

KYIV SCHOOL OF ECONOMICS

MASTER'S PROGRAM IN PUBLIC POLICY AND GOVERNANCE

MASTER'S THESIS

**“Populist Rhetoric and Electoral Performance: A Dictionary-Based
Analysis of Party Manifestos in Central and Eastern Europe”**

Student: Maksym Tsukan
Thesis advisor: Ivan Homza

Thesis submitted for the Master's degree
Major: 281 Public Governance and Administration

Kyiv 2026

CONTENTS

CONTENTS	2
ABSTRACT	3
INTRODUCTION	4
LITERATURE REVIEW	5
ANALYTICAL FRAMEWORK	9
Conceptualization	9
Theoretical Framework	11
Hypotheses	12
RESEARCH METHODOLOGICAL DESIGN	12
Research Approach and Justification	12
Data Sources	13
Sample	13
Operationalisation of Independent Variables	14
Dependent Variable	14
Estimation Strategy	14
Macroeconomic Controls	15
RESULTS	16
Descriptive Statistics and Validation	16
Rejection of the Null Hypothesis (Ho)	17
Testing Hypothesis 1	17
Testing Hypothesis 2	19
Testing Hypothesis 3	20
Wild Cluster Bootstrap Validation	22
Summary of Hypothesis Tests	23
EU-wide extension: geographical generalisation	24
Main Effects Across Regions	24
Populist and Mainstream Subsamples	26
Region Interaction Model	26
Bootstrap Validation (EU-wide)	27
DISCUSSION	28
IMPLICATIONS FOR PUBLIC POLICY	30
CONCLUSIONS	32

BIBLIOGRAPHY	33
SUPPLEMENTS	38
Appendix 1. Populism Dictionary	38
Appendix 2. R code implementation	38

ABSTRACT: This thesis examines whether the rhetorical content of party manifestos is associated with the electoral performance of populist parties in Central and Eastern Europe. Drawing on 305 party-election observations from ten countries over the period 2000–2023, it operationalises populist rhetoric along a five-dimensional framework (Antagonism, Morality, Nativism, Sovereignty, and Leadership) using dictionary-based content analysis of Manifesto Project texts. The five dimensions are tested as predictors of vote share in two-way fixed-effects OLS models with macroeconomic controls. In the CEE sample, the Nativism and Sovereignty dimensions are negatively and significantly associated with vote share; the remaining three dimensions produce null coefficients. Interaction models reveal that anti-elite rhetoric penalises populist parties, while leadership rhetoric benefits mainstream parties but penalises populist ones. An EU-wide extension confirms the Antagonism Penalty as the most robust pan-European finding and reveals that the Nativism Penalty and Leadership inversion are CEE-specific, while the Morality Penalty operates only in Western Europe. A wild cluster bootstrap validates the CEE results under finite-cluster conditions. The composite populism score is negative and significant across the full EU. These findings challenge the assumption that more populist rhetoric yields more votes and suggest that disaggregating populist rhetoric by dimension reveals patterns that composite scores conceal.

Keywords: Populism, ideational approach, dictionary analysis

Word count: 9440

INTRODUCTION

Populism has moved to the centre of political and academic attention in recent decades, with populist forces gaining footholds from Southeast Asia to the Americas. A large body of research has examined how populist discourse affects politics in host countries, but a gap persists: we lack systematic evidence on which specific rhetorical dimensions of populist rhetoric are most consistently associated with populist electoral performance in national elections. Demand-side determinants (economic shocks, immigration, institutional distrust) have received extensive attention (Norris & Inglehart, 2019; Swank, 2018), but the supply-side question of whether the substantive topics populist parties emphasise correlate with electoral success remains understudied. This thesis addresses that gap by testing whether the salience of specific rhetorical dimensions in party manifestos is associated with the electoral performance of parties in Central and Eastern European (CEE) national elections over the period 2000–2023.

Consequently, the analytical question guiding this study is: “What specific rhetorical dimensions in party manifestos are most strongly associated with the electoral success of populist parties in national elections in Central and Eastern Europe?”

The question matters because populism is a “thin” ideology that must attach itself to more substantive host themes (Stanley, 2008). The populist appeal depends as much on what parties choose to talk about as on the anti-elitist tone in which they say it: corruption, immigration, national identity, economic inequality (Mudde, 2004a; Olivas Osuna, 2021). If different substantive topics carry different electoral weight, then treating populist rhetoric as an undifferentiated whole, as most prior studies do, risks masking important variation. Despite the availability of large-scale textual datasets (the Manifesto Project, the PopuList, V-Dem), few studies have tested which specific rhetorical topics in manifestos correlate with populist parties’ electoral performance using regression analysis. Past research has relied on ideological classifications or measured only the presence of populist style, often ignoring the substantive content.

The study follows a confirmatory research design. Olivas Osuna (2021) multidimensional deconstruction of populism, together with the theoretical mechanisms of issue ownership (Petrocik, 1996) and rhetorical mimicry, provide the basis for three testable hypotheses about the association between manifesto rhetoric and electoral outcomes. A confirmatory approach is appropriate here because it allows for replicable inferences across a large set of parties and elections. Its limitation is that it cannot establish causal mechanisms; that would require qualitative research, such as process-tracing, on individual cases. The question also has a practical dimension. If certain topics are consistently linked to populist electoral performance, this information could help policymakers identify which

grievances or institutional shortcomings are being exploited politically, and focus their responses accordingly.

LITERATURE REVIEW

The scholarly debate on populism has converged toward an ideational approach, which conceptualizes populism as a set of ideas or a "thin-centered" ideology rather than merely a style or strategy (Hawkins, 2009; Mudde, 2004a). According to Mudde's widely cited definition, populism is "a thin-centered ideology that considers society to be ultimately separated into two homogeneous and antagonistic groups, 'the pure people' and 'the corrupt elite,' and which argues that politics should be an expression of the *volonté générale* (general will) of the people" (Mudde, 2004a, p. 543). This definition highlights two core rhetorical components of populist ideology: people-centrism (the exaltation of a virtuous, homogenous people as the rightful sovereign) and anti-elitism (the denigration of a corrupt establishment that betrays the people) (Mudde, 2004a; Mudde & Kaltwasser, 2017). In addition, Hawkins (2009) describes populism as a Manichaean discourse that identifies "good" with the unified will of the people and "evil" with a conspiring elite, underlining the moral dichotomy inherent in populist rhetoric. Populist rhetoric frames politics as a struggle between "the people" and "the elite," casting the latter as selfish or corrupt oppressors of the virtuous majority.

Under the ideational approach, populism is considered a "thin" ideology because it contains a limited set of core concepts and is therefore usually attached to more comprehensive "host" ideologies (Mudde & Kaltwasser, 2013; Stanley, 2008). This means that populist actors combine the generic populist ideas with other ideological commitments. On the right, often nativism or authoritarian nationalism, and on the left, versions of socialism or anti-neoliberalism (Mudde & Rovira Kaltwasser, 2013a; Mudde, 2019). For example, Europe's right-wing populist parties fuse populism with exclusionary nationalist and anti-immigrant appeals (hence the term "populist radical right"), whereas Latin America's classic left-wing populists marry populism with redistributive or inclusionary socio-economic agendas (Kazin, 2016; Mudde & Kaltwasser, 2017). All the same, every populist formation shares the same foundational rhetorical schema of elevating "the People" and vilifying "the Elite." In this thesis, populism is thus defined in line with the ideational approach: as a discourse or thin ideology that pits a morally pure people against a corrupt elite and advocates restoring popular sovereignty (Hawkins, 2009; Mudde, 2004a). By populist rhetoric, I refer to the communicative expression of these ideas – the language and arguments that emphasize popular will, attack the establishment, and often simplify politics into a moral binary (Hawkins, 2010).

Several other key terms require clarification. Here I specifically focus on populist rhetorical topics, i.e. manifesto content that reflects populist themes (such as anti-establishment rhetoric, people-centric appeals, and related policy issues often championed by populist parties). Finally, electoral success refers to a party's performance in national elections, measured by vote share. Consequently, a "populist party's electoral success" thus denotes its ability to attract voters and gain representation, while comparing relative success across parties involves examining their vote percentages. The literature has increasingly turned to examining how populist rhetoric might influence such electoral outcomes, which is the present study's focus.

The growth of populist parties across virtually every region has generated two competing lines of explanation (Mudde & Rovira Kaltwasser, 2018; Norris & Inglehart, 2019). Demand-side accounts locate the source in the electorate: economic grievances, cultural backlash, and institutional distrust. Supply-side accounts counter that structural discontent alone is insufficient; it is the rhetoric and programmatic appeal of populist parties that converts latent frustration into votes (Barr, 2009; Hawkins et al., 2012; Kriesi, 2014; Moffitt, 2016). The tension between these positions frames the present study: if supply-side rhetoric matters independently, we should observe a statistical association between the themes parties emphasise and their vote share, even after controlling for demand-side conditions.

A specific version of the supply-side argument holds that populist parties succeed by politicising issues that mainstream parties have neglected (Petrocik, 1996; Taggart, 2000). Mudde & Kaltwasser (2017) argue that populism can function as a corrective, pushing issues like immigration, corruption, or national sovereignty onto the agenda. By foregrounding these grievances and blaming unresponsive elites, populist parties tap into public dissatisfaction (Kriesi, 2014; Rooduijn, 2018). Survey evidence supports a link between populist attitudes and populist voting (Akkerman et al., 2014), suggesting that when a party's rhetoric stresses populist themes, it draws in voters who already hold those views. The question, however, is whether this link is driven by what parties say (supply) or by the pre-existing attitudes of voters (demand). The correlational design of the present study cannot fully resolve this, but it can test whether the association exists at all after structural demand-side factors are controlled for.

The debate is sharpened by conflicting evidence. Some studies find that populist discourse broadens a party's electorate by mobilising the disaffected, especially during crises (Bowler & Donovan, 2019; Caramani, 2017; Hawkins et al., 2018). Populist parties can act as issue entrepreneurs, forcing topics like EU sovereignty or immigration control onto the agenda and attracting voters who prioritise them (Hobolt & de Vries, 2015; Meguid, 2005). Against this, experimental work by (Dai & Kustov, 2024) finds that populist rhetorical features have no independent effect on

candidate choice in a U.S. conjoint experiment, with voters responding to policy positions rather than anti-elite framing. This suggests that the electoral payoff of populist rhetoric may be conditional on context, on the issues attached to it, and on whether the source is perceived as an authentic outsider (Barr, 2009). The present study tests precisely this: whether specific rhetorical dimensions are associated with vote share, and whether the association differs between populist and mainstream parties.

This thesis uses party manifestos as its primary data source. Manifestos are official campaign documents that lay out a party's issue positions, priorities, and appeals to voters, and they lend themselves well to quantitative textual analysis (Laver et al., 2003). The Manifesto Project Database (MARPOR) has been especially useful in comparative politics: it catalogues thousands of manifestos and codes the share of text devoted to predefined policy categories for each party in each election (Lehmann et al., 2025). Researchers have increasingly drawn on manifesto data to study party ideology, issue salience, and programmatic change. This approach has been extended to populism research as well, on the reasoning that populist ideology should show up in party texts, not just leaders' speeches (Meijers & Zaslove, 2021; Rooduijn & Pauwels, 2011).

A growing number of studies use manifestos to measure populist discourse or positioning. For example, Rooduijn & Pauwels (2011) conducted a content analysis of election manifestos in five Western European countries, coding for populist key phrases (e.g. references to "the people" versus "the elite"). They found that their manifesto-based populism scores corresponded well to expert perceptions of which parties are populist, validating manifestos as a source to detect populist rhetoric. Similarly, Jagers & Walgrave (2007) examined Belgian parties' communications (including manifesto sections) to distinguish "thin" populism (people-centric anti-elitism only) from "complete" populism (which also excludes out-groups like immigrants). Their analysis confirmed that manifestos of known populist parties contained substantially more people-vs-elite rhetoric than those of mainstream parties. More recently, machine learning approaches have emerged: (Hawkins et al., 2018, Chapter 1) propose supervised machine learning to classify sentences in manifestos as populist or not, producing a continuous populism score for each party's manifesto. Using hundreds of manifestos from the Manifesto Project, they demonstrate that this method can reliably rank parties by degree of populism in their written platforms. Manifestos are being recognized as valuable data for capturing the ideational content of party appeals, including populist themes. What these studies share is a focus on identifying which parties are populist and to what degree, rather than testing which rhetorical themes within manifestos correlate with electoral performance. That distinction is central to the present study.

Existing research using manifestos in populism has mostly focused on identifying which parties are populist and comparing overall levels of populist discourse (Rooduijn et al., 2014; Rooduijn & Pauwels, 2011). These studies confirm that manifestos of populist parties devote more attention to the evils of “the elite” and the virtue of “the people.” However, there is a relative scarcity of work examining how the specific content of populist manifestos relates to those parties’ electoral successes and losses. While we can measure what populist parties say in their manifestos, we know less about which particular manifesto themes may contribute to boosting their votes. This is a gap worth addressing: we can measure what populist parties say, but we do not know which of the things they say helps them win votes.

One problem with most existing manifesto-based research on populism is that it treats populism as a single thing. Studies typically produce either a composite score or a binary classification, without breaking the concept down into its thematic parts. Olivas Osuna (2021) takes issue with this, arguing that ontological disagreements about what populism really is have held the field back, and that moving from minimal definitions toward a multidimensional frame would make it easier to identify degrees and varieties of populism. He deconstructs populism into five connected dimensions: (1) antagonistic depiction of the polity, the Manichaeian split between “the pure people” and “the corrupt elite”; (2) moral interpretation, the normative framing of these groups as virtuous versus corrupt; (3) idealised construction of society, the invocation of a homogeneous “heartland” people; (4) absence of limits to popular sovereignty, expressed through demands for direct democracy and institutional bypass; and (5) charismatic leadership, the personalised bond between leader and people.

The implication for empirical work is that if populism has several distinct dimensions, lumping them into a single score could hide the fact that different dimensions have different, or even opposite implications, including the electoral effects. (Cossu, 2024), who applied NLP techniques to CEE manifestos from 2000 to 2022, found exactly this kind of variation: populist parties in the region share some thematic commonalities (anti-EU sentiment, anti-corruption rhetoric, ethno-centric language), but the specific emphases differ across national contexts. She also found that anti-elite discourse has spread widely across the CEE party spectrum, not remaining confined to populist parties. Disaggregating by dimension is therefore worth doing. The present study uses (Olivas Osuna, 2021) five-dimensional structure as its operationalisation framework for precisely this reason.

There are at least two plausible theoretical mechanisms connecting manifesto rhetoric to electoral results. The first is issue ownership. Parties that devote disproportionate attention to a topic in their campaign materials signal credibility on that topic. Populist parties tend to claim ownership of anti-establishment issues like

corruption, political renewal, or immigration control, where mainstream parties often lack credibility. By making these topics prominent in their manifestos, populists position themselves as the genuine champions of change, attracting voters who prioritise those issues. The second mechanism is activation of latent support. When populist rhetorical themes line up with public frustrations, they can mobilise voters who might otherwise have stayed home. Research shows that economic anxiety and political distrust increase receptivity to anti-elite messaging. According to (Hawkins et al., 2018), populist attitudes among citizens lie dormant until activated by frames that populist actors deploy. A manifesto heavy on anti-corruption or anti-globalisation themes could therefore convert more disaffected voters by speaking to their frustrations directly.

Two gaps emerge from this review. First, most existing studies treat populism as a single rhetorical object, without testing whether its constituent dimensions carry different weight. Second, the few studies that examine manifesto content and electoral outcomes do so with composite scores that, as the results of this thesis will show, can mask opposing effects across dimensions. The analytical framework that follows addresses both gaps by operationalising populism as five separate dimensions and testing their individual and conditional associations with vote share.

ANALYTICAL FRAMEWORK

Conceptualization

The dependent variable, independent variables, and moderating variables are operationalised as follows: (1) the dependent variable, electoral performance, measured as a party's percentage vote share in a given national election; (2) the five independent variables representing the rhetorical dimensions of populist discourse;

and (3) a binary moderating variable indicating whether a party is classified as populist according to the PopuList classification.

Electoral performance is measured continuously as the percentage of the national vote received by a party in a given election, sourced from the Manifesto Project Main Dataset (Lehmann et al., 2025). Vote share is used instead of the binary win/loss indicator because it captures variation in degrees of electoral success and allows for the identification of marginal effects of rhetorical salience on voter mobilisation. Alternative specification of using vote-share change compared to the previous election-year was considered but the simpler vote share in a specific election was preferred because many parties appear only once in the sample, making analysis impossible for a large portion of observations.

The five independent variables correspond directly to the five latent dimensions of populism proposed by (Olivas Osuna, 2021) in his multidimensional deconstruction of the concept. Olivas Osuna (2021, p. 836) argues that the scholarly field has been inhibited by ontological disputes concerning the essential nature of populism and proposes that shifting from minimal categorical definitions toward a multidimensional framework “may stimulate the generation of comparative data on a wider range of attributes and facilitate the identification of degrees and varieties within populism.” Accordingly, this study adopts this five-dimensional structure rather than treating populism as a monolithic rhetorical object.

Antagonistic Depiction of the Polity captures the Manichaeian framing of society as divided between “the pure people” and “the corrupt elite”. In practice, the dictionary captures lexemes related to elites, betrayal, systemic corruption, and adversarial conflict. Moral Interpretation captures the normative construction of these two groups: ordinary people as virtuous, the establishment as corrupt or deceitful. Nativism (corresponding to Olivas Osuna’s “Idealised Construction of Society”) captures the populist tendency to invoke a homogeneous, authentic “heartland” people, typically through exclusionary in-group language; the dimension is relabelled here because the dictionary operationalisation emphasises ethnic and cultural identity markers rather than the broader societal idealisation that Olivas Osuna describes. Absence of Limits to Popular Sovereignty covers the populist commitment to unmediated popular will, as reflected in references to referenda, direct democracy, and institutional bypass. Charismatic Leadership reflects the personalised relationship between a leader figure and the people, captured through references to strong leadership and embodiment of popular will.

Each dimension is measured as a normalised frequency score: the number of stemmed token matches per 1,000 tokens in the manifesto. This normalisation is standard practice in manifesto-based populism research (Cossu, 2024; Rooduijn & Pauwels, 2011) and eliminates the variation in manifesto length across parties and elections. In addition to the five individual dimension scores, a composite fullPopulism score is computed as their arithmetic sum to enable a robustness check.

The moderating variable is *Populist* is a binary indicator (1 = populist, 0 = mainstream) derived from the PopuList database (Rooduijn et al., 2024), which employs an expert-informed qualitative comparative classification procedure to categorise parties across European countries as populist, far-right, far-left, or mainstream. This variable is used to test Hypothesis 3, allowing the study to estimate whether the electoral payoff of rhetorical salience differs systematically between established populist parties and mainstream parties that may strategically deploy populist rhetoric.

The ten Central and Eastern European countries: Poland, Hungary, the Czech Republic, Slovakia, Estonia, Latvia, Lithuania, Bulgaria, Romania, and Slovenia were selected based on consistent Manifesto Project coverage and inclusion in the PopuList database (Froio et al., 2023). Croatia was excluded because the Manifesto Project corpus (v.2025-1) does not include English-translated manifesto texts for Croatian parties. The country selection spans the full range of post-communist party system trajectories and includes both cases of populist governance (Hungary, Poland post-2015) and countries with more episodic populist penetration (Bulgaria, Slovakia, the Baltic states), maximising within-region variation.

Theoretical Framework

The theoretical framework underlying this study is grounded in the ideational approach to populism. Within this tradition, populism is understood as a thin-centred ideology that conceives of society as divided between a virtuous and homogeneous people and a corrupt elite, and which holds that politics should be the expression of the general will of the former (Hawkins et al., 2018; Mudde, 2004b). Under the ideational approach, populism is not a binary attribute; parties may deploy more or less of it. Following Mudde's and (Olivas Osuna, 2021) multidimensional extension of this tradition, populism is treated as a matter of degree whose rhetorical manifestations are reflected in textual data. Dictionary-based analysis of party manifestos is therefore a workable operationalisation strategy.

The theoretical framework builds upon two mechanisms linking populist rhetoric, as measured at the manifesto level, to electoral outcomes.

The first mechanism is issue ownership. Following (Meguid, 2005; Petrocik, 1996), parties that devote disproportionate attention to a particular issue in their campaign materials signal credibility and competence on that issue to the electorate. Populist parties are argued to own anti-establishment issues, such as elite corruption, institutional failure, and national sovereignty, because they have cultivated credibility as “outsiders” (Barr, 2009). The greater the salience a party affords to these themes in its manifesto, the stronger its claim to issue ownership and, consequently, the more effectively it mobilises voters who prioritise those grievances. This mechanism generates Hypotheses 1 and 2.

The second mechanism concerns the conditional effectiveness of mimicking rhetoric. Mastropaolo (2008, pp. 30–48) and Rooduijn et al. (2014) document a process of

“programmatic contagion” under which the mainstream parties adopt populist rhetorical elements under competitive pressure. However, if voters perceive the strategy and deem it as inauthentic, the electoral payoff of deploying these themes may be lower, or even negative, for mainstream parties. The experimental findings of (Dai & Kustov, 2024) support this and demonstrate that populist talking points do not offer independent vote gains when deployed by non-populist actors (although their findings also suggest that the populist rhetoric may have no effect even for explicitly populist actors). This asymmetry provides the theoretical basis for Hypothesis 3.

Macroeconomic conditions at $t-1$ (GDP growth, unemployment, inflation) and institutional quality (control of corruption from the Worldwide Governance Indicators) are included as controls, following the demand-side framework of Norris & Inglehart (2019) and (Swank, 2018), which attributes a structural baseline of populist support to economic insecurity and institutional distrust. The government effectiveness indicator, included in earlier specifications, was dropped due to high collinearity with control of corruption (GVIF > 5).

Hypotheses

H₀: The salience of populist rhetorical dimensions in party manifestos has no statistically significant association with electoral vote share in national elections in Central and Eastern Europe.

H₁: The salience of populist rhetorical dimensions in party manifestos is associated with electoral vote share in national elections in Central and Eastern Europe.

H₂: The salience of specific rhetorical dimensions in party manifestos is positively associated with the electoral vote share of populist parties in national elections in Central and Eastern Europe.

H₃: The positive association between rhetorical salience and electoral success is stronger for parties classified as populist than for mainstream parties that employ populist rhetoric.

The omnibus test (H₀) establishes whether rhetorical variation is associated with electoral outcomes at all, while the directional and conditional tests (H₁, H₂) characterise the nature and heterogeneity of that association. A Bonferroni-corrected significance threshold of $\alpha = 0.01$ is applied across all five dimension-specific tests for H₂ and H₃ to control the family-wise error rate.

RESEARCH METHODOLOGICAL DESIGN

Research Approach and Justification

This study employs a quantitative research approach grounded in a confirmatory research design. Given that the three hypotheses derived above specify directional expectations regarding the relationships between clearly defined independent and

dependent variables, a confirmatory design is appropriate: the goal is to test whether the theoretically anticipated associations hold in the empirical data, rather than to discover or build new theory. The principal advantage of this approach is its capacity to generate replicable, generalisable inferences across a relatively large corpus of parties and elections. Its limitation is that it is less suited to explaining the causal mechanisms underlying observed associations, which would require qualitative design, including process-tracing, designs applied to individual cases. These trade-offs are explicitly acknowledged in the interpretation of findings.

Data Sources

The sole data source for texts is the Manifesto Project Database (MARPOR), version 2025 (Lehmann et al., 2025) accessed via the manifestoR API in R. The corpus consists of all available English-translated electoral manifestos from parties competing in national legislative elections in ten Central and Eastern European countries: Poland, Hungary, the Czech Republic, Slovakia, Estonia, Latvia, Lithuania, Bulgaria, Romania, and Slovenia for the period 2000-2023. Croatia was excluded because English translations of the manifesto texts were not available in the corpus used. The Manifesto Project Database is particularly well-suited for this study because it provides standardized coverage across a large number of countries and election years, enabling cross-national comparisons that would be impossible with single-country corpora. It also provides high quality English translations for most manifesto texts, which makes the corpus accessible to uniform dictionary-based analysis. Most importantly, the Manifesto Project has established itself as the reference corpus in comparative research on party rhetoric, with prior quantitative studies employing it for (Cossu, 2024; Rooduijn & Pauwels, 2011).

Party-level populist classifications are drawn from the PopuList database (Froio et al., 2023), which operationalises populist status through an expert-informed qualitative comparative classification. The party identifiers from this database are harmonised with the Manifesto Project via the Party Facts bridge file (Döring & Regel, 2019). Macroeconomic control variables are sourced from the World Bank Development Indicators (WDI) and the Worldwide Governance Indicators (WGI), retrieved for the period 1999-2025 using the WDI package in R.

Sample

The unit of analysis is the party-election observation: a single party's manifesto and corresponding vote share in a specific national election. The sample includes all party-election observations in the nine CEE countries for which an English-translated manifesto is available in the Manifesto Project corpus, a valid vote share is recorded in the main dataset, and the party is matched to the PopuList via the Party Facts bridge. Observations with missing values on all five rhetorical dimension scores or on the dependent variable are excluded listwise. The resulting analytical sample encompasses 305 party-election observations across 50 unique elections and 153 unique parties.

Operationalisation of Independent Variables

The five rhetorical dimensions are operationalised using dictionary-based content analysis within the Quanteda natural language processing framework in R (Benoit et al., 2018). Following the procedure recommended by Rooduijn & Pauwels (2011), manifesto texts are first tokenised and converted to lower case, stop words (sourced from the tidytext package and supplemented by a domain-specific suppression list of generic political terms) are removed, and all tokens are subject to Porter stemming to ensure that morphological variants of the same root are counted together. A Document-Feature Matrix is constructed from the resulting tokens.

The populism dictionary was constructed deductively from the conceptual definitions provided by [Olivas Osuna \(2021\)](#) and subsequently refined through an iterative empirical process. The initial dictionary contained terms drawn directly from the theoretical descriptions of each dimension. After preliminary validation revealed that several dimensions (particularly Antagonism and Leadership) failed to discriminate between populist and non-populist parties in the corpus, the dictionary was revised. High-frequency generic terms that appeared equally in populist and non-populist manifestos (e.g., system*, fight*, nation*) were removed, and terms with stronger discriminatory power were added based on inspection of the term-frequency ratios between populist and non-populist subcorpora. The full dictionary is presented in Appendix 1.

As a validation step prior to the main regressions, correlations are computed between each dimension score and the PopuList binary classification. Given that five simultaneous tests are conducted, a Bonferroni correction is applied, setting the significance threshold to $\alpha = 0.01$ ($0.05/5$). Dimensions that achieve significance at this corrected threshold are considered empirically validated as discriminating between populist and non-populist parties within the corpus.

Dependent Variable

The dependent variable is vote share: the percentage of the national vote received by a party in a given election, as recorded in the Manifesto Project Main Dataset. Vote share is treated as a continuous interval variable. This choice follows (Mudde & Rovira Kaltwasser, 2018), who argue that electoral performance is the most direct indicator of populist success, and is consistent with the regression-based approach adopted in comparable studies (Rooduijn, 2018).

Estimation Strategy

The primary estimator is ordinary least squares (OLS) with two-way fixed effects for country and election year, implemented via the `feols` function from the `fixest` package in R (Bergé et al., 2026). The country fixed effects absorb all time-invariant, country-specific confounders (e.g., party system institutionalisation, historical legacies of communism, size of the electorate), and election-year fixed effects control for common shocks affecting all parties simultaneously in a given year (e.g., EU accession cycles, the 2008 financial crisis, the COVID-19 pandemic). Standard errors

are clustered at the country level to account for within-country correlation across electoral cycles.

Three model specifications are estimated. Model 1 includes the five individual dimension scores and the macroeconomic control variables as predictors of vote share. Model 2 substitutes the five individual scores with the composite fullPopulism score to assess aggregate rhetorical salience. To test Hypothesis 3, five additional interaction models are estimated, one for each rhetorical dimension. Each model includes the dimension score, the isPopulist binary, their interaction, and the macroeconomic controls, with the same two-way FE structure and clustered standard errors.

Several robustness checks are conducted to assess the sensitivity of the findings. First, a wild cluster bootstrap (Cameron et al., 2008) is implemented using the `fwildclusterboot` package in R with six-point weights and $B = 9,999$ replications (Fischer & Roodman, 2021). This procedure is critical because the CEE sample contains only 10 country clusters, which is at the lower bound of reliability for cluster-robust variance estimation. Simulation evidence suggests that conventional clustered standard errors can over-reject the null hypothesis with fewer than approximately 20 clusters (Canay et al., 2021). The wild bootstrap with Webb weights performs adequately with as few as 6–10 clusters (MacKinnon & Webb, 2018).

Second, a Hausman specification test is conducted to compare the fixed-effects estimator with a random-effects GLS alternative (Hausman, 1978) using the `plm` package in R. The panel structure for this test uses party-within-country as the individual unit, observed across election years. This structure is unbalanced (not all parties appear in every election period). On the CEE sample, the Hausman test yields $\chi^2 = 10.01$ ($df = 9$, $p = 0.35$), failing to reject the random-effects specification. On the full EU sample (see the EU Extension chapter below), the test yields $\chi^2 = 25.01$ ($df = 9$, $p = 0.003$), rejecting random effects. The fixed-effects specification is retained as the primary estimator for both samples because it requires no assumptions about the independence of unobserved heterogeneity and regressors, and the EU-wide test confirms its appropriateness.

Third, variance inflation factor (VIF) diagnostics are computed for Model 1 to assess multicollinearity among the regressors. VIF values for the five rhetorical dimensions range from 1.09 to 1.24 ($GVIF^{1/(2 \cdot Df)}$), well below conventional concern thresholds (Table A4). Among the control variables, corruption exhibits an elevated $GVIF^{1/(2 \cdot Df)}$ of 3.89, which reflects moderate collinearity with the macroeconomic controls but does not reach levels that would invalidate inference.

Macroeconomic Controls

Four control variables are included in all specifications. GDP growth, unemployment rate, and inflation rate are sourced from the World Bank WDI and lagged by one year relative to the election date to capture the pre-election economic context that voters evaluate retrospectively (Stiers, 2022). The control of corruption indicator is sourced from the WGI dataset. Values for election years with no $t-1$ observation before 2000

are floored to 2000. For elections in 2002, the contemporaneous year value is used in the absence of valid 2001 data for all countries.

RESULTS

Descriptive Statistics and Validation

Descriptive statistics for all variables in the analytical sample are presented in Table A1. The sample comprises 305 party-election observations drawn from ten CEE countries across the period 2000–2023. The mean vote share is 14.4% (SD = 11.6), spanning a range from 0.1% to 67.9%, which reflects the diversity of the sample across both marginal parties and dominant governing formations. Approximately 29% of observations are classified as populist according to the PopuList database.

Table A1: Descriptive Statistics

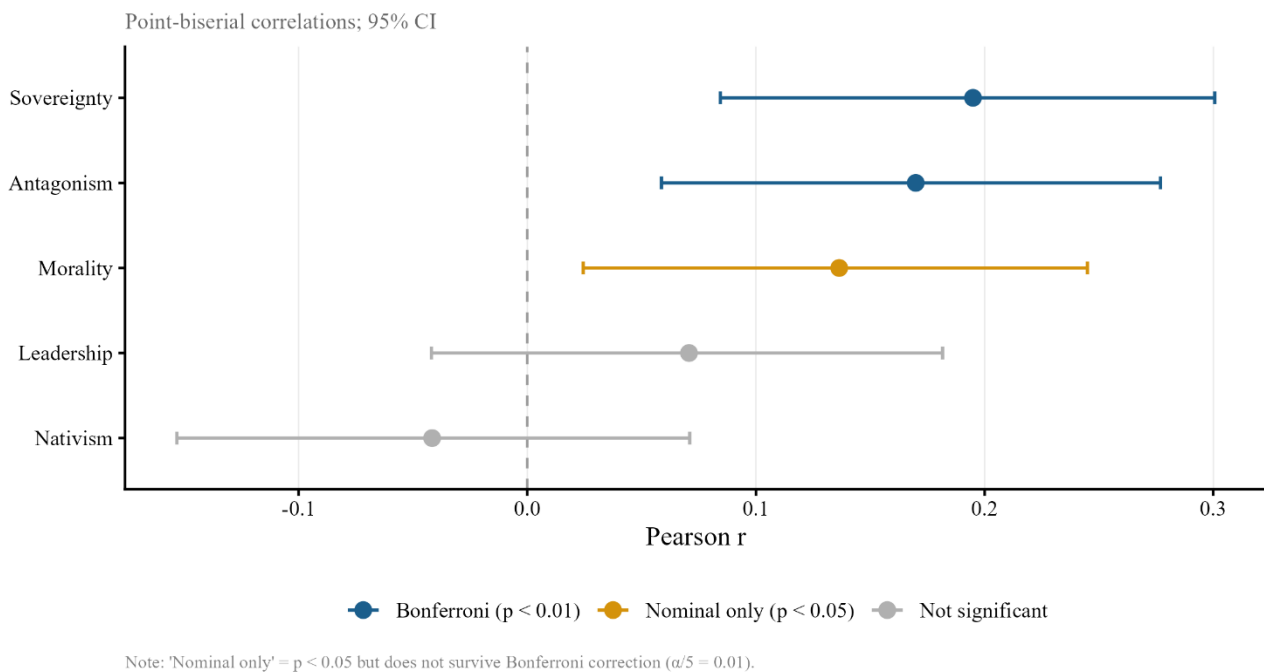
	MEAN	SD	MIN	MEDIAN	MAX
"Vote Share (%)	14.4	11.6	0.1	9.8	67.9
antagonism_norm	1.5	1.5	0.0	1.2	8.8
morality_norm	1.9	1.9	0.0	1.4	15.3
nativism_norm	2.7	3.0	0.0	2.2	27.7
sovereignty_norm	1.1	1.2	0.0	0.8	7.8
leadership_norm	0.4	0.7	0.0	0.3	4.6
fullPopulism_norm	7.7	4.7	0.0	6.7	36.7
isPopulist	0.3	0.5	0.0	0.0	1.0
GDP growth (annual %)	3.2	4.6	-16.0	3.4	11.6
Unemployment, total (% of total labor force) (modeled ILO estimate)	9.3	4.0	3.8	7.9	19.5
Inflation, consumer prices (annual %)	3.3	6.1	-1.1	2.3	45.7
corruption	0.4	0.4	-0.6	0.4	1.5

The distributional properties of the five rhetorical dimension scores reflect the expected pattern of rhetorical salience. Antagonism has a mean of 1.5 occurrences per 1,000 tokens (SD = 1.5), followed by Nativism (mean = 2.7, SD = 3.0) and Morality (mean = 1.9, SD = 1.9). Sovereignty (mean = 1.1, SD = 1.2) and Leadership (mean = 0.4, SD = 0.7) are considerably rarer. The lower mean frequencies for Sovereignty and Leadership indicate that the lexical items constituting these dimensions are sparse in manifesto text relative to the more diffuse nativist and moral vocabulary.

The validation correlations between each dimension score and the PopuList classification are displayed in Figure 3. Sovereignty achieves Bonferroni-corrected significance ($r = 0.20$, $p < 0.01$) and Antagonism likewise ($r = 0.20$, $p < 0.01$).

Morality attains nominal significance ($r = 0.14$, $p < 0.05$) but does not survive the Bonferroni correction. Leadership ($r = 0.07$) and Nativism ($r = -0.04$) are statistically indistinguishable from zero. The negative Nativism correlation is notable: it suggests that the nativist vocabulary captured by the dictionary is not disproportionately concentrated in populist manifestos, but rather is distributed broadly across the CEE party spectrum. This may reflect the normalisation of identity-based rhetoric in the region, consistent with Cossu's (2024) findings.

Figure 3. Dictionary Validation: Dimension Scores vs PopuList Label (CEE)



Rejection of the Null Hypothesis (H_0)

A joint Wald test conducted on Model 1, restricting all five dimensional coefficients to zero, yields a test statistic that is statistically significant ($F(5, 9) = 7.11$; $p < 0.001$), allowing rejection of H_0 . The Adjusted R^2 of 0.065 for Model 1, while modest, is positive and indicates that the rhetorical dimensions together with the controls account for a non-trivial share of variance in vote share beyond the fixed effects. With H_0 rejected, the analysis proceeds to characterise the nature and direction of the observed associations.

Testing Hypothesis 1

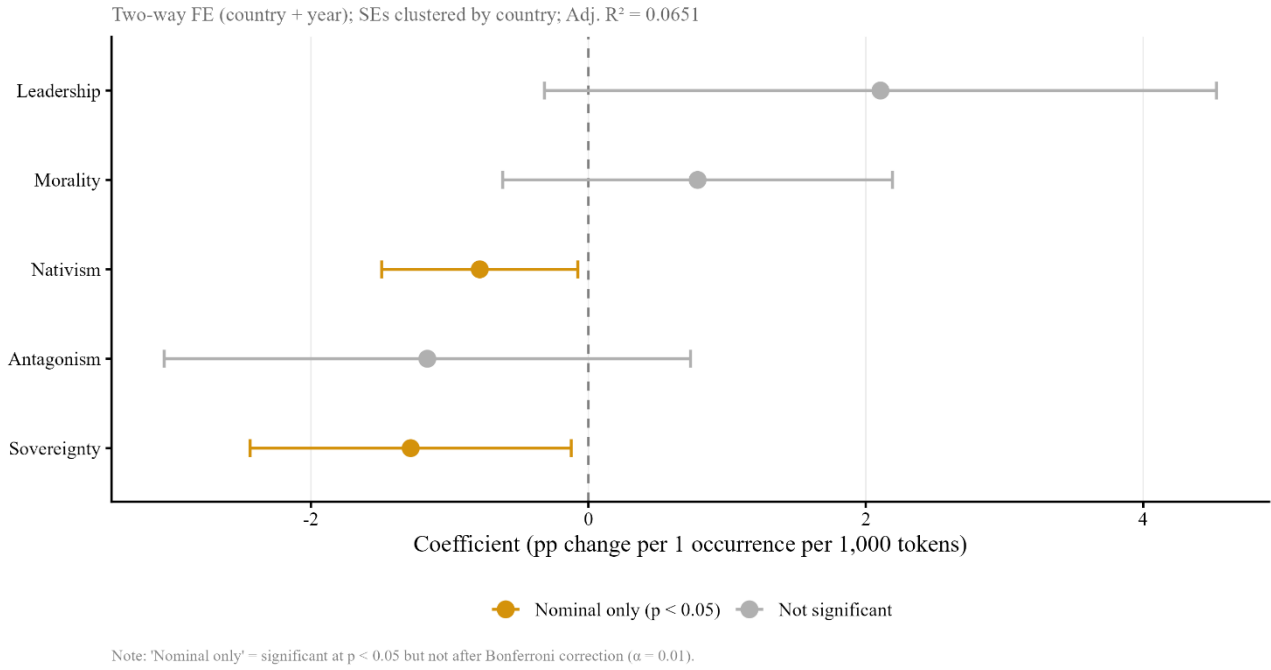
The coefficient estimates from the three model specifications are presented in Table 1 and visualised in Figure 1. The joint Wald test reported above established that the five rhetorical dimensions are collectively associated with vote share ($F(5, 9) = 7.11$; $p < 0.001$), confirming H_1 . The R^2 of 0.065 for Model 1, while modest, indicates that the rhetorical dimensions account for a non-trivial share of variance beyond the country and year fixed effects. H_1 is therefore supported: manifesto rhetoric, taken as a set of five dimensions, is associated with electoral performance.

Table 1: Main Effects on Electoral Vote Share (CEE)

	(1) Dimensions	(2) Composite
Antagonism	-1.161 (0.839)	
Morality	0.788 (0.621)	
Nativism	-0.783* (0.312)	
Sovereignty	-1.281* (0.512)	
Leadership	2.106 (1.071)	
Full Populism (composite)		-0.422 (0.194)
GDP Growth (t-1)	-0.537** (0.149)	-0.586** (0.154)
Unemployment (t-1)	0.199 (0.137)	0.070 (0.117)
Inflation (t-1)	-0.159*** (0.030)	-0.164** (0.035)
Control of Corruption (t-1)	-7.244 (3.449)	-6.420 (3.031)
Num.Obs.	305	305
R2 Adj.	0.065	0.010
• p < 0.05, ** p < 0.01, *** p < 0.001		

Two-way FE (country + election year). SEs clustered by country.

Figure 1. Main Effects of Rhetorical Dimensions on Vote Share (CEE)



Testing Hypothesis 2

Across the full CEE sample (Model 1, $N = 305$), two dimensions achieve statistical significance at the conventional $\alpha = 0.05$ level, and both are negative. The Sovereignty dimension yields a coefficient of $\beta = -1.282$ ($SE = 0.513$, $p = 0.034$), indicating that each additional occurrence of sovereignty-related rhetoric per 1,000 manifesto tokens is associated with a reduction in vote share of approximately 1.3 percentage points. The Nativism dimension yields $\beta = -0.783$ ($SE = 0.313$, $p = 0.034$). The remaining three dimensions, Antagonism ($\beta = -1.161$, $SE = 0.837$), Morality ($\beta = 0.787$, $SE = 0.621$), and Leadership ($\beta = 2.106$, $SE = 1.071$), are statistically indistinguishable from zero, although the Leadership point estimate is positive and substantively large. Neither Sovereignty nor Nativism survives the Bonferroni correction at $\alpha = 0.01$. Hypothesis 2 therefore receives partial and directionally contrary support: two dimensions are individually significant at conventional levels, but in the negative direction.

Model 2, which uses the composite fullPopulism score, yields a coefficient of -0.422 ($SE = 0.194$, $p = 0.057$), marginally non-significant at the conventional threshold. The VIF diagnostics presented in Table A4 confirm that multicollinearity among the five rhetorical dimension scores is negligible ($G\text{VIF}^{(1/(2 \cdot Df))}$ range: 1.09–1.24). Among the macroeconomic controls, GDP growth ($\beta = -0.538$, $SE = 0.149$, $p < 0.01$) and inflation ($\beta = -0.159$, $SE = 0.030$, $p < 0.001$) are negatively associated with vote share. The negative coefficient on GDP growth is consistent with retrospective economic voting theory (Stiers, 2022): when the economy grows, incumbent parties are rewarded and opposition parties receive correspondingly smaller vote shares.

Table A4. Variance Inflation Factors (CEE, Model 1)

Variable	GVIF	Df	GVIF ^{1/(2*Df)}
antagonism_norm	1.53	1	1.24
morality_norm	1.30	1	1.14
nativism_norm	1.22	1	1.11
sovereignty_norm	1.22	1	1.11
leadership_norm	1.19	1	1.09
gdp_growth	8.92	1	2.99
unemployment	6.32	1	2.51
inflation	3.18	1	1.78
corruption	15.16	1	3.89
<i>Country and year FE omitted. Threshold: GVIF^{1/(2*Df)} > 2.24 ≈ VIF > 5.</i>			

Testing Hypothesis 3

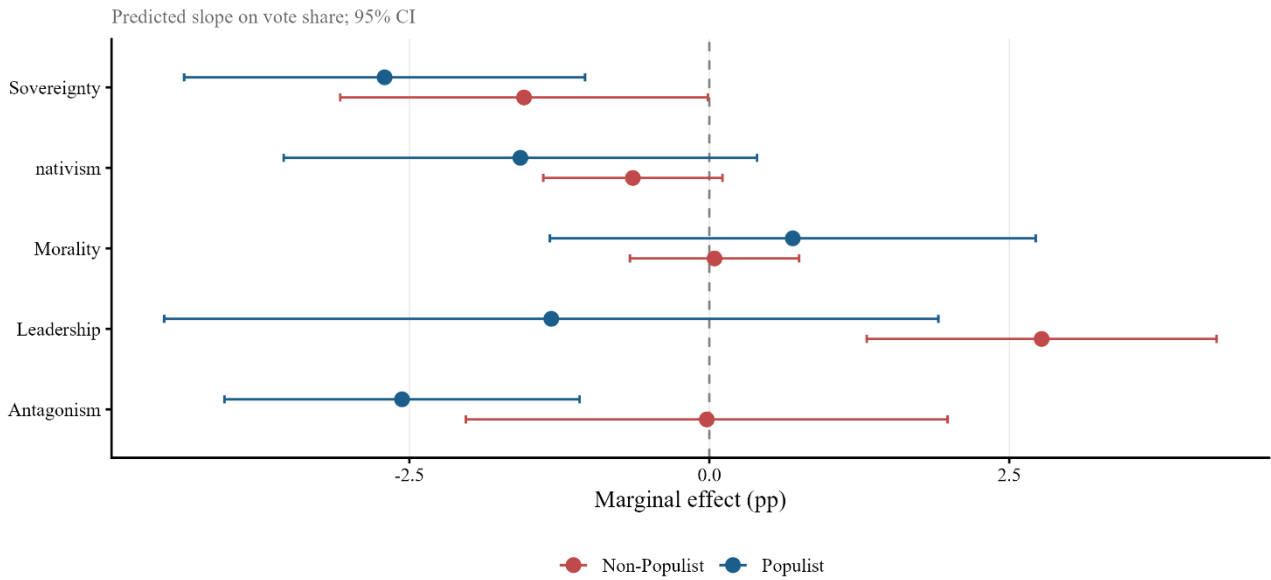
The interaction model estimates are presented in Table 2 and visualised in Figure 2. Hypothesis 3 predicted that the electoral payoff of rhetorical salience would be stronger for populist parties than for mainstream parties. Two of the five interactions achieve significance at the conventional $\alpha = 0.05$ level. The Antagonism $\times$ Populist interaction is negative and significant ($\beta = -2.540$, $SE = 0.883$, $p = 0.025$), indicating that anti-elite rhetoric penalises populist parties specifically: the marginal effect of Antagonism on vote share is near zero for mainstream parties but substantially negative for populist ones. The Leadership $\times$ Populist interaction is also negative and significant ($\beta = -4.087$, $SE = 1.766$, $p = 0.041$), with an inverted pattern: mainstream parties appear to benefit from leadership language ($\beta = 2.770$, $SE = 0.744$, $p < 0.01$), while populist parties are penalised. Hypothesis 3 is rejected: populist party status does not amplify the electoral return to rhetorical salience. For Antagonism and Leadership, the conditional effect is in the opposite direction.

Table 2. Interaction Effects: Dimension $\times$ Populist Status (CEE)

	Antagonism	Morality	Nativism	Sovereignty	Leadership
Antagonism	-0.021				
	(1.025)				
Morality		0.043			
		(0.360)			
Nativism			-0.638		
			(0.381)		
Sovereignty				-1.545	
				(0.782)	
Leadership					2.770**
					(0.744)

Populist Party (binary)	9.824*	3.151	7.081	7.024	6.476*
	(3.354)	(2.632)	(3.917)	(3.352)	(2.714)
Antagonism × Populist	-2.540*				
	(0.883)				
Morality × Populist		0.652			
		(1.212)			
Nativism × Populist			-0.938		
			(1.089)		
Sovereignty × Populist				-1.163	
				(0.869)	
Leadership × Populist					-4.087*
					(1.766)
GDP Growth (t-1)	-0.564*	-0.640*	-0.734**	-0.650**	-0.540*
	(0.178)	(0.210)	(0.207)	(0.183)	(0.172)
Unemployment (t-1)	0.061	0.026	0.015	-0.033	0.060
	(0.151)	(0.137)	(0.136)	(0.153)	(0.127)
Inflation (t-1)	-0.216***	-0.246*	-0.229**	-0.210**	-0.247***
	(0.043)	(0.082)	(0.059)	(0.058)	(0.050)
Control of Corruption (t-1)	-7.211	-5.064	-6.555	-5.128	-4.566
	(3.762)	(3.138)	(3.417)	(2.503)	(3.119)
Num.Obs.	305	305	305	305	305
R2 Adj.	0.056	0.015	0.048	0.056	0.028
• p < 0.05, ** p < 0.01, *** p < 0.001					
<i>Two-way FE (country + election year). SEs clustered by country.</i>					

Figure 2. Marginal Effect of Rhetoric by Party Type (CEE)



Wild Cluster Bootstrap Validation

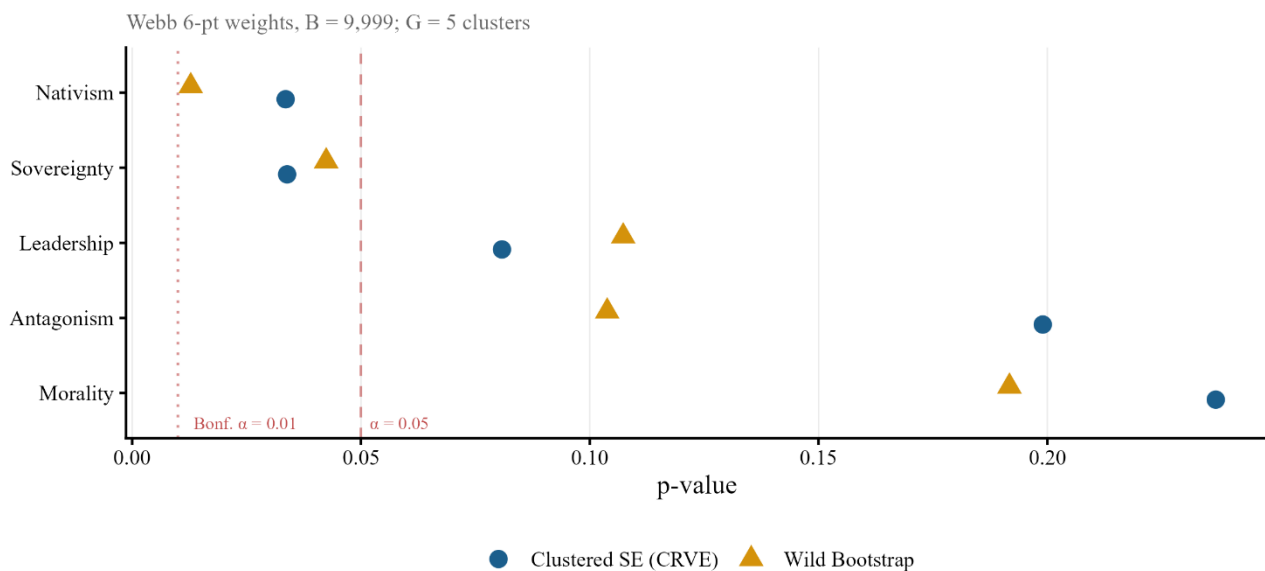
Given the small number of country clusters ($G = 10$), the main CEE results are validated using the wild cluster bootstrap with Webb (2014) six-point weights ($B = 9,999$). The results are presented in Table A2 and Figure 4. The two nominally significant dimensions from Model 1 both survive the bootstrap: Nativism (CRVE $p = 0.034$, bootstrap $p = 0.013$) and Sovereignty (CRVE $p = 0.034$, bootstrap $p = 0.042$). For Nativism, the bootstrap p -value is actually lower than the CRVE p -value, suggesting that the conventional clustered standard errors were slightly conservative for this coefficient. Antagonism, which was non-significant under CRVE ($p = 0.199$), remains non-significant under bootstrap ($p = 0.104$), though the point estimate is large ($\beta = -1.161$) and the directional pattern is consistent. The composite fullPopulism score achieves significance under bootstrap ($p = 0.025$).

Table A2: CRVE vs Wild Cluster Bootstrap (CEE)

Dimension	β	SE	p (CRVE)	p (WCB)	Bootstrap 95% CI		Significance	
					CI low	CI high	CRVE	WCB
Antagonism	-1.161	0.837	0.1990	0.1038	-3.310	0.227		
Morality	0.787	0.621	0.2368	0.1917	-0.128	3.671		
Nativism	-0.783	0.313	0.0335	0.0128	-1.899	-0.194	*	*
Sovereignty	-1.282	0.513	0.0339	0.0424	-2.794	-0.072	*	*
Leadership	2.106	1.071	0.0808	0.1073	-0.831	5.452		

*Webb 6-pt weights, B = 9,999. ** $p < 0.01$; * $p < 0.05$.*

Figure 4. Inference Comparison: CRVE vs Wild Cluster Bootstrap (CEE)



Summary of Hypothesis Tests

H_0 is rejected. The joint Wald test rejects the null that all five dimension coefficients are simultaneously zero. H_1 is supported: the five dimensions are jointly associated with vote share. H_2 receives partial and directionally contrary support: Sovereignty and Nativism are negatively associated with vote share ($p < 0.05$), but neither survives the Bonferroni correction; the remaining three dimensions produce null coefficients. H_3 is rejected: populist party status does not amplify the electoral return to rhetorical salience. For Antagonism and Leadership, the conditional effect is inverted, with populist parties penalised and mainstream

parties either unaffected or modestly benefiting.

EU-wide extension: geographical generalisation

To assess whether the CEE findings are region-specific or reflect a broader European pattern, the analysis is extended to the full European Union. The EU-wide corpus comprises 907 party-election observations from 25 countries (10 CEE and 15 Western European). Croatia and Malta are excluded due to the absence of English-translated manifesto texts in the Manifesto Project corpus (v.2025-1). The same dictionary, preprocessing pipeline, and estimation strategy are applied. The EU-wide sample provides two advantages over the CEE-only analysis: substantially more statistical power (25 country clusters vs. 10), and sufficient observations to estimate separate populist-only ($N = 191$) and mainstream-only ($N = 714$) subsamples with reasonable precision.

Main Effects Across Regions

The regional comparison results are presented in Table 3 and Figure 5. In the pooled EU sample, Antagonism is the strongest and most robust predictor ($\beta = -1.030$, $SE = 0.296$, $p = 0.002$), significant at the Bonferroni-corrected level. Sovereignty is negative and nominally significant ($\beta = -0.569$, $SE = 0.240$, $p = 0.026$) but does not survive the wild cluster bootstrap (bootstrap $p = 0.106$; Table A3). The composite fullPopulism score is negative and highly significant ($\beta = -0.432$, $SE = 0.112$, $p < 0.001$). The Adjusted R^2 for the EU-wide dimension model is 0.079.

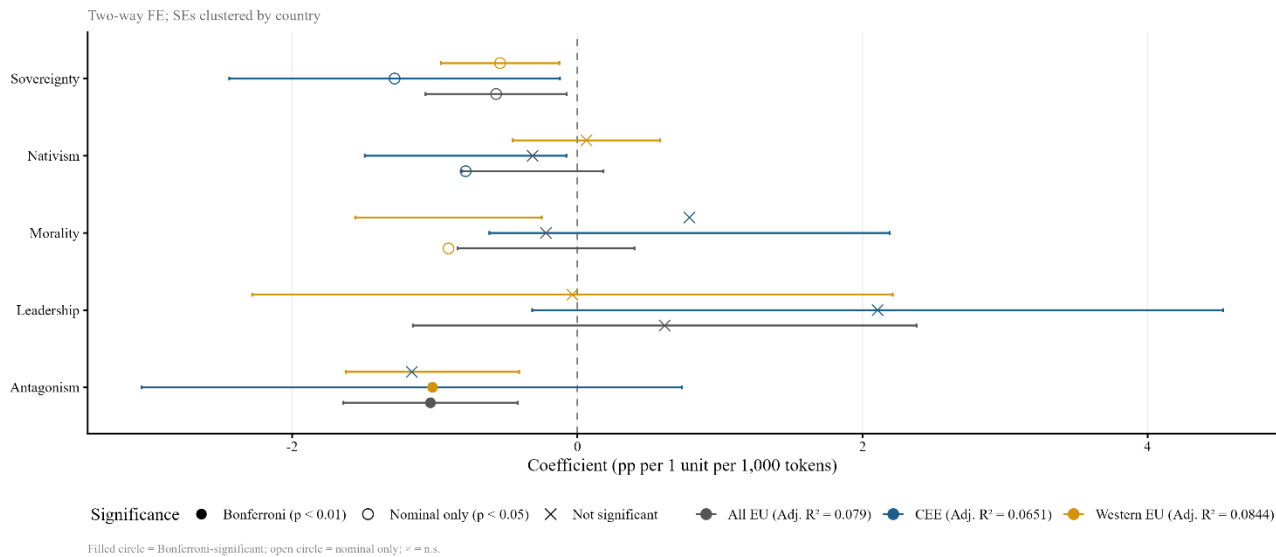
Decomposing by region reveals distinct patterns. In the Western EU subsample ($N = 602$), Antagonism ($\beta = -1.015$, $p = 0.003$) and Morality ($\beta = -0.904$, $p = 0.010$) are both significant and negative. Sovereignty is nominally significant ($\beta = -0.541$, $p = 0.014$) but does not survive the bootstrap ($p = 0.118$). Nativism and Leadership are null. In the CEE subsample ($N = 305$), the pattern matches the primary analysis: Nativism and Sovereignty are significant; Antagonism is large but non-significant due to the smaller cluster count.

Table 3. Regional Main Effects (EU Extension)

	All EU	All EU (comp.)	CEE	Western	Pop. Only	Mainst ream
Antagonism	- 1.030** (0.296)		-1.161 (0.837)	-1.015** (0.284)	- 1.503* * (0.423)	- 0.871** (0.288)
Morality	-0.220 (0.301)		0.787 (0.621)	-0.904* (0.305)	-0.044 (0.466)	-0.458 (0.304)

Nativism	-0.314 (0.241)		-0.783* (0.313)	0.064 (0.241)	-1.222 (0.672)	-0.281 (0.275)
Sovereignty	-0.569* (0.240)		-1.282* (0.513)	-0.541* (0.194)	-0.729 (0.454)	-0.527 (0.267)
Leadership	0.613 (0.856)		2.106 (1.071)	-0.034 (1.047)	-3.167* (1.435)	1.244 (1.049)
Full Populism (composite)		-0.432*** (0.112)				
GDP Growth (t-1)	-0.029 (0.105)	-0.040 (0.110)	- 0.538** (0.149)	0.144 (0.107)	0.044 (0.277)	-0.072 (0.137)
Unemployment (t-1)	-0.059 (0.094)	-0.072 (0.098)	0.199 (0.137)	-0.259 (0.188)	0.417 (0.276)	-0.151 (0.197)
Inflation (t-1)	-0.193* (0.090)	-0.195* (0.086)	- 0.159** * (0.030)	0.159 (0.452)	0.625 (0.472)	- 0.276** (0.084)
Control of Corruption (t-1)	-1.894 (1.814)	-1.512 (1.883)	-7.245 (3.444)	-8.842** (2.525)	-3.760 (4.900)	-0.466 (2.655)
Num.Obs.	907	907	305	602	191	714
R2 Adj.	0.079	0.073	0.065	0.084	0.253	0.100
• $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$						
<i>Two-way FE (country + election year). SEs clustered by country.</i>						

Figure 5. Regional Comparison: Rhetorical Dimensions



Populist and Mainstream Subsamples

The EU-wide sample permits a more reliable estimation of subsample models than the CEE-only data. Among populist parties ($N = 191$), Antagonism ($\beta = -1.503$, $p = 0.002$) and Leadership ($\beta = -3.167$, $p = 0.038$) are both negative and significant. Nativism is large and negative ($\beta = -1.222$) but marginally non-significant ($p = 0.082$). Among mainstream parties ($N = 714$), Antagonism remains the strongest predictor ($\beta = -0.871$, $p = 0.006$). The Adjusted R^2 for the populist subsample is notably higher (0.253) than for the mainstream subsample (0.100), indicating that the rhetorical dimensions explain more within-party variation among populist parties than among the broader party population.

Region Interaction Model

A region interaction model (Table 4) estimates whether the dimension coefficients differ between CEE and Western Europe. The isCEE main effect is dropped due to perfect collinearity with the country fixed effects; the interaction terms remain identified. Two interactions achieve significance. Morality $\times$ CEE is positive ($\beta = 1.620$, $p = 0.020$), indicating that the Morality Penalty observed in Western Europe does not apply in CEE, where the coefficient is near zero or slightly positive. Nativism $\times$ CEE is negative ($\beta = -0.783$, $p = 0.044$), confirming that the Nativism Penalty is a CEE-specific phenomenon absent in Western Europe. The Sovereignty $\times$ CEE interaction is negative ($\beta = -0.742$) but does not reach significance ($p = 0.133$), consistent with the finding that Sovereignty carries a penalty in both regions, though of larger magnitude in CEE.

Table 4. Region Interaction (Dimensions $\times$ CEE)

	DIMS $\times$ CEE
Antagonism	-1.027** (0.282)
Morality	-0.883**

	(0.295)
Nativism	0.067
	(0.238)
Sovereignty	-0.529**
	(0.183)
Leadership	-0.059
	(1.031)
Antagonism × CEE	-0.011
	(0.820)
Morality × CEE	1.620*
	(0.651)
Nativism × CEE	-0.783*
	(0.368)
Sovereignty × CEE	-0.742
	(0.477)
Leadership × CEE	2.000
	(1.406)
GDP Growth (t-1)	-0.018
	(0.108)
Unemployment (t-1)	-0.036
	(0.107)
Inflation (t-1)	-0.205*
	(0.087)
Control of Corruption (t-1)	-2.750
	(2.026)
Num.Obs.	907
R ² Adj.	0.097
• $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	
<i>Two-way FE (country + election year). SEs clustered by country. isCEE dropped due to collinearity with country FE; interaction terms remain identified.</i>	

Bootstrap Validation (EU-wide)

The wild cluster bootstrap on the full EU sample ($G = 25$) confirms Antagonism as the single most robust finding ($\beta = -1.030$, bootstrap $p = 0.003$; Table A3). The composite score also survives (bootstrap $p = 0.002$). Sovereignty, nominally significant under CRVE ($p = 0.026$), loses significance under bootstrap ($p = 0.106$). The Western EU bootstrap ($G = 15$) confirms Antagonism ($p = 0.008$) and Morality ($p = 0.018$) but not Sovereignty ($p = 0.118$).

Table A3. CRVE vs Wild Cluster Bootstrap (All EU)

Dimension	β	SE	p (CRVE)	p (WCB)	Bootstrap 95% CI		Significance	
					CI low	CI high	CRVE	WCB
Antagonism	-1.030	0.296	0.0020	0.0026	-1.716	-0.458	**	**
Morality	- 0.220	0.301	0.4718	0.4934	-0.835	0.452		
Nativism	-0.314	0.241	0.2043	0.2374	-1.042	0.173		
Sovereignty	- 0.569	0.240	0.0260	0.1058	-1.548	-0.250	*	
Leadership	0.613	0.856	0.4806	0.5127	-1.361	2.507		

*Webb 6-pt weights, B = 9,999. ** $p < 0.01$; * $p < 0.05$.*

DISCUSSION

The starting expectation of this thesis was that populist rhetoric in manifestos would be positively associated with vote share, and that this association would be stronger for populist parties than for mainstream ones. Neither expectation was confirmed. The five dimensions of populist rhetoric are jointly associated with electoral outcomes, but where individual dimensions reach significance, they do so in the negative direction. The broader populist rhetoric literature, grounded in issue ownership theory (Petrocik, 1996) and the ideational approach (Hawkins et al., 2018; Mudde, 2004b), would predict the opposite. The discussion that follows attempts to make sense of these results by situating them within the existing academic conversation.

The most consistent finding across both the CEE and EU-wide samples is that anti-elite rhetoric, the theoretical core of populism under Mudde's definition, is associated with lower vote share. This finding resonates with (Dai & Kustov, 2024) experimental conclusion that populist rhetorical features do not independently affect candidate choice. It also aligns with Elena Cossu's observation that anti-elite discourse has diffused broadly across the CEE party spectrum and is no longer confined to populist actors. If antagonistic framing is no longer distinctively populist, its marginal electoral value should be low or negative, which is what the data suggest. The fact that this result emerged clearly only in the larger EU sample, while the CEE point estimate was similar in size but non-significant, is best understood as a matter of statistical power rather than a substantive difference between regions.

Sovereignty rhetoric and nativist rhetoric carry negative associations in CEE, while producing no discernible effect in Western Europe. The sovereignty finding invites comparison with the literature on Euroscepticism and electoral positioning. In CEE countries that joined the EU relatively recently, explicit anti-EU sovereignty rhetoric may signal a position outside the electoral mainstream, particularly among

electorates for whom EU membership is associated with concrete economic gains. The parties most heavily invested in sovereignty rhetoric tend to be smaller radical formations whose electoral ceiling is constrained by factors other than what they write in their manifestos. The nativism result is harder to interpret. Nativist language, as captured by the dictionary, does not discriminate between populist and non-populist parties in the corpus (the validation correlation is near zero), suggesting that identity-based rhetoric is broadly distributed across the CEE party spectrum. This is consistent with the observation that ethno-nationalist framing in post-communist countries is not a populist monopoly but a shared feature of political competition.

Moralising language (fraud, scandal, shame, injustice) carries a penalty in Western Europe but produces no effect in CEE. One plausible reading is that Western European party systems, which are more programmatically structured, penalise normative condemnation in manifestos because voters there expect policy substance rather than moral indictment. In CEE, where anti-corruption rhetoric is a staple of democratic competition across the ideological spectrum, such language may be too common to carry electoral weight in either direction. The region interaction model confirms this pattern: the Morality coefficient flips sign between CEE and the West.

Anti-elite rhetoric appears to penalise populist parties specifically while leaving mainstream parties unaffected. If issue ownership theory were correct, populist parties should be rewarded for claiming anti-establishment themes as their own. The opposite result suggests a credibility dynamic: voters may discount populist anti-elite framing as performative, while mainstream parties deploying similar language are not held to the same standard, perhaps because their institutional track record provides a different interpretive frame. The leadership inversion tells a parallel story. Leadership-oriented language in manifestos appears to benefit mainstream parties, plausibly because it signals executive competence and decisiveness. For populist parties, the same rhetoric is associated with a lower vote share, possibly because it evokes authoritarian personalisation. (Moffitt, 2016) has argued that populist leadership is performative and media-driven rather than text-based, which would explain why written manifesto references to leadership are a poor proxy for the actual leadership appeal that populist voters respond to.

The estimation strategy uses country and election-year fixed effects but not party fixed effects. This means the results are identified primarily from cross-party variation: the model compares parties with different rhetorical profiles within the same country and election. It cannot distinguish between two possibilities. The first is that rhetoric has an electoral effect: a party that uses more sovereignty language gets fewer votes because of that rhetoric. The second is that rhetoric is a marker of party type: smaller, niche, radical parties use more sovereignty language and get fewer votes because they are niche parties, not because of what they write. Under the second interpretation, the negative coefficients reflect the structural disadvantage of niche parties rather than an effect of rhetoric on voters. Party fixed effects would resolve this by isolating within-party rhetorical shifts across elections, but the large number of parties in the sample appear in only one election, making such a

specification infeasible without severe sample loss. The results should therefore be read as documenting associations between rhetorical profile and electoral position in the party system. They show that certain kinds of rhetoric are concentrated among parties that perform worse electorally. Whether this concentration is a cause or a consequence of those parties' electoral position is a question this design cannot answer.

The dictionary-based operationalisation introduces its own limitations. The method counts word frequencies without regard for context: a manifesto invoking sovereignty in a pro-European argument scores identically to one that invokes it in a Eurosceptic polemic. This context-blindness probably weakens whatever true associations exist. Machine translation from ten source languages adds further noise. More semantically sensitive approaches, such as supervised classifiers or transformer-based embeddings, would offer a better approximation of rhetorical intent.

Two lines of future research follow from these limitations. First, within-party panel analysis, tracking the same party across multiple elections, would enable party fixed effects and credibly test whether rhetorical shifts within a party are followed by changes in vote share. This is the design that would address the misspecification concern raised above. Second, qualitative case studies of specific party campaigns, particularly cases where a party visibly shifted its rhetorical strategy between elections, could illuminate the causal pathways that the regression estimates can only suggest.

IMPLICATIONS FOR PUBLIC POLICY

The policy implications of these findings are limited by the correlational nature of the evidence and the constraints of dictionary-based measurement, but a few observations are worth making.

The most consistent finding across the CEE and EU-wide analyses is that more populist rhetoric in manifestos is associated with lower, not higher, vote share. The composite populism score is negative and significant across the full EU ($\beta = -0.43$, $p < 0.001$). This general pattern carries a practical implication: mainstream parties attempting to compete with populist challengers by imitating their rhetorical style are unlikely to benefit electorally. The interaction analysis confirms this directly: anti-elite rhetoric does not penalise mainstream parties, but doubling down on it does not help them either. For populist parties, the penalty is active. The policy advice that follows is straightforward: do normal politics, and failing that, don't try to speak as the populists do.

The Sovereignty and Nativism penalties in CEE suggest that sovereignty-maximalist and ethno-nationalist rhetoric do not translate into broad electoral support. This may reassure adherents to the principles of liberal democracy, but it should not breed complacency. Sovereignty rhetoric remains concentrated in populist manifestos, and even if it does not win votes on average, it may consolidate a committed ideological base and shift the parameters of acceptable public discourse. Policymakers would do

better to address the underlying frustrations (perceived loss of autonomy, uneven EU integration benefits) than to rely on counter-messaging or mimicry.

The inverted Leadership finding, confirmed across the CEE and EU-wide populist subsamples, suggests that voters respond better to signals of policy competence and decisiveness than to personalized charismatic appeals. Democratic parties competing with populist challengers may gain more from projecting institutional competence than from personality-driven mobilisation.

More broadly, the modest explanatory power of manifesto rhetoric reinforces what the demand-side literature has long argued: economic conditions, institutional quality, and historical legacies matter more for populist success than what parties write in their programmes. Policy responses to populism should prioritise structural interventions over rhetorical counter-messaging.

CONCLUSIONS

This thesis tested whether the rhetorical content of party manifestos, measured along five dimensions of populism is associated with electoral performance across ten CEE countries from 2000 to 2023. The dimensions were operationalised through dictionary-based content analysis and tested with two-way fixed-effects OLS on 305 party-election observations. An EU-wide extension tested the geographical generalisability of the findings.

H₀ was rejected: manifesto rhetoric is associated with vote share. H₁ was supported: the five dimensions are jointly significant. H₂ received partial and directionally contrary support: Sovereignty and Nativism are negatively associated with vote share in CEE; both survive the wild cluster bootstrap. The remaining dimensions are null. H₃ was rejected: populist party status does not amplify the electoral return to any dimension. For Antagonism and Leadership, the conditional effect is inverted.

The EU-wide extension reveals that Antagonism is the most robust pan-European penalty, a finding that lacked power to emerge in the 10-cluster CEE sample. The Morality penalty operates only in Western Europe; the Nativism penalty is CEE-specific. The composite populism score is negative and significant across the full EU. In the EU-wide populist subsample, Antagonism and Leadership are both significant and negative, confirming that populist rhetoric in manifestos is not just ineffective but actively penalising for populist parties specifically.

The limitations of this study are substantial. Dictionary-based analysis counts words without understanding context. Machine translation adds noise. The Adjusted R² values are modest. The CEE sample has only 10 country clusters, which limits the reliability of cluster-robust inference even with bootstrap correction. A further limitation concerns the unit of analysis. The model uses country and election-year fixed effects but not party fixed effects, meaning the results are identified from cross-party comparisons within the same electoral context rather than from within-party rhetorical shifts across elections. The negative associations observed for Sovereignty and Nativism (and Antagonism in the EU-wide analysis) may therefore reflect the structural electoral disadvantage of niche parties that happen to use this rhetoric, rather than an effect of the rhetoric on voters. The findings should accordingly be interpreted as associations between rhetorical profile and electoral position, not as evidence that rhetoric causally depresses vote share. A within-party panel design would be needed to make stronger claims.

The data are sufficient to support the main takeaway: analysing populist rhetoric dimension by dimension reveals patterns that a composite score misses entirely. The full Populism composite was borderline in CEE but significant in the EU sample; the individual dimensions showed the Sovereignty Penalty, the Nativism Penalty, and the Leadership inversion. Olivas Osuna's multidimensional approach is vindicated: future work on populist rhetoric and elections should disaggregate the concept rather than relying on unitary scores.

BIBLIOGRAPHY

- Akkerman, A., Mudde, C., & Zaslove, A. (2014). How Populist Are the People? Measuring Populist Attitudes in Voters. *Comparative Political Studies*, 47(9), 1324–1353.
<https://doi.org/10.1177/0010414013512600>
- Barr, R. R. (2009). Populists, Outsiders and Anti-Establishment Politics. *Party Politics*, 15(1), 29–48.
<https://doi.org/10.1177/1354068808097890>
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. <https://doi.org/10.21105/joss.00774>
- Bergé, L. R., Butts, K., & McDermott, G. (2026). *fixest: A fast and feature-rich framework for econometric estimations in R* (arXiv:2601.21749). arXiv.
<https://doi.org/10.48550/arXiv.2601.21749>
- Bowler, S., & Donovan, T. (2019). Perceptions of Referendums and Democracy: The Referendum Disappointment Gap. *Politics and Governance*, 7(2), 227–241.
<https://doi.org/10.17645/pag.v7i2.1874>
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3), 414–427.
<https://doi.org/10.1162/rest.90.3.414>
- Canay, I. A., Santos, A., & Shaikh, A. M. (2021). The Wild Bootstrap with a “Small” Number of “Large” Clusters. *The Review of Economics and Statistics*, 103(2), 346–363.
https://doi.org/10.1162/rest_a_00887
- Caramani, D. (2017). Will vs. Reason: The Populist and Technocratic Forms of Political Representation and Their Critique to Party Government. *The American Political Science Review*, 111(1), 54–67.
- Cossu, E. (2024). Application of Natural Language Processing to the electoral manifestos of parties characterised by populist rhetoric in Central and Eastern Europe. *Intersections. East European Journal of Society and Politics*, 10(4), 6–33. <https://doi.org/10.17356/ieejsp.v10i4.1230>

- Dai, Y., & Kustov, A. (2024). The (in)effectiveness of populist rhetoric: A conjoint experiment of campaign messaging. *Political Science Research and Methods*, 12(4), 849–856.
<https://doi.org/10.1017/psrm.2023.55>
- Döring, H., & Regel, S. (2019). Party Facts: A database of political parties worldwide. *Party Politics*, 25(2), 97–109. <https://doi.org/10.1177/1354068818820671>
- Fischer, A., & Roodman, D. (2021). *fwildclusterboot: Fast Wild Cluster Bootstrap Inference for Linear Regression Models (Version 0.13.0)*. <https://cran.r-project.org/package=fwildclusterboot>
- Froio, C., Rooduijn, M., de Lange, S., Halikiopoulou, D., Taggart, P., Pirro, A., van Kessel, S., & Mudde, C. (2023). *The PopuList 3.0*. <https://doi.org/10.17605/OSF.IO/2EWKQ>
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251–1271.
<https://doi.org/10.2307/1913827>
- Hawkins, K. A. (2009). Is Chávez Populist?: Measuring Populist Discourse in Comparative Perspective. *Comparative Political Studies*, 42(8), 1040–1067. <https://doi.org/10.1177/0010414009331721>
- Hawkins, K. A. (Ed.). (2010). Measuring the Populist Discourse of Chavismo. In *Venezuela's Chavismo and Populism in Comparative Perspective* (pp. 50–85). Cambridge University Press.
Cambridge Core. <https://doi.org/10.1017/CBO9780511730245.004>
- Hawkins, K. A., Carlin, R. E., Littvay, L., & Kaltwasser, C. R. (Eds.). (2018). *The Ideational Approach to Populism: Concept, Theory, and Analysis*. Routledge. <https://doi.org/10.4324/9781315196923>
- Hawkins, K. A., Riding, S., & Mudde, C. (2012). *Measuring Populist Attitudes*. Committee on Concepts and Methods. <https://books.google.com.ua/books?id=gpJJAQAAIAJ>
- Hobolt, S. B., & de Vries, C. E. (2015). Issue Entrepreneurship and Multiparty Competition. *Comparative Political Studies*, 48(9), 1159–1185. <https://doi.org/10.1177/0010414015575030>
- Jagers, J., & Walgrave, S. (2007). Populism as political communication style: An empirical study of political parties' discourse in Belgium. *European Journal of Political Research*, 46(3), 319–345.
<https://doi.org/10.1111/j.1475-6765.2006.00690.x>

- Kazin, M. (2016, October 6). Trump and American Populism. *Foreign Affairs*, 95(6).
<https://www.foreignaffairs.com/articles/united-states/2016-10-06/trump-and-american-populism>
- Kriesi, H. (2014). The Populist Challenge. *West European Politics*, 37(2), 361–378.
<https://doi.org/10.1080/01402382.2014.887879>
- Laver, M., Benoit, K., & Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review*, 97(2), 311–331.
<https://doi.org/10.1017/S0003055403000698>
- Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Regel, S., Riethmüller, F., Volkens, A., Weßels, B., Zehnter, L., Wissenschaftszentrum Berlin für Sozialforschung (WZB), & Institut für Demokratieforschung Göttingen (IfDem). (2025). *Manifesto Project Dataset* (Version 2025a) [Dataset]. Manifesto Project.
<https://doi.org/10.25522/MANIFESTO.MPDS.2025A>
- MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2), 114–135. <https://doi.org/10.1111/ectj.12107>
- Mastropaolo, A. (2008). Politics against Democracy: Party Withdrawal and Populist Breakthrough. In D. Albertazzi & D. McDonnell (Eds.), *Twenty-First Century Populism: The Spectre of Western European Democracy* (pp. 30–48). Palgrave Macmillan UK.
https://doi.org/10.1057/9780230592100_3
- Meguid, B. M. (2005). Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success. *American Political Science Review*, 99(3), 347–359.
<https://doi.org/10.1017/S0003055405051701>
- Meijers, M. J., & Zaslove, A. (2021). Measuring Populism in Political Parties: Appraisal of a New Approach. *Comparative Political Studies*, 54(2), 372–407.
<https://doi.org/10.1177/0010414020938081>

- Moffitt, B. (2016). The Audience: Populism and 'The People.' In B. Moffitt (Ed.), *The Global Rise of Populism: Performance, Political Style, and Representation* (p. 0). Stanford University Press.
<https://doi.org/10.11126/stanford/9780804796132.003.0006>
- Mudde, C. (2004a). The Populist Zeitgeist. *Government and Opposition*, 39(4), 541–563.
<https://doi.org/10.1111/j.1477-7053.2004.00135.x>
- Mudde, C. (2004b). The Populist Zeitgeist. *Government and Opposition*, 39(4), 541–563.
<https://doi.org/10.1111/j.1477-7053.2004.00135.x>
- Mudde, C., & Kaltwasser, C. R. (2013). Exclusionary vs. Inclusionary Populism: Comparing Contemporary Europe and Latin America. *Government and Opposition*, 48(2), 147–174.
<https://doi.org/10.1017/gov.2012.11>
- Mudde, C., & Kaltwasser, C. R. (2017). (page 42) Populism and mobilization. In C. Mudde & C. Rovira Kaltwasser (Eds.), *Populism: A Very Short Introduction* (p. 0). Oxford University Press.
<https://doi.org/10.1093/actrade/9780190234874.003.0003>
- Mudde, C., & Rovira Kaltwasser, C. (2018). Studying Populism in Comparative Perspective: Reflections on the Contemporary and Future Research Agenda. *Comparative Political Studies*, 51(13), 1667–1693. <https://doi.org/10.1177/0010414018789490>
- Norris, P., & Inglehart, R. (2019). *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press. <https://doi.org/10.1017/9781108595841>
- Olivas Osuna, J. J. O. (2021). From chasing populists to deconstructing populism: A new multidimensional approach to understanding and comparing populism. *European Journal of Political Research*, 60(4), 829–853. <https://doi.org/10.1111/1475-6765.12428>
- Petrocik, J. R. (1996). Issue Ownership in Presidential Elections, with a 1980 Case Study. *American Journal of Political Science*, 40(3), 825–850. <https://doi.org/10.2307/2111797>
- Rooduijn, M. (2018). What unites the voter bases of populist parties? Comparing the electorates of 15 populist parties. *European Political Science Review*, 10(3), 351–368.
<https://doi.org/10.1017/S1755773917000145>

- Rooduijn, M., de Lange, S. L., & van der Brug, W. (2014). A populist Zeitgeist? Programmatic contagion by populist parties in Western Europe. *Party Politics*, 20(4), 563–575.
<https://doi.org/10.1177/1354068811436065>
- Rooduijn, M., & Pauwels, T. (2011). Measuring Populism: Comparing Two Methods of Content Analysis. *West European Politics*, 34(6), 1272–1283.
<https://doi.org/10.1080/01402382.2011.616665>
- Rooduijn, M., Pirro, A. L. P., Halikiopoulou, D., Froio, C., Kessel, S. V., Lange, S. L. D., Mudde, C., & Taggart, P. (2024). The PopuList: A Database of Populist, Far-Left, and Far-Right Parties Using Expert-Informed Qualitative Comparative Classification (EiQCC). *British Journal of Political Science*, 54(3), 969–978. <https://doi.org/10.1017/S0007123423000431>
- Stanley, B. (2008). The thin ideology of populism. *Journal of Political Ideologies*, 13(1), 95–110.
<https://doi.org/10.1080/13569310701822289>
- Stiers, D. (2022). Performance voting, retrospective voting, and economic voting. Conceptual clarity and empirical testing. *Social Science Quarterly*, 103(2), 399–408.
<https://doi.org/10.1111/ssqu.13125>
- Swank, D. (2018, September 21). *Globalization, Institutions of Social Solidarity, and Radical Right-Wing Populism in Western Europe*.
- Taggart, P. A. (2000). *Populism* (1 edition). Open University Press.

SUPPLEMENTS

Appendix 1. Populism Dictionary

Dimension	N	Stemmed Terms
D1: Antagonism	23	elit*, corrupt*, enem*, underdog*, oligarch*, croni*, betray*, traitor*, puppet*, plutocrat*, parasit*, loot*, caste, rigg*, swamp*, cabal*, cartel*, sell-out*, conspir*, exploit*, oppress*, plunder*, mafia
D2: Morality	25	virtuous*, honest*, truth*, noble, deceit*, fraud*, dishonest*, scandal*, propagand*, immoral*, shame*, liar*, hypocrisi*, sincere*, righteous*, wicked*, greed*, treacher*, villain*, dignit*, courag*, coward*, sacred*, justic*, injustic*
D3: Nativism	20	homogenous*, authent*, heartland*, identit*, volk, ordinar*, grassroot*, folk*, kinship, homeland*, patriot*, ancestr*, heritage*, tradition*, rootedness, communiti*, nativ*, belong*, motherland*, fatherland*
D4: Sovereignty	18	sovereign*, referend*, bypass*, unmediat*, self-determin*, veto*, plebiscit*, independen*, autonom*, supranation*, diktat*, imposit*, majorit*, accountab*, recall*, petition*, grassroot*, emancip*
D5: Leadership	14	leader*, charismat*, saviour*, savior*, guardian*, protector*, strongman*, decisive*, visionar*, command*, liberator*, heroic*, defender*, crusad*

Appendix 2. R code implementation

Populist Rhetoric and Electoral Performance: A Dictionary-Based Analysis of Party Manifestos in Central and Eastern Europe

Master's Thesis — Public Policy and Governance, KSE

Author: Maksym Tsukan

SECTION 0: PACKAGE LOADING

library(manifestoR)

library(tidyverse)

library(quanteda)

library(tidytext)

library(fixest)

library(WDI)

library(modelsummary)

library(ggplot2)

```

library(ggdist)
library(marginaleffects)
library(patchwork)
library(scales)
library(gt)
library(car)
library(plm)
library(readxl)

# SECTION 1: DATA ACQUISITION

mp_setapikey("manifesto_apikey.txt")
manifestoR::mp_use_corpus_version("2025-1")

# Nine CEE countries
ee_countries <- c("Poland", "Hungary", "Czech Republic", "Slovakia",
                  "Estonia", "Latvia", "Lithuania", "Bulgaria", "Romania", "Croatia", "Slovenia")

# Pull all elections ≥ 2000
main_df <- mp_maindataset() %>%
  filter(countryname %in% ee_countries,
         edate >= as.Date("2000-01-01"))
main_df_backup <- main_df

# Retrieve full-text corpora with English translations
corp <- mp_corpus(main_df, translation = "en")
q_corp <- quanteda::corpus(corp)
corp_backup <- corp

# SECTION 2: TEXT PRE-PROCESSING

# Merge tidytext stop-words with a domain-specific list
domain_stopwords <- c("party", "government", "poland", "hungary", "czech",
                      "slovak", "estonia", "latvia", "lithuania", "bulgaria",
                      "romania", "european", "union", "percent", "shall",
                      "will", "also", "must", "within", "among", "however",
                      "croatia", "slovenia")

all_stopwords <- union(tidytext::stop_words$word, domain_stopwords)

toks <- tokens(q_corp) %>%
  tokens_tolower() %>%
  tokens_remove(pattern = all_stopwords) %>%

```

```

tokens_remove(pattern = "[[:punct:]]", valuetype = "regex") %>%
tokens(remove_numbers = TRUE) %>%
tokens_wordstem()

# Build Document-Feature Matrix and attach manifesto ID as docname
clean_dfm <- dfm(toks)
docnames(clean_dfm) <- docvars(clean_dfm, "manifesto_id")

# SECTION 3: FIVE-DIMENSIONAL POPULISM DICTIONARY
#improved dictionary
dict_populism_5dim <- dictionary(list(
  antagonism = c(
    "elit*", "corrupt*", "enem*", "underdog*", "oligarch*", "croni*",
    "betray*", "traitor*", "puppet*", "plutocrat*", "parasit*", "loot*", "caste",
    "rigg*", "swamp*", "cabal*", "cartel*", "sell-out*", "conspir*", "exploit*",
    "oppress*", "plunder*", "mafia"
  ),
  morality = c(
    "virtuous*", "honest*", "truth*", "noble", "deceit*", "fraud*", "dishonest*",
    "scandal*", "propagand*", "immoral*", "shame*", "liar*", "hypocrisi*",
    "sincere*", "righteous*", "wicked*", "greed*", "treacher*", "villain*",
    "dignit*", "courag*", "coward*", "sacred*", "justic*", "injustic*"
  ),
  nativism = c(
    "homogenous*", "authent*", "heartland*", "identit*", "volk",
    "ordinar*", "grassroot*", "folk*", "kinship", "homeland*",
    "patriot*", "ancestr*", "heritage*", "tradition*", "rootedness", "communiti*",
    "nativ*", "belong*", "motherland*", "fatherland*"
  ),
  sovereignty = c(
    "sovereign*", "referend*", "bypass*", "unmediat*", "self-determin*",
    "veto*", "plebiscit*", "independen*", "autonom*", "supranation*", "diktat*",
    "imposit*", "majorit*", "accountab*", "recall*", "petition*", "grassroot*", "emancip*"
  ),
  leadership = c(
    "leader*", "charismat*", "saviour*", "savior*", "guardian*", "protector*", "strongman*",

```

```

"decisive*", "visionar*", "command*", "liberator*", "heroic*", "defender*", "crusad*"
)))
# SECTION 4: DICTIONARY SCORING & NORMALISATION
dict_dfm <- dfm_lookup(clean_dfm, dictionary = dict_populism_5dim)
doc_lengths <- ntoken(clean_dfm)
# Proportional scores per 1,000 tokens
populism_scores <- convert(dict_dfm, to = "data.frame") %>%
  mutate(across(antagonism:leadership,
    ~ (. / doc_lengths) * 1000,
    .names = "{.col}_norm"))
# Retain original counts alongside normalised scores for robustnes
populism_scores <- populism_scores %>%
  mutate(
    fullPopulism_norm = antagonism_norm + morality_norm +
      nativism_norm + sovereignty_norm + leadership_norm
  )
# SECTION 5: MERGE WITH MANIFESTO METADATA & POPULIST LABELS
main_df <- main_df %>%
  mutate(manifesto_id = paste0(party, "_", date))
analysis_ready <- populism_scores %>%
  mutate(manifesto_id = doc_id) %>%
  left_join(main_df, by = "manifesto_id")
# Party-Facts Bridge
pf_bridge <- read_csv(
  "https://partyfacts.herokuapp.com/download/external-parties-csv/"
) %>%
  filter(dataset_key == "manifesto") %>%
  select(dataset_party_id, partyfacts_id) %>%
  mutate(party = as.numeric(dataset_party_id))
analysis_ready <- analysis_ready %>%
  left_join(pf_bridge, by = "party")
# PopuList classification from PopulisTree (local file, API not functioning properly, retrieved at
https://populistree.org/dataset-on-populist-parties-in-national-elections/)
pop_tree_df <- read_csv("ok_dataset_national-elections.csv") %>%

```

```

select(Partyfacts_id, Type) %>%
mutate(
  isPopulist = 1L,
  partyfacts_id = Partyfacts_id
)
analysis_ready <- analysis_ready %>%
  left_join(pop_tree_df,
    join_by("partyfacts_id" == "partyfacts_id"),
    multiple = "first")
# Replace NA in isPopulist with 0 (non-populist parties)
analysis_ready <- analysis_ready %>%
  mutate(isPopulist = replace_na(as.numeric(isPopulist), 0))
analysis <- analysis_ready %>%
  select(countryname, partyname, edate, pervote,
    antagonism_norm, morality_norm, nativism_norm,
    sovereignty_norm, leadership_norm,
    fullPopulism_norm, isPopulist) %>%
  mutate(
    election_year = as.numeric(format(edate, "%Y"))
  )
# SECTION 6: MACROECONOMIC CONTROLS (WDI)
indicators <- c(
  "gdp_growth" = "NY.GDP.MKTP.KD.ZG",
  "unemployment" = "SL.UEM.TOTL.ZS",
  "inflation" = "FP.CPI.TOTL.ZG"
)
ee_iso <- c("PL", "HU", "CZ", "SK", "EE", "LV", "LT", "BG", "RO", "HR", "SI")
options(timeout = 300)
Wb_data_raw <- WDI(
  country = ee_iso,
  indicator = indicators,
  start = 1999,
  end = 2025
)

```

```

# Harmonise country name spelling
analysis <- analysis %>%
  mutate(
    countryname = case_when(
      countryname == "Czech Republic" ~ "Czechia",
      countryname == "Slovakia" ~ "Slovak Republic",
      TRUE ~ countryname
    )
  )

# One-year lag of macroeconomic variables (t-1 relative to election year),
analysis <- analysis %>%
  mutate(
    lag_year = case_when(
      election_year == 2002 ~ 2002L,
      (election_year - 1) == 1999 ~ 2000L,
      TRUE ~ election_year - 1L
    )
  ) %>%
  left_join(Wb_data_raw,
    by = c("countryname" = "country", "lag_year" = "year"))
analysis <- analysis %>%
  filter(!is.na(pervote), !is.na(antagonism_norm))

# CORRUPTION CONTROL from WGI source file (replaces WDI API call)
wgi_cc <- read_excel("wgidataset_with_sourcedata-2025.xlsx", sheet = "cc") %>%
  select(
    iso3 = `Economy (code)`,
    year = Year,
    corruption = `Governance estimate (approx. -2.5 to +2.5)`
  ) %>%
  mutate(year = as.integer(year),
    corruption = as.numeric(corruption)) %>%
  filter(!is.na(corruption))

# Map ISO3 codes to the country names used in analysis
iso3_to_country <- c(

```

```

"POL" = "Poland", "HUN" = "Hungary", "CZE" = "Czechia",
"SVK" = "Slovak Republic", "EST" = "Estonia", "LVA" = "Latvia",
"LTU" = "Lithuania", "BGR" = "Bulgaria", "ROU" = "Romania",
"SVN" = "Slovenia", "HRV" = "Croatia"
)
wgi_cc <- wgi_cc %>%
  filter(iso3 %in% names(iso3_to_country)) %>%
  mutate(countryname = iso3_to_country[iso3])
# Merge into analysis using lag_year (t-1)
analysis <- analysis %>%
  left_join(wgi_cc %>% select(countryname, year, corruption),
    by = c("countryname" = "countryname", "lag_year" = "year"))
# SECTION 7: VALIDATION — POINT-BISERIAL CORRELATIONS (WITH BONFERRONI)
bonferroni_alpha <- 0.05 / 5
pop_dimensions <- c("antagonism_norm", "morality_norm", "nativism_norm",
  "sovereignty_norm", "leadership_norm", "fullPopulism_norm")
validation_summary <- pop_dimensions %>%
  map_df(~{
    test <- cor.test(analysis[.[x]], analysis$sisPopulist)
    tibble(
      dimension    = .x,
      correlation_r = test$estimate,
      p_value      = test$p.value,
      ci_lower     = test$conf.int[1],
      ci_upper     = test$conf.int[2]
    )
  }) %>%
  mutate(
    significant_raw    = p_value < 0.05,
    significant_bonferroni = p_value < bonferroni_alpha
  )
print(validation_summary)
# SECTION 8: FIXED-EFFECTS REGRESSION MODELS (H1, H2) and Wald test (H0)
controls_formula <- "gdp_growth + unemployment + inflation + corruption"

```

```

# Model 1a: Individual dimensional scores
m_dims <- feols(
  as.formula(paste(
    "pervote ~ antagonism_norm + morality_norm + nativism_norm +",
    "sovereignty_norm + leadership_norm +", controls_formula,
    "| countryname + election_year"
  )),
  data = analysis,
  cluster = ~ countryname # cluster SEs at country level
)

# Model 1b: Composite populism score
m_full <- feols(
  as.formula(paste(
    "pervote ~ fullPopulism_norm +", controls_formula,
    "| countryname + election_year"
  )),
  data = analysis,
  cluster = ~ countryname
)

# Wald test: H0 = all five dimension coefficients = 0
wald(m_dims, keep = c("antagonism_norm", "morality_norm",
  "nativism_norm", "sovereignty_norm", "leadership_norm"))

summary(m_dims)
summary(m_full)

# SECTION 9: INTERACTION MODELS (H2 — STRATEGIC MIMICRY TEST)
m_interact_antagonism <- feols(
  as.formula(paste(
    "pervote ~ antagonism_norm * isPopulist +", controls_formula,
    "| countryname + election_year"
  )),
  data = analysis,
  cluster = ~ countryname
)

m_interact_morality <- feols(

```

```

as.formula(paste(
  "pervote ~ morality_norm * isPopulist +", controls_formula,
  "| countryname + election_year"
)),
data = analysis,
cluster = ~ countryname
)
m_interact_nativism <- feols(
  as.formula(paste(
    "pervote ~ nativism_norm * isPopulist +", controls_formula,
    "| countryname + election_year"
  )),
  data = analysis,
  cluster = ~ countryname
)
m_interact_sovereignty <- feols(
  as.formula(paste(
    "pervote ~ sovereignty_norm * isPopulist +", controls_formula,
    "| countryname + election_year"
  )),
  data = analysis,
  cluster = ~ countryname
)
m_interact_leadership <- feols(
  as.formula(paste(
    "pervote ~ leadership_norm * isPopulist +", controls_formula,
    "| countryname + election_year"
  )),
  data = analysis,
  cluster = ~ countryname
)
# Collect all interaction models into a named list
interaction_models <- list(
  "Antagonism" = m_interact_antagonism,

```

```

"Morality" = m_interact_morality,
"nativism" = m_interact_nativism,
"Sovereignty" = m_interact_sovereignty,
"Leadership" = m_interact_leadership
)
print(interaction_models)

# SECTION 10: BONFERRONI CORRECTION FOR REGRESSION COEFFICIENTS
coef_table_raw <- broom::tidy(m_dims, conf.int = TRUE) %>%
  filter(str_detect(term, "_norm")) %>%
  mutate(
    p_bonferroni_sig = p.value < bonferroni_alpha,
    label_sig = case_when(
      p.value < 0.001 ~ "****",
      p.value < bonferroni_alpha ~ "***",
      p.value < 0.05 ~ "**",
      TRUE ~ ""
    )
  )
print(coef_table_raw)

# COEFFICIENT PLOT — Main Effects (H2)
dim_labels <- c(
  "antagonism_norm" = "Antagonism",
  "morality_norm" = "Morality",
  "nativism_norm" = "nativism",
  "sovereignty_norm" = "Sovereignty",
  "leadership_norm" = "Leadership"
)
p_coef_main <- coef_table_raw %>%
  mutate(
    term_label = case_match(term,
      "antagonism_norm" ~ "Antagonism",
      "morality_norm" ~ "Morality",
      "nativism_norm" ~ "nativism",
      "sovereignty_norm" ~ "Sovereignty",

```

```

    "leadership_norm" ~ "Leadership"),
  bonf_label = if_else(p_bonferroni_sig, "Bonferroni-significant", "Not significant")
) %>%
ggplot(aes(x = estimate, y = fct_reorder(term_label, estimate),
  colour = bonf_label)) +
  geom_vline(xintercept = 0, linetype = "dashed", colour = "grey60") +
  geom_errorbar(aes(xmin = conf.low, xmax = conf.high),
    height = 0.15, linewidth = 0.7) +
  geom_point(size = 3) +
  scale_colour_manual(
    values = c("Bonferroni-significant" = "#1B5E8C",
      "Not significant" = "#C04A4A"),
    name = NULL
  ) +
  labs(
    title = "Figure 1. Estimated Effect of Populist Rhetoric Dimensions on Vote Share",
    subtitle = "Two-way FE (country + year); SEs clustered by country; Bonferroni  $\alpha = 0.01$ ",
    x = "Coefficient (percentage-point change in vote share\nper 1 additional occurrence per 1,000
tokens)",
    y = NULL
  ) +
  thesis_theme
ggsave("Figure1_MainEffects_CoefPlot_.png", p_coef_main,
  width = 8, height = 5, dpi = 300)
# — 12: EXTRA
# Multicollinearity test
m_dims_lm <- lm(pervote ~ antagonism_norm + morality_norm + nativism_norm +
  sovereignty_norm + leadership_norm +
  gdp_growth + unemployment + inflation +
  factor(countryname) + factor(election_year),
  data = analysis)
vif_raw <- vif(m_dims_lm)
print(vif_raw)

```

```

as.data.frame(vif_raw) %>%
  rownames_to_column("Variable") %>%
  filter(!str_detect(Variable, "factor\\(")) %>%
  gt() %>%
  tab_header(title = "Table A2. Variance Inflation Factors (Model 1)") %>%
  fmt_number(columns = c(GVIF, `GVIF^(1/(2*Df))`), decimals = 2) %>%
  tab_source_note(
    source_note = "Country and election-year fixed effects omitted. Threshold:  $GVIF^{1/(2*Df)} > 3.16$ 
indicates high collinearity."
  ) %>%
  gtsave("TableA2_VIF_2.docx")
# Hausman test
# Prepare panel structure
analysis_plm <- analysis %>%
  mutate(panel_id = paste0(countryname, "_", partyname))
pdata <- pdata.frame(analysis_plm,
  index = c("panel_id", "election_year"),
  drop.index = FALSE)
# Fixed effects (for Hausman comparison)
fe_plm <- plm(pervote ~ antagonism_norm + morality_norm + nativism_norm +
  sovereignty_norm + leadership_norm +
  gdp_growth + unemployment + inflation + corruption,
  data = pdata, model = "within", effect = "individual")
# Random effects
re_plm <- plm(pervote ~ antagonism_norm + morality_norm + nativism_norm +
  sovereignty_norm + leadership_norm +
  gdp_growth + unemployment + inflation + corruption,
  data = pdata, model = "random", effect = "individual")
# The test
hausman <- phptest(fe_plm, re_plm)
print(hausman)
# WILD CLUSTER BOOTSTRAP ROBUSTNESS CHECKs
library(fwildclusterboot)
library(fixest)

```

```

library(tidyverse)

set.seed(8417652)

dqrng::dqset.seed(8417652)

B <- 9999

# PRE-PROCESSING: convert FE vars to factors, remove singletons
analysis <- analysis %>%

  mutate(countryname = as.factor(countryname),

         election_year = as.factor(election_year))

controls_formula <- "gdp_growth + unemployment + inflation + corruption"

dims_formula <- paste(

  "pervote ~ antagonism_norm + morality_norm + nativism_norm +",

  "sovereignty_norm + leadership_norm +", controls_formula,

  "| countryname + election_year"

)

m_detect <- feols(as.formula(dims_formula), data = analysis, cluster = ~countryname)

kept_rows <- obs(m_detect)

cat("Full sample:", nrow(analysis), "->", length(kept_rows), "after singleton removal\n")

analysis_clean <- analysis[kept_rows, ] %>%

  mutate(countryname = droplevels(countryname),

         election_year = droplevels(election_year))

# RE-ESTIMATE ON CLEANED DATA

m_dims <- feols(as.formula(dims_formula), data = analysis_clean, cluster = ~countryname)

stopifnot(is.null(m_dims$fixef_removed))

m_full <- feols(

  as.formula(paste("pervote ~ fullPopulism_norm +", controls_formula,

                  "| countryname + election_year")),

  data = analysis_clean, cluster = ~countryname)

# BOOTSTRAP HELPER

run_boot <- function(model, param) {

  boot <- boottest(model, param = param, clustid = "countryname",

                  fe = "countryname", B = B, type = "webb",

                  impose_null = TRUE, p_val_type = "two-tailed")

  tibble(term = param, estimate = boot$point_estimate,

         t_stat = boot$t_stat, p_boot = pval(boot),

```

```

ci_lower = confint(boot)[1], ci_upper = confint(boot)[2])
}

dim_params <- c("antagonism_norm", "morality_norm", "nativism_norm",
               "sovereignty_norm", "leadership_norm")

# MODEL 1: Five dimensions
cat("\n--- Model 1: Five Dimensions ---\n")
boot_m1 <- map_df(dim_params, ~{ cat(.x, ""); run_boot(m_dims, .x) })
print(boot_m1)

# MODEL 2: Composite
cat("\n--- Model 2: Composite ---\n")
boot_m2 <- run_boot(m_full, "fullPopulism_norm")
print(boot_m2)

# CRVE vs BOOTSTRAP COMPARISON
cat("\n--- CRVE vs Bootstrap ---\n")
comparison <- broom::tidy(m_dims, conf.int = TRUE) %>%
  filter(term %in% dim_params) %>%
  select(term, estimate, se_crve = std.error, p_crve = p.value) %>%
  left_join(boot_m1 %>% select(term, p_boot, ci_lower, ci_upper), by = "term") %>%
  mutate(sig_crve = ifelse(p_crve < 0.01, "***", ifelse(p_crve < 0.05, "**", "")),
         sig_boot = ifelse(p_boot < 0.01, "***", ifelse(p_boot < 0.05, "**", "")))
print(comparison)

# CONTROLS
cat("\n--- Controls ---\n")
boot_controls <- map_df(c("gdp_growth", "unemployment", "inflation", "corruption"),
                      ~ run_boot(m_dims, .x))
print(boot_controls)

# ALL-EU AND EAST-VS-WEST EXTENSION
# COUNTRY DEFINITIONS
cee_countries <- c("Poland", "Hungary", "Czech Republic", "Slovakia",
                  "Estonia", "Latvia", "Lithuania", "Bulgaria", "Romania", "Croatia", "Slovenia")
weu_countries <- c("Austria", "Belgium", "Denmark", "Finland", "France",
                  "Germany", "Greece", "Ireland", "Italy", "Luxembourg",
                  "Netherlands", "Portugal", "Spain", "Sweden",
                  "Cyprus", "Malta")

```

```

all_eu <- c(cee_countries, weu_countries)

# — DATA ACQUISITION —

main_df_eu <- mp_maindataset() %>%
  filter(countryname %in% all_eu, edate >= as.Date("2000-01-01"))
corp_eu <- mp_corpus(main_df_eu, translation = "en")
q_corp_eu <- quanteda::corpus(corp_eu)

# TEXT PREPROCESSING

domain_stopwords <- c("party", "government", "european", "union",
  "percent", "shall", "will", "also", "must",
  "within", "among", "however",
  tolower(all_eu))

all_stopwords <- union(tidytext::stop_words$word, domain_stopwords)

toks_eu <- tokens(q_corp_eu) %>%
  tokens_tolower() %>%
  tokens_remove(pattern = all_stopwords) %>%
  tokens_remove(pattern = "[[:punct:]]", valuetype = "regex") %>%
  tokens(remove_numbers = TRUE) %>%
  tokens_wordstem()

clean_dfm_eu <- dfm(toks_eu)

docnames(clean_dfm_eu) <- docvars(clean_dfm_eu, "manifesto_id")

# SCORING

dict_dfm_eu <- dfm_lookup(clean_dfm_eu, dictionary = dict_populism_5dim)
doc_lengths_eu <- ntoken(clean_dfm_eu)

scores_eu <- convert(dict_dfm_eu, to = "data.frame") %>%
  mutate(across(antagonism:leadership,
    ~ (. / doc_lengths_eu) * 1000,
    .names = "{.col}_norm")) %>%
  mutate(fullPopulism_norm = antagonism_norm + morality_norm +
    nativism_norm + sovereignty_norm + leadership_norm)

# MERGE METADATA & POPULIST LABELS

main_df_eu <- main_df_eu %>%
  mutate(manifesto_id = paste0(party, "_", date))

analysis_eu <- scores_eu %>%
  mutate(manifesto_id = doc_id) %>%

```

```

left_join(main_df_eu, by = "manifesto_id")
pf_bridge <- read_csv(
  "https://partyfacts.herokuapp.com/download/external-parties-csv/"
) %>%
  filter(dataset_key == "manifesto") %>%
  select(dataset_party_id, partyfacts_id) %>%
  mutate(party = as.numeric(dataset_party_id))
analysis_eu <- analysis_eu %>%
  left_join(pf_bridge, by = "party")
pop_tree_df <- read_csv("ok_dataset_national-elections.csv") %>%
  select(Partyfacts_id, Type) %>%
  mutate(isPopulist = 1L, partyfacts_id = Partyfacts_id)
analysis_eu <- analysis_eu %>%
  left_join(pop_tree_df,
    join_by("partyfacts_id" == "partyfacts_id"),
    multiple = "first") %>%
  mutate(isPopulist = replace_na(as.numeric(isPopulist), 0))
analysis_eu <- analysis_eu %>%
  select(countryname, partyname, edate, pervote,
    antagonism_norm, morality_norm, nativism_norm,
    sovereignty_norm, leadership_norm,
    fullPopulism_norm, isPopulist) %>%
  mutate(election_year = as.numeric(format(edate, "%Y")),
    region = if_else(countryname %in% c(
      "Poland", "Hungary", "Czech Republic", "Czechia",
      "Slovakia", "Slovak Republic",
      "Estonia", "Latvia", "Lithuania",
      "Bulgaria", "Romania", "Croatia", "Slovenia"
    ), "CEE", "Western"))
# MACROECONOMIC CONTROLS
eu_iso <- c("PL", "HU", "CZ", "SK", "EE", "LV", "LT", "BG", "RO",
  "AT", "BE", "DK", "FI", "FR", "DE", "GR", "IE", "IT",
  "LU", "NL", "PT", "ES", "SE", "CY", "MT", "HR", "SI")
indicators <- c("gdp_growth" = "NY.GDP.MKTP.KD.ZG",

```

```

      "unemployment" = "SL.UEM.TOTL.ZS",
      "inflation" = "FP.CPI.TOTL.ZG")
options(timeout = 300)
wb_eu <- WDI(country = eu_iso, indicator = indicators,
             start = 1999, end = 2025)
analysis_eu <- analysis_eu %>%
  mutate(countryname = case_when(
    countryname == "Czech Republic" ~ "Czechia",
    countryname == "Slovakia" ~ "Slovak Republic",
    TRUE ~ countryname
  ))
analysis_eu <- analysis_eu %>%
  mutate(lag_year = case_when(
    election_year == 2002 ~ 2002L,
    (election_year - 1) == 1999 ~ 2000L,
    TRUE ~ election_year - 1L
  )) %>%
  left_join(wb_eu, by = c("countryname" = "country", "lag_year" = "year")) %>%
  filter(!is.na(pervote), !is.na(antagonism_norm))
cat("All-EU N =", nrow(analysis_eu), "\n")
cat("CEE N =", sum(analysis_eu$region == "CEE"), "\n")
cat("Western N =", sum(analysis_eu$region == "Western"), "\n")
cat("Populist N =", sum(analysis_eu$isPopulist == 1), "\n")
# CORRUPTION CONTROL from WGI source file
wgi_cc <- read_excel("wgidataset_with_sourcedata-2025.xlsx", sheet = "cc") %>%
  select(
    iso3 = `Economy (code)` ,
    year = Year,
    corruption = `Governance estimate (approx. -2.5 to +2.5)`
  ) %>%
  mutate(year = as.integer(year),
         corruption = as.numeric(corruption)) %>%
  filter(!is.na(corruption))
# Map ISO3 codes to the country names used in analysis

```

```

iso3_to_country <- c(
  "AUT" = "Austria",    "BEL" = "Belgium",    "BGR" = "Bulgaria",
  "HRV" = "Croatia",    "CYP" = "Cyprus",      "CZE" = "Czechia",
  "DNK" = "Denmark",    "EST" = "Estonia",    "FIN" = "Finland",
  "FRA" = "France",     "DEU" = "Germany",    "GRC" = "Greece",
  "HUN" = "Hungary",    "IRL" = "Ireland",    "ITA" = "Italy",
  "LVA" = "Latvia",     "LTU" = "Lithuania",  "LUX" = "Luxembourg",
  "MLT" = "Malta",      "NLD" = "Netherlands", "POL" = "Poland",
  "PRT" = "Portugal",   "ROU" = "Romania",    "SVK" = "Slovakia",
  "SVN" = "Slovenia",   "ESP" = "Spain",      "SWE" = "Sweden"
)

wgi_cc <- wgi_cc %>%
  filter(iso3 %in% names(iso3_to_country)) %>%
  mutate(countryname = iso3_to_country[iso3])

wgi_cc <- wgi_cc %>%
  mutate(countryname = ifelse(countryname == "Slovakia", "Slovak Republic", countryname))

# Merge into analysis using lag_year (t-1)
analysis_eu <- analysis_eu %>%
  left_join(wgi_cc %>% select(countryname, year, corruption),
    by = c("countryname" = "countryname", "lag_year" = "year"))

analysis_eu_backup <- analysis_eu

# VALIDATION: POINT-BISERIAL CORRELATIONS
bonferroni_alpha <- 0.05 / 5

dims <- c("antagonism_norm", "morality_norm", "nativism_norm",
  "sovereignty_norm", "leadership_norm", "fullPopulism_norm")

run_validation <- function(df, label) {
  cat("\n===", label, "(N =", nrow(df), ") ===\n")
  res <- dims %>%
    map_df(~ {
      test <- cor.test(df[.[x]], df$isPopulist)
      tibble(dimension = .x, r = test$estimate, p = test$p.value)
    }) %>%
    mutate(sig_bonf = p < bonferroni_alpha)
  print(res)

```

```

res
}

val_all <- run_validation(analysis_eu, "All EU")
val_cee <- run_validation(filter(analysis_eu, region == "CEE"), "CEE")
val_west <- run_validation(filter(analysis_eu, region == "Western"), "Western EU")

# — MODELS —

controls <- "gdp_growth + unemployment + inflation + corruption"
run_models <- function(df, label) {
  cat(label)

  m_dims <- feols(
    as.formula(paste("pervote ~ antagonism_norm + morality_norm + nativism_norm +",
      "sovereignty_norm + leadership_norm +", controls,
      "| countryname + election_year")),
    data = df, cluster = ~countryname)

  m_full <- feols(
    as.formula(paste("pervote ~ fullPopulism_norm +", controls,
      "| countryname + election_year")),
    data = df, cluster = ~countryname)

  m_pop <- feols(
    as.formula(paste("pervote ~ antagonism_norm + morality_norm + nativism_norm +",
      "sovereignty_norm + leadership_norm +", controls,
      "| countryname + election_year")),
    data = filter(df, isPopulist == 1), cluster = ~countryname)

  m_main <- feols(
    as.formula(paste("pervote ~ antagonism_norm + morality_norm + nativism_norm +",
      "sovereignty_norm + leadership_norm +", controls,
      "| countryname + election_year")),
    data = filter(df, isPopulist == 0), cluster = ~countryname)

  wald(m_dims, keep = c("antagonism_norm", "morality_norm",
    "nativism_norm", "sovereignty_norm", "leadership_norm"))

```

```

  cat("\n--- Dimensions model ---\n")
  print(summary(m_dims))
  cat("\n--- Composite model ---\n")
  print(summary(m_full))
  cat("\n--- Populist only (N =", nrow(filter(df, isPopulist == 1)), ") ---\n")
  print(summary(m_pop))
  cat("\n--- Mainstream only (N =", nrow(filter(df, isPopulist == 0)), ") ---\n")
  print(summary(m_main))

  list(dims = m_dims, full = m_full, pop = m_pop, main = m_main)
}

models_all <- run_models(analysis_eu, "ALL EU")
models_cee <- run_models(filter(analysis_eu, region == "CEE"), "CEE ONLY")
models_west <- run_models(filter(analysis_eu, region == "Western"), "WESTERN EU ONLY")

# INTERACTION MODELS (ALL EU)
dim_names <- c("antagonism", "morality", "nativism", "sovereignty", "leadership")
interact_all <- map(dim_names, ~{
  feols(as.formula(paste0("pervote ~ ", .x, "_norm * isPopulist + ",
    controls, " | countryname + election_year")),
    data = analysis_eu, cluster = ~countryname)
}) %>% set_names(dim_names)

cat("\n\n=== INTERACTION MODELS (ALL EU) ===\n")
walk2(interact_all, names(interact_all), ~{
  cat("\n---", .y, "---\n"); print(summary(.x))
})

# REGION INTERACTION (does rhetoric work differently in CEE vs West?)
analysis_eu$isCEE <- as.numeric(analysis_eu$region == "CEE")
m_region_interact <- feols(
  as.formula(paste(
    "pervote ~ (antagonism_norm + morality_norm + nativism_norm +",
    "sovereignty_norm + leadership_norm) * isCEE +", controls,
    "| countryname + election_year"
  )),
  data = analysis_eu, cluster = ~countryname)

```

```

cat("\n\n=== REGION INTERACTION MODEL (dims × CEE) ===\n")
print(summary(m_region_interact))

# Hausman test

# Prepare panel structure
analysis_plm <- analysis_eu %>%
  mutate(panel_id = paste0(countryname, "_", partyname)) %>%
  group_by(panel_id, election_year) %>%
  slice_max(pervote, n = 1, with_ties = FALSE) %>%
  ungroup()

pdata <- pdata.frame(analysis_plm,
  index = c("panel_id", "election_year"),
  drop.index = FALSE)

pdata <- pdata.frame(analysis_plm,
  index = c("panel_id", "election_year"),
  drop.index = FALSE)

# Fixed effects (for Hausman comparison)
fe_plm <- plm(pervote ~ antagonism_norm + morality_norm + nativism_norm +
  sovereignty_norm + leadership_norm +
  gdp_growth + unemployment + inflation + corruption,
  data = pdata, model = "within", effect = "individual")

# Random effects
re_plm <- plm(pervote ~ antagonism_norm + morality_norm + nativism_norm +
  sovereignty_norm + leadership_norm +
  gdp_growth + unemployment + inflation + corruption,
  data = pdata, model = "random", effect = "individual")

# The test
hausman <- phtest(fe_plm, re_plm)
print(hausman)

#Wald test
wald(m_dims, keep = c("antagonism_norm", "morality_norm",
  "nativism_norm", "sovereignty_norm", "leadership_norm"))

# WILD CLUSTER BOOTSTRAP - EU EXTENSION
library(fwildclusterboot)
library(fixest)

```

```

library(tidyverse)

set.seed(8417652)

dqrng::dqset.seed(8417652)

B <- 9999

controls_formula <- "gdp_growth + unemployment + inflation + corruption"

dims_formula <- paste(
  "pervote ~ antagonism_norm + morality_norm + nativism_norm +",
  "sovereignty_norm + leadership_norm +", controls_formula,
  "| countryname + election_year"
)

full_formula <- paste(
  "pervote ~ fullPopulism_norm +", controls_formula,
  "| countryname + election_year"
)

dim_params <- c("antagonism_norm", "morality_norm", "nativism_norm",
  "sovereignty_norm", "leadership_norm")

# BOOTSTRAP module
run_boot <- function(model, param) {
  boot <- boottest(model, param = param, clustid = "countryname",
    fe = "countryname", B = B, type = "webb",
    impose_null = TRUE, p_val_type = "two-tailed")
  tibble(term = param, estimate = boot$point_estimate,
    t_stat = boot$t_stat, p_boot = pval(boot),
    ci_lower = confint(boot)[1], ci_upper = confint(boot)[2])
}

# FULL PIPELINE running a sub-sample
run_wcb <- function(df, label) {
  cat("\n\n=====", label, "=====\n")
  df <- df %>%
    mutate(countryname = as.factor(countryname),
      election_year = as.factor(election_year))

  m_detect <- feols(as.formula(dims_formula), data = df, cluster = ~countryname)
  kept <- obs(m_detect)

```

```

cat("N:", nrow(df), "->", length(kept), "after singleton removal\n")
cat("Clusters:", nlevels(droplevels(df$countryname[kept])), "\n")

df_clean <- df[kept, ] %>%

mutate(countryname = droplevels(countryname),
       election_year = droplevels(election_year))

m_dims <- feols(as.formula(dims_formula), data = df_clean, cluster = ~countryname)
m_full <- feols(as.formula(full_formula), data = df_clean, cluster = ~countryname)

cat("\n--- Dimensions ---\n")

boot_dims <- map_df(dim_params, ~{ cat(.x, " "); run_boot(m_dims, .x) })
cat("\n")

print(boot_dims)

cat("\n--- Composite ---\n")

boot_full <- run_boot(m_full, "fullPopulism_norm")
print(boot_full)

cat("\n--- CRVE vs Bootstrap ---\n")

comparison <- broom::tidy(m_dims, conf.int = TRUE) %>%

filter(term %in% dim_params) %>%

select(term, estimate, se_crve = std.error, p_crve = p.value) %>%

left_join(boot_dims %>% select(term, p_boot, ci_lower, ci_upper), by = "term") %>%

mutate(sig_crve = ifelse(p_crve < 0.01, "***", ifelse(p_crve < 0.05, "*", "")),
       sig_boot = ifelse(p_boot < 0.01, "***", ifelse(p_boot < 0.05, "*", "")))

print(comparison)

list(dims = boot_dims, full = boot_full, comparison = comparison)
}

# RUN

wcb_all <- run_wcb(analysis_eu, "ALL EU")

wcb_cee <- run_wcb(filter(analysis_eu, region == "CEE"), "CEE ONLY")

wcb_west <- run_wcb(filter(analysis_eu, region == "Western"), "WESTERN EU ONLY")

```