm.2026.10138

final 12 months prior to termination ($0 \leq \tau \leq 11$). The dependent variable, *diff_z*, is the z-score standardization of the maximum monthly differentiation observed within each cabinet. Standardization ensures comparability across countries and over time by placing all observations on the same scale.

The key explanatory variable is a binary measure that indicates different types of cabinet termination and is derived from metadata provided in PAGED (Bergman *et al.*, 2021). Specifically, I rely on an indicator capturing whether a cabinet ended through discretionary or nondiscretionary termination. Discretionary terminations refer to cases in which cabinet dissolution occurs at the strategic initiative of one or more of the governing parties, whereas nondiscretionary terminations reflect constitutionally scheduled or otherwise exogenously determined endings of the term in office.[7]

To model temporal dynamics, I approximate the pattern of differentiation prior to termination using cubic B-splines in τ. This captures the likely nonlinear changes in differentiation over the final year of a cabinet's life cycle. I interact each spline basis term with the binary variable indicating discretionary termination to assess whether these patterns differ by termination type. The model is estimated using ordinary least squares with standard errors clustered at the cabinet level. Formally, this specification can be written as:

$$diff_z_{ct} = \sum_{k=1}^{K} \beta_k S_k(\tau_{ct}) + \sum_{k=1}^{K} \gamma_k \left[S_k(\tau_{ct}) \times D_c \right] + \varepsilon_{ct} \tag{1}$$

where $S_k(\tau)$ denotes the spline basis functions, D_c is an indicator for discretionary termination, and errors are clustered by cabinet c.

For visualization, I generate predicted values over a continuous grid of event-time values from 0 to 12 months before termination, separately for cabinets characterized by discretionary and nondiscretionary terminations. To account for sampling uncertainty, I compute bootstrapped confidence intervals based on 1,000 cabinet-level resamples. The results are plotted in Figure 8.

The visualization suggests that differentiation tends to rise in the months preceding terminations. This holds for both cabinets ending in discretionary and nondiscretionary termination, albeit the latter shows a slightly later trend. Substantively, this pattern is consistent with the theoretical expectation. Theory prescribes that differentiation is expected to increase in the months leading up to a planned election, as parties commence campaigning (Martin and Vanberg, 2008; Sagarzazu and Klüver, 2017). A similar trend can be expected if coalition parties prepare for or strategically precipitate an early cabinet dissolution as well. At the same time, the confidence intervals around these estimates overlap substantially. This is not wholly surprising, given the limited number of cabinets in each group (approximately 20) and the flexibility introduced by the spline–interaction specification.

Thus, to further examine whether differentiation accelerates in the months preceding cabinet termination, I also calculate the change in differentiation between an early and a late pre-termination window. Specifically, I define two event-time windows: an *early* window covering months $\tau = 5$ to $\tau = 8$ prior to termination and a *late* window covering months $\tau = 0$ to $\tau = 4$.[8] For each cabinet, I compute the mean maximum differentiation within each window and then calculate a late-term acceleration measure as the difference between the late- and early-window means:

$$\Delta_c = \overline{diff_z_{c,\text{late}}} - \overline{diff_z_{c,\text{early}}} \tag{2}$$

[7] This operationalization should not be conflated with terminations due to coalition conflict. Because the number of cabinets terminated by conflict is comparatively small in my sample, I instead opt for the broader discretionary versus nondiscretionary distinction. Appendix 5 re-estimates the models using the narrower set of conflictually terminated cabinets, which generates substantively similar, albeit smaller effects.

[8] Appendix 5 includes a robustness test where the acceleration test is specified with alternative windows.

https://doi.org/10.1017/psrm.2026.10138

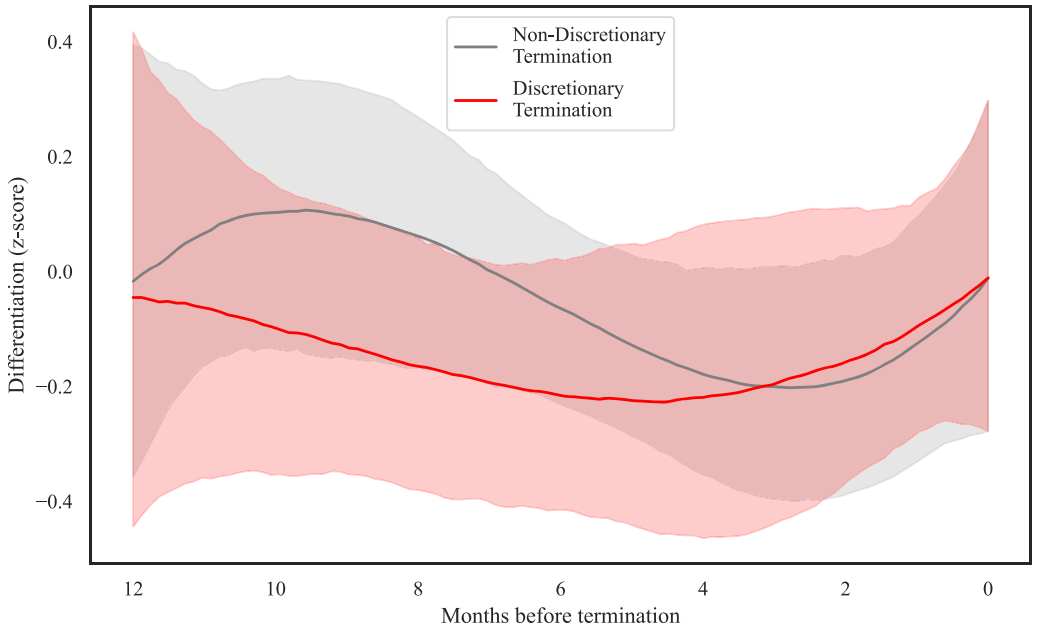


Figure 8. Spline-estimated differentiation patterns with bootstrapped 95% confidence intervals in the 12 months preceding cabinet termination.

This summarizes the temporal trajectory of each cabinet into a single measure indicating change in differentiation during the final months of incumbency. In order to assess whether late-term acceleration differs by termination type, I again divide cabinets according to discretionary termination and compare the mean acceleration across these two groups. Given the small number of cabinets in each category, I implement a permutation test to evaluate the statistical significance of the observed difference in late-term acceleration. Specifically, I calculate the observed difference in group means:

$$\Delta_{\text{obs}} = \bar{\Delta}_{D_c=1} - \bar{\Delta}_{D_c=0} \tag{3}$$

then randomly permute the termination labels 10,000 times, recomputing the mean difference in each permutation.

I plot this mean late-term acceleration for each group, along with bootstrapped 95% confidence intervals in Figure 9. This visualization shows that mean differentiation rises in the months before discretionary terminations, while cabinets completing their planned term display a slight decrease over the same period. Specifically, the mean late-term change in differentiation for cabinets ending in a discretionary termination is $\Delta = 0.266$ compared to $\Delta = -0.064$ for cabinets with a nondiscretionary termination, resulting in an observed difference of $\Delta_{\text{obs}} = 0.330$. This shows that while both types of cabinets experience a descriptive increase in differentiation prior to termination, this increase is more marked for discretionary terminations. Although the difference is not statistically significant, the observed effect does suggest a potential increase in differentiation prior to discretionary terminations.

In sum, the validation test reported above provides partial support for interpreting these scores as rhetorical differentiation in parliamentary speech, as the findings align with theoretical expectations, although some are weak once aggregated to cabinet-level averages.

https://doi.org/10.1017/psrm.2026.10138

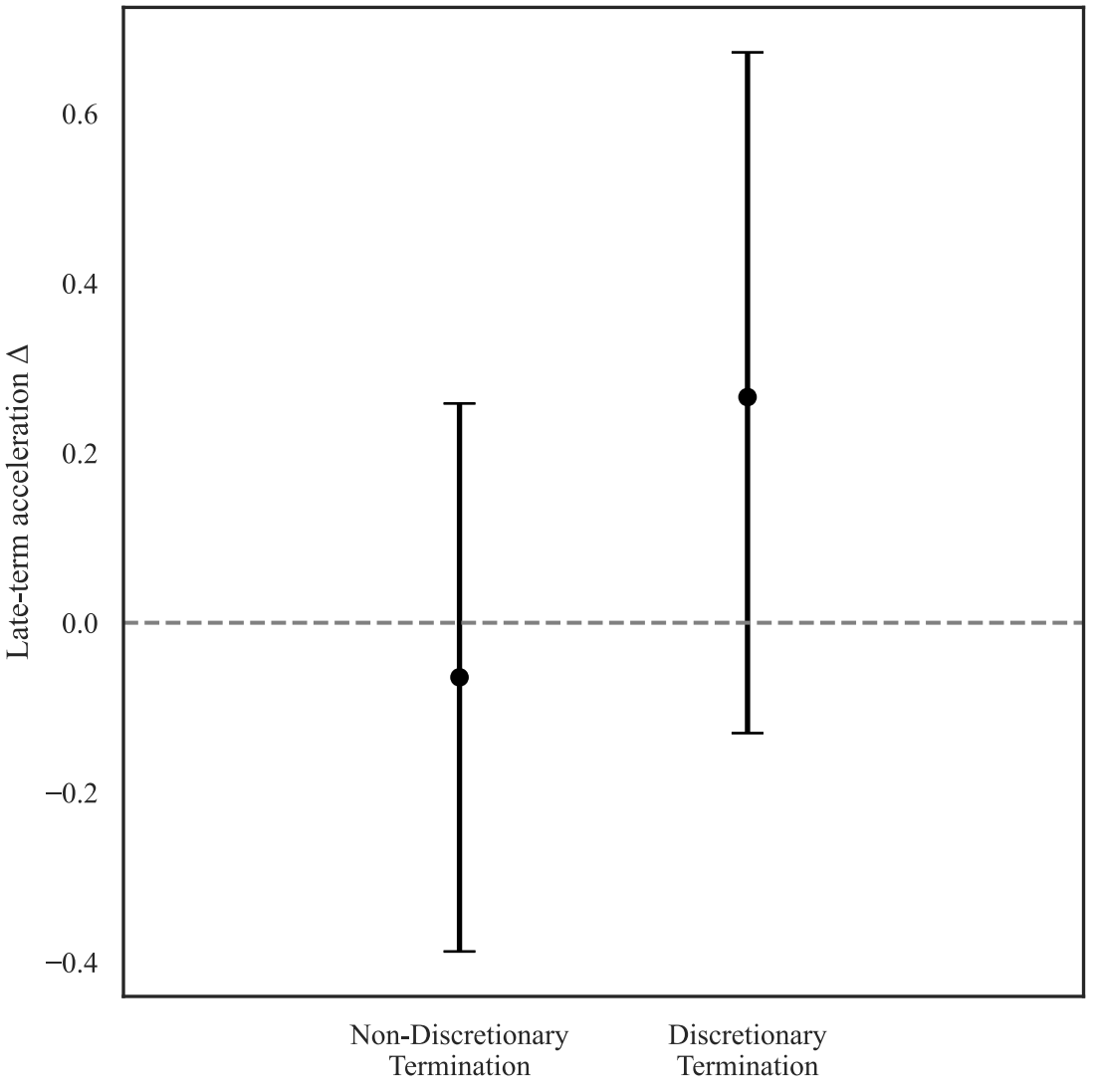


Figure 9. Late-term acceleration of differentiation by cabinet termination type. The figure shows the mean change in differentiation (Δ) between the early pre-termination window (months $\tau = 5$–8) and the late pre-termination window (months $\tau = 0$–4) for cabinets with discretionary and nondiscretionary terminations. Whiskers represent 95% bootstrap confidence intervals. A value above zero indicates an increase in differentiation in the months immediately preceding termination.

7. Conclusion

Using supervised classification, the measure presented in this paper captures the degree to which coalition parties present distinct rhetoric in parliament for every month a coalition is in office. With 3,066 monthly party scores covering close to 50 governments across six European democracies, this approach enables detailed, time-varying analysis of intra-coalition rhetorical differences. It allows researchers not only to observe coalition-level behavior but also to examine dynamics at the party level, highlighting each party's contribution to overall differentiation and variation over time.

Because of its granularity, this measure allows researchers to examine coalition governance with far greater temporal and party-level detail than previously possible. It captures how differences in

https://doi.org/10.1017/psrm.2026.10138

party rhetoric unfold over the course of incumbency, enabling analyses of the (de-)escalation of disagreement, which parties drive or absorb differentiation, and how public differentiation relates to private disagreements. For instance, differentiation scores can be used to investigate whether rhetorical divergence accumulates over time and whether heightened differentiation serves as an early warning signal of coalition instability. The pre-termination trends presented above suggest that differentiation increases prior to discretionary cabinet terminations. This would be an interesting avenue for future research.

The measure also enables analyses of party-level behavior by investigating which parties exhibit greater variation, how differentiation differs across different institutional settings, and how rhetorical differentiation relates to other forms of legislative behavioral differentiation, such as applause and amendments. In turn, this measurement facilitates exploration of how differentiation behaviors relate to other outcomes such as electoral performance or legislative output. Recent work suggests that the electoral benefits of differentiation might be smaller than previously assumed (So, 2024; Rieger, 2025). This approach provides a way to test these claims in a dynamic manner. Moreover, the scores offer a tool to analyze whether coalition responses to public demand are unified or internally contested.

There are, however, important limitations to consider. The performance of the supervised classification is influenced by country-specific characteristics, including the size and composition of parliamentary speech corpora and legislative procedures that affect speech quantity. As a result, the measure is not suited for direct cross-country comparison without accounting for these factors. Instead, the inclusion of country fixed effects or the standardization of scores is recommended when analyzing multiple countries. The measure's comparative value lies primarily in cross-cabinet and intra-coalition party comparisons within countries. Additionally, the measure assumes a degree of party discipline in the legislature. In political systems with weakly institutionalized parties or less centralized legislative behavior, the measure is likely to capture more noise than signal, as indicated in Appendix 3.

This classification-based approach provides a useful tool for researchers to study the temporal dynamics of coalition incumbency through individual parties' rhetorical divergence and subsequently its effects on legislative outcomes. Moreover, the classification method is easily extendable to other party-labeled coalition corpora. By bridging the gap between static and dynamic measures of coalition differentiation, this measurement allows for further, more detailed investigation into the strategic and institutional mechanisms that shape coalition governance.

Acknowledgements. I would like to thank Petra Schleiter, Hanna Bäck, Tarik Abou-Chadi and Slava Jankin for their valuable comments and feedback on this work.

Data availability statement. The full dataset and all replication code is available on Dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DKKIP4.

Declaration of interests. None.

Funding statement. This research was conducted without financial support from any external funding body.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2026.10138.

References

Bäck H, Baumann M, Debus M and Müller J (2019) The unequal distribution of speaking time in parliamentary-party groups. *Legislative Studies Quarterly* **44**(1), 163–193. https://doi.org/10.1111/lsq.12222

Bergman T, Bäck H and Hellström J. (2021) *Coalition Governance in Western Europe*. Oxford: Oxford University Press. ISBN: 0198868480. https://doi.org/10.1093/oso/9780198868484.001.0001

Bird S, Klein E and Loper E (2009) *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, CA: "O'Reilly Media, Inc."

https://doi.org/10.1017/psrm.2026.10138

Boston J and Bullock D (2010) Multi-party governance: Managing the unity-distinctiveness dilemma in executive coalitions. *Party Politics* **18**(3), 349–368. https://doi.org/10.1177/1354068810382937

Carroll R and Cox GW (2012) Shadowing ministers. *Comparative Political Studies* **45**(2), 220–236. https://doi.org/10.1177/0010414011421309

Dörrenbächer N, Mastenbroek E and Toshkov DD (2015) National parliaments and transposition of EU law: A matter of coalition conflict? *Journal of Common Market Studies* **53**(5), 1010–1026. https://doi.org/10.1111/jcms.12242

Fortunato D (2019) Legislative review and party differentiation in coalition governments. *American Political Science Review* **113**(1), 242–247. https://doi.org/10.1017/S000305541800062X

Fortunato D (2021) *The Cycle of Coalition: How Parties and Voters Interact under Coalition Governance*. Cambridge: Cambridge University Press.

Gentzkow M, Shapiro JM and Taddy M (2019) Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica* **87**(4), 1307–1340. https://doi.org/10.3982/ECTA16566

Goet ND (2019) Measuring polarization with text analysis: Evidence from the UK House of Commons, 1811-2015. *Political Analysis* **27**(4), 518–539. https://doi.org/10.1017/pan.2019.2

Grimmer J (2010) A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis* **18**(1), 1–35. https://doi.org/10.1093/pan/mpp034

Herzog A and Mikhaylov S (2017) *Database of Parliamentary Speeches in Ireland, 1919-2013*. V. V2. https://doi.org/10.7910/DVN/6MZN76.

Høyland B, Godbout J-F, Lapponi E and Velldal E (2014) Predicting Party Affiliations from European Parliament Debates. *ACL 2014 Workshop on Language Technologies and Computational Social Science*. Baltimore, MD: Association for Computational Linguistics, pp. 56–60. https://doi.org/10.3115/v1/W14-2516.

Imre M, Ecker A, Meyer TM and Müller WC (2022) Coalition mood in European parliamentary democracies. *British Journal of Political Science* **53**(1) (April), 1–18. https://doi.org/10.1017/S0007123421000739

Ishima H (2024) Talking like opposition parties? Electoral proximity and language styles employed by coalition partners in a mixed member majoritarian system [in Eng]. *Legislative Studies Quarterly (Iowa City)* **49**(3), 721–740.

Klüver H and Bäck H (2019) Coalition agreements, issue attention, and cabinet governance. *Comparative Political Studies* **52**(13-14), 1995–2031. https://doi.org/10.1177/0010414019830726

Klüver H and Spoon -J-J (2017) Challenges to multiparty governments: How governing in coalitions affects coalition parties' responsiveness to voters. *Party Politics* **23**(6), 793–803. https://doi.org/10.1177/1354068815627399

Klüver H and Spoon J-J (2020) Helping or hurting? How governing as a junior coalition partner influences electoral outcomes. *The Journal of Politics* **82**(4), 1231–1242. https://doi.org/10.1086/708239

König T, Lin N, Xiao L, Silva TN, Yordanova N and Zudenkova G (2022) Agenda control and timing of bill initiation: A temporal perspective on coalition governance in parliamentary democracies. *American Political Science Review* **116**(1), 231–248. https://doi.org/10.1017/S0003055421000897

Lupia A and Strøm K (2010) *Bargaining, Transaction Costs, and Coalition Governance*. Strøm K, Müller WC and Bergman T (eds). Oxford: Oxford University Press, pp. 51–82.

Martin L and Vanberg G (2008) Coalition government and political communication. *Political Research Quarterly* **61**(3), 502–516. https://doi.org/10.1177/1065912907308348

Martin L and Vanberg G (2011) *Parliaments and Coalitions: The Role of Legislative Institutions in Multiparty Governance*. Oxford: Oxford University Press.

Megahed FM, Ying-Ju Chen LAJ-F, Rigdon SE, Krzywinski M and Altman N (2024) Comparing classifier performance with baselines. *Nature Methods* **21**(4), 546–548. https://doi.org/10.1038/s41592-024-02234-5

Meyer TM, Sieberer U and Schmuck D (2023) *Rebuilding the Coalition Ship at Sea: How Uncertainty and Complexity Drive the Reform of Portfolio Design in Coalition Cabinets West European Politics*, ahead-of-print (ahead-of-print). 1–22. https://doi.org/10.1080/01402382.2023.2169512

Mitchell P (2021) Ireland: Coalition politics in a fragmenting party system. Bergman T, Bäck H and Hellström J (eds), *Coalition Governance in Western Europe*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198868484.003.0011

Müller WC (2021) Austria: Phasing-out grand coalition government. Bergman T, Bäck H and Hellström J (eds), *Coalition Governance in Western Europe*. Oxford: Oxford University Press, pp. 41–80. https://doi.org/10.1093/oso/9780198868484.003.0003

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay É (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830.

Peterson A and Spirling A (2018) Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems. *Political Analysis* **26**(1), 120–128. https://doi.org/10.1017/pan.2017.39

Plescia C and Kritzinger S (2022) When marriage gets hard: Intra-coalition conflict and electoral accountability. *Comparative Political Studies* **55**(1), 32–59. https://doi.org/10.1177/00104140211024307

Proksch S-O and Slapin JB(2015) *The Politics of Parliamentary Debate: Parties, Rebels and Representation*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/psrm.2026.10138

Rainio O Teuho J and Klén R (2024) Evaluation metrics and statistical tests for machine learning. *Scientific Reports* **14**, 6086. https://doi.org/10.1038/s41598-024-56706-x

Rauh C and Schwalbach J (2020) *The ParlSpeech V2 Data Set: Full-text Corpora of 6.3 million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies*. V. V1. https://doi.org/10.7910/DVN/L4OAKN

Rieger P (2025) From delegation to defiance: Voter support for party differentiation in coalition governments [in Eng]. *Party Politics*, 1–7. https://doi.org/10.1177/13540688251388917

Sagarzazu I and Klüver H (2017) Coalition governments and party competition: Political communication strategies of coalition parties. *Political Science Research and Methods* **5**(2), 333–349. https://doi.org/10.1017/psrm.2015.56

Slapin JB and Proksch S-O (2008) A scaling model for estimating time-series party positions from texts. *American Journal of Political Science* **52**(3), 705–722. https://doi.org/10.1111/j.1540-5907.2008.00338.x

So F (2024) Serious conflicts with benign outcomes? The electoral consequences of conflictual cabinet terminations [in Eng]. *The American Political Science Review (New York, USA)* **118**(3), 1449–1463.

Wäckerle J and Silva BC (2023) Distinctive voices: Political speech, rhetoric, and the substantive representation of women in European parliaments. *Legislative Studies Quarterly* **48**(4), 797–831. https://doi.org/10.1111/lsq.12410

Warwick P (1994) *Government survival in parliamentary democracies*. ISBN: 9780511528132.

Yu B, Kaufmann S and Diermeier D (2008) Classifying party affiliation from political speech. *Journal of Information Technology Politics* **5**(1), 33–48. https://doi.org/10.1080/19331680802149608

Zubek R (2015) Coalition government and committee power. *West European Politics* **38**(5), 1020–1041. https://doi.org/10.1080/01402382.2015.1045322

Cite this article: Sheldon C (2026) Coalition differentiation at scale: measuring intra-coalition rhetorical distinctiveness with supervised classification. *Political Science Research and Methods*, 1–20. https://doi.org/10.1017/psrm.2026.10138

https://doi.org/10.1017/psrm.2026.10138