Do Elections Moderate or Polarize Political Rhetoric?*

Tito Boeri† Nina Nikiforova‡ Guido Tabellini§

July 27, 2026

Abstract

This paper studies how elections shape political communication. We formulate a theory of electoral competition with costly communication. If opposite candidates reach different audiences with their messages, they diverge and seek to polarize the electorate. We then analyze Twitter/X communication by 367 political leaders in 21 countries during 2013-2022, focusing on competition between populists and non-populists. Following Gentzkow, Shapiro and Taddy (2019), we estimate a structural model of speech and show that political rhetoric becomes more polarized before the election. This reflects two main theoretical mechanisms: politicians increasingly emphasize issues distinctive of their type, and within each issue, they use more extreme language.

Keywords: Electoral competition, Populism, Partisanship, Polarization.
JEL classification: P, H

*We thank Fondazione Rodolfo Debenedetti for access to Twitter data, Leonardo Bianchi, Ginevra Casini, Marco Cerundolo and Gabriele Nespoli for outstanding research assistance, and Elliott Ash, Jim Fearon, Torsten Persson, Andrea Prat, Jesse Shapiro and seminar participants at the Barcelona School of Economics, Bocconi University, Cambridge, CEMFI, King's College, Oxford, the IOG meeting at Berkeley and the CEPR media conference at Bocconi for helpful comments. Tabellini gratefully acknowledges financial support from the Italian Ministry of University and Research, D.D. n. 1802 del 21/11/2024 BANDO FIS 3 - progetto FIS-2024-05869 - CUP J53C25002480001

†Department of Economics and IGIER, Bocconi University; IZA; CEP-LSE
‡Department of Economics, Bocconi University
§Department of Economics and IGIER, Bocconi University; CEPR; CES-Ifo

1 Introduction

Do elections induce moderation or polarization between competing politicians? If competing candidates seek to persuade the same voters, as in standard theories of Downsian or probabilistic voting, then most theories of electoral competition predict policy convergence (Persson and Tabellini, 2002). In this approach, divergence is due to intrinsic partisan ideologies that make candidates accept vote losses to be closer to their preferred positions (Alesina, 1988 and Besley and Coate, 1997). An alternative view is that elections have a polarizing effect. As discussed below, this can happen in a variety of situations, but the most compelling one is that candidates' communication reaches different constituencies. If so, elections create incentives to diverge, as each candidate seeks to please or mobilize their exclusive audiences. (Glaeser, Ponzetto and Shapiro, 2005 and Gennaioli and Tabellini, 2025 and Prummer, 2020) Despite a large empirical literature discussed below, the evidence on which view is closer to the truth is still inconclusive.

To shed light on this question, we study how political communication on Twitter/X is affected by the imminence of elections. During an electoral campaign, communication focuses on the short run goal of persuading and mobilizing voters. In off-election periods, instead, politicians can afford to let their tweets be shaped by their ideology and personal traits, or to react to external events. To guide the empirical analysis, we develop a simple model of electoral competition over two issues and show that, when opposite politicians reach different voters with their messages, elections create an incentive to diverge and polarize the electorate. If the exclusive audiences are sufficiently large, the model generates two main testable predictions: as the election approaches and the goal of winning the election becomes more important, polarization increases both *between* and *within* issues. Between issues, politicians prioritize the topic more distinctive of their political type and more salient for their core supporters; within the same issue, they use more extreme language, both as a signal of their positions and with the goal of polarizing the electorate.

We then take these predictions to the data. We study about 3.4 million tweets sent by 367 political leaders in 21 Western democracies during 2013-2022, and exploit staggered election dates across countries. Messages on Twitter/X, unlike electoral programs or speeches, provide high-frequency data, which allows us to study time variation as the election approaches. Tweets offer more freedom of expression than traditional media, and provide a free platform for politicians challenging the establishment, as with most populist leaders. For these reasons, platforms like Twitter/X have been key arenas of political competition, and affect the salience of topics such as crime, illegal migration or mass shootings, as well as how these topics are discussed (Battiston, 2022 and Zhang et al., 2025). Finally, compared to party manifestos

or political speeches and to other media, tweets can be more easily targeted to specific audiences, and politicians can monitor the feedback received and by which type of voter. As discussed below, this can contribute to explain why our findings differ from those in other similar studies that have focused on candidate websites and manifestos.

Our units of observation are the bigrams (i.e., two-word sequences) used by each politician during a quarter. The main challenge is that the sample of observed bigrams is small, relative to the very large number of bigrams that could be used. As explained by Gentzkow, Shapiro and Taddy (2019) (henceforth GST), simple measures of word counts over-estimate deliberate rhetorical polarization, because many bigrams are randomly chosen. We follow their approach, and estimate a structural model of speech that corrects for finite-sample bias, accounts for differences in verbosity, and allows bigram usage to depend on political type and other observables.

We start by measuring rhetorical polarization by *partisanship*, namely the ease with which one can correctly predict a politician's type based on their words, as in GST. For each politician, partisanship provides a quarterly measure of how polarized their tweets are, relative to politicians with the same observed characteristics but of the opposite political type. Second, using the model's recovered structural objects, we develop an approach to directly test the model's predictions.

Our main focus is on populist vs non-populist politicians, but we also consider left vs right, as well as a four-way partition along both dimensions (left- and right-wing populists, and similarly for non-populists). These classifications are defined ex-ante, based on Funke, Schularick and Trebesch (2023) and by human coders assisted by ChatGPT and party affiliations. There are several reasons for focusing on populist vs non-populist competition. This rivalry is arguably the main dimension of political competition in most Western democracies in this period. Populist leaders are or have been in government in the US, Brazil and Mexico, they are part of governing coalitions in the Netherlands, Austria, Italy and Sweden, influence politics from the outside in France, and command large majorities in Eastern German states. The rise of populist challengers is the other side of the coin of the decline of mainstream parties. The centrality of populist vs non-populist competition is confirmed by our data. Throughout the entire sample, populists vs non-populist rhetorical polarization is about 30% higher than between left vs right. Moreover, populists are more connected across countries (measured by Twitter links), compared to other political groups, and international connections among same-type politicians is greater along the populist vs non-populist divide than among the left vs right divide. Unlike most of the literature on populism, we focus on leaders rather than parties. This seems appropriate when studying the supply of populism, which is often based on the emergence of a charismatic leader.

To assess how proximity to elections shapes political communication, we estimate an event study, comparing rhetorical polarization in a window of 7 quarters around national (legislative and presidential) elections with quarters outside this window. Our main result is that the rhetoric of populist and non-populist politicians becomes significantly more polarized as elections approach. Polarization rises two quarters before the election and persists through the election quarter, relative to its average in more distant quarters, before falling back in post-election quarters. The effect is robust and large: during the election quarter, the predictability of a politician's type exceeds its level in distant quarters by about 15% of a standard deviation of predictability over the whole sample. This is comparable to the peak effect of a major aggregate shock, such as the 2015 European refugee crisis. By contrast, for the left vs right divide we find no significant pre-electoral cycle.

We then unpack the increase in polarization between populists and non-populists and test the more specific predictions. To remain as close as possible to the theory, we measure: (i) two dimensions of competition between populist vs non-populist leaders, each dimension distinctive of one political type; (ii) ideological extremism of bigrams within each dimension.

To construct the two dimensions, we classify tweets among 19 topics chosen a priori, and then assign bigrams to each topic based on the frequency of use in tweets on that topic. Next, we estimate the posterior that the politician is populist, conditional on the bigram's topic. Topics for which the average posterior over the entire sample is above (below) 1/2 are distinctive of populist (non-populist) politicians. As expected, non-populist politicians are more likely to tweet about policy issues, with the exception of immigration and, after COVID, health policy, which instead are distinctively populist. On the other hand, populists are more likely to tweet against the establishment, engage in self-promotion, or to mention specific parties, politicians and elections. To measure bigrams' ideological extremism, we estimate a time invariant measure of each bigram's populist distinctiveness over the entire sample – namely by how much, and in which direction, a neutral observer would change their posterior belief that a politician who used this bigram is populist.[1]

With these two measures, we test the theoretical predictions of between- and within-issue divergence. We cannot identify which political type is responsible for the divergence, but the differences between their language allow us to test the theory. Our results support the theory's implications: closer to the elections, politicians (i) speak more frequently about the issue more distinctive of their type, and (ii) use ideologically more extreme and more distinctive language within each dimension.

Overall, these findings suggest that elections have a polarizing effect on the rhetoric of

[1]This is similar to the measure of bigram distinctiveness used by GST, except that our indicator is time invariant.

populist vs non-populist leaders on Twitter/X. Could this be due to changes in the audience? The number of politicians' followers on Twitter/X does not fluctuate around election dates (Van Kessel et al., 2020). We don't observe the audience composition, but it is not clear why more extreme voters would be more attentive before the election than in off-election periods, compared to moderate voters. A more plausible interpretation is that electoral incentives exert a stronger influence on political communication in pre-election periods. This is consistent with the observed rise in politicians' verbosity before the election – in Section 8 we discuss why verbosity is unlikely to be driving our results. Although there are several reasons why stronger electoral incentives may induce polarization, a prominent one is that populist and non-populist leaders reach different groups of voters, and they seek to mobilize their constituency before the elections. Indeed, the more precise targeting of core supporters afforded by Twitter can explain why our results differ from those on other forms of political communication - cf. Barberá (2015).

We contribute to several lines of research. First, we apply and extend GST's methodology to study different data and a novel question - GST did not study the effect of elections on political rhetoric. Polarization on Twitter between populist and non-populist politicians is larger and more volatile than in congressional speeches of US Republicans and Democrats.[2]

Second, we contribute to the literature on whether elections induce political convergence or divergence by both providing a theory and developing an empirical approach to test it. Besides the research discussed above, on swing voters vs endogenous turnout and on voters' informational asymmetries, other theoretical papers have shown that elections create incentives to diverge, if voters have convex policy preferences (Kamada and Kojima, 2014), or candidates are risk averse and care about their vote share (Bonomi, 2024), or districts are heterogeneous and there is within party disagreement (Polborn and Snyder Jr, 2017).

Empirical research on this is fraught with difficulties, and has not reached conclusive results. In line with our divergence findings, Ash, Morelli and Van Weelden (2017) show that floor speeches of US senators devote more time to divisive issues when they are up for election. On the other hand, using data of candidate websites in US House elections and manifestos for French elections, Di Tella et al. (2025) find evidence of convergence in ideology and rhetorical complexity between primaries / first round elections and final elections. The difference with our results could be due to Twitter enabling better targeting of voters, or to primaries/first round elections inducing even more divergence than final elections. This second interpretation is in line with the findings by Brady, Han and Pope (2007) and Fowler and Fu (Forthcoming).

[2]Fiva, Nedregård and Øien (forthcoming) apply GST's methods to study different dimensions of heterogeneity in legislative speeches of Norwegian MPs.

Our results also speak to a related literature in political science, which has evaluated empirically the tradeoff in vote shares between converging to the center to capture swing voters, vs diverging to mobilize more extreme voters. Here too, results are inconclusive. An earlier literature suggests that pre-electoral divergence in Congressional roll call votes is punished at the election (Canes-Wrone, Brady and Cogan, 2002). Similarly, Hall (2015) and Hall and Snyder Jr (2015) find that, when a more extremist candidate wins the primaries, he is penalized in the general election. On the other hand, Bonica, Rhee and Studen (2025) argue that these findings are confounded by endogenous turnout and ballot composition effects. When these are taken into account, they find that the small electoral penalty for divergence is swamped by the large electoral gains of mobilizing core voters. Hill (2017) reaches similar conclusions with a different methodology.

Third, our paper contributes to the literature on populism and social media. Most of it seeks to explain voters' behavior - cf. Guriev and Papaioannou (2022) and Guriev, Melnikov and Zhuravskaya (2021) and Manacorda, Tabellini and Tesei (2026). But the supply side is equally important, and much less is known about it, particularly on the communication strategies of populists in proximity of elections, and on how non-populist politicians react to these strategies.[3]

Last, but not least, we contribute to research on political polarization. A consensus is emerging that politicians' messages and propaganda influence voters' beliefs and fuel polarization among citizens (see Zhang et al. (2025) and Klein (2020)). Nevertheless, it is less clear why politicians behave in this way, whether for opportunistic reasons or due to their ideological convictions. Our results are consistent with the predictions of Gennaioli and Tabellini (2025) and Bonomi (2024), who show that opportunistic politicians may find it optimal to polarize the electorate.

The paper is structured as follows. Section 2 develops a model of political competition with asymmetric audiences that motivates the empirical analysis. Section 3 describes the methodology. Section 4 describes the data. Section 5 presents aggregate patterns in the data. Section 6 presents the results on partisanship over the electoral cycle, and Section 7 unpacks those results by testing more specific predictions. Section 8 discusses the robustness of our results. Section 9 concludes. The Online Appendix contains additional details.

[3]Research in political science has studied populist communication strategies, but without a focus on the electoral cycle - cf. Cassell (2023) and Aslanidis (2018).

2 Theory

This section develops a model of political competition with costly communication to guide the empirical analysis. The central mechanism is that, when opposing candidates reach different groups of voters with their messages, electoral competition creates incentives to diverge and polarize the electorate.

2.1 The model

There are two political candidates, $P = A, B$, and two voter types, $v = b, a$, respectively with given primitive positions and primitive bliss points in a symmetric two-dimensional space indexed by $k = 1, 2 : \hat{g}_k^A = -\hat{g}_k^B = \varpi > 0$ and $\hat{g}_k^a = -\hat{g}_k^b = \varpi + \beta > \varpi$. Thus, voters are more extreme than politicians. The two dimensions can be interpreted as policies, or attitudes towards elites or institutions, or candidate features that differ between the two candidates. Each voter type has a continuum of voters of size $1/2$, so total population is 1.

Voters' preferences and perceived candidates' positions are affected by political communication, described below. After communication, voters' preferences are (vectors in boldface):

$$V(\mathbf{g}^v, \mathbf{g}^P) = -\frac{1}{2}[\sigma_1^v(g_1^v - g_1^P)^2 + \sigma_2^v(g_2^v - g_2^P)^2]$$

where g_k^v is the voter bliss point, g_k^P is candidate $P's$ position and σ_k^v is dimension k's relative salience, with $\sigma_1^v + \sigma_2^v = 1$. $(g_k^P, g_k^v, \sigma_k^v)$ are affected by candidates' communication.

Political communication Before the election, politicians engage in costly communication of three types: (i) $x_k^P \gtreqless 0$, aimed at changing own perceived positions; (ii) $y_k^P \gtreqless 0$, aimed at changing voters' bliss points; (iii) $s^P \gtreqless 0$, aimed at changing issue salience. We assume:

$$\begin{aligned} g_k^P &= \hat{g}_k^P + x_k^P && (1)\\ g_k^v &= \hat{g}_k^v + \lambda^v(y_k^A + y_k^B), \qquad \lambda^a = -\lambda^b = 1 && (2)\\ \sigma_1^v &= \bar{\sigma}^v + s^A + s^B \qquad \bar{\sigma}^a = 1 - \bar{\sigma}^b = \frac{1}{2} + \epsilon, \quad 0 \leq \epsilon < \frac{1}{2} && (3) \end{aligned}$$

Thus P's perceived positions are only affected by his/her own communication, x_k^P, whereas voters' preferences are affected by communication of both politicians. y_k^P aimed at voters' bliss points, however, shifts them in opposite directions depending on their type. This is due to a backlash on opposite voter types, as in Gennaioli and Tabellini (2025) where propaganda enhances partisan stereotypes. Finally, if $\epsilon > 0$, the model is not perfectly symmetric: with

no effort on salience-communication, issue 1 is more salient for a voters and issue 2 is more salient for b voters. These assumptions are depicted in Figure 1, for a single dimension.

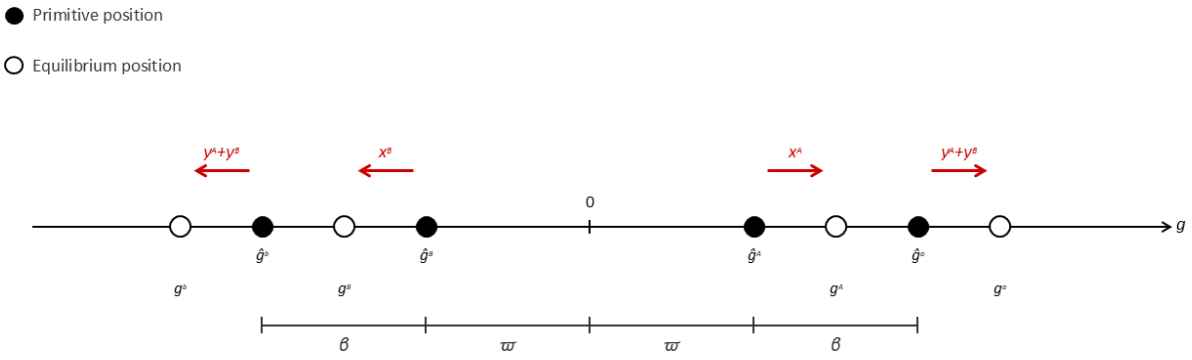


Figure 1: The one-dimensional model. Filled circles: primitive positions of parties ($\hat{g}^A = -\hat{g}^B = \varpi$) and voter bliss points ($\hat{g}^a = -\hat{g}^b = \varpi + \beta$). Open circles: equilibrium perceived party positions $g^P = \hat{g}^P + x^P$ and voter bliss points after communication in the divergent and polarizing regime. Red arrows show the direction of communication efforts.

The cost of communication is $C = \sum_k [C(x_k^P) + C(y_k^P)] + C(s^P)$ where $C(.)$ is strictly convex, twice continuously differentiable and symmetric with a minimum at $C(0) = 0$, and sufficiently convex that the equilibrium is always an interior optimum and salience weights are always between 0 and 1, so that the equilibrium is unique. We distinguish between different types of communications to be transparent on the mechanisms, although in principle a single communication effort could affect voters' beliefs on all objects at once.

Critically, we assume that all voters of type a (resp. b) are reached by the communication of candidate A (resp. B), while only a fraction $0 \leq \alpha \leq 1$ of voters of type b (resp. a) are reached by the communication of candidate A (resp. B). Thus, a fraction $(1 - \alpha)$ of a voters is an exclusive audience of candidate A, and a corresponding fraction of b voters is an exclusive audience of candidate B. The remaining fraction α of voters is reached by both parties. All types of communication obey this assumption. Voters who are not reached by the communication of candidate P behave as if that communication was 0.

Probabilistic voting After communications, voter i of type v votes for candidate A iff:

$$\Delta^v \equiv V(\mathbf{g}^v, \mathbf{g}^A) - V(\mathbf{g}^v, \mathbf{g}^B) > \varepsilon^{iv} \tag{4}$$

where ε^{iv} is independently and uniformly distributed over $[-1/2\phi, 1/2\phi]$, with large enough support that we only consider interior equilibria. Hence, $A's$ vote share is:

$$\pi^A = \frac{1}{2} + \frac{\phi}{2} \sum_v \Delta^v \tag{5}$$

Candidates simultaneously choose their communication strategies $(\mathbf{x}^P, \mathbf{y}^P, s^P)$ to maximize their vote share net of communication costs,

$$U^P(\mathbf{x}^P, \mathbf{y}^P, s^P) = \gamma \pi^P - C \tag{6}$$

taking the opponent's communication as given. Parameter $\gamma > 0$ denotes how much politicians care about their vote share.

Results would be the same if we add independent voters, or if exclusive audiences only choose between voting for their favored candidate or abstaining (rather than voting for either one or the other candidate, as here). The key assumption is informational asymmetry, not the voters' choice margin.

2.2 Equilibrium

Appendix A proves that there is a threshold $0 < \hat{\alpha} < 1$ such that, in equilibrium:

Proposition 1. *(i) If $\alpha = 1$ and $\epsilon = 0$, then candidates moderate their positions ($x_k^A = -x_k^B < 0$) and do not change voters' preferences ($y_k^P = s^P = 0$).*
(ii) If $\alpha < \hat{\alpha}$ and $\epsilon = 0$, then candidates polarize their positions ($x_k^A = -x_k^B > 0$) as well as voters' preferences ($y_k^P > 0$), and do not seek to change issue salience ($s^P = 0$).
(iii) If $\alpha < \hat{\alpha}$ and $\epsilon > 0$ sufficiently small, then candidates: (1) polarize their positions as well as voters' preferences, more so in the dimension that is more salient for their exclusive audiences ($y_1^A = y_2^B > y_2^A = y_1^B > 0$ and $x_1^A = -x_2^B > x_2^A = -x_1^B > 0$); (2) increase the salience of the issue that is already more salient at baseline for their exclusive audiences ($s^A = -s^B > 0$).
(iv) If ϵ is sufficiently small, all non 0 communication effort about candidates' positions and voters' bliss points increases with γ ($\frac{\partial |z_k^P|}{\partial \gamma} > 0 \;\; iff \; |z_k^P| \neq 0$, $for \; z = y, x$). If α is sufficiently small and the cost function C(.) is quadratic, salience effort also increases in γ ($\frac{\partial |s^P|}{\partial \gamma} > 0$ iff $|s^P| \neq 0$).

Consider first uniform information and perfect symmetry ($\alpha = 1$ and $\epsilon = 0$). Because the same voters are contested by both parties, each candidate has an incentive to moderate its perceived position (standard Downsian logic), more so if electoral stakes are higher (larger γ).

There is no incentive to change voters' preferences: with a shared audience and symmetric voter types, any communication effort exactly cancels across groups (since $\lambda^a + \lambda^b = 0$).

Suppose instead that a fraction $(1-\alpha) > 0$ of voters is reached exclusively by their closest candidate. Polarizing this exclusive audience is a one-sided electoral gain: it pushes them further away from both candidates, but with concave voters' preferences this benefits their closest candidate, without any offsetting loss elsewhere. Both candidates therefore invest in voter polarization ($y_k^A = y_k^B > 0$), and more so when electoral incentives are stronger (higher γ). This corresponds to the identity politics mechanism in Gennaioli and Tabellini (2025), where communication exacerbates partisan stereotypes. Whether candidates also diverge in perceived positions depends on a further trade-off. A more extreme position attracts the closest voters (who are more extreme than the candidate), but repels the more distant ones. If the exclusive audience is sufficiently large ($\alpha < \hat{\alpha}$), the first effect dominates and candidates diverge. With perfect symmetry ($\epsilon = 0$), there is no gain in changing issue salience.

If voters also initially disagree on which dimension is more relevant ($\epsilon > 0$), asymmetric communication has two additional effects (if the exclusive audience is sufficiently large).[4] First, communication has stronger electoral effects in the more salient dimension. Hence, both candidates exert greater polarizing effort — on perceived positions and bliss points — along the dimension favored by their exclusive audiences. As a result the equilibrium is no longer symmetric: voters and candidates are more extreme in the dimension which is more salient for captive voters. Second, each candidate benefits from making issue salience even more asymmetric across opposite voter types. Hence, the two candidates champion different issues, generating between-issue divergence in communication.[5]

Since these effects reflect how communication affects vote shares, they are all amplified by parameter γ capturing the strength of electoral motives.

Predictions To take these predictions to our Twitter/X data, we make two further assumptions. First, that the share $(1-\alpha)$ of exclusive audiences is sufficiently large. This is consistent with evidence showing that political communication on Twitter/X is exchanged primarily between individuals with similar political preferences (Barberá, 2015), and that politicians know a lot about their followers.[6] Second, that parameter γ rises in the immi-

[4] As shown in the proof, when $\epsilon > 0$ there are multiple thresholds, one for each aspect of communication, and $\hat{\alpha}$ corresponds to the lowest threshold.

[5] Although not covered by Proposition 1, it can be shown that, if $\alpha = 1$ and $\epsilon > 0$, most statements in part (iii) of the Proposition are reversed: in equilibrium candidate positions converge, they seek to moderate salience asymmetries, and on average voters' bliss points polarization is unaffected by communication. Thus, the critical parameter that induces communication divergence is the size of the exclusive audience, $1-\alpha$, not the asymmetric salience parameter ϵ.

[6] Until 2023 Twitter analytics allowed to recover breakdowns of followers by gender, location, interests (with categories such as "conservative politics"), and device usage. Thus, in addition to relying on self-

nence of elections. This is consistent with the observed pre-electoral rise in verbosity (cf. Appendix Figure H10), and with evidence that during an electoral campaign communication is more focused on the goal of gaining votes, relative to other motives such as expressing personal convictions or reacting to external events.[7]

Under these assumptions, Proposition 1 then predicts that political rhetoric polarizes in pre-electoral periods, compared to post-election periods or dates distant from the election. Specifically, the rest of this paper tests two predictions of Proposition 1. Before the election:

1) *Between-issue* divergence rises. This corresponds to s^A and s^B moving in opposite directions: each candidate speaks more about issues that are distinctively salient to its constituency and over which it enjoys an electoral advantage with its supporters;

2) *Within-issue* divergence rises. This corresponds to $\mathbf{x}^A$ and $\mathbf{x}^B$ moving in opposite directions and $\mathbf{y}^P$ rising. As candidates put more effort to polarize voters and their perceived positions, the way they speak about each issue becomes more distinctive of their political ideology, and more extreme in each dimension.

Of course, the model is very simple: it assumes only two competing candidates, two dimensions and perfect symmetry within each dimension, and some restrictions on parameter values to prove some results. Thus, a rejection can be interpreted that either the share $(1-\alpha)$ of the exclusive audience is small, or parameter γ remains constant over time, or a more general failure of the model.

3 Empirical methodology

This section describes how we take the theory to the data. We rely on the methodology proposed by GST. Relative to other approaches, this method of measuring polarization in political communication has several advantages.[8] First, it is based on a structural model of

selection of followers, leaders could assess whether their communication campaigns were effective in targeting specific audiences.

[7] Severin-Nielsen et al. (2025) interviewed Danish MPs about how their communication strategies differ between election and routine periods. They highlight that, while in routine periods communication is often developed on an ad hoc basis, during an electoral campaign it is much more thoroughly and coherently thought out, also because more resources are devoted to it. This is how a Danish MP described his communication during a routine (as opposed to election) period: "It is a bit more random than you would think. (. . .) It's very focused on issues of the day (. . .). And there are a lot of posts that are based on an impulse".

[8] Alternative methods of text analysis are discussed by Ash and Hansen (2023). Gauthier, Widmer and Ash (2025) propose a more general method for measuring speech polarization based on deep latent variable models, and show that their results are in line with Gentzkow, Shapiro and Taddy (2019). We adopt the latter, whose structural interpretability is central to the tests of our model's predictions in Section 7. Alternative methods based on AI are less transparent, and could attribute observed differences in speech to

speech, that also takes into account other determinants of word usage besides political type, such as nationality, gender, age, or institutional position. This is particularly important in our multi-country setting. Moreover, it allows us to dig deeper into the causes of rhetorical polarization, and directly test the implications of the model, as we discuss below. Second, as explained by GST, this approach substantially reduces finite sample bias and is robust to changes in verbosity across politicians over time. This allows us to exploit all observed communication, without restricting the analysis to a much smaller subset of speech to ensure comparability.

From theory to data We observe the tweets sent over time by several politicians indexed by i. As in the theory, we assume that there are only two political types, say populist ($P_i = 1$) and non-populist ($P_i = 0$), although individual politicians are allowed to differ in other observable features which may also influence communication (below we also allow for four political types). By equation (6), politicians' utility is a function of their communication. Here, following GST and much of the literature, we model communication as a choice of bigrams.[9] Thus, politician i's utility from using bigram j in quarter t is:

$$U_{ijt}^{P_i} = \alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i \tag{7}$$

where α_{jt} is a fixed effect for bigram j's popularity in quarter t, γ_j and ϕ_{jt} are parameters to be estimated, X_{it} are dummy variables for other features (not all time varying) of i, namely gender, higher education, being in government in quarter t (directly or by supporting the governing coalition), being a candidate in the closest election, age, and fixed effects for country-by-quarter and for the election window (defined as all quarters closest to the same election) - the latter capture election-specific bigrams.

The resulting probability that i uses bigram j in quarter t is:

$$q_{ijt}^{P_i} = \frac{exp(U_{ijt}^{P_i})}{\sum_k exp(U_{ikt}^{P_i})}, \qquad P_i = 0, 1 \tag{8}$$

We observe the number of times that each bigram j is used by politician i in calendar quarter t, $\mathbf{c_{it}} = \{c_{ijt}\}$, and assume that this count variable has a multinomial distribution:

$$\mathbf{c_{it}} \sim MN(m_{it}, \mathbf{q_{it}^{P_i}}),$$

confounding variables correlated with political type. See Vafa, Athey and Blei (2025) and Gordon et al. (2025) among others for a discussion.

[9] See Gentzkow, Kelly and Taddy (2019) for a discussion of the economics literature using n-grams.

where $m_{it} = \sum_j c_{ijt}$ is the verbosity of i in quarter t and $\mathbf{q_{it}^{P_i}} = \{q_{ijt}^{P_i}\}$.

Thus, a communication strategy $(\mathbf{x}^P, \mathbf{y}^P, s^P)$ is the choice of the entire distribution of bigram-usage probabilities $\mathbf{q_{it}^{P_i}}$ during period t. This formulation entails two simplifying assumptions. First, the utility of using bigram j is allowed to depend on i's observed features, including political type, but not on which other bigrams he/she uses. Second, i's utility from bigram choice depends on the communication of other speakers only through the bigram popularity fixed effects, α_{jt}, and the country by quarter fixed effects. These fixed effects thus capture equilibrium interactions between different politicians.

Estimation For each bigram j we estimate parameters $\{\alpha_{jt}, \gamma_j, \phi_{jt}\}_{t=1...T}$ by minimizing the following penalized objective function, as in GST:

$$\sum_j \Bigg[\sum_t \sum_i m_{it} \exp(\alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i) - c_{ijt}(\alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i) + \psi(|\alpha_{jt}| + ||\gamma_j||) + \lambda_j|\phi_{jt}| \Bigg] \tag{9}$$

Two aspects of this procedure are worth mentioning. First, in order to make the model computationally feasible, we approximate the likelihood of the multinomial logit with the likelihood of a Poisson model $c_{ijt} \sim Pois(exp(\mu_{it} + u_{ijt}))$ and the plug-in estimator $\hat{\mu_{it}} = log(m_{it})$ for μ_{it} is used. Second, we make use of Lasso shrinkage for populist loadings, $\lambda_j|\phi_{jt}|$. The value of λ_j is chosen separately for each bigram so as to minimize the corrected Akaike information criterion. Overall, this forces the parameter of interest ϕ_{jt} towards 0, reducing finite sample error.[10] For the other parameters, we set the penalty at $\psi = 10^{-5}$. This value controls the extent to which variation in bigram usage across individuals or over time is attributed to covariates rather than being absorbed P_i. This penalty also facilitates numerical convergence. We discuss the choice of the penalty in Appendix B.[11]

Using (7)-(8), we can estimate the probability that a politician of type (P_i, X_{it}) uses bigram j in quarter t, $\hat{q}_{ijt}^{P_i}$, for all j in our sample. The set of covariates is sufficiently rich that this estimated probability is specific to each individual politician.

[10] In our baseline estimates, the estimate of ϕ_{jt} is non-zero for about 2.5% of bigrams per quarter.

[11] Penalizing non-zero coefficients γ_j implies that around 30% of all the country by calendar quarter fixed effects are included on average (the rest are estimated to be 0). Given the importance of distinguishing between different countries and quarters, we thus also include in X_{it} separate fixed effects for each country and each calendar quarter (without the penalty, this would be redundant, since these fixed effects would all be absorbed by the country -by-calendar-quarter fixed effect). The estimates of the event studies described below are not significantly affected by this inclusion (results available upon request).

Partisanship Using $\hat{q}_{ijt}^{P_i}$, the posterior that an observer with neutral priors assigns to politician i having type $P_i = 1$, conditional on i having used bigram j in quarter t, is:

$$\hat{\rho}_{ijt} = \frac{\hat{q}_{ijt}^{1}}{\hat{q}_{ijt}^{1} + \hat{q}_{ijt}^{0}} \tag{10}$$

Following GST, our core measure of rhetorical polarization by politician i in quarter t is *partisanship*, defined as:

$$\hat{\pi}_{it} = \frac{1}{2} \sum_{j} [\hat{q}_{ijt}^{1} \hat{\rho}_{ijt} + \hat{q}_{ijt}^{0} (1 - \hat{\rho}_{ijt})] \tag{11}$$

This variable measures the average predictability of i's type, given a single bigram. It is the average posterior probability with which an observer with neutral priors expects to correctly classify politician i's type, conditional on a single bigram. The average is computed over all possible bigrams, weighted by the estimated probabilities of their use by i in quarter t.

Any specific politician is either $P_i = 1$ or $P_i = 0$. The variable $\hat{\pi}_{it}$ averages the true type P_i and his "clone" $1 - P_i$ with neutral priors. In other words, partisanship measures the average predictability of two specific politicians with features (X_{it}) who are identical in everything but their political type. If opposite types always use the same bigrams, then $\hat{\pi}_{it} = 1/2$. If instead they always use different bigrams, then $\hat{\pi}_{it} = 1$.

We also consider a more granular partition of four political types, defined on two dimensions: populist / non-populist and right / left. Appendix C provides more details on a construction of this measure.

Despite its useful properties, partisanship has two limitations. First, it measures rhetorical divergence, but not who is responsible for it. The reason is that the model of bigram choice is identified only up to *differences* in utilities between the two political types, $U_{ijt}^{1} - U_{ijt}^{0}$, within a given quarter.[12] As a result, the difference in bigram frequencies between opposite political types in a given quarter is a well-defined object, $\mathbf{q}_{it}^{1} - \mathbf{q}_{it}^{0}$, whereas the evolution of $\mathbf{q}_{it}^{P_i}$ over time conflates the effects of political types and of the other covariates entering equation (9). In other words, the specification in (7) does not allow us to tell apart whether one type uses a given bigram more or the other uses it less.

Second, partisanship is silent about the form taken by rhetorical divergence. Yet, the theory emphasizes that divergence is due to both political types: i) speaking more frequently about the issues which are more salient for their core supporters; ii) using ideologically more extreme and more distinctive language within each issue. To overcome this gap, in Section 7 we show that both points can be formulated as testable predictions about specific components of $\mathbf{q}_{it}^{1} - \mathbf{q}_{it}^{0}$, which allows us to test the more precise predictions of the theory.

[12] This is a well-known feature of discrete choice models; see, e.g., Gandhi and Nevo (2021) for a discussion.

4 Data

4.1 Tweets and Politicians

Data on tweets of politicians were collected through Twitter Academic API, version 2. The sample covers the period 2013-22 in 21 democracies selected also based on the prominence of populist political leaders (Australia, Austria, Belgium, Brazil, Czech Republic, Denmark, Finland, France, Germany, Hungary, Italy, Mexico, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, United Kingdom, United States).[13]

Politicians were selected based on their leadership positions, according to the following criteria: (i) All candidates for Prime Minister/President in general elections between 2001 and 2022. (ii) Leaders of main political parties with vote share above 5% in at least one general election between 2010 and 2022. (iii) Influential political leaders at the local level or leaders of niche parties, even if the national vote share of their party is below 5 % (e.g., Nicola Sturgeon in the UK, Basque and Catalan political leaders in Spain, Giorgia Meloni in Italy). These criteria identified 496 leaders, of whom 367 had an active Twitter account, for a total of about 3.4 million tweets written between January 2013 and June 2022. In addition to the text of each tweet, we have the exact date and time in which they were issued.

A distinguishing feature of our paper is that we focus on political leaders rather than parties. We classify politicians along two dimensions: populist (P)/non-populist (NP) and left (L)/right (R).

4.1.1 Populist vs. Non-Populist

To identify populist leaders, we draw on the classification of Funke, Schularick and Trebesch (2023) for those few leaders in their dataset who were active during the period covered by our analysis. For the remaining politicians, we conducted a manual classification informed by GPT-4o mini responses to the following question: *Is leader X from country Y commonly considered a populist leader?* A human coder then interpreted the lengthy and detailed answer produced by ChatGPT and used it to classify the leader as populist, non-populist or ambiguous. Key identifiers of populist leaders included the cleavage between the people and the elite, an anti-establishment stance, and a strong emphasis on national sovereignty. The ambiguous category includes genuinely unclear cases, situations where ChatGPT was uncertain, or cases where leaders appeared to shift over time between

[13]We thank the Rodolfo De Benedetti Foundation for giving us access to these data that they had collected.

populist and non-populist positions.[14]

After this step, 29 politicians were classified as *ambiguous*. We classify 16 of these as populist because their party is considered populist by at least one of Norris (2019), Rooduijn (2019) and Guriev and Papaioannou (2022). For the remaining 13 politicians we check that our results are robust to classifying them either way, and in our main analysis we include them among populist types. Examples of leaders belonging to this residual group are Lula da Silva in Brazil, Nick Xenophon in Australia, and Marjan Sarec in Slovenia. Importantly, there are at most four politicians that, according to ChatGPT, may have switched from NP to P or vice versa in the period covered by our data. For these four politicians, we retain throughout the most recent classification. Appendix F.1 describes how we established that. Appendix F.6 lists all politicians and their political type. In total, we classify 84 politicians as populist and 283 as non-populist.

We validate our classification by comparing it to two benchmarks: (a) 150 random binary divisions of politicians and (b) a fully automated classification of politicians into populists and non-populists produced by the Claude Sonnet 4.6 model, with 84 out of 367 populist politicians in all cases. For each partition we computed a fixation index based on a χ^2 measure of speech heterogeneity. The fixation index measures the share of total variation in the language patterns among all politicians due to between group variation (cf. Desmet, Ortuño-Ortín and Wacziarg (2025)). Our classification outperforms 99% of all random partitions, while the classification by Claude outperforms 87% of them. This suggests that our baseline classification captures meaningful latent structures in politicians' language, and does so better than a fully automated AI-based classification. Appendix F.2 provides more details.[15]

Our leader-based classification of populism differs from a party-based classification, and adds relevant information to it. About 86 % of populist leaders according to our definition operate in parties classified as populist in the Global Party Survey (Norris, 2020). However, there are also several *NP* politicians belonging to populist parties, such as Mitt Romney, Mike Pence and George Bush in the US, Gordon Brown, David Cameron, and Theresa May in the UK. Instead, Alexandria Ocasio-Cortez in the US and Mark Latham in Australia are examples of populist leaders in parties that are classified as non-populist.

Figure 2 illustrates the number of active politicians in each quarter disaggregated by their type, and the number of tweets they sent. There are politicians of both types always active

[14]We also validated our classification against the fully automated classification produced by ChatGPT. There are only nine cases of complete disagreement between the two approaches – that is, politicians whom we classify as populists but whom the automated ChatGPT classification identifies as non-populists. Further details on this comparison are provided in Appendix F.1.

[15]Among the politicians unambiguously classified by both Claude and us, only 6 politicians are classified differently by us and Claude.

in all quarters (the minimum number of active populist politicians is always above 40).

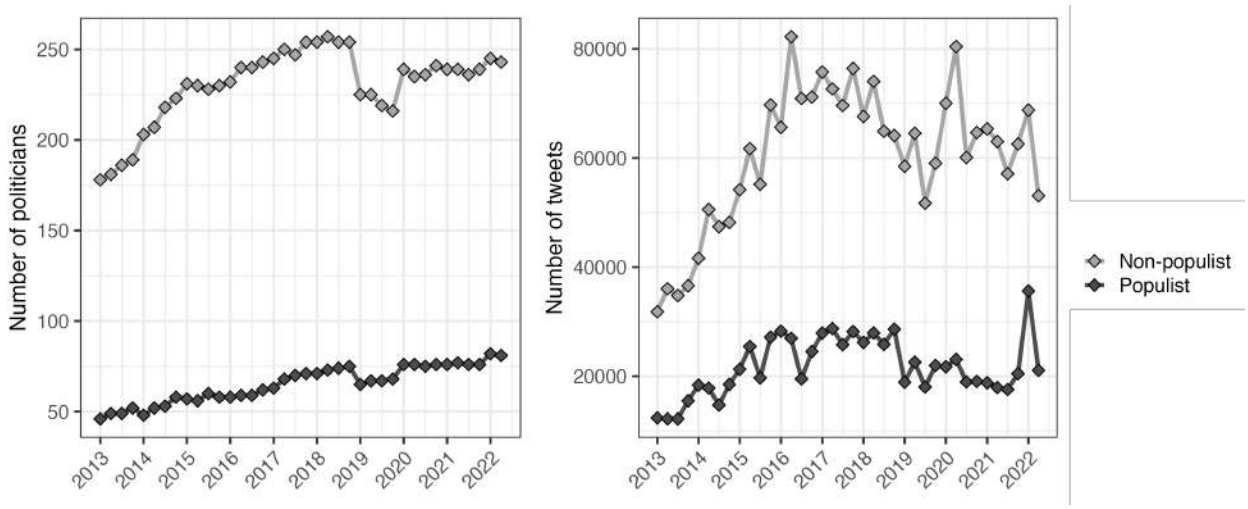


Figure 2: Number of politicians active on Twitter per calendar quarter, and number of tweets

The first three columns of Table 1 provide information on the characteristics of the two sets of politicians. P and NP leaders have approximately the same age, 3 out of 4 are males, and, on average, they began tweeting in 2012. What distinguishes P vs NP is their experience (P entered the national legislature at a later stage) and exposure to government positions (higher for NP leaders).[16] Some 80 % of NP leaders belong to mainstream parties, defined as parties that were in government (or lent support to a government) between 1990 and 2010.[17] NP leaders also display, on average, higher educational attainments than P leaders, but the difference is not statistically significant at conventional levels. Almost 50 % of populist leaders belong to niche parties, based on a GPT-4o model classification that defines as niche parties those emphasizing a limited set of issues that do not coincide with the predominant economic Left/Right division (Meguid, 2005) (see Appendix F.3 for the ChatGPT prompt). Overall, compared to NP leaders, several P leaders are *challengers*, because either they have not been in Government or they belong to parties representing issues not typically dealt with by the establishment. This is why, to avoid confounding political types with this feature, in estimating $\mathbf{q}_{\mathbf{it}}^{\mathbf{P_i}}$ we always control for being in government or supporting the governing coalition.

The analysis that follows implicitly assumes that the distinctive rhetorical features of politicians of the same type are similar across countries. To gauge the plausibility of this assumption, we study how connected political leaders are with same-type leaders in other coun-

[16]The variable years in Govt measures the number of years in which the leader was either in a party in Government, or supporting the Government, or was herself a Minister

[17]A leader is defined as mainstream even without government experience if she belonged to a party for a short period of time and that party was in government in a different period.

tries. Appendix F.4 shows that cross-border ties are homophilous along the populist/non-populist divide (that is, ties form disproportionately between same-type leaders), and that this homophily is stronger for populist leaders. Although populism often endorses nationalism and hostility to multiculturalism, its leaders behave as part of an international movement. Ties are also homophilous along the left–right divide described below, although more weakly than along the populist/non-populist divide.

Table 1: Differences between Populists and Non-Populists, and Right and Left (means; difference with p-value in parentheses).

	Populist vs Non-Populist			Right vs Left		
Variable	P	NP	Δ(p)	R	L	Δ(p)
Age (in 2010)	44.43	48.32	-3.89	46.31	48.15	-1.85
			(0.01)			(0.13)
Male	0.73	0.76	-0.03	0.78	0.74	0.05
			(0.59)			(0.28)
Higher Education	0.86	0.91	-0.05	0.91	0.89	0.02
			(0.26)			(0.50)
Year of First Tweet	2013.0	2012.4	0.58	2012.7	2012.4	0.39
			(0.08)			(0.14)
Years in Govt	6.44	9.79	-3.35	8.56	9.28	-0.72
			(0.00)			(0.48)
Legislative Experience	0.26	0.39	-0.14	0.40	0.34	0.07
			(0.00)			(0.04)
Mainstream	0.58	0.82	-0.24	0.83	0.72	0.10
			(0.00)			(0.02)
Niche Party	0.43	0.17	0.26	0.28	0.20	0.08
			(0.00)			(0.07)

Notes. Legislative Experience: years passed in 2010 since they first entered the national legislature. *Mainstream*: equals 1 if the leader has at any point in their career belonged to a party which was in government or supported the government coalition in the years 1990-2010. *Niche parties* classification is produced by GPT-4o, and based on the definition proposed in Meguid (2005). See Appendix F.3 for more details.

4.1.2 Left vs Right

We relied on the GPT-4o mini model to classify leaders along the L/R divide. We performed 4 iterations with the prompt displayed in Appendix F.5. Politicians belong to the "Right" group if GPT always unambiguously classified the politician in this category,

and the same with "Left". For 82 politicians, one of the 4 GPT iterations said that the classification is ambiguous. To classify these 82 ambiguous cases, we relied on their party affiliations, using the most recent RILE index of the Manifesto project, which measures how much a party covers left-wing or right-wing issues.[18] Based on this, we assigned 64 out of 82 leaders to the left and the remaining 18 to the right. Appendix F.5 discusses the differences between our classification and the one that would result from only using party affiliations. Appendix F.6 lists all politicians along the L-R divide. L vs R have similar features - cf. the three right-most columns in Table 1.

Appendix Figure H1 depicts the number of active L and R politicians in our sample and their activity in each quarter. Left-wing politicians are more numerous than those on the right – 223 politicians are classified as L and 144 classified as R. We also classify leaders along the two dimensions simultaneously, to obtain four political types (LP, RP etc.). Appendix Table H1 reports the cross tabulation of the four types. Non-populist politicians in our sample tend to be left-wing, while populists are about equally split between left and right.

4.2 Bigrams

All tweets were translated into English using the open-access Python library googletrans, and then decomposed into bigrams.[19] To do so, all special symbols were removed (punctuation marks, hyphens and apostrophes, emojis, user mentions and other hyperlinks, while hashtags were kept) and words were reduced to their stems.

The final dataset comprises around 40 mln bigrams, of which there are 15 mln unique bigrams. The vast majority of bigrams were used just once. Estimations drawing on such a large number of bigrams would be computationally infeasible and in the remainder we restrict the analysis to the most frequent bigrams only. For our benchmark specification, only bigrams that were spoken in at least 10 calendar quarters and that have more than 10 occurrences in at least one quarter were used. We refer to this dataset restriction as 10-10 (Appendix Figure F2 presents the histograms for these variables). This left us with about 25,000 unique bigrams and about 65% of the tweets. In section 8 we assess the robustness to different thresholds (5-10, 5-20, 10-25).

[18] There are no iterations when GPT-4o mini model classification switched from classifying a politician from right to left or vice-versa, but just observations where the classification switched between right and ambiguous or between left and ambiguous or vice versa. The source for the manifesto data is: https://manifesto-project.wzb.eu

[19] To assess the quality of English translations, a sample of 50 tweets per country was translated back into their original languages. The similarity between each back-translated tweet and its original version was then evaluated using multiple text similarity metrics. Cosine similarity is above .9 for all languages, denoting a high degree of semantic similarity between the original and the translated texts. See Appendix F.9 for the results of this exercise.

4.3 Topic classification

We classified bigrams using GPT and BERT into 18 topics chosen a priori (business & industry, climate & environment, energy policy, EU, foreign policy, public health, immigration, inequality, inflation, labor market, rights, security, taxation, trade policy, anti-establishment, connection & personal communication[20], elections, self-promotion). These topics were chosen to achieve a good balance between granularity and scale. Too broad topics would be uninformative, while too granular classifications would not provide a sufficiently large number of tweets per topic for estimation.

We then proceeded in four steps: 1) we used GPT-5 mini to label 100,000 tweets into these 18 topics (see Appendix F.8 for the prompt used in instructing ChatGPT); 2) we then trained BERT on these labeled tweets; 3) next we used the trained BERT to classify all 3.4 million tweets into topics with a threshold probability of .4;[21] 4) finally, we classified bigrams based on their relative recurrence in tweets belonging to the different categories, calculating a simple measure of bigram specificity for each topic: bigrams exceeding the 20 % threshold in the ratio of the number of tweets containing bigram j on topic k to the total number of tweets containing bigram j were assigned to the respective topics.[22] In total, 22,627 bigrams (out of 25,819) were assigned to at least one out of these 18 topics.

We additionally created a 19th manually coded topic, *Specific Parties & Politicians*, which consists of bigrams that include the names of individual politicians or parties. As robustness check, we carried out all our estimations ignoring tweets belonging uniquely to this topic. Results were very similar, and throughout we only report estimates for all bigrams, including those unique of this 19th topic.

As a check of the validity of the classification provided by BERT, we asked two master students to manually (and blindly) classify a sample of 510 tweets, which includes tweets on all the topics. Appendix Table H4 displays the accuracy and F1 score of the different topics compared to the manual classification.

4.3.1 Populist and Non-populist Topics

In the model of Section 2 there are only two issues, each distinctive of one political type. To bring the data as close as possible to this structure, we group these 19 topics into two sets, distinctive of non-populist and populist speech respectively.

[20] Tweets only about personal communication such as greetings, weather or holidays

[21] For the anti-establishment topic, we retained only tweets with negative sentiment.

[22] Relatively low thresholds allow us to capture the fact that some bigrams belong to several topics. For example, "sustainable economy" belongs to both climate and business; "world economy" belongs to business, trade, and foreign policy. Higher thresholds miss this multidimensionality across topics.

To do so, let $\hat{q}_{ikt}^{P_i} = \sum_{j \in k} \hat{q}_{ijt}^{P_i}$ denote the probability that i speaks about issue k in quarter t. Let ρ_{ikt} be the posterior that speaker i is populist, conditional on having spoken about topic k, computed with topics (rather than bigrams) as units of speech. Thus, ρ_{ikt} is defined as in equation (10), except that on the right hand side we have $\hat{q}_{ikt}^{P_i}$ rather than $\hat{q}_{ijt}^{P_i}$. A topic is considered populist if the average ρ_k of ρ_{ikt} over the entire sample exceeds 0.5, and non-populist otherwise. We allow for one exception, public health, which exhibits a clear break during the COVID epidemic. We thus split health into two separate topics, before and after COVID (i.e. after Q1 2020).

Table 2 reports, for each topic, the average posterior ρ_k together with the number of bigrams and the standard deviation of ρ_{ikt} across politicians and quarters. Apart from immigration and post-COVID health, the topics most typical of populist politicians are non-policy oriented: mentions of specific parties and politicians, self-promotion, anti-establishment (the idea that the current political system is broken, corrupt, rigged and undemocratic), and elections (topics related to elections, voting or political campaigns). By contrast, the topics most distinctive of non-populist politicians are predominantly policy-oriented, with climate and environment, trade and foreign policy, and inflation among the most pronounced. These patterns are consistent with the idea that populism is a *thin-centered ideology*, with few clearly defined policy positions other than opposition to immigration and globalization (Mudde and Kaltwasser, 2012), and they map onto the model: k_1 topics are plausibly those more salient to populist voters, and k_0 to non-populist voters. The heterogeneity of ρ_{ikt} within a topic is non-trivial, so the precise ranking of individual topics should not be over-interpreted; what matters for the analysis below is only the binary split into a populist set k_1 ($\rho_k > 0.5$) and a non-populist set k_0 ($\rho_k < 0.5$).[23]

This aggregation is also substantively natural. Boundaries between narrowly defined topics are often blurry, and topics within each set can be viewed as close substitutes: politicians may choose bigrams from one topic or another with similar considerations in mind. Within the non-populist set, for instance, topics such as climate and energy policy, or taxation and inequality, often share messages and bigrams. The same applies to the populist set: self-promotion may substitute for election-related content, while immigration often overlaps with anti-establishment rhetoric, as it frequently invokes similar attitudes.

[23]Looking at median ρ_{ikt}, rather than average, gives very similar results. Also, omitting seven quarters around the election (those that are the focus of the event studies below) when computing posterior means yields a similar classification of topics, with one exception: the security topic, currently just below the 0.5 threshold, lies just above it. Results reported in Section 7 are robust to this change.

Table 2: Average partisanship by topic: number of bigrams, mean, and standard deviation

Populist topics

Topic	N bigrams	Mean ρ_J	SD
Specific parties & politicians	764	0.544	0.126
Immigration	503	0.522	0.083
Anti-establishment	127	0.520	0.099
Health (post-COVID)	1718	0.513	0.060
Elections	5488	0.513	0.056
Self-promotion	7282	0.512	0.046

Non-populist topics

Topic	N bigrams	Mean ρ_J	SD
Security	2206	0.498	0.062
Rights	2934	0.494	0.057
Inequality	804	0.492	0.066
Labour market	1217	0.489	0.064
European Union	1642	0.488	0.065
Health (pre-COVID)	1718	0.487	0.067
Taxation	2499	0.487	0.061
Business & industry	2438	0.485	0.062
Energy policy	937	0.485	0.063
Connection	2284	0.483	0.061
Inflation	358	0.480	0.060
Foreign policy	2741	0.480	0.063
Trade policy	427	0.477	0.061
Climate & environment	1357	0.476	0.060

Note: Each row corresponds to a topic. *N bigrams* is the number of distinct bigrams assigned to the topic. *Mean* ρ_k is the average ρ_{ikt} for all politician-quarter observations. *SD* is the standard deviation of ρ_{ikt} across politician-quarter observations within each topic. The left panel reports the six *populist* topics ($\rho_k > 0.5$); the right panel reports the remaining *non-populist* topics ($\rho_k < 0.5$).

5 Partisanship over time

This section illustrates the main properties of our measures of polarization, and describes how they evolved over time.

5.1 Most distinctive bigrams

We begin by examining which words are most indicative of populist versus non-populist rhetoric in each quarter t. Following GST, we assess the distinctiveness of each bigram based on its contribution to our partisanship measure. Specifically, define the populist partisanship of bigram j in quarter t for politician i as the extent to which an observer with neutral priors would change her expected posterior that politician i is populist if that bigram was removed

from the vocabulary, unbeknownst to the observer.[24] The populist partisanship of bigram j in quarter t is the median of this measure across all politicians active in quarter t. Positive values of this measure denote bigrams typically used by populist politicians, negative values by non-populists.

Table 3 below lists the 15 most distinctively populist and non-populist bigrams (i.e. with the largest positive and negative median values) in 2015 Q3 and 2020 Q1. These dates roughly correspond to the inflow of refugees from Syria to Europe and the COVID shock. During the refugee crisis (2015 Q3), non-populist politicians were more likely to refer to immigrants as *refugees*, and mention topics such as *"climate change"* and *"sustainable development"*. Populist politicians instead frequently emphasized phrases such as *"asylum seekers"*, explicitly mentioned that immigration is *illegal*, or mentioned *"border control"*. During the COVID-19 epidemic (2020 Q1), the bigram *"public health"* was distinctive of non-populist discourse, while populist politicians emphasized connection with people (*"take care"*) and continued to focus on immigration. More generally, populists use direct language and emphasize communication ("*good morning*", "*via youtube*"), whereas non-populists refer to the establishment ("*Prime Minister*", "*The Anniversary*").

5.2 Average partisanship

Next, consider average partisanship in calendar quarter t, defined as:

$$\bar{\pi}_t = \frac{1}{I_t} \sum_i \hat{\pi}_{it} \tag{12}$$

where I_t is the number of politicians active in quarter t.

The solid pink and green lines in Figure 3 display average partisanship for P vs NP ($\bar{\pi}_{t,PNP}$) and L vs R ($\bar{\pi}_{t,LR}$) respectively.[25] P/NP partisanship peaks during the European immigration crisis (2014-2016), and it is much higher than for L vs R. This could suggest that competition between P and NP is more central than the L vs R divide. It could also be due to significant variation in how L – R positions are defined and understood across national contexts, providing further evidence of the dominant importance of the populist

[24] Following GST, the populist partisanship of bigram j used by politician i in quarter t is:

$$\xi_{jit} = \frac{1}{2} - \frac{1}{2} \sum_{h \neq j} \left(\frac{q^1_{iht}}{1 - q^1_{ijt}} + \frac{q^0_{iht}}{1 - q^0_{ijt}} \right) \rho_{iht}$$

[25] Appendix Figure H2 reports the results with estimated confidence intervals for average partisanship. Since standard nonparametric bootstrap is invalid for lasso regression, we use subsampling instead, though it is considerably less efficient when the number of units (politicians) is relatively small.

Table 3: 15 most partisan populist and non-populist bigrams in 2015 Q3 and 2020 Q1

2015 Q3		*2020 Q1*	
Non-Populists	**Populists**	**Non-Populists**	**Populists**
Young People	Asylum Seeker	Young People	Take Care
Prime Minister	Greek People	Prime Minister	Good Morning
Climate Change	Good Morning	Will Continue	European Parliament
Sustainable Development	Via Youtube	Public Health	Fake News
The Anniversary	Press Release	Climate Change	Will Never
Economic Growth	Will Never	The Anniversary	Million People
Good Luck	Take Care	Continue Work	Health Care
Will Continue	Border Control	Rule of Law	Asylum Seeker
Work Together	Illegal Immigration	Work Together	People Will
Will Work	Today PM	President Republic	Bernie Sanders
Developed Countries	Jeremy Corbyn	Will Work	People Want
Well Done	I'm Going	Member State	Help You
Civil Society	Don't Want	Protect Health	Stay Home
Syrian Refugee	Common Sense	Economic Growth	Via Youtube
Renewable Energy	Get Rid	Last Year	Work People

Notes. The table reports the bigrams with the largest and smallest median values (taken over i) of the bigram partisanship measure ξ_{jit} for $t \in \{2015\ \text{Q3}, 2020\ \text{Q1}\}$ (footnote 24 defines ξ_{jit}). Larger median values correspond to populist bigrams, smaller median values – to non-populist bigrams.

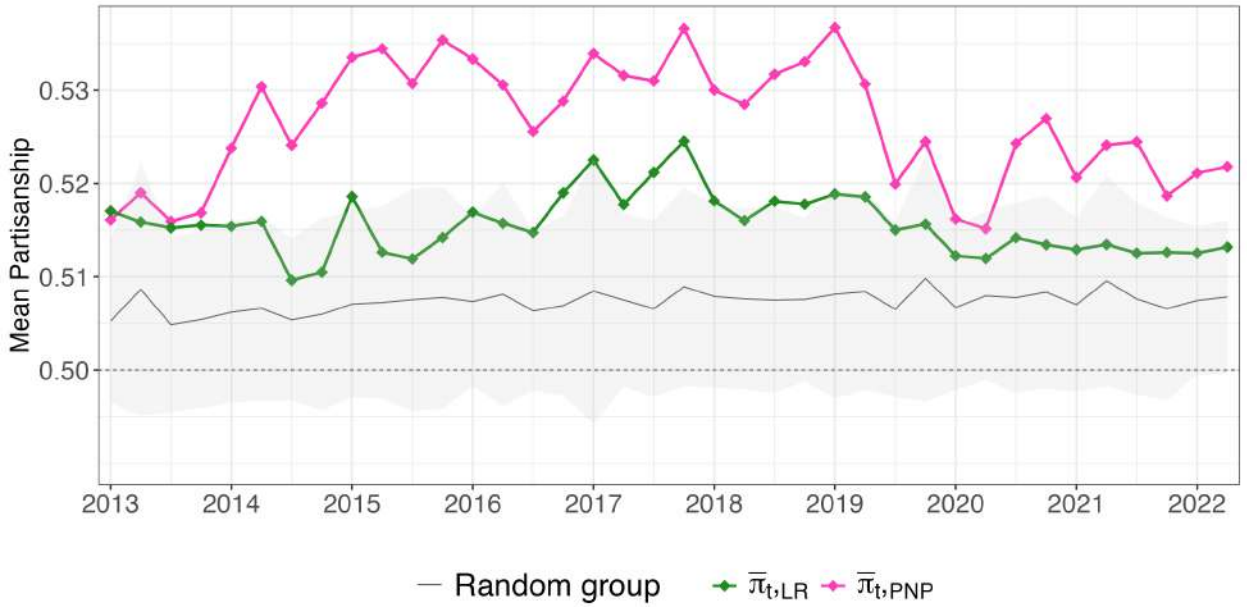
divide in our international sample of politicians.

Our estimate of populist partisanship is notably higher and more volatile than in GST's analysis of U.S. Congressional speeches over the past 35 years. As discussed in Appendix Section E, our estimates of $\hat{\pi}_t$ imply that, upon observing 5 bigrams (the typical content of a single tweet), the posterior of a politician being populist reaches about 0.6 during the refugee crisis. By contrast, the posterior of a speaker being Republican rather than Democrat in GST's sample, conditional on 5 bigrams, peaks at 0.55.[26] The larger and more volatile partisanship in our sample is likely to reflect multiple factors. Tweets represent a high-frequency, low-cost, and highly flexible medium of communication, allowing politicians to react immediately to unfolding events and tailor messages to specific audiences. In contrast, Congressional speeches are more static, tend to follow more standardized formats, with topics and rhetoric constrained by institutional norms. Additionally, our estimates are computed at the quarterly level, rather than every two years as in GST, and politicians differ in their levels of activity across time. These factors jointly contribute to the elevated and more volatile patterns of populist partisanship observed in our setting.

Finally, populist partisanship is substantially higher than that estimated when we randomly divide politicians into two groups reflecting the observed frequencies of true P /NP

[26] In GST's data, partisanship based on their preferred penalized estimator, which is comparable to the one we use, peaks at approximately 0.512 in 2010.

Figure 3: Average partisanship and χ^2 by calendar quarter



Notes. $\bar{\pi}_{t,LR}$ refers to average partisanship defined in the model with two political types: left and right; $\bar{\pi}_{t,PNP}$ refers to average partisanship defined in the model with two political types: populist and non-populist. Random group refers to average partisanship between two randomly defined political types, keeping the relative group sizes similar to those of the populist and non-populist groups; the shaded area is the 95% confidence interval. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.

types in our sample, repeating this procedure 100 times (the black line is the average, the shaded area is the 95% confidence interval). This random assignment serves as a benchmark to quantify the extent of remaining finite-sample bias in our measure of average partisanship. By construction, partisanship within random groups should be around 0.5. In our case, average random partisanship values are generally just below 0.51, although never significantly above 0.5. This is probably due to the relatively small number of politicians in our sample, and to the possible presence of unobserved political types. We discuss this point more extensively in the next section and in the online Appendix G.

Figures H3 and H4 in the online annex also document a peak in populist partisanship within the topic of immigration for the sample of European countries beginning in the fourth quarter of 2014, coinciding with the onset of the mass inflow of Syrian refugees; and a steep increase in populist partisanship within the topic of public health at the onset of the COVID-19 pandemic. The magnitude of partisanship within the immigration topic significantly exceeds that observed for aggregate partisanship, confirming that immigration is a highly salient and mobilizing topic, especially for populist politicians.

6 Electoral Cycles in Partisanship

In this section we exploit the fact that election dates differ across countries, as shown in Appendix Table H3, to describe how partisanship evolves over the electoral cycle. The election date refers to national parliamentary elections for parliamentary regimes. For presidential and mixed systems, the election date refers to all parliamentary elections and presidential elections in which more than two candidates are present in our sample of politicians (France, Mexico, the US, Slovenia, Portugal, Finland, Czech Republic).[27]

6.1 Average partisanship and distance from the election

The left hand panel of Figure 4 plots P/NP partisanship averaged across politicians by quarterly distance d from the election, d=0 being the election quarter. On average, populist vs non-populist partisanship increases in the run-up to the election, reaching a peak in the election quarter, and drops fast right after. In other words, as the election date approaches, the language of populist and non-populist politicians becomes more distinctive. Appendix Figure H5 shows that, when conditioning on calendar-quarter fixed effects, the pre-electoral increase in partisanship is even more pronounced. The shaded light pink area represents 90% point-wise confidence intervals of populist vs non-populist partisanship based on subsampling. Appendix D discusses how the confidence intervals are constructed.

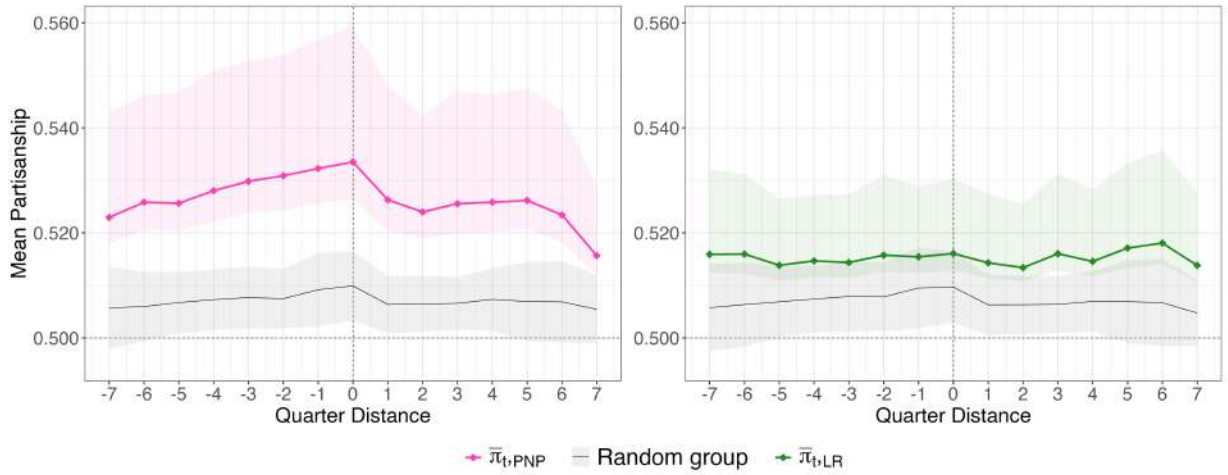
The right panel of Figure 4 plots average Left-Right partisanship over the electoral cycle, and the random group benchmark with L/R proportions. There is no evidence of electoral dynamics. This confirms the dominance of the populist vs non-populist dimension of electoral competition in our sample.

Note that all these patterns refer to aggregate measures of polarization, so they are also affected by changes in the composition of active politicians near the elections. We deal with this issue in Subsection 6.2 below, where we estimate an event study at the level of individual politicians.

The solid line at the bottom of the figure depicts partisanship for randomly grouped politicians, and the shaded gray area is the 95% confidence interval. P/NP partisanship is always well above the random group benchmark. Importantly, the difference between populist partisanship and the random group benchmark increases before the election and peaks at the election quarter, suggesting that the populist dimension of rhetorical polarization has a significant electoral component.

[27] We exclude elections at levels different than national. In our politician-election panel, the mean of the indicator for participating in an election (either directly as a candidate, or indirectly as the member of a party participating in the election) is 0.6.

Figure 4: Average estimated π_{it} per quarter difference from elections.



Notes. $\bar{\pi}_{t,PNP}$ refers to average partisanship defined in the model with two political types: populist and non-populist. $\bar{\pi}_{t,LR}$ – in the model with Left and Right political types. Shaded colored areas behind $\bar{\pi}_{t,PNP}$, $\bar{\pi}_{t,LR}$ are 90% pointwise CI based on subsampling. Random group refers to average partisanship between two randomly defined political types, with group sizes matching those of the corresponding political groups; the shaded grey area behind the random group is the 95% confidence interval. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.

When looking at the average random-group benchmark over the electoral cycle, it lies significantly above 0.5 and exhibits some pre-electoral dynamics. To explain these patterns, Appendix G runs Monte Carlo simulations using synthetic samples of politicians and bigram usage. We explore two issues: we vary the number of politicians in the sample, and we allow for the presence of unobserved political types. Both aspects are relevant. As the number of politicians increases, the random group benchmark converges towards 0.5. But the presence of electorally relevant unobserved groups also matters and can explain the pre-electoral patterns. The reason is that, in the presence of prominent unobserved types, their bigram usage may be spuriously attributed to random groups by the penalized estimator, resulting in partisanship estimates that systematically deviate from the 0.5 benchmark (although we find this effect to be small in practice). Appendix G provides further details on the simulation design and results.[28]

[28]The penalized estimate of partisanship is bounded below by 0.5, but could exceed 0.5 in the random sample of politicians due to its non-linearity and convexity. The reason is that, despite the Lasso penalization, in a relatively small sample some bigrams will, by pure chance, correlate with the random political label. We verify that this is indeed the case by implementing the leave-out estimate of partisanship proposed by GST, which is an out-of-sample estimator, and as such it is much less subject to the positive bias due to Jensen's inequality. The leave-out estimate recovers the 0.5 benchmark for random-group partisanship even when unobserved types are present, in line with the findings of GST.

Magnitudes Recall that partisanship measures the average predictability of a politician's type, conditional upon observing a single bigram. According to Figure 4, on average for populist vs non-populist politicians this predictability increases from 0.522 seven quarters before the election to 0.535 in the election quarter, and then declines by about the same amount two quarters after the election. How can this magnitude be interpreted?

To address this question, Appendix E computes speech informativeness conditional on observing n bigrams, for different values of n, following GST. These computations imply that, upon reading a single tweet two years before the election, an observer with neutral priors would raise its posterior probability of correctly classifying the politician as P or NP by about 7 percentage points (from 0.5 to about 0.57). During the election quarter, instead, the increase in predictability from reading one tweet jumps to almost 12 percentage points (from 0.5 to 0.62), almost doubling the informativeness of a single tweet. After reading two tweets (10 bigrams) during the election quarter, predictability increases by 17 percentage points (from 0.5 to 0.67), in contrast to an increase of 12 percentage points (from 0.5 to 0.62) if two tweets are read two years before the election, an increase of about 50%. Note that the largest marginal increases in informativeness are obtained after just a few bigrams.

We can compare these estimates to the informativeness of text in 2015 Q3, during the first acute phase of an immigration crisis and when partisanship reached a maximum in our sample. The texts from 2015 Q3 are slightly less informative about populist partisanship than the texts written during the election quarter.

6.2 Event study

The estimates displayed in Figure 4 could reflect the coincidence of elections with other relevant events, or changes in the composition of active politicians. To control for these potential confounders, we exploit staggered election dates across countries and estimate an event study on i's partisanship, π_{it}, with fixed effects for politician (γ_i), calendar date (δ_t) and country by election-window (defined as country-quarters closest to each election, El_{it}):

$$\pi_{it} = \sum_{d(i,t)} \beta_{d(i,t)} D_{d(i,t)} + \gamma_i + \delta_t + El_{it} + \epsilon_{it}, \tag{13}$$

where $D_{d(i,t)}$ are indicators for quarterly distance from the election - which in our baseline specification include 3 quarters before and after the election as well as the election quarter. Country by election-window fixed effects, El_{it}, are included to capture country specific shocks occurring during the election-window, but results are robust to dropping El_{it}. Thus, the estimated coefficients $\beta_{d(i,t)}$ measure the average effect on rhetorical polarization of being at

distance $d(i,t)$ from the election, relative to the average quarters at a greater distance from the election within each election-window, for the same politician. Errors are clustered at the (closest) election level - the treatment variable of interest.[29]

The event study for P/NP partisanship, depicted in Figure 5, confirms that the language used by politicians becomes more polarized two quarters before the election and remains so until the election date. Once the election is over, however, rhetorical heterogeneity declines to its average level outside of the election window. In terms of magnitudes, the estimated rise in populist partisanship during the election quarter relative to non-election quarters is about 15% of one standard deviation of the populist partisanship measure across the full sample (0.008/0.054).

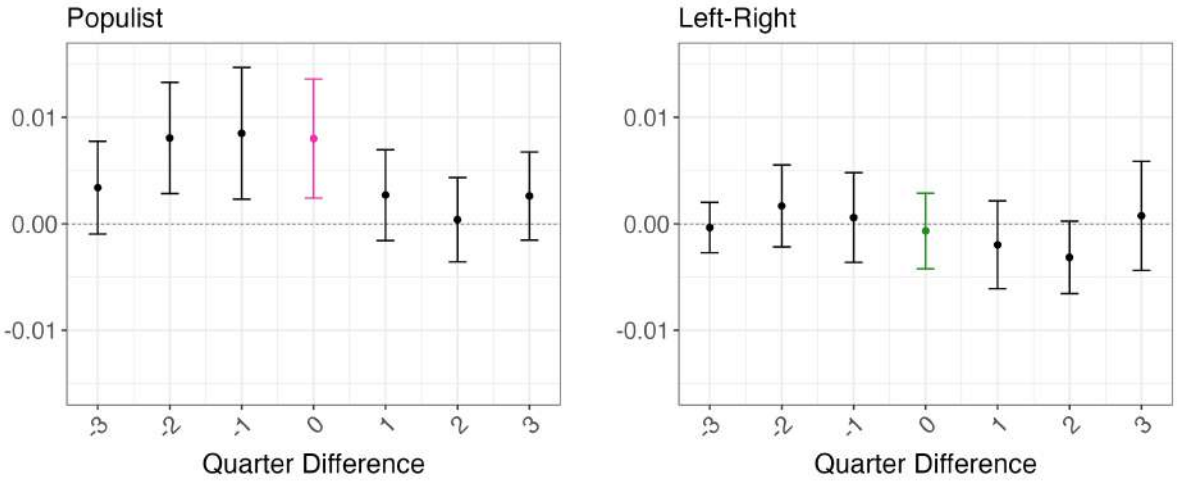
In Appendix Table H2, we report a Wald test for the null that the estimated coefficient β for the election quarter ($d(i,t) = 0$) is equal to those after the election. The null is rejected with p-values of 0.05, 0.03 and 0.3 for 1, 2 and 3 quarters after the elections respectively.

The right panel of Figure 5 depicts the event study for L vs R partisanship. Rhetorical polarization between these two groups does not rise before the elections. Appendix Figure H6 depicts the event study on partisanship when the probabilities q are estimated allowing for four political types, as in Appendix C. The left-most panel refers to populist vs non-populist partisanship, the center panel to left-right partisanship, and the right-most panel refers to partisanship between four types (note that the scale here is smaller, since with four types partisanship ranges between 0.25 and 1). The pre-election pattern is present only for populists vs non-populists and for the four types, but not for the left-right division.

Overall, these findings suggest that the pre-electoral increase in rhetorical polarization is only present for the populist dimension of political competition. From here on we focus in more detail on this dimension only.

[29] That is, clustering is among observations sharing the same fixed effect El_{it}. Thus, we allow arbitrary correlations in residuals among observations closest to the same country-election date. To define El_{it}, for each country and calendar quarter, we find which national election is closest to the middle day of that quarter. This creates an electoral-cycle fixed effect by country, with no ties because closest is defined using the exact middle day of the quarter.

Figure 5: Partisanship event study



Notes. Left panel plots the event study results for the populist partisanship estimates; Right panel – the Left-Right estimates. Each figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13. Outcome variable is populist/Left-Right partisanship π_{it}. Errors are clustered at the elections level, 95% CI are reported. Average populist partisanship is 0.523, average Left-Right partisanship is 0.516. Partisanship is defined as in Equation 11.

7 Between-issue and Within-issue Divergence

So far we have established that, as predicted by the theory, rhetorical polarization between P and NP politicians rises in the quarters before an election and reverts afterwards (Section 8 shows that this pattern is very robust). This result establishes *that* competing politicians diverge, but not *how*. In this section we test the two sharper predictions about the anatomy of this divergence. To do so, we focus on the electoral cycle in specific components of $\mathbf{q}^1_{it} - \mathbf{q}^0_{it}$, measuring the difference in the probability of bigram usage between opposite political types.

7.1 Between-issue divergence

Prediction 1 says that, as the election approaches, both political types speak more often about the issues more salient to their constituency, and hence distinctive of their type—the *extensive margin* of bigram choice across topics. To test it, we exploit the two sets of topics listed in Table 2, namely we pool all populist topics (with $\rho_k > 0.5$ in Table 2) into a single broader populist set indexed by k_1, and all non-populist topics (with $\rho_k < 0.5$) into a broader non-populist set indexed by k_0.[30]

[30] In principle the analysis could be run separately for a larger number of topics, but in our setting many individual topics are too small to support this. See however the robustness analysis in section 8.

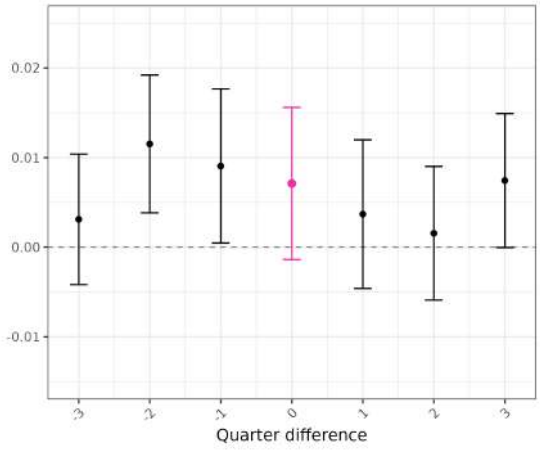
Let $q^1_{i,k_1,t} = \sum_{j \in k_1} q^1_{ijt}$ and $q^1_{i,k_0,t} = \sum_{j \in k_0} q^1_{ijt}$ respectively denote the probability that politician i uses bigrams drawn from populist (resp. non-populist) topics, if i is of type P. Variables $q^0_{i,k_1,t}$ and $q^0_{i,k_0,t}$ are similarly defined, for type NP. Because the type-specific shares are not separately identified, we work with their difference,

$$\Delta q_{i,k_1,t} = q^1_{i,k_1,t} - q^0_{i,k_1,t},$$

namely the gap in the frequency with which P and NP politicians (with the same observable characteristics) discuss the populist set k_1. We normalize bigram frequencies as if speech only consisted of classified bigrams, so that by construction $\Delta q_{i,k_1,t} = -\Delta q_{i,k_0,t}$.[31]

The prediction is that $\Delta q_{i,k_1,t}$ increases before the election and then reverts to the off-election average. We test it with the event-study specification of Equation (13), using $\Delta q_{i,k_1,t}$ as the outcome. Figure 6 presents the results. In the quarters preceding the election, P types talk more frequently about populist topics, compared to NP types with the same features—and less frequently about non-populist topics. When electoral stakes rise, politicians tilt the agenda toward the issues that are more distinctive of their type over the entire sample, as predicted by the theory. The effect is sizable: the pre-electoral coefficient estimates are around 15% of the standard deviation of the dependent variable.

Figure 6: Predicted difference in frequencies of populist-topic speech between populist and non-populist politicians



Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13 with $\Delta q_{i,k_1,t} = q^1_{i,k_1,t} - q^0_{i,k_1,t}$. $\Delta q_{i,k_1,t}$ is a predicted difference in frequencies of speech for the populist topic between populist and non-populist politicians. Errors are clustered at the elections level, 95% CI are reported.

[31] About 12% of bigrams remain unclassified, meaning that the two dimensions do not cover the entire vocabulary.

7.2 Within-issue divergence

Prediction 2 concerns the *intensive margin*: as the election approaches, politicians use bigrams more extreme and more distinctive of their type *within* a given issue. To test it, we use a time-invariant measure of bigram populist distinctiveness, $\xi_j = \text{median}_{i,t}\, \xi_{ijt}$, where ξ_{ijt} is the bigram-level contribution to partisanship defined in footnote 24 and obtained from a specification of (7) in which we constrain ϕ_j to be constant over time.[32] The index ξ_j has both sign and magnitude: positive values identify populist bigrams, negative values non-populist ones, and the absolute value measures how strongly the bigram is associated with one type. Because ξ_j is fixed over time, it provides a measure of bigram *extremism* over the entire sample period in the ideological dimensions distinguishing populist and non-populist politicians.

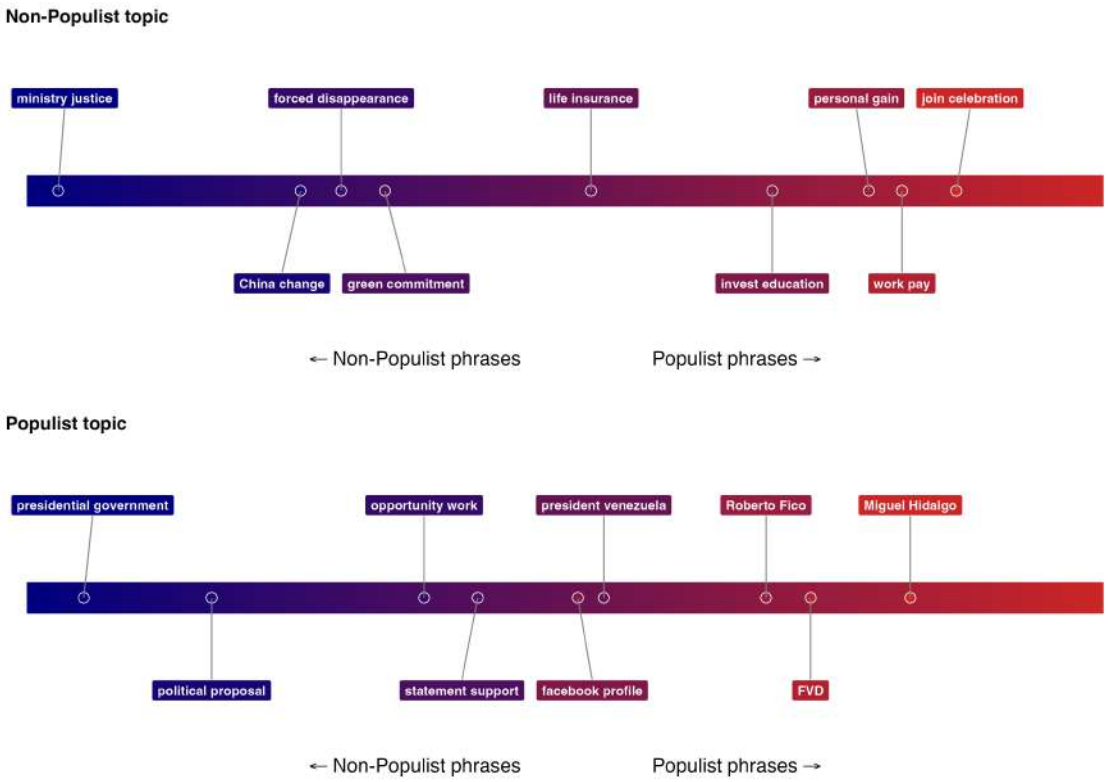
Figure 7 plots examples of bigrams with different values of ξ_j, within the set of populist (k_1) and non-populist topics (k_0) listed in Table 2. While these dimensions do not map onto the traditional Left-Right cleavage, they meaningfully organize the competition between populist and non-populist politicians. Non-populist topics are more policy-oriented, with distinctively non-populist bigrams referring to foreign policy or the environment, among others, while populist bigrams focus on issues and rhetoric closer to that of populist politicians. By contrast, within populist topics, distinctively populist bigrams refer to populist politicians or their parties, such as Roberto Fico or FvD, and invoke national historical figures, such as Miguel Hidalgo, whereas non-populist bigrams concern democratic institutions and bodies of the establishment. Overall, the figure highlights substantial differences in language across political types within each set of topics, and provides guidance for interpreting the within-topic divergence results that follow.

We normalize $\mathbf{q}^0_{iKt}$ and $\mathbf{q}^1_{iKt}$ to sum to one within each set of topics $K = k_1, k_0$, quarter t and politician i, so that the analysis isolates the composition of language within a set of topics. As before, only the difference between the two political types is identified. Define $Y_{iKjt} = q^1_{i,K,j,t} - q^0_{i,K,j,t}$. This variable measures the gap between P and NP types in the frequency of use of bigram j assigned to topics K. Prediction 2 says that, before the election, Y_{iKjt} *increases* for bigrams more distinctively populist (i.e. with higher values of ξ_j) within each set of topics $K = k_0, k_1$, and *decreases* for less distinctively populist bigrams (i.e. with lower ξ_j) within each K.

To test this without imposing a functional form on how the gap varies with distinctiveness, we partition the bigrams of each set of topics K into overlapping 20-percentile windows of ξ_j,

[32] This enables us to measure how *distinctive* is bigram j of each political type over the entire sample period and not just in quarter t, allowing a comparison of bigram usage over time. The results are robust to taking median over ξ_{ijt} using only quarters outside the election window, or a random subsample of quarters.

Figure 7: Examples of bigrams by ideological extremism across both sets of topics



Notes. The figure shows the original, non-stemmed versions of randomly selected bigrams from topics distinctive of non-populist and populist politicians, based on their distinctiveness ξ_j, defined as in footnote 24. Bigrams to the left (dark blue) are more distinctive of non-populist politicians, while bigrams to the right (bright red) are more distinctive of populist politicians.

stepped by 10 percentiles, yielding 9 percentile windows: $[0, 20\%)$, $[10, 30\%)$, ..., $[80, 100\%]$. Letting $\mathcal{Q}_{K,b}$ be the set of bigrams in window b within topics K, we define, for each politician i, quarter t, set of topics K and window b,

$$Y_{iKbt} = \sum_{j \in \mathcal{Q}_{K,b}} Y_{ijt},$$

and estimate the event-study specification of Equation (13) separately for each window b, with Y_{iKbt} as the outcome. We obtain a set of estimated coefficients, $\hat{\beta}_{Kdb}$, for each set of topics K, each quarterly distance d from the election and for each window of populist distinctiveness b. These estimated coefficients capture how frequently bigrams in window b and topics K are used by politicians of type P (relative to type NP) at elections distance d, compared to their relative frequency of use in quarters more distant from the election.

Figure 8 plots these estimated coefficients $\hat{\beta}_{Kdb}$ and their 95% confidence intervals, for each quarterly distance to the election d and for each quantile window b ordered from the

most non-populist bigrams (dark blue, left) to the most populist bigrams (bright red, right), following the same color scale as in Figure 7. Darker colors denote statistically significant estimates $\hat{\beta}_{Kdb}$. Panel (a) refers to non-populist topics, panel (b) to populist topics. The pattern matches the prediction. Within both sets of topics, starting two or three quarters before the election and until the election date, P politicians use more frequently the more distinctively populist bigrams—and less frequently the less distinctive ones— relative to NP politicians, compared to quarters distant from the election. The estimated values of $\hat{\beta}_{Kdb}$ in pre-electoral quarters are about 10-30% of the standard deviation of the dependent variable, depending on windows and distance from election.

Results are especially pronounced for non-populist bigrams within populist topics, which, as noted above, mostly concern democratic institutions and bodies of the establishment. While we cannot tell whether this is due to P types using these bigrams less or NP types using them more, the imminence of elections clearly widens the pre-existing gap in their usage (the mean of Y_{iKbt} for the bigrams in the lowest five windows b is below 0). Paradoxically, a byproduct of this particular form of rhetorical divergence during an electoral campaign could be to erode trust in democratic institutions and undermine their perceived legitimacy.

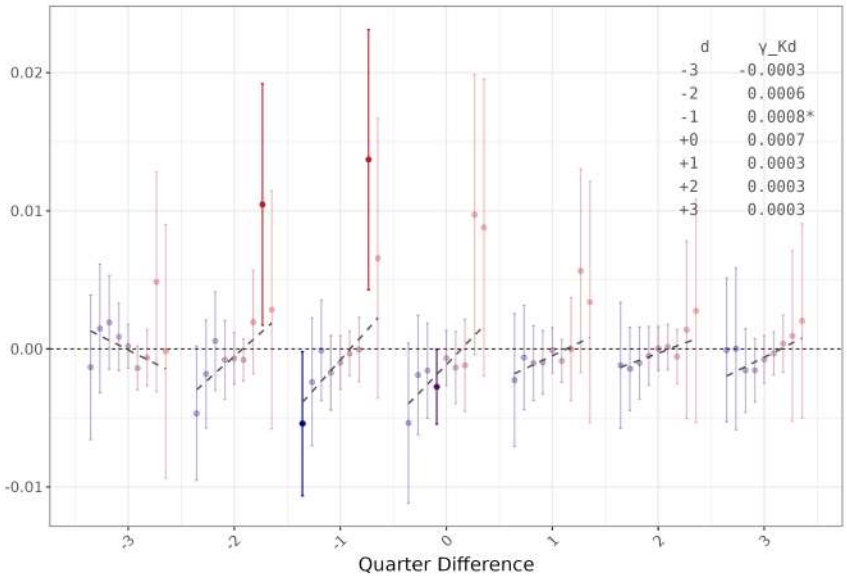
The gradient of how the event-study estimates $\hat{\beta}_{Kdb}$ vary with populist distinctiveness (i.e. with window b) is broadly monotonic: $\hat{\beta}_{Kdb}$ is lowest for the bottom windows (more non-populist bigrams) and highest for the top windows (more populist bigrams). After the election, the pattern in the $\hat{\beta}_{Kdb}$ coefficient is much less pronounced and the estimated $\hat{\beta}_{Kdb}$ are never statistically significant, for all windows b.

To summarize this gradient while accounting for estimation uncertainty, we fit a linear trend across windows by weighted least squares. Namely, for each quarterly distance $d \in \{-3, \ldots, 3\}$ we estimate the cross-window slope

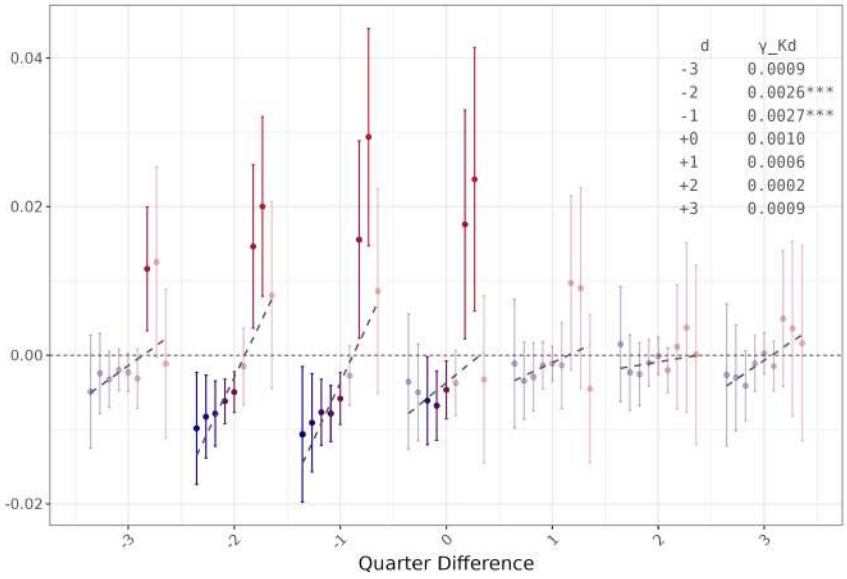
$$\hat{\beta}_{Kdb} \;=\; \alpha_d + \gamma_{Kd}\, b + u_{db}, \qquad \text{weights} = \frac{1}{\widehat{SE}^{\,2}_{Kdb}}.$$

where $\widehat{SE}_{Kdb}$ is the estimated standard error of $\hat{\beta}_{Kdb}$. The standard error of $\hat{\gamma}_{Kd}$ accounts for both the estimation uncertainty in the $\hat{\beta}_{Kdb}$ and their correlation across windows. For each electoral distance d, the slope coefficient γ_{Kd} captures how strongly the relative tendency of P to use a bigram rises with its populist distinctiveness. The gray dashed line in each panel plots the fitted trend, and the slope γ_{Kd} and its significance are reported in the upper-right corner. For both sets of topics the slope is largest in the two quarters immediately preceding the election, and the effect is more pronounced for populist topics. The latent measure ξ_j, recovered from overall speech, thus shapes communication over the electoral cycle in the way the theory predicts, and especially so on the issues that define populist identity.

Figure 8: Event-study estimates for within-topics divergence separately for populist and non-populist dimensions.



(a) Non-Populist Topics, $K = k_0$



(b) Populist topics, $K = k_1$

Notes. Colors in the output plots correspond to windows, ranging from dark blue (most non-populist bigrams) to bright red (most populist bigrams). Non-significant coefficients ($p > 0.05$) are shown in faded colors. The figure plots the event study estimates from the specification 13 with an outcome being differences in bigram usage between populist and non-populist politicians, $Y_{ijt} = q^1_{i,j,t} - q^0_{i,j,t}$, across bigrams with varying distinctiveness ξ_j. $\mathbf{q}^0_{iKt}$ and $\mathbf{q}^1_{iKt}$ are normalized to sum to one within each set of topics $K = k_1, k_0$, quarter t and politician i. Bigrams are grouped into overlapping 20-percentile windows of distinctiveness values ξ_j, with a 10-percentile overlap, separately for non-populist and populist topics (J_0 and J_1).

Taken together, the two tests are in line with both predictions of the model. As the election approaches, populists and non-populists diverge along the *extensive* margin—each type tilts its agenda toward the issues that favor its own constituency (Prediction 1)—and along the *intensive* margin—each type frames those issues in increasingly distinctive language (Prediction 2). This is the anatomy behind the aggregate rise in partisanship of Section 6: the pre-electoral divergence is not confined to topic selection or to slogans and party names, but extends to how policy issues themselves are discussed.

8 Robustness

We now discuss a number of possible concerns with our estimates of partisanship, and show their robustness.

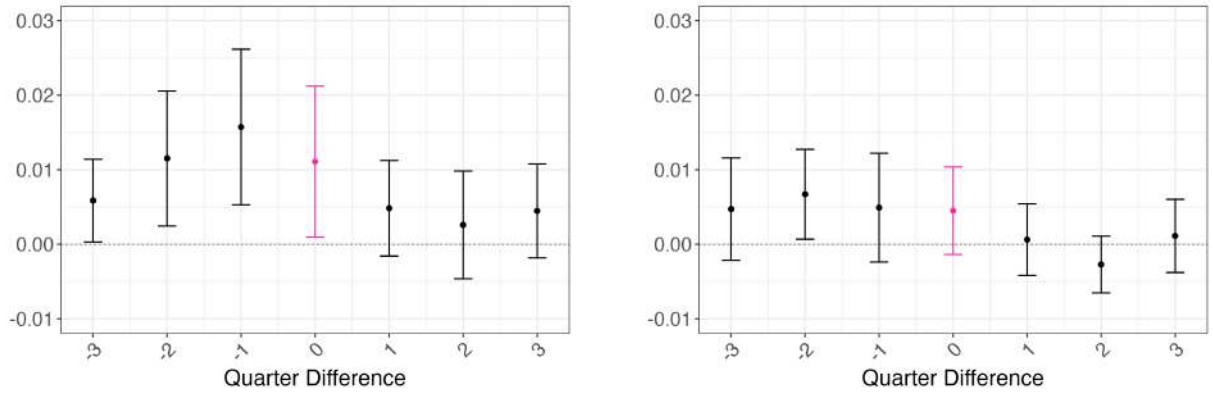
Different electoral systems In standard models of electoral competition where candidates announce their policy positions ahead of the elections, like Downs' or probabilistic voting, candidates' incentives to converge depend on the number of competing candidates. Two vote maximizing candidates who compete for the same voters have strong incentives to converge. With more than two candidates or with the threat of entry by a third candidate, however, convergence is not assured (e.g. Palfrey (1984)). Although we have emphasized the distinction between only two political types, several countries in our sample have more than two parties. Could this account for the observed divergence?

To address this question, we split the sample in two groups of elections: (i) all national elections held in countries with plurality rule and in presidential regimes plus all presidential elections, where competition is typically between two serious parties or candidates; (ii) parliamentary elections in proportional systems, that typically have several competing parties and candidates in the same districts.[33]

Figure 9 reproduces the event study separately for these two groups of elections. While the pre-electoral polarizing effect is present in both groups of elections, it is much larger for plurality rule and presidential elections. Hence, it is unlikely to be driven by competition between more than two parties.

[33]Group (i) includes all elections held in Australia, the UK, the US, France, Brazil, Mexico and Hungary, plus presidential elections with more than one active candidate in our sample of political leaders in Finland, Poland and Slovenia. In Brazil presidential and parliamentary elections are always held in the same dates. In Mexico three congressional elections in our sample were held without a contemporaneous presidential election; we nevertheless include all Mexican elections in group (i) because they tend to be dominated by electoral competition for president. Hungary has a mixed system that strongly favors the largest party.

Figure 9: Partisanship event study estimates for elections under different electoral systems.



(a) Plurality rule and presidential elections

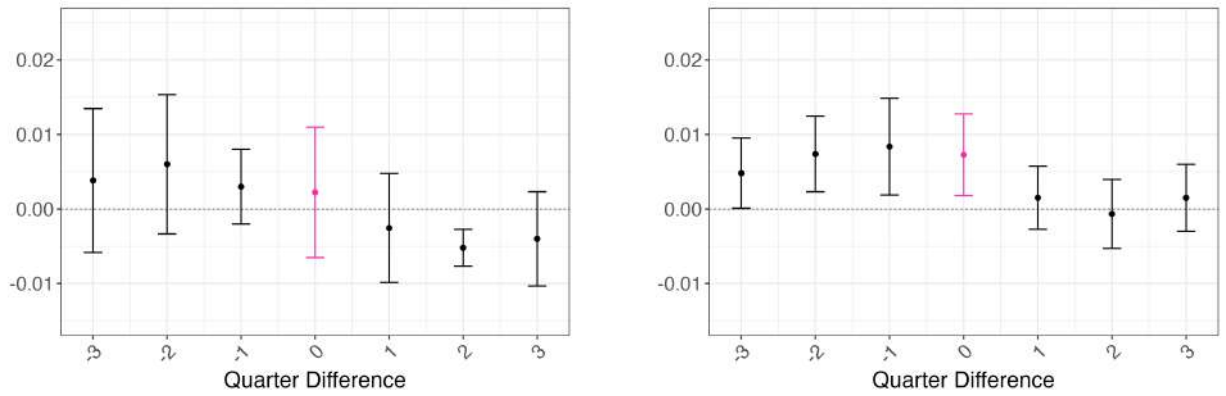
(b) Proportional rule

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13 separately for the two subsamples of elections held under different institutional rules. Left Panel plots the results for all national elections held in countries with plurality rule and in presidential regimes, as well as for all presidential elections, where competition is typically between two serious parties or candidates; Right Panel – for parliamentary elections in proportional systems, that typically have several competing parties and candidates in the same districts. Outcome variable is partisanship, π_{it}, as defined in Equation 11. Errors are clustered at the elections level, 95% CI are reported.

Snap elections About 20% of elections in our data are snap elections - i.e. called prematurely before the end of the legislature. Although snap elections do not happen without premonitions, they may have a shorter pre-electoral cycle in communications. Moreover, unobserved confounding events could cause both a rise in rhetorical polarization and the occurrence of snap elections. To allow for this, we estimate the specification separately for elections held on time and the elections held ahead of time. Figure 10 shows that indeed snap elections induce a shorter and less precisely estimated cycle of polarization (also because of the smaller sample), which is more focused on the electoral quarter itself. Nevertheless, results remain unaffected for the regular election dates.

Being a candidate Although all politicians in our sample are prominent political leaders with strong stakes in the election outcomes, not all of them stand as candidates in all elections. On the one hand, one would expect that the leaders who are also candidates have sharper electoral incentives, and so their rhetoric should exhibit a stronger pre-electoral cycle. On the other hand, rhetorical polarization could be the result of politicians diverging in their immediate goals – some of them are seeking reelection, while others are continuing with their usual agenda, or supporting running candidates. To explore these conjectures, we analyze how the results differ among politicians who are candidates in given elections, and those who are not. Figure 11 shows that the pre-election cycle is starker among the running

Figure 10: Event study estimates for π_{it} for snap elections (left panel) and elections held in standard time (right panel)

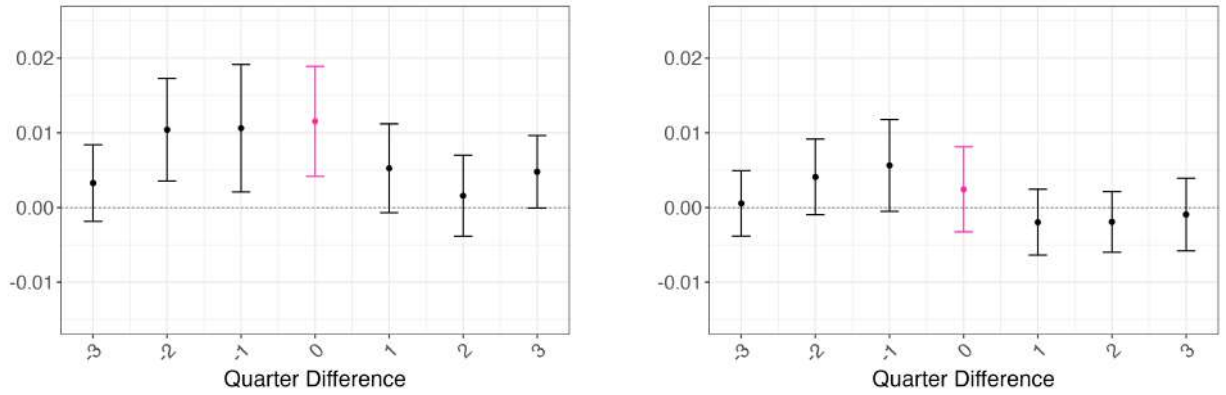


(a) Snap elections (b) Not snap elections

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the subsample of snap elections (left panel) and elections held at a regular time (right panel). Errors are clustered at the elections level, 95% confidence intervals are reported. Partisanship is defined as in Equation 11. Snap elections have been manually classified, with the use of both ChatGPT and Wikipedia. We consider elections to be "snap" if they are held at least 6 months before the natural end of the legislature. Out of 97 unique elections in our sample, 20 elections were classified as "snap".

candidates, in line with the idea that it reflects their stronger electoral incentives, rather than different goals of politicians in more heterogeneous situations.

Figure 11: Event study estimates for π_{it} separately for politicians running for reelection in given elections (candidate, left panel), and those who are not running (not candidate, right panel)



(a) Candidate (b) Not candidate

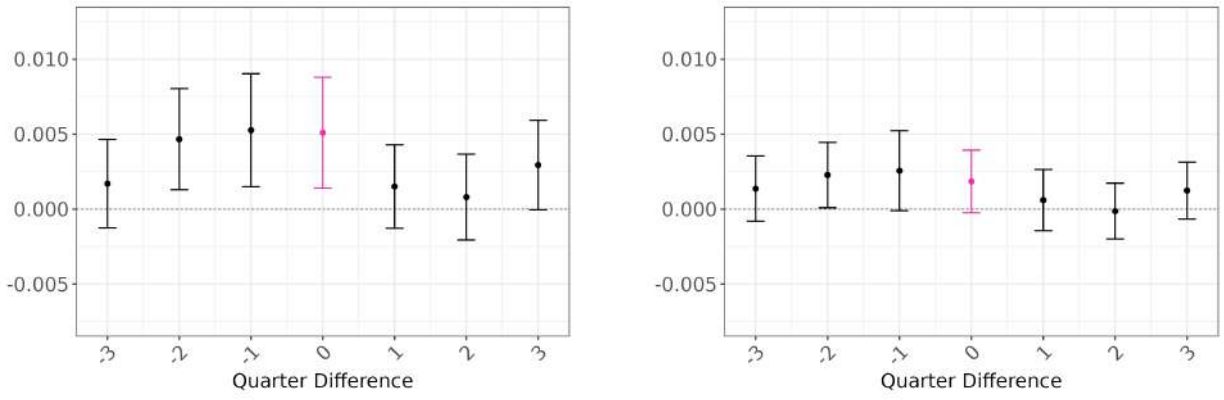
Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the subsample of politicians running for office in a given electoral cycle (left) and those not running (right). Errors are clustered at the elections level, 95% confidence intervals are reported. Partisanship is defined as in Equation 11.

Partisanship between and within topics Our classification of topics into two categories, each one distinctive of a political type, may be subject to measurement error. An alternative approach is to study between- and within-topic divergence on the primitive 19 topics, without attempting to classify them as distinctive of P vs NP types. Although we lose the ability to test the more specific theoretical predictions, we can still explore whether increased rhetorical divergence before the election is due to politicians talking about different issues, or framing each issue differently, or both.

Following GST, define partisanship *within* topic k in quarter t, $\hat{\pi}^k_{it}$ as in (11), except that $\hat{q}^{P_i}_{ijt}$ is averaged only over bigrams that are used to speak about topic k. Within-topic partisanship of i for all 19 topics is the average of $\hat{\pi}^k_{it}$ over topics, weighting each topic k by its frequency of use by i in quarter t. Partisanship of i *between* topics, instead, is constructed using $\hat{q}^{P_i}_{ikt} = \sum_{j \in k} \hat{q}^{P_i}_{ijt}$ computed with topics (rather than bigrams) as units of speech - i.e., using the probability that i speaks about issue k in quarter t. In all cases, $\hat{q}^{P_i}_{ijt}$ is always estimated over the entire sample of bigrams. The between-topic component measures how predictable a politician's type is from the topics he chooses to discuss; the within-topic component measures how predictable his type is from the specific bigrams he uses, given the topic.

Figure 12 reports the event studies. Both between- and within-topic partisanship between P and NP increase before the election, with stronger results within topics. Thus, both the extensive and intensive margin contribute to explain the electoral pattern in partisanship, as predicted by the theory.

Figure 12: Event studies for the between- and within-topics partisanship



(a) Within-topics (b) Between-topics

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13, using within-topic partisanship (left panel) and between-topic partisanship (right panel) as outcome variables. Errors are clustered at the elections level, 95% CI are reported. Section 8 defines within- and between-topic partisanship.

English speaking countries Given that our analysis is in a multinational context and we translate all the tweets to English, one might suspect that our results are driven more by English-speaking countries. This could also potentially explain why the electoral cycles are more pronounced in the group of plurality rule and presidential elections, where the fraction of English speaking countries is higher. But Appendix Figure H7 shows that this is not the case: partisanship increases in both groups of countries, with the effect actually being more pronounced in non-English-speaking countries. As discussed in Section 4, our back translation metrics also point to a high quality of our translations (in terms of similarity in semantic meaning between the original and the back translated tweets).

Time window We also perform robustness checks, with regard to the length of the time window of the event study. In Appendix Figure H8 we consider longer windows of 4, 5, 6 quarters (rather than 3) around the elections. In principle this could matter because it changes the non-treated observations. But the main results are confirmed, and the election cycle generally begins two quarters before the election, as in the baseline estimates.

Heterogeneous treatment effects As in any two-way fixed-effects design, our event-study estimates could be contaminated by treatment effects that are heterogeneous across elections. Appendix Figure H9 shows that our results are robust to this concern. We re-estimate the specification with leader $\times$ election-window fixed effects, restrict the reference group to a narrow set of pre-election quarters, and implement the interaction-weighted estimator of Sun and Abraham (2021), which interacts the quarter-to-election indicators with election-cycle indicators to recover a separate effect for each cycle at each relative quarter and then aggregates these with non-negative weights.

Verbosity Politicians are more verbose as the election approaches, and verbosity reverts to its average afterwards (Appendix Figure H10). This too is an implication of the theory. Could this electoral pattern in verbosity, rather than the language used, explain our divergence results? We don't think so. First, partisanship measures predictability conditional on a single bigram, not on the entire quarterly sequence of bigrams. Second, verbosity is a parameter in the Poisson model used to estimate bigram frequency. So, unless there is a direct link between verbosity and speech polarization, this should not be a concern (Gentzkow, Shapiro and Taddy, 2019). Third, the model is estimated at the level of calendar quarters, and the political-type coefficient ϕ_{jt} is not individual specific by design. Hence, verbosity of individual politicians cannot directly translate into higher partisanship estimates.

Self-selection into being active A potentially more relevant issue is the extensive (rather than intensive) margin of politicians' activity. Quarters closer to the election could have a larger number of active politicians. By construction, our measure of segregation is based on estimated non-zero coefficients ϕ_{jt} from the Lasso-like optimization problem. Hence, all else equal, the larger is the number of speakers active during the quarter, the more there are non-zero estimated bigram-specific loadings on the populist dummy variable, and the larger is the estimated value of partisanship.

This issue too is not a concern in our sample, however, for several reasons. First, we estimate partisanship over calendar quarters. Hence, the number of speakers could be causing a problem only if those calendar quarters with a larger number of speakers are also more likely to be election quarters. This is hardly the case (see figure H11 in the Appendix).

Second, while our results indicate the profound effect on partisanship in the electoral quarter and in the two quarters immediately preceding it, the number of active speakers does not differ much within a six months window around election time (Appendix Figure H12). The reason is that our sample consists almost exclusively of high-ranking politicians, most of whom use Twitter on a regular basis to communicate with the electorate. This is illustrated in Appendix figure H13, which provides the distribution of the number of politicians corresponding to each maximum absolute quarterly distance from the elections. There are almost no politicians who tweet exclusively around electoral quarters.

Third, in the event study we explicitly control for all quarters in the election window, thus capturing mean partisanship levels during these quarters. Hence, indirectly, this captures the popularity of the quarter too.

Fourth, to make sure that our results are not driven by different numbers of speakers per calendar quarters, we perform a robustness check, keeping 220 speakers[34] in each calendar quarter at random. Note that most quarters in our sample have more than 220 active speakers, as shown in Figure H11, so this simulation could give less precise and smaller estimates of partisanship. The event study results of such robustness analysis are in Appendix Figure H14 and confirm the main results.

Sample of bigrams We also applied different thresholds for bigram selection. Appendix Figure H15 shows that the event studies are unaffected.

Populist classification Our findings remain robust under an alternative classification of populist politicians, in which ambiguous cases are coded as non-populists rather than as

[34] 220 is the minimum observed number of politicians (2013 Q1) that allows us to perform selection without replacement.

populists, as in the main specification – Appendix Figure H16.

Mentions of specific politicians and parties One concern is that our results may be driven by mentions of specific parties and politicians closer to elections. Appendix Figure H17 shows that the results are robust to excluding bigrams of the specific parties & politicians topic. Moreover, if this was driving the result, we should also observe pre-electoral divergence for L vs R politicians, but we don't.

9 Concluding Remarks

This paper studies how electoral competition shapes political communication. First, we formulate a theory of electoral competition with exclusive audiences. Second, using high-frequency data from Twitter/X for 367 leaders in 21 democracies over 2013 - 2022, we document a rise in rhetorical polarization among populist vs non-populist politicians before the elections, and unpack it. The populist dimension plays a central role in this process. Compared to the conventional left-right divide, populist versus non-populist polarization is both more pronounced and more responsive to the electoral cycle.

This electoral cycle suggests that rhetorical divergence is not driven solely by fixed ideological differences. Rather, it reflects an opportunistic response to electoral incentives, as politicians strategically differentiate their communication when electoral stakes are highest. Guided by the theory, we show that this pre-electoral differentiation involves politicians: (i) speaking more about their distinctive issues, and (ii) choosing more extreme and more distinctive vocabulary within each issue. These patterns are consistent with a setting in which populist vs non-populist politicians do not compete for the same swing voters, but instead reach and seek to mobilize different groups of voters.

These findings point to two important directions for future research on polarization. First, is pre-electoral differentiation a general feature of political communication, or is it only present on Twitter/X? The convergence results by Di Tella et al. (2025) on candidates' official positions suggest that social media may play a special role. But this question deserves further attention, both with regard to other social media, and because the convergence result in Di Tella et al. (2025) reflects a comparison of final elections with US primaries or French first round elections, whereas we compare pre-election quarters with post-election or more distant quarters.

Second, greater attention should be paid to how platform design and targeting technologies affect political competition. Much of the literature on social media has emphasized that echo chambers or the diffusion of emotional content may fuel political extremism, although

the evidence points in both directions (Melnikov, 2024 and Allcott et al., 2025). Our results highlight another important mechanism, operating on the supply side. Some social media enable politicians to select their audience. For instance Allcott et al. (2025) show that, in the 2020 Presidential election, political ads on Facebook and Instagram predominantly reached the parties' own supporters. This targeting in turn may profoundly alter candidates' incentives, encouraging differentiation and polarizing propaganda.

References

Alesina, Alberto. 1988. "Credibility and policy convergence in a two-party system with rational voters." *The American Economic Review* 78(4):796–805.

Allcott, Hunt, Matthew Gentzkow, Ro'ee Levy et al. 2025. The Effects of Political Advertising on Facebook and Instagram before the 2020 US Election. Technical report National Bureau of Economic Research.

Ash, Elliott, Massimo Morelli and Richard Van Weelden. 2017. "Elections and Divisiveness: Theory and Evidence." *The Journal of Politics* 79(4):1268–1285.

Ash, Elliott and Stephen Hansen. 2023. "Text algorithms in economics." *Annual Review of Economics* 15:659–688.

Aslanidis, Paris. 2018. "Measuring populist discourse with semantic text analysis: an application on grassroots populist mobilization." *Quality & Quantity* 52(3):1241–1263.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political analysis* 23(1):76–91.

Battiston, Giacomo. 2022. "Rescue on Stage: Border Enforcement and Public Attention in the Mediterranean Sea." *Available at https://www.dropbox.com/sh/ct8qe4x71l7hjry/AABxUJxsYmQy1Tea2* .

Besley, Timothy and Stephen Coate. 1997. "An economic model of representative democracy." *The Quarterly Journal of Economics* 112(1):85–114.

Bonica, Adam, Kasey Rhee and Nicolas Studen. 2025. "The Electoral Consequences of Ideological Persuasion: Evidence from a Within-Precinct Analysis of US Elections." *Available at SSRN* .

Bonomi, Giampaolo. 2024. "Divide and diverge: Polarization incentives." *Available at https://ssrn.com/abstract=4990250* .

Brady, David W., Hahrie Han and Jeremy C. Pope. 2007. "Primary Elections and Candi-

date Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1):79–105.

Canes-Wrone, Brandice, David W Brady and John F Cogan. 2002. "Out of step, out of office: Electoral accountability and House members' voting." *American Political Science Review* 96(1):127–140.

Cassell, Kaitlen J. 2023. "The Populist Communication Strategy in Comparative Perspective." *The International Journal of Press/Politics* 28(4):725–746.

Desmet, Klaus, Ignacio Ortuño-Ortín and Romain Wacziarg. 2025. Latent Polarization. NBER Working Paper 34229 National Bureau of Economic Research.

Di Tella, Rafael, Randy Kotti, Caroline Le Pennec and Vincent Pons. 2025. "Keep your enemies closer: strategic platform adjustments during US and French elections." *American Economic Review* 115(8):2488–2528.

Fiva, Jon, Oda Nedregård and Henning Øien. forthcoming. "Group Identity and Parliamentary Debates." *Journal of Politics* .

Fowler, Anthony and Shu Fu. Forthcoming. "Do primary elections exacerbate congressional polarization?" *Journal of Politics* .

Funke, Manuel, Moritz Schularick and Christoph Trebesch. 2023. "Populist leaders and the economy." *American Economic Review* 113(12):3249–3288.

Gandhi, Amit and Aviv Nevo. 2021. Empirical models of demand and supply in differentiated products industries. In *Handbook of Industrial Organization*. Vol. 4 Elsevier pp. 63–139.

Gauthier, Germain, Philine Widmer and Elliott Ash. 2025. Deep Latent Variable Models for Unstructured Data. Technical report mimeo.

Gennaioli, Nicola and Guido Tabellini. 2025. "Identity Politics." *Econometrica* 93(6):1937–67.

Gentzkow, Matthew, Bryan Kelly and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature* 57(3):535–574.

Gentzkow, Matthew, Jesse M Shapiro and Matt Taddy. 2019. "Measuring group differences in high-dimensional choices: method and application to congressional speech." *Econometrica* 87(4):1307–1340.

Glaeser, Edward L., Giacomo Ponzetto and Jesse Shapiro. 2005. "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values." *The Quarterly Journal of Economics* 120(4):1283–1330.

Gordon, Matthew, Megan Ayers, Eliana Stone and Luke Sanford. 2025. "Remote Control:

Debiasing Machine Learning Predictions for Causal Inference." *Working paper* .

Guriev, Sergei and Elias Papaioannou. 2022. "The Political Economy of Populism." *Journal of Economic Literature* 60(3):753–832.

Guriev, Sergei, Nikita Melnikov and Ekaterina Zhuravskaya. 2021. "3g internet and confidence in government." *The Quarterly Journal of Economics* 136(4):2533–2613.

Hall, Andrew B. 2015. "What happens when extremists win primaries?" *American Political Science Review* 109(1):18–42.

Hall, Andrew B and James M Snyder Jr. 2015. "Candidate ideology and electoral success." *Manuscript, Stanford University* .

Hill, Seth J. 2017. "Changing votes or changing voters? How candidates and election context swing voters and mobilize the base." *Electoral Studies* 48:131–148.

Kamada, Yuichiro and Fuhito Kojima. 2014. "Voter Preferences, Polarization, and Electoral Policies." *American Economic Journal: Microeconomics* 6(4):203–36.

Klein, Ezra. 2020. *Why We're Polarized*. Simon & Schuster, Inc.

Manacorda, Marco, Guido Tabellini and Andrea Tesei. 2026. "Mobile internet and the rise of political tribalism in Europe." *Review of Economic Studies, forthcoming* .

Meguid, Bonnie M. 2005. "Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success." *American Political Science Review* 99(3):347–359.

Melnikov, Nikita. 2024. "Mobile internet and political polarization." *Available at SSRN 3937760* .

Mudde, Cas and Cristóbal Rovira Kaltwasser. 2012. *Populism in Europe and the Americas: Threat or corrective for democracy?* Cambridge University Press.

Norris, Pippa. 2019. "Varieties of populist parties." *Philosophy & Social Criticism* 45(9-10):981–1012.

Norris, Pippa. 2020. "Measuring populism worldwide." *Party Politics* 26(6):697–717.

Palfrey, Thomas R. 1984. "Spatial Equilibrium with Entry." *The Review of Economic Studies* 51(1):139–156.

Persson, Torsten and Guido Tabellini. 2002. *Political economics: explaining economic policy*. MIT press.

Polborn, Mattias K and James M Snyder Jr. 2017. "Party polarization in legislatures with office-motivated candidates." *The Quarterly Journal of Economics* 132(3):1509–1550.

Prummer, Anja. 2020. "Micro-Targeting and Polarization." *Journal of Public Economics*

188:1–21.

Rooduijn, Matthijs. 2019. "State of the field: How to study populism and adjacent topics? A plea for both more and less focus." *European Journal of Political Research* 58(1):362–372.

Severin-Nielsen, Majbritt Kappelgaard, Sanne Kruikemeier, Jakob Ohme and Kristian Gade Kjelmann. 2025. "From permanent campaign to permanent communication: parliamentarians' cross-media practices in routine and election times." *Journal of Information Technology Politics* 14:1–17.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics* 225(2):175–199.

Vafa, Keyon, Susan Athey and David M Blei. 2025. "Estimating wage disparities using foundation models." *Proceedings of the National Academy of Sciences* 122(22):e2427298122.

Van Kessel, Patricia, Regina Widjaya, Sono Shah, Aaron Smith and Adam Hughes. 2020. "The congressional social media landscape." *Available at https://www.pewresearch.org/internet/2020/07/16/1-the-congressional-social-media-landscape/* .

Zhang, Lehan, Gabriele Gratton, Pauline Grosjean and Hasin Yousaf. 2025. "Adding Fuel to the (Gun) Fire: How Politicians Polarize the Public Debate.".