Abstract

Why do some politicians employ populist rhetoric and others do not within the same electoral context? To address this question, we present a formal model of populism in which two competing actors decide whether to allocate their effort to conventional campaigning in an attempt to mobilize their supporters or to populist campaigning in an attempt to demobilize their opponent’s supporters. In line with the empirical literature, however, the use of populism is assumed to backfire by demobilizing the candidate’s own support to a various extent. We demonstrate that, despite its potential electoral risks, the politician with the lower level of pre-existing support is more likely to use populism to have a chance of winning. To test the implications of our theoretical model, we construct the most comprehensive corpus of U.S. presidential campaign speeches from 1952 to 2016 and then estimate the prevalence of populist rhetoric across these speeches using a novel automated text analysis method combining supervised learning and advanced natural language processing models. In line with our theoretical expectations, we show the robustly greater use of populist rhetoric among the presidential candidates with the initially lower polling numbers.
Introduction

Over the past two decades, the world has been witnessing the gradual emergence of a “populist zeitgeist” (Mudde, 2004) with populist parties and candidates gaining increasing electoral support across countries. To understand this change, scholars examined various economic and cultural factors that could have potentially increased the demand for populism among voters. Regardless of their preferred explanation of this change, however, many view the supply of populism simply as an opportunistic campaign strategy employed by politicians to exploit widespread grievances. While the increasing demand for populism can explain its electoral success in general, it is still unclear why some politicians are more likely to employ populist rhetoric than others, especially within the same electoral context when the popular demand is arguably fixed. After all, the upsurge of populist rhetoric across regions is evidently concentrated among particular political actors rather than being spread out throughout the political system. For instance, one can easily observe the recent populist campaigns such as of Donald Trump and Bernie Sanders in the 2016 U.S. presidential elections. But one can also observe even a greater number of non-populist campaigns within the same elections. So, when do–and don’t–politicians strategically decide to use populism?

To address this question, we present a formal model of populism as a campaign gamble with the two as-if identical political candidates who decide whether to allocate their effort to conventional or populist campaigning in an attempt to improve their electoral chances. While the use of conventional campaigning is assumed to primarily mobilize additional support, populist campaigning is assumed to primarily demobilize the existing support of the opponent. Importantly, the use of populism can backfire such that there is some chance that it can demobilize the candidate’s own support and mobilize more support for the opponent. We demonstrate theoretically that, despite its potential electoral risks, the candidate with the lower level of pre-existing support is more likely to use populism to have at least some chance of winning.
To test the implications of our theoretical model, we construct the most comprehensive corpus of U.S. presidential campaign speeches from 1952 to 2016 and then estimate the prevalence of populist rhetoric across these speeches. In doing so, we use a novel automated text analysis method which combines random forests, a supervised machine learning method, and Doc2vec, an advanced natural language processing model. Compared to the dictionary method commonly used for measuring populist rhetoric, our method is better at capturing complex language features and the underlying concept. In line with our theoretical expectations, we show the greater use of populist rhetoric among the presidential candidates with the initially lower polling numbers. These results are further robust to a number of alternative specifications.

Our contribution is thus two-fold. Theoretically, we bridge the previously disconnected ideational and game theoretic approaches to the study of populism by providing a formal model of populist rhetoric as a risky campaign game, elucidating when political actors may strategically decide to be populist. Empirically, we provide the most comprehensive estimates of populism across U.S. presidential campaigns, corroborating the intuition of the unprecedented use of populist rhetoric by Donald Trump in 2016 elections. In doing so, we introduce a new algorithm that significantly improves upon the dictionary-based measurements from previous populism research.

**Populism as strategic campaigning**

Its popularity notwithstanding, populism has been a highly contested term. To improve its conceptual clarity and ensure generalizability across contexts, over the last decade a growing number of scholars have gradually adopted the “ideational” definition of populism. The ideational approach views populism as a set of ideas depicting society as divided into two homogeneous and antagonistic groups—the pure people and the corrupt few—and emphasizing that politics should reflect the general will of the people
(Mudde, 2004; Mudde and Kaltwasser, 2018). To that end, this approach constructs a minimalist definition highlighting the shared core of populist ideas and separating the concept of populism from its possible causes and consequences. Since, by design, such conceptualization has little specific ideological or policy content, populist rhetoric can in principle be utilized by politicians and parties from both the left and the right of the ideological spectrum (Gidron and Bonikowski, 2013; Akkerman et al., 2013).

Although the demand for populist rhetoric certainly varies across space and time, many scholars of the ideational approach usually take it as given and explore how this demand is then exploited by opportunistic politicians. In line with this reasoning, it has been increasingly shown that some of the major sources of populism, such as related to anti-immigration sentiments, are rather stable on the individual level and robust to various economic and demographic shocks (for a review, see Kustov et al., 2019).\(^1\) Furthermore, not only these underlying attitudes, but the individual propensity for populist voting itself, may withstand major economic and political crises (Gidron and Mijs, 2019). Although these (non)findings may seem counter-intuitive, they dovetail well with the evidence that populist attitudes are largely a function of stable personality traits such as openness to experience and agreeableness (Bakker et al., 2016), as well as authoritarianism and ethnocentrism (Assche et al., 2019).

But while the ideational literature on populism rightly assumes that opportunistic politicians can strategically exploit stable and widespread popular anti-immigration and anti-elitist sentiments, there is very little elaboration on why populist rhetoric is relatively infrequent or why it is used by some actors more than others even within the same electoral context. Unfortunately, scholars of populism rarely specify what games politicians are actually playing, what other campaign strategies are available, and why populist rhetoric may or may not constitute someone’s best response in equilibrium. One of the biggest omissions in the literature is perhaps the relative neglect of the potential costs incurred by politicians who employ populist rhetoric.

\(^1\)For some evidence of the stability of anti-elitist attitudes, see Motta (2018).
What are the potential electoral benefits and costs of employing populist rhetoric? The often cited “benefit” of populism is the increased turnout of politically disaffected citizens who may find such rhetoric appealing (Huber and Ruth, 2017). Populist rhetoric is thus conventionally viewed as one of the top-down strategies for voter mobilization (Weyland, 2001). These purely mobilizational effects of populism, however, have been recently questioned both theoretically and empirically (e.g., Ardag et al., 2019). Most prominently, populist rhetoric can also be deliberately used to demobilize voters for mainstream parties by amplifying negativity in politics and potentially triggering popular distrust in democracy (see Ansolabehere and Iyengar, 1995). Consistent with this idea, one of the most comprehensive empirical tests of the populism effects on turnout finds that the emergence of successful populist parties may indeed demobilize a substantial share of new voters (Immerzeel and Pickup, 2015).

Immerzeel and Pickup (2015), however, also find that populism may mobilize those who vote for mainstream, non-populist parties. In other words, besides its potential benefits, populist rhetoric can also impose significant costs by repelling those voters who reject the associated moralizing and anti-pluralist views (Hameleers et al., 2019). All in all, similar to findings on negative campaigning (Lau and Rovner, 2009; Krupnikov, 2011), populist rhetoric does not appear to be always effective at (de)mobilizing voters, especially when potential alternatives are considered (Bornschier, 2017). Accordingly, while many scholars try to explain the relative success of radical parties by emphasizing the electoral effectiveness of populism, this raises the follow-up question of why all politicians would not use populist rhetoric even more often.

Meanwhile, the existing game theoretic treatment of populism has so far been quite disconnected from these conceptual and empirical debates by treating supply-side populism as a “distorted” left-wing ideology (Acemoglu et al., 2013), anti-institutionalism (Acemoglu et al., 2013), or low experience and high charisma (Serra, 2018). While insightful, these formal conceptualizations of populism do not speak to the latest the-
oretical understanding of populism as a strategic campaign rhetoric as defined by the ideational approach. These models are thus not well-suited for explaining why political actors of similar ideology may differ in their use of populist rhetoric while facing the same demand (or similarly “populist-friendly” electorate) within the same electoral context. In the same vein, the models that treat populism—similar to ideology—as a stable characteristic of political actors also fail to explain why the same politicians may differ in their use of populist rhetoric under different electoral circumstances.

To bridge the divide between these two literatures, our theoretical model of populist rhetoric instead builds on the earlier game theoretic literature exploring the use of various campaign strategies. As elaborated above, especially relevant for our purposes are the models of negative campaigning such as that by Skaperdas and Grofman (1995). We believe models of negative campaigning are in particular fruitful in elucidating the strategic logic of gloom-ridden populist rhetoric as a campaign gamble within a certain electoral context. In line with this idea, empirical literature confirms the pervasive negativity of populist campaigning (Nai, 2018).

Model of populism as a campaign gamble

Building on Skaperdas and Grofman (1995), our model assumes a standard office-seeking two-candidate race where each candidate has a certain level of pre-existing electorate support among the decided (or henceforth “mobilized”) voters, while the rest of the (potential) electorate is undecided (or henceforth “unmobilized”). We further stipulate that candidates can use a combination of “conventional” and “populist” campaign rhetoric to improve their electoral chances. While conventional rhetoric is assumed to help mobilize additional support among the unmobilized, populist rhetoric is assumed to demobilize the opponent’s pre-existing support. The use of populism, however, can backfire such that there is some chance that it can demobilize the can-
candidate’s own pre-existing support (or, equivalently, mobilize more votes for the other candidate). Overall, we show that, despite the potential risk, the ex-ante losing candidate is more likely to use populism to have at least some chance of winning.

Our basic model of populist rhetoric as a campaign gamble is of imperfect information. There are two candidates (or parties), $A$ and $B$. Both candidates observe each others’ level of pre-existing support $\alpha_i$ and the share of unmobilized electorate $\bar{\alpha} = 1 - \alpha_A + \alpha_B$. Then, each candidate simultaneously decides to allocate its effort to populist ($p_i$) or non-populist, conventional ($c_i$) rhetoric so that $p_i + c_i = 1$.

While political campaigning can have a number of aims including changing voter’s preferences over candidates, we assume that the primary function of conventional campaign rhetoric is mobilizing electoral support among the (currently) unmobilized. Put formally, let $m^A(c_A,c_B)$ and $m^B(c_A,c_B)$ indicate the share of unmobilized electorate ultimately attracted by candidates $A$ and $B$ such that, for any given combination pair of conventional campaign strategies pursued by both candidates, $m^i(c_i,c_j)$ is increasing in $c_i$ and decreasing in $c_j$. To that end, we also assume that all of the unmobilized are equally and ultimately susceptible to mobilization by either candidate such that $m^A(c_A,c_B) = m^B(c_A,c_B)$ and $m^A(c_A,c_B) + m^B(c_A,c_B) = 1$. As a result, the candidates attract the same share of the unmobilized electorate when they decide to allocate the same amount of effort to conventional campaigning. Since the function $m$ is symmetric ($m^i(c_i,c_j) = 1 - m^j(c_j,c_i)$), we can simply denote $m^A(c_A,c_B)$ by $m$ and $m^B(c_A,c_B)$ by $1 - m$. Finally, we assume that conventional campaigning has diminishing returns (so that $m$ is a concave function: $d^2m/dc_i^2 \leq 0$).

Unlike conventional campaigning to attract the unmobilized, we assume that the primary function of populist rhetoric is demobilizing the opponent’s pre-existing support. In line with some of the empirical literature described above, however, we also assume that populist campaigning can backfire by demobilizing the candidate’s own current supporters. Put formally, for any given combination pair of populist campaign...
strategies $p_i$ pursued by both candidates, let $\alpha_i(p_i + E p_j)$ be the resulting decrease in pre-existing support share for candidate $i$, where $E$ indicates the relative effectiveness of populist campaigning (or to what extent the opponent is hurt more than the candidate). In our base model, we assume that $E > 1$. Importantly, at least in terms of the relative electoral advantage, the backfire effect of populist campaigning that demobilizes one’s own support is equivalent to the one that mobilizes the electorate to vote for the opponent. In other words, although the model focuses on populism as primarily a tool for demobilization, the main distinctive feature of populist rhetoric is ultimately assumed to be its greater riskiness (relative to non-populist rhetoric).

We can now summarize the final overall support that each candidate gets after deciding on their use of conventional and populist rhetoric. To simplify, given that $p_i + c_i = 1$, we can represent the resulting support ($\alpha_i'$) as just a function of $c_i$:

$$\alpha_i' = \alpha_i + \alpha_m - \alpha_i(1 - c_i + E(1 - c_j)) \quad (1)$$

Similar to other campaign strategy literature, we assume that candidates ultimately care about maximizing their winning margin or electoral advantage. Consequently, given equation 1, we can define the utility function for each candidate as follows:

$$u_i(c_i, c_j) = \alpha_i' - \alpha_j' = \pi(2m(c_i, c_j) - 1) - \alpha_i(1 - c_i + E(1 - c_j)) + \alpha_j(1 - c_j + E(1 - c_i)) + \alpha_i - \alpha_j \quad (2)$$

After formulating the strategic form of our game, we can now proceed with determining the possible Nash equilibria. We can say that a campaign strategy pair $(c_A^*, c_B^*)$ is an equilibrium if $u_i(c_i^*, c_j^*) \geq u_i(c_i, c_j)$ for all $c_i, i \neq j$. Let $u_i' = du_i(c_i, c_j)/dc_i$, $m_i' = dm/dc_i$, and assume that $m_i'' = d^2m/dc_i^2 \leq 0$. Then, we can find the first derivative and characterize the marginal benefits of putting extra effort into conventional and populist campaigning as follows:
\[ u'_i(c_i, c_j) = \alpha 2m'_i - (E\alpha_j - \alpha_i). \tag{3} \]

We can then similarly derive \( u''_i = \alpha 2m''_i \leq 0 \). Conditional on the assumptions above being satisfied, we can now show that candidates’ utility function \( u_i \) is concave in their own strategy \( c_i \) and thus that there exists a pure-strategy Nash equilibrium.

Given equation 3, both candidates would only devote a non-zero effort to both campaign strategies \((0 < c^*_A < 1 \text{ and } 0 < c^*_B < 1)\) if and only if their marginal benefits and costs are equalized. Quite naturally, this implies that \( E\alpha_A > \alpha_B \) and \( E\alpha_B > \alpha_A \) simultaneously, which necessarily requires that the derivatives in equation 3 are equal to zero for both candidates:

\[
\alpha 2m'_A(c^*_A, c^*_B) - (E\alpha_B - \alpha_A) = 0 \implies m'_A(c^*_A, c^*_B) = (E\alpha_B - \alpha_A)/(2\pi) \\
-\alpha 2m'_B(c^*_A, c^*_B) - (E\alpha_A - \alpha_B) = 0 \implies -m'_B(c^*_A, c^*_B) = (E\alpha_A - \alpha_B)/(2\pi) \tag{4}
\]

Now suppose that one of the candidates has more pre-existing support \( \alpha_A > \alpha_B \). We can then show that \( E\alpha_A - \alpha_B > E\alpha_B - \alpha_A \) and, by equation above, \( m'_A(c^*_A, c^*_B) < -m'_B(c^*_A, c^*_B) \) so that \( m(c^*_A, c^*_B) > 1/2 \) under our assumptions. In turn, this is equivalent to \( c^*_A > c^*_B \) or \( p^*_A < p^*_B \), which gives us our main result: “the candidate with a lower pre-existing support is expected to use more populist campaign rhetoric relative to his opponent” (Proposition 1).

In addition to this general result it may also be instructive to examine two special cases where one of the candidate allocates all effort to either conventional or populist campaigning \((c^*_i \text{ is equal to } 0 \text{ or } 1)\). First, suppose that the pre-existing support is lower for one of the candidates \((\alpha_i > \alpha_j)\) and that the effectiveness of populist rhetoric is relatively low \((E\alpha_j \leq \alpha_i)\). Then, in line with equation 3, \( u'_i(c_i, c_j) > 0 \) given that \( \alpha_i \geq E\alpha_j \). Consequently, candidate \( i \) would only do conventional campaigning in
equilibrium \((c^*_i = 1\) is the optimal choice regardless of \(c_j\)). Second, consider a function \(m\) with a finite derivative \(m'_i(0, c_j)\) and sufficiently low \(\alpha_i\) (or sufficiently high \(E\)). Then, regardless of \(c_j\), it must be true that \(u_i(0, c_j) = \overline{\alpha}m'_i(0, c_j) - (E\alpha_j - \alpha_i) \leq 0\).

In other words, candidate \(i\) would only do populist campaigning in equilibrium \((c^*_i = 0\) is the optimal choice regardless of \(c_j\)). In sum, although this is less realistic than the general proposition 1, *if the candidate’s pre-existing support is sufficiently low (high) or populist rhetoric is sufficiently (in)effective, then the candidate is expected to fully engage in populist (conventional) campaigning.* Importantly, the results hold even if we introduce some uncertainty about the (in)effectiveness of populism and relax the assumption that \(E > 1\), i.e., that populist rhetoric is necessarily hurting the opponent more than the candidate instigator of such rhetoric (not shown).

**Empirical strategy and design**

**Populist rhetoric in U.S. presidential campaigns**

There can be several ways to test the empirical implications of our model, but the well-documented universe of U.S. presidential campaigning with normally two major contestants arguably provides a great case for these purposes. Although populism has been largely explored as a prominent feature of Latin American and European politics, there has been a growing attention to populism in the United States (Hawkins and Littvay, 2019). This has been especially true since the recent rise of the Tea Party and, subsequently, Donald Trump as a part of the Republican party, as well as Bernie Sanders as a part of the Democratic party. Indeed, as Lee (2019) recently argued, while the U.S. political system offers much less opportunity for organized populist parties, it still provides ample opportunities for populist candidacies. In line with this, Bonikowski and Gidron (2015) confirm that populist rhetoric has historically been a rather common feature of U.S. presidential campaigning across all parties.
Data

Despite the growing interest in studying populism in U.S. presidential elections, existing research has either focused on the recent elections (see Hawkins and Littvay, 2019; Lacatus, 2019) or more historical cases (Bonikowski and Gidron, 2015). We expand the scope of this research by building a comprehensive U.S. presidential campaign corpus of 4,324 speeches from 1952-2016. The speeches are collected from two major data sources: The Annenberg/Pew Archive of Presidential Campaign Discourse (Annenberg, 2000) and The American Presidency Project (Woolley and Peters, 2008). The Annenberg/Pew Archive of Presidential Campaign Discourse includes transcripts of campaign speeches delivered by the Democratic and Republican presidential nominees between September 1st and the election day, as well as their nomination acceptance speeches. Overall, it covers 12 elections and 21 presidential campaigns from 1952 to 1996 with 2,406 speeches which have been previously used to examine populist rhetoric by Bonikowski and Gidron (2015). The American Presidency Project is an online database of presidential documents hosted at the University of California, Santa Barbara. We use the American Presidency Project to expand on the Annenberg/Pew Archive data by adding 5 most recent elections from 2000 to 2016 and incorporating all speeches delivered during presidential campaigns from January to the election day.

After utilizing our original machine learning algorithm to measure populist rhetoric in all of these presidential speeches (see below), we merge these estimates with the monthly polling results for all major candidates throughout their respective campaigns (Gallup, 2013). We use the monthly polling results to measure electoral advantage by coding whether and what extent each major candidate was leading in the polls across campaign months. While our speech data and populism measurement include speeches and populism score by all candidates, we only include the final candidates.

1The polling results include the earliest possible candidate polling for each month. For some campaign months with no available polling, we impute an estimate based on the closest available polls. Our results are robust to the complete exclusion of these extrapolated monthly estimates (see Table A3).
from the two parties to test our theoretical model, which results in 3,435 speeches. In addition to the level of populism rhetoric and electoral advantage, we also include several candidate and speech characteristics in our data, such as party membership, party incumbency and speech length. For summary statistics of all major variables, see Table A1 in Appendix.

**Measurement of populist rhetoric**

Populist rhetoric in politician speeches has previously been measured using either human-coded content analysis (e.g., Jagers and Walgrave, 2007; Hawkins, 2009; Hawkins and Kaltwasser, 2018) or dictionary-based automated text analysis (e.g., Rooduijn and Pauwels, 2011; Bonikowski and Gidron, 2015). Human-coded methods generally have a high content validity, but they are also costly and time-consuming. Given that we have over 4,000 lengthy speeches, we use automated text analysis to code populist rhetoric. Instead of using the common dictionary-based method, however, we use a novel method that combines random forests, a supervised machine learning method, and Doc2vec, a word embedding method. As we demonstrate below, this new method is potentially better at capturing complex language features, as well as the necessary and sufficient conditions of the underlying theoretical concept.

In conceptualizing populism, we follow its definition by the ideational approach as a discursive style that portrays society in moral terms as being divided into two homogeneous and antagonistic groups (the “good” people and the “corrupt” few) while emphasizing that politics should reflect the general will of the people (Mudde and Kaltwasser, 2018). A text will thus be considered populist if and only if it (1) recognizes the people as the only legitimate source of power (people-centric); (2) creates separation between us/people and them/enemy (anti-pluralist); and, in doing so, (3) stipulates the separation of us and them on moral grounds (good versus evil) (Dai, 2019).
Because the average speech length is 2,281 words and populist rhetoric may only appear in some speech parts, we divide all speeches into sub-speeches of 10 paragraphs, resulting in 17,453 sub-speeches. Dividing the speech data into sub-speeches also improves the performance of our classification model, because there is less noise and irrelevant information on the sub-speech level relative to the much longer speech level. Therefore, we treat each sub-speech as a document for our supervised classification task. We classify sub-speech documents as populist (as opposed to non-populist) if and only if they contain all of the necessary and sufficient conditions described above. After applying the supervised learning algorithm, each sub-speech document has a score of either 1 (populist) or 0 (non-populist). We then aggregate this sub-speech level classification into a speech level measurement of populism by simply calculating the proportion of populist sub-speeches in a given speech.

As in any supervised approach, there are three steps in our classification method (D’Orazio et al., 2014). First, we hand-code a small sample of the corpus. We randomly select 65 out of 4,363 speeches and hand-code all of the 419 sub-speeches in the selected speeches. 62 (14.8%) sub-speeches are categorized as populist. The same set of 69 sub-speeches are coded by a second coder to evaluate inter-coder reliability. 388% of the time the 2 coders agree with each other: 12 sub-speeches are coded as populist by both coders, 49 sub-speeches are coded as non-populist by both coders, and the two coders only disagree with each other on 8 sub-speeches. The two coders’ hand coding highly correlate with each other: the Pearson correlation between the two coders’ hand coding is 0.72. The inter-coder reliability is similar to the inter-coder reliability reported by other papers using hand-coding method to code populist rhetoric (Hawkins, 2009; Hawkins and Silva, 2018).

Second, we train a classification algorithm that learns the rules from the hand-
coded sample in the first step to predict the (non-)populist document class as close as possible to the human coder. In vectorizing the documents and words, we use Doc2vec, an advanced neural network based on the word embedding model from Natural Language Processing (Le and Mikolov, 2014). Unlike the common “bag-of-words” approach, which represents documents using the simple counts of as-if independent words, Doc2vec learns to maintain the semantic and syntactic relationships by vectorizing words and documents in a dense vector space. As a result, words similar in their meaning and documents similar in their contexts are positioned close to each other in the vector space.\(^4\) Furthermore, Doc2vec reduces the high dimensionality of the raw text data and significantly improves the model performance relative to the same classification algorithm that uses a document-term matrix constructed by the “bag-of-words” approach. After vectorizing all documents, we train a random forest classifier to separate populist and non-populist documents. To evaluate the classifier and to avoid over-fitting, we use cross validation and only train the model on 80% of the hand-coded data (training set) and test it on the rest of the sample (test set).\(^5\) Overall, our model achieves an 87% accuracy in the test set (i.e., making the same prediction as the human coder).

Because populist rhetoric is a rare event in our data that only 14% of sub-speeches are coded as populist in the training data, we also plot the ROC curve and examine the corresponding area under the curve (AUC) in Figure 1. As can be seen from the plot, our model achieves a high AUC of 0.9 while suffering only a minor false positive rate to obtain a high true positive rate in out-of-sample prediction. For example, to correctly identify 80% of populist documents, our model makes only less than 10% of false positive predictions. In other words, less than 10% of the documents that are predicted as populist are not populist.

\(^4\)For more details on the algorithm and how we trained our Doc2vec model, see Online Appendix B.

\(^5\)During the training phrase, we also use probability calibration with 2-fold cross-validation, i.e., the training set is further split into two folds and probabilities for each fold are then averaged for prediction.
Figure 1: Out-of-sample Performance of Classification Model

Note: The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) in out-of-sample prediction for all thresholds. The TPR captures the proportion of populist documents correctly classified as populist, while the FPR captures the proportion of out-of-sample documents that were incorrectly classified as populist. A ROC curve for a classifier that perfectly predicts class membership (populist and non-populist) would form a 90 degree angle in the upper left corner of the plot—the TPR would be 1 and the FPR would be 0. A classifier with no predictive power, one that performs randomly, would have a ROC curve that follows the diagonal dashed gray line in Figure 1.

After training the classifier in the second step, the final step is simply applying the classifier to the full corpus in order to predict whether each sub-speech is populist or not. A total of 623 sub-speeches (4%) are classified as being populist. To better illustrate the model performance or what a populist sub-speech may look like, below we discuss two randomly drawn sub-speeches from either party which were classified as populist by the algorithm (outside of the hand-coded sample). Because the sub-speeches are still quite long, we only include the highlighted parts of these sub-speeches for demonstration. In the first example of Barack Obama’s campaign speech in 2008, he creates a separation between Main Street (us) and Wall Street (them). While the Main Street is innocent, Wall Street is greedy, irresponsible, and corrupted. Furthermore, Wall Street is the reason behind the economic suffering of the Main Street. To that end, Obama also claims that, while he represents the millions of innocent people, the political establishment in Washington represents special interests.
Barack Obama 2008: Third, I said that we cannot and will not simply bailout Wall Street without helping the millions of innocent homeowners who are struggling to stay in their homes. ... I said that I would not allow this plan to become a welfare program for the Wall Street executives whose greed and irresponsibility got us into this mess. ... We don’t just need a plan for bankers and investors, we need a plan for autoworkers and teachers and small business owners. ... That means taking on the lobbyists and special interests in Washington. That means taking on the greed and corruption on Wall Street ... It is time to reform Washington. (Remarks in Detroit, Michigan. September 28th, 2008.)

Similarly, in Donald Trump’s 2016 campaign speech he creates a moralized separation between us (the American people) and them, while claiming to represent all Americans. In his narrative, the American people have been failed by the corrupt status quo. While the exact identity of the corrupt few here is somewhat ambiguous, it includes all the people who disagree with his campaign and policies because of their vested interests.

Donald Trump 2016: Change is coming. All the people who’ve rigged the system for their own personal benefit are trying to stop our change campaign because they know that their gravy train has reached its last stop. It’s your turn now. This is your time ... We are fighting for all Americans ... who’ve been failed by this corrupt system. We’re fighting for everyone who doesn’t have a voice. Hillary Clinton is the candidate of the past. Ours is the campaign of the future. In this future, we are going to pursue new trade policies that put American workers first – and that keep jobs in our country ... The era of economic surrender is over. (Remarks at a Rally at the Pensacola Bay Center in Pensacola, Florida. September 9th, 2016.)

So far, our measurement of populism has been at the sub-speech level: a sub-speech is either coded/predicted as being populist with a score of 1 or not being populist with a score of 0. To measure the level of populist rhetoric at the speech level, we then calculate the proportion of populist sub-speeches in a given speech. Similarly, we calculate the proportion of populist sub-speeches among candidate speeches to create the level of populist rhetoric at the candidate level.
Compared to the dictionary-based approach previously used to measure populist rhetoric (e.g., Bonikowski and Gidron, 2015; Rooduijn and Pauwels, 2011), our measurement has several advantages. The dictionary-based method measures populist rhetoric by counting the words associated with populism in each document using a pre-defined dictionary. First, this essentially word-level measurement is based on the “bag-of-words” assumption and cannot take the words’ meaning and contexts into account. A populist speech with similar but not identical phrases as in the populism dictionary would be categorized as non-populist. In fact, our previous example from the Trump’s 2016 speech would not be considered populist by the dictionary approach such as in Bonikowski and Gidron (2015), because it does not contain any words from their populism dictionary. It can be easily seen, however, that many phrases in that Trump’s speech are synonymous with the words in their dictionary. Instead of “special interest”, Trump uses “personal benefit”; and instead of “forgotten Americans”, Trump claims to fight for “everyone who doesn’t have a voice.” With the Doc2vec component, our method is able to learn the semantic relationships between such words and phrases. Therefore, our model will still be able to make accurate prediction when a document uses similar but not identical words as in the populist documents used to train the model.

Second, the dictionary-based method assumes that all dictionary words have equal importance in measuring the level of populist rhetoric, which cannot capture the idea of necessary and sufficient conditions in defining populism. As a result, a non-populist speech with many references to “the people” without criticizing the corrupt few would have a high populist score. In contrast, our decision tree based random forest algorithm can learn complex non-linear relationships between the document features and its class, which resembles the necessary and sufficient conditions more closely.\footnote{For example, when a document mentions “American people” a lot, the algorithm will not immediately give it a high probability for being populist. Instead, it will trigger the next decision rule, such as whether the document also uses phrases like “special interests” and “betrayed.” For a more detailed and accurate explanation of random forests, see Online Appendix C.}
Analysis and results

We start by presenting our measure descriptively and verifying some of the previous stylized findings on populist rhetoric in U.S. presidential elections (Bonikowski and Gidron, 2015; Hawkins and Kaltwasser, 2018; Lacatus, 2019) using our expanded data and new methodology. In Figure 2, we show the average annual level of populist rhetoric from 1952 to 2016 based on our measure and the average level of populist rhetoric from 1952 to 1996 from Bonikowski and Gidron (2015). To match and to compare with Bonikowski and Gidron (2015), we create the annual average level of populist rhetoric by calculating the proportion of populist sub-speeches in a year. As can be seen, both measures have similar trends between 1952 and 1996 with 1972 and 1988 as the most populist elections while 1960 and 1980 as the least populist elections.

Figure 2: Average Share of Populist Rhetoric across Years

Note: The bar plot indicates the average share of populist sub-speeches in U.S. presidential campaigns across years based on our measure.

Since we have a sub-speech level measure and Bonikowski and Gidron (2015) measure is at word level, the two measures are on different scales. To make it more comparable, their measure is re-scaled to have the same lower and upper limit as our 1952 to 1996 measure in Figure 2.
In Figure 3, we show the average level of populist rhetoric by candidate and campaign. Similarly as Bonikowski and Gidron (2015), between 1952 and 1996 the top three most populist Democratic campaigns are the 1988 Dukakis campaign, the 1972 McGovern campaign and the 1992 Clinton campaign. The top two most populist Republican campaigns from Bonikowski and Gidron (2015), the 1952 Eisenhower campaign and the 1996 Dole campaign, are also the most populist Republican campaigns between 1952 and 1996 based on our measure. At the same time, based on our measure the 1988 Bush campaign and 1992 Bush campaign utilize more populist rhetoric than the 1968 Nixon campaign, the third most populist Republican campaign in Bonikowski and Gidron (2015).

Figure 3: Average Share of Populist Rhetoric across Campaigns

Note: The plot indicates the average share of populist sub-speeches across U.S. presidential campaigns based on our measure.

Finally, it is important to note the extremely high usage of populist rhetoric in 2016 elections by the Republican candidate Donald Trump. While this is in line with the intuition of many political observers, our research provides the first comparative quantitative assessment of his outlier populist status in U.S. general elections. Of
course, our theoretical model could not predict the sheer extent of Trump’s populism, but it is in line with our general expectations based on his lower support initially and throughout the campaign.

**Populist rhetoric as a function of electoral advantage**

The main proposition derived from our model predicts that candidates are more likely to use populist rhetoric when they are confronted with a lower pre-existing support (or electoral disadvantage) relative to their opponent. While most of the variation in electoral advantage is across candidate-years, there have been significant monthly fluctuations in support within each election. Therefore, as an initial test of our theory, we visualize the average share of populist rhetoric in speeches depending on whether the speech was given by a candidate who was leading in the most recent poll in Figure 4. As can be seen, campaign speeches are indeed generally more likely to contain populist rhetoric under electoral disadvantage.

**Figure 4: Electoral Advantage and Populist Rhetoric in U.S. Presidential Speeches**

\[\text{Note: The figure indicates the average share of populist rhetoric in speeches depending on whether a certain presidential candidate is experience an electoral advantage in the most recent polls. The bars indicate 95\% confidence interval. For details, see Table 1 (1).}\]

\[8\text{Since we do not explicitly model the temporal dynamic within campaigns, we replicate our analysis by ignoring the monthly variation with no change in the substantive results (see Table A3).}\]
To further test our theory, we estimate and report the results from four OLS regressions with different specifications in Table 1. Model 1 includes the Electoral Advantage only, while Model 2 to 4 also control for variables that are likely to influence the use of populist rhetoric based on previous studies. Party incumbency is a dichotomous variable with 1 indicating that candidate is from the incumbent party and 0 otherwise. Speech length is standardized to be between 0–1. Given that some candidates and elections appear to be much more populist than others as seen in the descriptive analysis, we also include year, month, and candidate fixed effects in Model 3 and 4. Overall, the results reported in Table 1 provide strong support for our theory: the speeches of winning candidates contain almost twice less populist rhetoric, which also translates into a substantial difference of nearly two standard deviations. This difference persist even after taking account a number of important characteristics including incumbency, partisanship, as well as speech length and month. Furthermore, the expected electoral advantage can even explain variation within a particular campaign after including candidate and year fixed effects.

Robustness checks

Our main findings are robust to a number of additional tests with no change in their substantive interpretation. First, we consider a polling advantage in percentage points as an alternative, continuous indicator of pre-existing support (Table A2). Second, given that we do not model the temporal dynamics in electoral campaigns, we omit monthly variation in polling support and only consider the average use of populist rhetoric as a function of one’s advantage in the first available campaign poll (Table A3). Since it is unlikely that a candidate’s decision to use populism throughout the campaign can cause poor polling performance early in the campaign, such analysis also
Table 1: Populist Rhetoric as a Function of Electoral Advantage

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Advantage</td>
<td>$-0.018^{***}$</td>
<td>$-0.018^{***}$</td>
<td>$-0.018^{***}$</td>
<td>$-0.016^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Party Incumbency</td>
<td>$-0.022^{***}$</td>
<td>$-0.023^{***}$</td>
<td>$-0.007$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Partisanship (GOP)</td>
<td>0.004</td>
<td>0.004</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.018)</td>
<td></td>
</tr>
<tr>
<td>Speech Length</td>
<td></td>
<td>0.026</td>
<td></td>
<td>$-0.064^{***}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.015)</td>
<td></td>
<td>(0.014)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.031^{***}</td>
<td>0.040^{***}</td>
<td>0.032$^*$</td>
<td>$-0.037^*$</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.016)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Month FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Candidate FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* p<0.05; ** p<0.01; *** p<0.001 (two-tailed)

Note: All models are OLS regressions of the average share of populist rhetoric in U.S. presidential speeches on electoral advantage in the recent monthly polls and other candidate or speech characteristics. The standard errors are given in parentheses.

Helps addressing the issue of reverse causality. Third, since the variation in our main independent variable of electoral advantage primarily comes from candidates (and, to a less extent, candidate-year-months), we also replicate our analysis at the monthly level—as opposed to the speech level as in previous analyses (Table A4). Fourth, we constrain our analysis to three last months of the campaign, which excludes all primary campaign speeches with potentially different electoral incentives. Finally, our results are robust to the exclusion of particular candidates and elections (not shown).

Nonetheless, we cannot eliminate the possibility of omitted variable bias, such as related to unobserved candidates’ characteristics affecting both their electoral performance and rhetorical choices.
Discussion

Many people appear to believe that politics is about the righteous Manichean fight between “the good people” and “the corrupt few.” Similarly, various anti-elitist and anti-pluralist sentiments are rather widespread in the electorate. As a result, scholars often use this popular demand of populism to explain the rise and fall of populism: politicians can exploit these sentiments in an opportunistic fashion to win elections when the demand is particularly high. However, the demand-side explanations cannot explain the puzzle of why populist rhetoric is not omnipresent when the demand is rather fixed in the same election. Of course, populism is more credible for some than others. It is also only one possible rhetorical device and one set of popularly appealing ideas among many (Hawkins and Rovira Kaltwasser, 2017). Nonetheless, as of now the ideational approach can hardly explain the likely strategic rather than just principled non-use of populist rhetoric among the majority of political actors most of the time. To remedy this omission, we introduce a new formal model of populism as a campaign gamble and argue that politicians are more likely to employ populism under the condition of (as-if) exogenous electoral disadvantage. We then corroborate the empirical implications of this proposition by measuring populist rhetoric in U.S. presidential campaign speeches using an original supervised machine learning algorithm and modeling the use of populism as a function of the initial polling results.

By formally defining the potential costs and benefits of populist rhetoric in terms of voter (de)mobilization, our model of populism provides a fruitful way to explain when political actors may decide (not) to be populist as a part of the empirically testable equilibrium candidate strategies. The model follows a minimal conceptualization of populism that is independent of politicians’ attributes and policy positions which can bridge the previously disconnected ideational and game theoretic approaches to the study of populism. At the same time, our further empirical examination of populism in U.S. presidential rhetoric allows testing this and other related theories by deriving
precise speech-level estimates of populist rhetoric. In doing so, we improve upon the dictionary-based methods from previous research (Bonikowski and Gidron, 2015) by introducing a new, more advanced measurement algorithm that takes into account the words’ contextual meaning. In turn, this gives us comprehensive, comparable measures of populism across U.S. presidential campaigns, corroborating the intuition of the likely unprecedented use of populist rhetoric by Donald Trump in 2016 elections.

Of course, our research is not without limitations. First, our model can only speak to a limited set of considerations regarding the strategic use of populism. For instance, we do not address why there is a high popular demand for populist rhetoric or, relatedly, why it can be sometimes effective in the first place. In that sense, we cannot explain variation in populist rhetoric that is unrelated to electoral support such as based on candidates’ personality or electoral institutions. Second, we cannot explain why candidates may choose to use other non-ideological types of political rhetoric such as related to clientelism (Hawkins and Rovira Kaltwasser, 2017). Third, while our new measurement method is able to achieve high accuracy in out-of-sample predictions in the training phrase, the training data is relatively small compared to the full corpus. In turn, this can potentially lead to less precise predictions and more false negatives than currently assumed.10 Finally, despite a large number of analyzed speeches, the available variation in candidates’ polling results across campaign months is rather limited. However, none of these limitations likely challenge our main result regarding the greater strategic use of populist rhetoric among the initially losing candidates.

While further theoretical and empirical additions are beyond the scope of this paper, one can easily expand on our results in the future. When it comes to the theory, the model can be generalized to multiple actors and time periods, which would allow to derive hypotheses about the use of populism rhetoric across a variety of election types throughout the entire campaign. It would be also useful to examine the role

10Our measurement can also be easily improved, however, by annotating a bigger training sample.
of mobilization as opposed to persuasion, as well as the uncertainty of populism effectiveness in more detail. Finally, one can complement our account by considering alternative strategic, formal conceptualizations of populism that are still in line with the ideational approach. For instance, it may be fruitful to model populism as a special symbolic type of turnout and vote buying (Nichter, 2008).

When it comes to further empirical implications, one of our follow-up studies aims at analyzing campaign rhetoric in U.S. congressional races, which could provide a much larger and arguably more comparable sample within the same electoral cycle. Combined with a multi-player model extension, one can also consider the use of populist rhetoric in primary elections, as well as two-party and multi-party elections outside of the United States. We believe it may be especially fruitful to devise empirical comparisons of the strategic (non-)use of populist campaign rhetoric across countries and different election types. In doing so, one can build on some of the recent research expanding text analysis methods to non-English languages and to multi-lingual corpora (Dai, 2019; Dai and Radford, 2018).

References


Le, Q. and T. Mikolov (2014). Distributed representations of sentences and documents.


Online Appendix A:

Tables and Figures

Table A1: Summary Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Pctl(25)</th>
<th>Pctl(75)</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Populism Rhetoric (Average Share)</td>
<td>0.02</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Electoral Advantage (Binary)</td>
<td>0.53</td>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Electoral Advantage (Percent)</td>
<td>0.44</td>
<td>6.40</td>
<td>−29.50</td>
<td>−4.00</td>
<td>5.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Party Incumbency</td>
<td>0.46</td>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Party Membership (Republican)</td>
<td>0.42</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Speech Length (Standardized)</td>
<td>0.19</td>
<td>0.12</td>
<td>0.00</td>
<td>0.10</td>
<td>0.26</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Full Speech-level Data (n = 3,435). The mean speech length is 2,167 words.
Table A2: Populist Rhetoric as a Function of Electoral Advantage (Continuous)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Advantage (Percent)</td>
<td>-0.001*</td>
<td>-0.0004</td>
<td>-0.0003</td>
<td>-0.002**</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Party Incumbency</td>
<td>-0.021***</td>
<td>-0.022***</td>
<td>-0.009</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Partisanship (GOP)</td>
<td>0.007†</td>
<td>0.007†</td>
<td>0.053*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Speech Length</td>
<td></td>
<td></td>
<td>0.025†</td>
<td>-0.065***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.015)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.022***</td>
<td>0.029***</td>
<td>0.021</td>
<td>-0.049**</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.015)</td>
<td>(0.018)</td>
</tr>
</tbody>
</table>

Month FE | No | No | Yes | Yes  
Year FE  | No | No | No  | Yes  
Candidate FE | No | No | No  | Yes  

All models are OLS regressions of the average share of populist rhetoric in U.S. presidential speeches on electoral advantage in the most recent monthly polls and other characteristics. The standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.
Table A3: Populist Rhetoric as a Function of Electoral Advantage (Initial Polls Only)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Advantage</td>
<td>−0.008*</td>
<td>−0.010**</td>
<td>−0.011**</td>
<td>−0.013***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Party Incumbency</td>
<td>−0.021***</td>
<td>−0.023***</td>
<td>−0.021***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>Partisanship (GOP)</td>
<td>0.009*</td>
<td>0.010**</td>
<td>0.007*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>Speech Length</td>
<td></td>
<td>0.032*</td>
<td>−0.034*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.025***</td>
<td>0.032***</td>
<td>0.027***</td>
<td>0.022***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Year FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Candidate FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>3,487</td>
<td>3,487</td>
<td>3,487</td>
<td>3,487</td>
</tr>
</tbody>
</table>

All models are OLS regressions of the average share of populist rhetoric in U.S. presidential speeches on electoral advantage in the initial campaign polls and other candidate or speech characteristics. The standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.
Table A4: Populist Rhetoric as a Function of Electoral Advantage (Candidate-Month)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Advantage</td>
<td>−0.022*</td>
<td>−0.020*</td>
<td>−0.021**</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Party Incumbency</td>
<td>−0.021*</td>
<td>−0.026**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>Partisanship (GOP)</td>
<td>−0.003</td>
<td>−0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.037***</td>
<td>0.047***</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Year FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>189</td>
<td>189</td>
<td>189</td>
</tr>
</tbody>
</table>

All models are OLS regressions of the average share of populist rhetoric among U.S. presidential candidates on electoral advantage in the most recent monthly polls and other characteristics. The standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.
Table A5: Populist Rhetoric as a Function of Electoral Advantage (Sep.-Nov. Only)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electoral Advantage</td>
<td>−0.020***</td>
<td>−0.019***</td>
<td>−0.019***</td>
<td>−0.037**</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Party Incumbency</td>
<td>−0.019***</td>
<td>−0.020***</td>
<td>−0.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>Partisanship (GOP)</td>
<td>0.005</td>
<td>0.005</td>
<td>0.038</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>Speech Length</td>
<td></td>
<td></td>
<td></td>
<td>−0.042**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.017)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.031***</td>
<td>0.037***</td>
<td>0.034***</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Month FE</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Candidate FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>2,777</td>
<td>2,777</td>
<td>2,777</td>
<td>2,777</td>
</tr>
</tbody>
</table>

All models are OLS regressions of the average share of populist rhetoric in U.S. presidential speeches on electoral advantage in the most recent monthly polls and other characteristics. The standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.
Online Appendix B:

Word Embedding Models

Word embedding is a type of language model that maps words or sentences and documents into vectors of real numbers. A more common way of vectorizing words and documents is the ‘bag-of-words’ approach, which represents each word as a unique dimension (one-hot vector with 1 in that word’s dimension, 0 in all other word dimensions) and represents documents as counts of each unique word (sum of one-hot vectors). The order of words is assumed to be irrelevant to the analysis. The semantic and syntactic meaning of words are also lost in ‘bag-of-words’ approach.

Unlike the ‘bag-of-words’ method of vectorization, in which one word is one dimension, word embedding represents words and documents in a dense continuous vector space with many fewer dimensions and positions semantically and syntactically similar words close to each other in this vector space. The method of word embedding is based on a distributional hypothesis in linguistics theory, which states that the meaning of a word is a function of its contexts or surrounding words. Unlike the ‘bag-of-words’ assumption, which treats words as independent atomic units, the distributional hypothesis aims to model the meaning of a word and assumes that the meaning of a word is given, and can be approximated, by the sets of contexts in which the word appears. In effect, the underlying idea is that words that frequently appear in same contexts are likely to have a similar meaning.

There are several different ways to train word embedding. In this paper, we use a Doc2vec model (Le and Mikolov, 2014), which is based on the more foundational word2vec model (Mikolov et al., 2013). We begin by describing the word2vec model. The word2vec model is a neural network based model that takes each unique word in the vocabulary of a corpus as an input. The input word, represented as a one-hot vector, is then multiplied by a dense, real-valued weights matrix of size \( V \times d \), where
$V$ is the length of the vocabulary in the corpus and $d$ is the chosen size of the hidden layer or ‘embedding’. By multiplying the $1 \times V$ input vector for a word with the $V \times d$ weights matrix, a $1 \times d$ vector is generated; this is the word’s vector representation, $v_{\text{word}}$. The model then uses this vector representation of the input target word as the input to a softmax classifier to predict which of the $V$ words in the vocabulary are likely to be the context words of the input word. Context words are those that appear in a certain range of words before and after the current/target word. The model learns the embedding or the parameters in the hidden layer by finding the parameters that maximize the predicted probability of true context words. In other words, the word2vec model seeks to set parameters $\theta$ to maximize the conditional probability of contexts $C$ when observing the target word $T$: \( p(C|T; \theta) \) for all words in the vocabulary (Mikolov et al., 2013; Goldberg and Levy, 2014). Therefore, mathematically, the model assigns similar parameters to words that are used interchangeably in the same contexts.

Because maximizing $p(C|T; \theta)$ for all target and possible contexts is expensive to compute and there are more words that do not appear together than words that often appear together, we adopt negative sampling skip-gram in training the model. In negative sampling skip-gram, the input layer contains target-context word pairs. The target-context pairs are generated by taking the target word at index $i$ and pairing it with all context words from $i - k$ to $i + k$ given a window size $k$. For every true target-context word pair, we generate $s$ negative samples; these are target-context word pairs that are not observed in the actual text corpus. The output layer contains dummy values 1 and 0 indicating whether the input pair is a true target-context pair.

1. We choose $d = 100$, in keeping with standard practice.
2. Word2vec encompasses two different, related neural-network based models, including the continuous bag-of-words (CBOW) and skip-gram (SG) models (Mikolov et al., 2013). The SG model, which is used and explained here, inputs a target word from a text and attempts to predict the target word’s likely context words. The CBOV model does the reverse. Given a set of context words, CBOW attempts to predict the context’s target word.
3. The actual window size is randomly sampled, per target word, from 1 to $k$. This effectively causes more distant context words to be sampled with less frequency than nearer context words. In our model we use $k = 10$.
4. In our model, we used $s = 10$. 

36
that co-locates in the texts (1) or a negative/fake pair that does not appear together in the texts (0). The predicted value given an input pair is computed by taking the dot product of the target word vector (target embedding) and the context word vector (context embedding) and then applying the logistic function, $\sigma(\cdot)$. The model uses small non-zero random values as the initial parameters in the hidden layer to produce the embedding/word vector. Stochastic gradient descent is then used to optimize the parameters through back-propagation to minimize the logarithmic loss between $\sigma(v_{\text{target word}} \cdot v_{\text{context word}})$ and the true value $[0, 1]$.

Expanding the word2vec model to the document level is simple; each document is labeled with an ID, and treated as one unit (like a word). This document ID is positioned within the text in the document. For example, suppose we have a one-sentence document labeled as Doc1: “We are fighting for the forgotten Americans.” The document ID is treated as one unit and positioned within its text: “We are fighting for Doc1 the forgotten Americans.” 5 The negative sampling algorithm can now be applied to both the target word and the document, which is treated as a target word. In this way, the documents sharing similar texts or content are positioned close to each other in the vector space (Le and Mikolov, 2014).

---

5 In practice, the model is adjusted so that the Doc1 token occurs in all of document 1’s words’ contexts and all of the document 1’s words appear in the Doc1 token’s context.
Online Appendix C:

Random Forests Classification Algorithm

In this paper, we treat the problem of measuring populism as a classification problem, thereby making it a supervised learning problem. Some discourses are populist and others are not. To measure populism, therefore, we simply need to sort the populist discourses from the non-populist discourses. In the main text, we outline the general steps of the supervised method. Here, we explain the Random Forest algorithm used to perform the supervised classification tasks in more detail.

The random forests algorithm is a kind of ensemble of decision trees. To understand how the random forests algorithm works, we first need to understand decision trees. Intuitively, we can consider the decision tree algorithm as mimicking the human decision-making process. For example, when deciding whether a certain document is populist or not, a human coder’s decision process might follow a decision tree similar to the one shown in Figure C1. First, the human coder decides whether the document depicts the people as the only legitimate source of power. If this is the case, then the human coder needs to decide whether the people are pluralist or not. If the people are pluralist, then the document would be classified as liberal democratic. If the people are not pluralist, then the human coder needs to identify whether the document structures politics along moralized lines. If politics is structured along moralized lines, then the human coder will classify the document as populist. The decision tree algorithm attempts to learn a hierarchy of if/else questions, like the one shown in Figure C1, that leads it to classification decisions that match those made by a human coder. The researcher does not need to identify the rule used in each node of the hierarchy for the machine. Instead, the decision tree algorithm learns those rules from the labeled data in the training set, which in our case comprises the documents that have been

\[6\] In addition to classification, the other supervised learning problem is regression.
hand-coded into populist and non-populist categories using holistic grading.

But how does the decision tree algorithm learn to make the rules? To grow a single decision tree in the random forest, the algorithm first randomly bootstraps a sub-sample of the training data. Second, it randomly selects a sample of features. The features are the 100 dimensions used to represent documents in the Doc2vec word embedding approach or the unique words in the bag-of-words approach. Third, it finds the best pair \{feature, feature threshold\} to split the data into two subsets, in our case populist and non-populist, that are the ‘purer’ among all the possible splits by minimizing the cost function of the ‘impurity’ measure shown in Equation 5.

\[
J(k, t_k) = \frac{m_a}{m} G_a + \frac{m_b}{m} G_b
\]  

(5)
where \[ G_a = 1 - \sum_{t=1}^{n}(p_{a,k})^2 \] is the gini impurity measure at subsample/node \( a \)

\( m_a \) is the number of instances/cases at subsample \( a \) after the split.

After a split is conducted, the data (documents) are sorted into two different nodes. In each ‘daughter’ node, the algorithm repeats steps two and three to continue splitting the data until a stop-point criteria is met, such as when the training set is perfectly sorted (into populist or non-populist classes) or a pre-determined maximum depth of the tree is reached (Strobl et al., 2009; Hill and Jones, 2014).

At the end of this process, we have a decision tree in which each document has been classified into one of the possible classes, in our case populist or non-populist. This whole process is then repeated many, many times, resulting in a large number of decision trees. Together these trees make a ‘random forest’. Because each decision tree is only trained on a random sample of the training set and each split feature node is selected from a random sample of the features, this algorithm is less likely to overfit the training data. Each decision tree is a ‘weak learner’ meaning that it is only slightly better than a random guess. However, the ensemble of them results in powerful classifiers. Once we build the random forest using the training data, we can use it to classify new documents. To classify a new document, each decision tree in the random forest makes a prediction based on its hierarchical rules and the document’s features. The final prediction on the class of the new document is made using a majority rule, which is the most commonly predicted class among all the decision trees. Thus, if the new document ends up in the populist class in more than half of the trees in the random forest, then it is predicted to be a populist document by the random forest. Because the predicted class membership for a given observation is an ensemble of a large number of decision trees or weak classifiers, we can also calculate the uncertainty of the prediction based on the distribution of the predictions from all the decision trees.

\(^7\)In the model, we used 5,000 trees.
Figure C2: An Example of a Decision Tree in a Random Forest

One advantage of the random forests algorithm is that it can be visualized and is easier to interpret than other supervised learning algorithms. However, since our model represents the words and texts in our training set using word embedding, which has 100 dimensional real value vectors, it is hard to interpret. Therefore, to demonstrate the random forests algorithm, we present an example of a decision tree in a random forest trained using the ‘bag-of-words’ approach. The decision tree in Figure C2 is trained using a random sample of hand-coded training data. In the beginning, 54% documents in the bootstrapped training sample are not populist. The task is to sort the populist documents into the populist class and the non-populist documents into
the non-populist class. The features that the algorithm uses to do the sorting are the words/terms frequencies in each document. The first node ‘reality’ with a threshold of 0.5 is selected to split the data, because among the random sample of features and possible thresholds, ‘reality’ with a threshold of 0.5 results in the purest sub-samples. The sub-sample of documents with a value on ‘reality’ above the 0.5 threshold are all populist in the training sample, which is a pure populism group. Therefore, there is no need for further splitting. For the documents that have a value on ‘reality’ that is below the threshold, there is still a mixture of populist and non-populist documents. A daughter node, ‘special_interest’ with a threshold of 1 gives the purest sub-samples and is selected to split the data. The documents that have values on ‘special_interest’ above the threshold of 1 are all populist documents. The process is repeated two more times, and a stopping criteria is reached. In this case, after splitting at the end node, the remaining data splits into a pure non-populist group and a pure populist group; there is no need to split the data further. This tree built using the training sample is then used to predict new documents. If the new document mentions ‘reality’ above the 0.5 threshold, it is classified as populist by this tree. If not, the tree will move the next node until a decision is made. A random forest consists of thousands of decision trees like the one shown in Figure C2. To classify a single case, the random forest algorithm predicts the class using the most commonly predicted class for that case among all of the decision trees (majority vote).