Corruption and the Rise of Political Outsiders:

Evidence from Audits and Local Elections in Brazil[*]

Natalia Garbiras-Díaz[†]

European University Institute, Max Weber Programme

November 3, 2022

Latest version here

DRAFT. *Please do not circulate*

Abstract

How does corruption salience influence the entry of political outsiders? While revelations about corruption are known to shape citizens' votes, it is unclear whether outsiders take advantage of corruption scandals to contest elections. I argue that corruption is a valence issue that allows outsider candidates to draw support from disaffected voters across the ideological spectrum. First, I construct a novel measure of candidates' use of anti-corruption rhetoric using manifestos registered by over forty-nine thousand mayoral candidates in Brazil who ran for election between 2012 and 2020. These data reveal that outsiders, especially those who adopt more centrist positions, are more likely to resort to this valence appeal. Second, I use random annual audits conducted by the federal government as a source of exogenous variation in corruption revelations in Brazilian mayoral accounts and show that municipalities exposed to these shocks experience a greater entry of outsider candidates. However, outsiders do not obtain better electoral results. The evidence suggests that this may be due to mainstream candidates adjusting their campaign strategies in response to outsider entry. In particular, I find that non-outsiders, when exposed to this shock, mimic outsiders by adopting their anti-corruption rhetoric. This paper demonstrates how an increase in the salience of corruption can spur the emergence of outsiders, who in challenging the existing system and establishment may contribute to democratic backsliding. As such, it highlights consequential and understudied indirect effects of corruption on the persistence of democracy.

1 Introduction

In both young and advanced democracies, citizens have become disillusioned with political and party systems, fueled by corruption scandals that have tainted politicians across the ideological spectrum (Engler 2016; Droste 2021).[1] This has allowed for the breakthrough of political outsiders, i.e. candidates running outside existing parties or without previous experience in office, who use these sentiments to mobilize disenchanted voters.[2] In Latin America, home to some of the world's most populous democracies, the presence of outsiders in presidential elections over the past four decades has increased over time (see Figure 1): from the early 1980s to the four-year period from 2016 to 2020, this percentage has nearly doubled. Not only has the number of these candidates grown but so too has their vote share (from 20% to 61% over the study period). This has naturally produced both increased media coverage of their campaigns and, more surprisingly, victories (e.g., Jair Bolsonaro in Brazil, Nayib Bukele in El Salvador, and Mauricio Macri in Argentina) or near victories (e.g., Rodolfo Hernández in Colombia and Beatriz Sánchez in Chile).

What explains the entry and electoral success of political outsider candidates? To answer this question, I propose a theory of outsider candidate entry based on the salience of valence issues, broadly defined as those on which there is some consensus among voters. I study the case of corruption as a crucial valence issue. In particular, I argue that voters agree on their preferences for clean/honest candidates and rejection of those tainted with corruption allegations. Thus, given its multivalent nature and combined with voters' widespread concerns about political corruption and desire for clean politicians, anti-corruption discourse offers the opportunity to mobilize a wide range of voters who are spread evenly across the ideological spectrum. I provide empirical testing of this claim using public opinion data from Latin America.
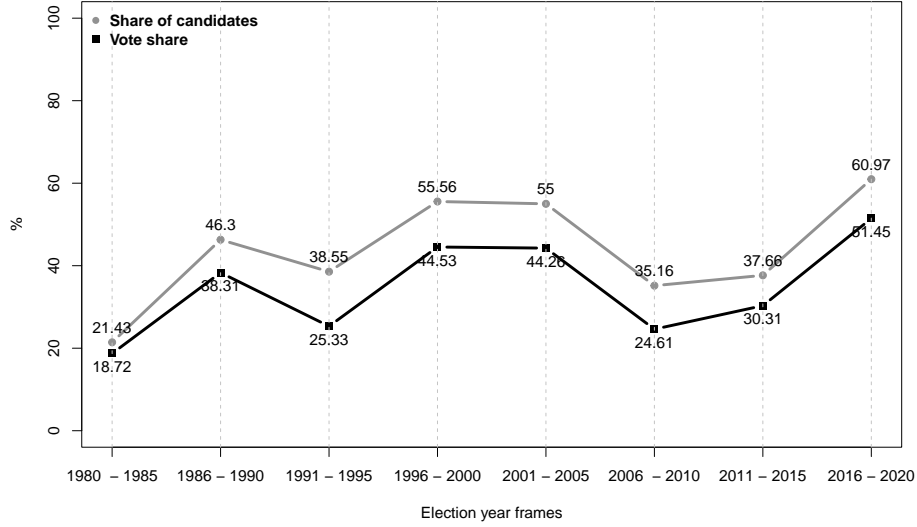
Building on the above, and the empirical observation that outsiders tend to focus their campaigns on valence appeals, I theorize that outsiders' use of these appeals has a relative advantage compared to other campaign strategies.[3] Specifically, I hypothesize that these candidates should be more likely than traditional politicians to mobilize voters based on anti-corruption appeals for two reasons. First, they enjoy a comparative advantage over

---

[1]Data from Colombia corroborates this. In a survey conducted around the 2022 elections for an ongoing project with Loreto Cox, when we asked what makes citizens angry with politicians, the most frequent responses pointed to corruption (see Appendix I).

[2]According to a 2021 Ipsos study, 70% of citizens believe that "*The main divide in our society is between ordinary citizens and the political and economic elite.*" (Ipsos 2021)

[3]In Latin America, this trend began as early as the 1980 presidential elections in Brazil, where Collor de Mello, who would later be impeached in a corruption scandal, became known as the bandit hunter (*caçador de marajás*), in reference to his promise to fight corrupt politicians in the Brazilian political class.

Figure 1: Share of outsiders in Latin American presidential elections 1980-2020

candidates from established parties, who often cannot credibly anchor their campaigns in such appeals. Second, outsiders often enter an electoral arena where established parties already populate and may even claim "ownership" over some spatial issues, e.g., those located on the left-right spectrum. This yields the first empirical implication of my theory: *outsiders are more likely to use anti-corruption appeals.*

Although the fact that outsiders are more credible in mobilizing voters with anti-corruption appeals provides them with a comparative advantage, this issue is not always a principal concern of voters. As such, their success depends on the salience that this valence appeal has among citizens. Outsiders therefore should measure the weight that voters give to corruption, or to valence issues more generally, relative to programmatic policies—e.g., security, taxes, employment, and redistributive programs—in deciding their vote. Mobilizing voters around anti-corruption appeals will be an effective strategy for outsiders as long as this issue is sufficiently salient. This leads to the second empirical implication of my theory: *shocks that increase the salience of corruption lead to an increase in outsider candidate entry.*

I study the case of Brazil, leveraging several characteristics of the local context. First, with more than 5,000 municipalities, Brazil allows for a rich analysis of the electoral cam-

paigns of more than forty-nine thousand mayoral candidates running between 2012 and 2020. Moreover, it offers substantial variation both among municipalities and elections in terms of the presence of outsiders and the relevance of corruption. From a substantive standpoint, the country has been characterized by widespread corruption scandals, peaking with the infamous *Operation Car Wash*, which created an information environment in which corruption poses a credible threat to political accountability and to democracy itself. While this likely renders corruption a national concern, local-level dynamics may make this issue less of a priority as other concerns, such as crime and unemployment, might be first-order priorities for citizens.[4] Finally, the weakness of local party systems and partisanship in Brazil allows for the emergence of outsiders (Samuels and Zucco 2016, 2014; Novaes 2018).[5]

To test my theory, I use a twofold empirical strategy. First, I create an original dataset of electoral programs or *manifestos* of mayoral candidates. Using text analysis, I characterize these manifestos in terms of the intensity with which they rely on anti-corruption appeals. I combine this information with detailed candidate-level administrative data to assess how outsider candidate strategies differ from those of mainstream candidates. Next, I exploit random audits by the federal government of mayors' accounts to obtain exogenous variation in the salience of corruption at the municipal level. Specifically, I test whether audited municipalities experience an increase in the supply of outsider candidates in post-audit elections.

The empirical evidence supports my theory. I show that being a political outsider is associated with increased use of anti-corruption appeals, a relationship that holds after accounting for other characteristics possibly correlated with both variables. I also find that centrist outsiders are the most likely to rely on anti-corruption appeals, which is consistent with evidence from Central and Eastern European anti-corruption parties (e.g., Engler 2020*a*) and with theoretical models of valence competition. Second, I show that exogenous changes in corruption salience lead to greater entry of outsider candidates at the municipal level. These results hold across specifications and robustness checks.

While these shocks spur the entry of outsiders, I find no effect on outsiders' vote share or their likelihood of winning elections. One plausible explanation is that these election outcomes incorporate mainstream candidates' strategic responses to the entry of outsiders to counter their threatened electoral prospects.[6] In support of this interpretation, I show

---

[4]For example, public opinion data available in Chile from the 2016 wave of the periodic survey conducted by the "*Centro de Estudios Públicos*" indicate that 18.7% of respondents mentioned corruption as the main problem that should be addressed by the national government, while only 8% ranked the fight against corruption as the priority problem for the next mayor (CEP 2016).

[5]The emergence of outsiders or so-called third parties is less frequent or even unfeasible in places where partisan identification is one of the key predictors of vote choice, as in the United States.

[6]Another plausible explanation is voter coordination.

that non-outsider candidates running in exposed municipalities mimic the outsider campaign by adopting anti-corruption rhetoric to a greater degree.

This paper contributes in at least five ways to the literature. First, a growing body of research has examined what produces the rise of outsider candidates. These studies have highlighted the role of institutional arrangements (e.g., Carreras 2017, 2012), party system variables (Abedi 2004; Pop-Eleches 2010), and economic and other structural factors (Hanley and Sikk 2016; Rodrik 2020; Carreras 2017). Taking a step back, I instead examine the *decisions* of outsiders to contest elections. By demonstrating the importance of the salience of valence issues in explaining the entry of outsiders into politics, this paper contributes to this growing scholarship.

Second, it speaks to the literature on electoral accountability. While much of this agenda has focused on how voters respond to information shocks about corruption (e.g., Ferraz and Finan 2008; Klašnja, Tucker and Deegan-Krause 2016),[7] less is known about how candidates respond to such shocks. This paper delves into the reasons why these shocks can help overcome some barriers to entry that outsider candidates face when challenging established parties. To the extent that the entrance of outsiders to the political arena enhances competition, this may enhance democratic accountability (Berliner and Erlich 2015; Weitz-Shapiro 2014; Binzer Hobolt and Klemmensen 2008). My theoretical arguments and empirical findings, therefore, have broader implications for understanding what improves the quality of democracy.

Third, this paper also contributes to the literature on electoral competition, specifically work that studies issues that cross-cut the spatial dimension. Some theoretical models incorporate candidates' valence into strategic models of party competition, in addition to the standard policy dimension (see, for instance, Ansolabehere and Snyder 2000; Ashworth and Bueno de Mesquita 2009). Empirically, recent work has assessed these models' predictions and tested how voters respond to changes in candidates' perceived valence (Kendall, Nannicini and Trebbi 2015), or how they trade-off between the policy and valence dimensions (Uscinski et al. 2021; Graham and Svolik 2020). This paper contributes to these debates by marshalling novel empirical evidence showing that certain types of candidates are more likely to campaign on valence issues and strategically respond to shocks that increase their salience.

Fourth, this paper is related to an increasing number of studies on the politicization of corruption. Most work studying parties' incorporation of either the fight against corruption or explicit rejection of existing parties focuses on Europe (Bågenholm 2013; Engler, Pytlas and Deegan-Krause 2019; Engler 2020b). By casting corruption as a valence issue, the paper

---

[7]See Grossman and Slough (2021) for an exhaustive review.

4

explores the conditions under which this issue is likely to be politicized. Moreover, Brazilian mayoral candidates' heavy reliance on anti-corruption appeals, especially among outsiders, shows that this phenomenon is present well beyond the European cases generally studied. In doing so, it also makes a methodological contribution by building one of the only datasets on candidates' use of anti-corruption appeals in Latin America.

Finally, the paper contributes to the literature on democratic backsliding. While the entry of outsiders is thought to be desirable for democracies—stimulating political competition and accountability—less is known about their likely negative effects in contexts of political representation crises. The elections of anti-establishment candidates such as Jair Bolsonaro, Nayib Bukele, or Andrés Manuel López Obrador have shown a particularly poor democratic record. Some emerging work in this area has shown that the election of outsiders can deepen executive-legislative crises (e.g., Carreras 2014) and that the emblematic charisma of these candidates provides incentives to cultivate a personal vote, rather than building party brands (Hollyer, Klašnja and Titiunik 2021), or to adopt more extremist positions (Serra 2018). The irruption of outsiders, thus, calls for future research on their relationship with increased risks of democratic backsliding.

Understanding how outsiders overcome barriers has important implications. First, such candidates have already succeeded in getting elected as mayors of some of Latin America's major cities.[8] Second, mayors in the region, but also in other contexts such as Italy and France, are key actors in local politics, with strong decision-making power to allocate resources and control access to public goods and service delivery (Falleti 2010).[9] Although the evidence in this paper focuses on *local* outsiders, the findings have broader implications. The study of how local outsiders gain prominence provides broader lessons for our understanding of outsiders at the national level. Becoming known at the local level provides a credible path to break through at the national level, as Mauricio Macri (former mayor of Buenos Aires province, Argentina), López Obrador (former head of government of Mexico City), Klaus Iohannis (former mayor of Sibiu, Romania) and Mateo Renzi, Italy (former mayor of Florence) have all shown.

---

[8]Some examples in Brazil include, Marcelo Crivella, bishop, and João Doria, businessman and host of a TV show, who won the 2016 mayoral elections in Rio de Janeiro and São Paulo, respectively, or Romeu Zema, former CEO of a large consortium of companies, who won in 2018 the gubernatorial race in Mina Gerais (see more about the former two cases in this newspaper article). Outside Brazil, some cases in Colombia, for example, include the mayors of some of the country's largest cities (e.g., Claudia López in Bogotá, or Daniel Quintero in Medellín).

[9]This has been shown to be particularly important for politicians' career path in Brazil (Samuels 2003).

2 Theory

I flesh out a theory about the optimal strategies adopted by outsider candidates and how they respond to changes in the informational environment. First, I discuss in detail the concept of valence competition and argue that anti-corruption appeals are a seminal case of a valence issue.[10] I provide empirical support that corroborates this claim. I then theorize why outsider candidates have an electoral advantage over traditional parties when competing on these types of issues. Building on this, I posit that shocks that increase the salience of corruption help to overcome the barriers that outsider candidates often encounter when challenging established parties. Finally, I derive a number of observable implications of my theory, which I test empirically in the following sections, using the case of mayoral candidates running in local elections in Brazil.

2.1 Defining Outsiders

Before delving into the theory of outsider campaign strategies and their electoral prospects, I provide a definition of outsiders. I draw on the typology (of presidential candidates) developed by Carreras (2012), which is defined in relation to two variables: first, the candidate's location vis-à-vis the existing party system; and second, the candidate's prior experience in office. With respect to the first variable, candidates can run on the ticket of established parties (e.g., Donald Trump who ran with the Republican Party in the U.S.), as independent candidates (e.g., Sergio Fajardo who collected signatures to register his candidacy in Colombia), or with a new party or movement (e.g., Emmanuel Macron who ran with En Marche! in France). In terms of prior experience in office, candidates could have had no previous career in politics or in public administration (e.g., Alberto Fujimori who ran with Cambio90 in Peru) or some (e.g. Yanni Varoufakis who ran with MɛPA25 in Greece).

In this paper, I study separately three types of *outsider* candidates: (i) those running without existing parties (either with a new party or as independents); (ii) those running with no prior experience in office; and "full outsiders," defined as those satisfying (i) and (ii).[11]

2.2 Valence Competition

One of the most influential theories about party competition and elections is issue politics. Within this literature, in his seminal model of spatial competition, Downs (1957) originally

---

[10]Some existing accounts on valence competition focus on when and why established parties rely on valence issues (e.g., Curini 2018).

[11]Brazil's weakly institutionalized party system at the local level makes it especially important to study both dimensions.

argued that voters' preferences could be located along a one-dimensional space and that they would vote for the party with the closest policy stance to their ideal point. In this context, parties—characterized as office-seekers—would try to maximize their vote share by setting the policy position that minimized such distance.[12]

However, the evidence shows that parties often compete on issues that cannot be located on a positional dimension but rather reflect some consensus among voters in terms of their desirability. These so-called "valence issues" refer to cases in which candidates "(...) are differentiated not by what they advocate but by the degree to which they are linked in the public's mind with conditions or goals or symbols of which almost everyone approves or disapproves" (Stokes 1992, p. 143).[13] In this sense, voters hold a similar position on a valence issue, and candidates strive to signal that they are better equipped to handle it or are more endowed with it (i.e., vertical differentiation).[14]

Typical examples of valence issues are economic prosperity, crime reduction, a scandal-free administration, and the fight against corruption. Some empirical works have illustrated the importance of valence issues in voting behavior for elections in the U.S. and other European countries (Green 2007), as well as outside western democracies (Lupu et al. 2013). Furthermore, experimental evidence shows that, in some cases, candidates' messaging based on valence issues can prove even more persuasive than on their ideological stances (Kendall, Nannicini and Trebbi 2015).

2.3 Outsiders' Valence Strategies

How can outsiders mobilize voters who are disenchanted with existing political parties and are therefore willing to seek untested alternatives? I start from the empirical observation that, both in Latin America and in other regions, outsiders have increasingly begun to incorporate issues of valence into their political discourse. The most recent trend has been the use of an anti-corruption appeal to compete in elections against traditional parties; although, more recently, anti-establishment appeals have also gained prominence among such candidates (Bågenholm 2013; Engler 2020b).[15]

---

[12]This theory has provoked a series of critiques. There is a large body of research in American politics, and more recently, in other fields, that argue against this view. In the former, most work highlights the role of party ID, past performance evaluations, and attachments to societal groups in predicting vote choice.

[13]To be sure, valence issues can hide position-issues. Notwithstanding, for the purposes of the foregoing argument, what ultimately matters is how candidates frame these issues during the campaign.

[14]Likewise, when it comes to issues for which voters have a negative valence, candidates will attempt to convey the minimum endowment.

[15]A large body of research on European politics have studied the use of anti-corruption and anti-establishment appeals by different types of parties. See, for instance, Bågenholm and Charron (2015), Polk et al. (2017), Engler (2020a) and Engler, Pytlas and Deegan-Krause (2019).

I posit that there are strategic reasons why outsiders may find it appealing to focus their campaigns on valence issues as a way to cultivate new supporters as they enter the political race. At a minimum, valence issues allow outsider candidates to appeal to a broader range of voters across the ideological spectrum and thus garner enough votes to challenge the established parties.

Valence issues also help voters solve collective action problems when it comes to punishing bad outcomes. To illustrate this point, consider a scenario with two dominant parties, each located on different sides of the positional spectrum. If one or both of the parties perform poorly on any dimension (e.g., irresponsible management of the economy, accusations of corruption, etc.), even if voters wanted to punish them and vote for alternative (but smaller) parties that also compete on positional issues, doing so would be very risky. Voters will prefer to remain in a "bad equilibrium"—i.e., not punish their parties—to reduce the likelihood that the opposition party (on the other side of the positional spectrum) will win the election. Outsider candidates who rely on valence issues solve this collective action problem, or at least in part, because they capture disaffected voters on both sides of the positional spectrum. Thus, voters will view them as credible challengers to the existing parties from a strategic standpoint, thus making them a viable option to vote for and hold the existing parties accountable. Note that this feature should play especially in favor of outsiders, considering that, in some cases, established parties cannot compete on certain valence issues, mainly because they have been guilty of poor performance in trying to govern on this issue (Keefer 2007).

In addition to the advantages that outsiders enjoy when campaigning using valence issues, in some contexts, competing on positional issues may simply not be feasible.[16] In some cases, existing parties have already fully populated the spatial dimension and can more credibly claim ownership of certain positional issues. This is especially true in the presence of some electoral rules that make it difficult for new parties to emerge (e.g., plurality rules). Moreover, as societies become richer (as is increasingly the case in the democracies of the Global South), and so does the relative size of the middle-class (Ferreira et al. 2012), class-based appeals (which can be mapped onto the position space) may become less powerful in mobilizing voters. In such contexts, valence issues may offer a new repertoire of appeals to new parties or to outsider candidates, more broadly, who are beginning to create their own "brand."

Taken together, these reasons underpin the argument that, when competing with existing parties forged around positional issues, outsiders can overcome barriers to entry by cham-

---

[16]Outsiders could rely on direct appeals, such as clientelistic inducements, but they often lack the infrastructure required to do so. Furthermore, in some cases, incumbents may even monopolize the market of votes (Blattman et al. 2019).

pioning valence issues. However, the extent to which they can successfully mobilize voters around valence issues will depend on the institutional and informational environments in which outsider candidates run, as I describe in the following subsections.

2.4   Corruption as a Valence Issue

Because of its multivalent nature, candidates who rely on anti-corruption appeals can mobilize a broad coalition that encompasses voters on the left, center, and right of the political spectrum. What makes the anti-corruption appeal a kind of valence issue is the fact that voters agree in preferring honest politicians to corrupt ones, which does not imply unanimous willingness of citizens to punish or support corrupt politicians.
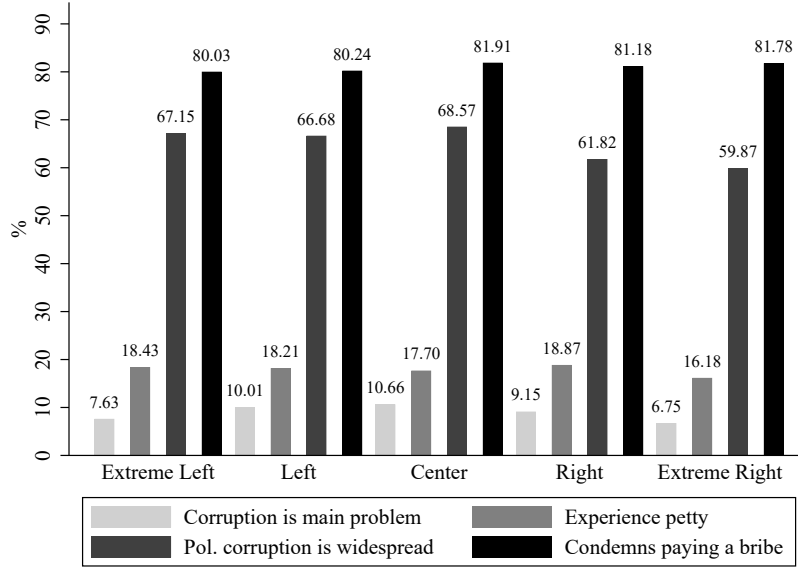
The anti-corruption rhetoric emphasizes the need to eradicate corruption from the political system. Thus, these calls range from directly singling out certain existing parties for their involvement with corruption to urging the purging of the state of corruption. Regarding the former, anti-corruption stances can be used in conjunction with an anti-establishment appeal, as parties attacking the political establishment often politicize corruption, claiming that the political elite is dishonest and only driven by self-interest. However, anti-corruption rhetoric can also feature prominently as part of a "good governance framework" without attacks on the political elite. Along these lines, candidates can also approach the fight against corruption as a means to achieve economic prosperity, promoting values such as state efficiency and transparency (Engler 2020*b*).

To validate this argument, in Figure 2, I plot different measures of citizens' attitudes towards corruption according to their ideology, using data from LAPOP surveys. Notice that citizens across the ideological spectrum condemn paying a bribe, just as they perceive corruption as the main problem to be solved in their country. Moreover, from right to left, respondents consider political corruption to be widespread. Interestingly, people's propensity to have experienced petty corruption (soliciting or receiving a bribe) follows this same pattern.

2.5   Anti-corruption Appeals and Outsiders' Campaign Strategies

An extensive literature on corruption has examined how political corruption influences citizens' voting behavior. Some studies are interested in the effects that perceived levels of corruption have on vote choice, a phenomenon described as "corruption voting." For example, Klašnja, Tucker and Deegan-Krause (2016) distinguish two distinct channels through which corruption influences voting behavior: personal experience, termed *pocketbook*, and perceptions of the incidence of corruption in the context of individuals, which they label

Figure 2: Exposure to and attitudes towards different types of corruption by ideology

sociotropic. They find evidence that the pocketbook corruption vote is stable, and that the latter is contingent on corruption being salient.

A related strand has focused on electoral accountability. Specifically, it asks whether citizens punish politicians when they learn of their involvement in corruption. Despite numerous studies focusing on this question, there is still no consensus on whether voters hold politicians accountable. While some previous studies find that voters punish corrupt incumbents (Ferraz and Finan 2008), others have found less empirical support for this. For example, Dunning et al. (2019) reports null results in similar experiments conducted in several countries on the effect of informing citizens about politicians' misbehavior. However,

the evidence so far suggests that voters are willing to punish corrupt politicians when there are alternatives that are perceived as "clean" (Agerberg 2020); depending on voters' priors about politicians' involvement in corruption (Arias et al. 2021); and that certain institutions, such as the re-election provision, might help (e.g., Ferraz and Finan 2011).[17]

Beyond voters, corruption also shapes candidate behavior. The phenomenon of politicization of corruption has increasingly occupied the research agendas of scholars of European politics (e.g., Engler 2020b), with a particular focus on Central and Eastern Europe. Recent work has examined the factors associated with this phenomenon. For example, Curini (2018) shows that ideological convergence of established parties leads to a greater propensity of the latter to incorporate references to corruption in their manifestos. Descriptive evidence also indicates that new and opposition parties are more electorally successful when they politicize corruption than established parties (Bågenholm and Charron 2014).

Bridging both literatures, I hypothesize that outsider candidates are more likely to make anti-corruption appeals when mobilizing voters than other types of candidates. Based on previous findings, I argue that this is so because outsiders anticipate that politicizing corruption is electorally profitable as long as two conditions are met. First, such candidates must be perceived as a clean alternative, and second, corruption must be sufficiently important or a real concern to outweigh other issues that normally occupy voters' priorities.

Candidates who run with existing parties, or who have been in politics before, may have to internalize the reputational costs of past involvement in corruption scandals by other party colleagues or may themselves show a poor track record in terms of their effort and competence to fight corruption. By not having to bear that burden, I argue that outsiders may be better equipped to satisfy the first condition.

As for the second condition, during the campaign, there are multiple ways in which outsiders can strive to increase the relative importance that voters assign to valence issues and assert their perceived competence on the latter. However, their ultimate decision to enter the electoral race depends on the likelihood that valence strategies will turn voters away from traditional candidates. There is no doubt that campaigning involves the investment of large amounts of resources to pay for staff, rallies, etc., as well as opportunity costs. In this context, shocks that increase the weight voters give to valence issues should favor the electoral prospects of outsiders (Klašnja, Tucker and Deegan-Krause 2016), producing a greater expected benefit from running for office.

---

[17]For a complete review of this literature, see Grossman and Slough (2021).

2.6 Testable Implications of the Theory

I have argued that it is optimal for outsider candidates to rely on valence issues when running for office. I have also pointed out that anti-corruption appeals can be considered a case of such issues since citizens across the ideological spectrum tend to reject corruption equally. It follows that outsider candidates' campaigns should be more likely to contain anti-corruption appeals than those of establishment candidates.

Shocks that increase the salience of corruption among voters might also change the weight they place on candidates' valence appeals (at the expense, for example, of that placed on candidates' programs). An observable implication is that latent outsider candidates should be more likely to run in places exposed to such shocks.

Observable implications:

H1 - Outsider's strategies: When campaigning on valence issues, outsider candidates have an electoral advantage over establishment candidates. Therefore, I expect:
Outsider candidates are more likely to rely on anti-corruption appeals when running for office than other types of candidates.

H2 - Valence shocks: Shocks that raise the salience of valence issues create a window of opportunity for candidates whose campaigns rely more heavily on those types of issues. From *H1*, I thus expect:
Places experiencing shocks increasing the salience of corruption should also have more outsider candidates running for elections.

The foregoing discussion yields some secondary implications.

H1$_b$ - Ideology and valence: Seminal theoretical models of valence competition show that candidates with a valence advantage are expected to take a moderate position with respect to the policy dimension (Ansolabehere and Snyder 2000; Macdonald and Rabinowitz 1998).[18] Observationally, thus, this would imply:
Centrist outsiders are more likely to rely on anti-corruption appeals.

H2$_b$ - Outsiders' electoral performance: Valence shocks may increase the weight voters place on candidates' valence vis-á-vis other considerations (e.g., ideology). I then expect:

---

[18]In this context, having a high valence score is interpreted as a higher probability of being perceived as a clean political candidate, which, I argue, should be stronger among outsiders.

Outsiders running in places experiencing the shock should perform better than those in places without it.[19]

3 Institutional Context: The Brazilian Case

The Brazilian case offers an ideal setting to study these questions. First, since 2009, upon registration of their candidacy, every single candidate running in the more than 5,000 municipalities is required to upload a document stipulating the electoral platform and the policies they commit to implementing if elected in office (which I will refer to as the candidate's "manifesto"). Second, it provides an exogenous source of variation in the salience of corruption. Specifically, since 2003, the federal government has run a program that annually selects a random set of municipalities to audit federal funds transferred to municipal mayors. Together, these two features provide the elements required for an empirical assessment of my theory. In this section, I describe in detail the aspects of Brazil's informational and institutional contexts important for my argument.

3.1 Municipal Elections and Electoral Rules

Elections for local government occur every four years in Brazil, with both the mayor and local legislators being elected. Since 1997, incumbent mayors are eligible for one-time reelection (Constitutional Amendment n.16/1997). Voting is mandatory in Brazil for citizens aged 18 through 64. In municipalities with at least $200,000$ voters, a runoff election can happen if no mayor candidates candidate receives more than 50% of valid votes in the first round. In municipalities below the threshold, the candidate is elected by a simple plurality rule.

The 1988 Constitution includes an article (art. 14) with the eligibility criteria for citizens to run for office, one of which stipulates partisan affiliation. In other words, citizens cannot put forth their candidacies through signatures nor with social movements as in contexts such as Colombia or Chile (see also Law no. 9504 of 1997, which updates the articles related to this provision). Furthermore, the requirements to create parties in Brazil restrict legal status to those able to prove a national character by showing support of voters not affiliated with a political party, corresponding to at least 0.5% of the votes cast in the past general election for the House of Representatives, and with one third or more states accruing at least 0.1% of the electorate that voted in each one. This institutional environment poses

---

[19]Although this expectation is derived from theory, the research design described below does not allow for a full test of it since the observed electoral outcomes correspond to general equilibrium results, which include the strategic response of non-outsiders to the valence shock. For example, they might adapt their campaign strategies in response to the anticipated entry of outsiders.

some challenges to the identification of outsider candidates in Brazil—a task that would be facilitated in countries allowing independent candidates to run—and thus obliges a stricter definition of outsiders, at least in terms of the dimension that refers to candidates' location vis-à-vis the party system. In Section 4.2, I go over the operationalization I resort to given these constraints. Appendix F provides further details on Brazil's electoral rules.

3.2 Corruption in Brazil

Corruption is widespread in Brazil. According to recent waves of the AmericasBarometer (2016 and 2018), Brazilians perceive that the vast majority of the politicians in the country are involved in corruption (see Figure 3). Moreover, 18% of the respondents state that corruption is the most important problem to tackle.

Figure 3: Proportion of Brazilian politicians that voters deem corrupt
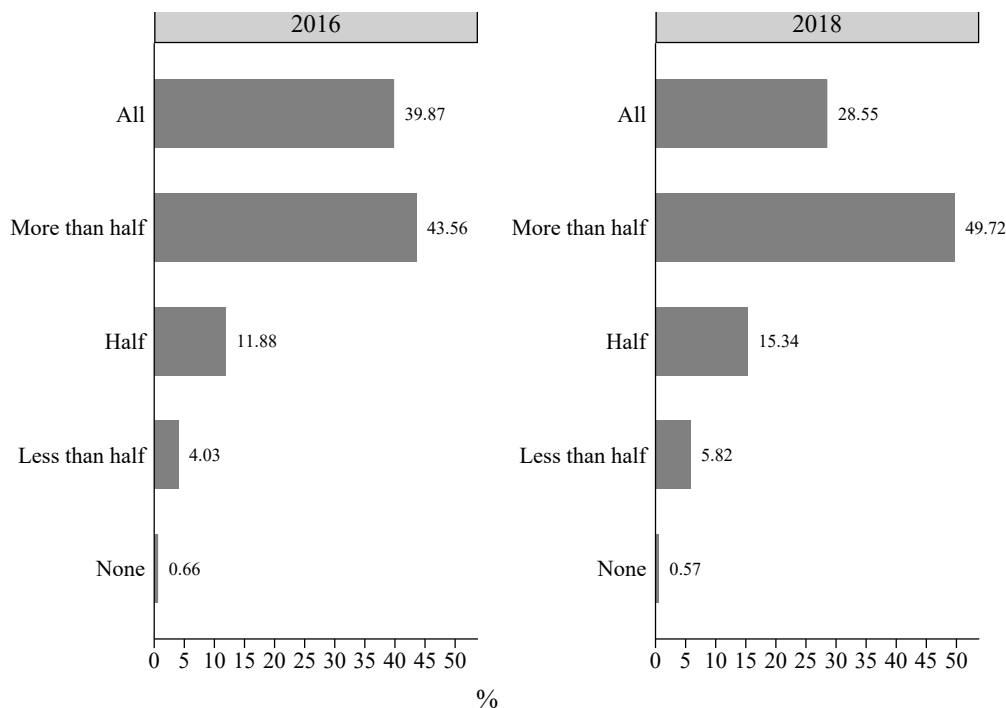
Source: Data from the 2016-2018 LAPOP surveys conducted in Brazil. Data are publicly available here.
Note: See the note in Figure 2 for the exact wording of the question asked.

This high perception of political corruption among citizens has been paralleled with a series of scandals that have been widely covered on the news, tainting political parties across the board. One of the most recent renowned cases of corruption in Brazil involved kickbacks received by top-level politicians in exchange for public contracts related, mainly, to the state-owned oil company, Petrobras. The case was revealed in 2014 (although investigations and

14

prosecutions are still ongoing, as of 2021) in what became known as *Operação "Lava Jato,"* and involved the major parties in the country, located on both the right and the left of the ideological spectrum. Such a context is well suited to test the theory proposed in this paper, in that Brazilian voters are presumably seeking new alternatives, i.e., "untried" ones (Pop-Eleches 2010), beyond existing challengers, and also because latent levels of corruption allow candidates' politicization of this phenomenon. To be sure, however, these conditions may also render it a hard case to test the theory as, presumably, outsiders may be already taking full advantage of the sufficiently high salience of corruption in the country.

3.3   CGU's Randomized Audits

As already previewed, I rely on the Brazilian federal government's audits to operationalize municipal-level shocks increasing the salience of corruption. This audits program was initiated by the Brazilian federal government in 2003. The program was implemented by the *Controladoria-Geral da União* (henceforth referred to as the CGU, for its acronym in Portuguese), the Office of the Comptroller General, an institution created in the same year by the federal government as an independent agency, with a mandate to increase transparency and combat corruption at all levels of government administration.

The program randomly selects municipalities for audits on their use of federal funds.[20] In each draft, a given number of municipalities from within a state are selected, and population thresholds are used to determine the universe of eligible municipalities. Each round was regulated by specific legislation, yet the cutoff most often used was 500,000 inhabitants: municipalities below this threshold were included in the draft. The lotteries also excluded state capitals. Typically, municipalities that had been audited in any of the three lotteries immediately before, or in other special auditing operations, were subject to a "grace period" and therefore were not eligible for selection in the next draft.

A total of 40 lotteries took place between June 2003 and February 2015. Details on the areas of government that were subject to inspection and eligibility of municipalities are provided in the Appendix.

Each selected municipality receives a set of inspections, called "service orders," that investigates the use of federal funds in different areas of government and programs implemented in the municipality. The inspections are conducted by a team of auditors that are federal public servants and receive competitive pay, and there has been no evidence of manipulation in the reports (Ferraz and Finan 2008).

At the end of the inspection period, the auditors send the information to the CGU

---

[20]See Appendix G.4 for covariate balance.

and the results are compiled in a report detailing all irregularities found in the municipality. Their severity is then classified by the CGU's team of experts as: high ('alto'), very high ('muito alto'), medium ('médio'), severe irregularity ('falha grave'), medium irregularity ('falha média), low ('baixo'), formal irregularity ('falha formal), and information ('informação').

These reports are made publicly available and also sent to several governmental institutions, including the Federal Court of Accounts (TCU), the Federal Public Prosecutor's Office (MPF), and the municipal local legislature. Lawsuits and judicial punishments eventually depend on the decision of prosecutors to further investigate the irregularities found in the audits.

The publication of the audit reports has provided important input to the prosecution and conviction of corrupt politicians, as shown in Avis, Ferraz and Finan (2018). The release of these results has also had an impact on the re-election rates of mayors, with electoral punishments being applied to corrupt politicians (Ferraz and Finan 2011 and Ferraz and Finan 2008).

Until 2015, the program was called *Fiscalização por Sorteios Públicos* and the selection of municipalities for audit inspections was based exclusively on random lotteries. In 2015, the program was subject to changes and renamed *Programa de Fiscalização em Entes Federativos* (FEF). Since then, it incorporates two additional methods of selecting municipalities, other than the random lotteries: a 'vulnerability matrix' and census criteria. In the first method, municipalities are ranked according to several socioeconomic indicators in order to identify potential shortcomings and identify those municipalities most at risk for corruption; in the second, all administrative entities of a certain geographic level are audited.

Figure 4a shows the total number of municipalities audited per year, from 2006 onward. It shows that the number of audits per year has decreased over time. As Figure 4b shows, beginning in 2015, the CGU started conducting non-random audits. The non-random selection of municipalities seems to be the predominant method used by the CGU lately in the allocation of audit inspections.

Underpinning my empirical strategy is the assumption that the random audits carried out by the CGU are consequential and tend to lead to corruption revelations. As for the first point, as documented by (Avis, Ferraz and Finan 2018), audits have increased the likelihood of legal action against public officials by 20% in selected municipalities. Regarding the latter, audited accounts usually entail findings of some type of fund mishandling. This has been largely documented by published articles that use this empirical case to study questions related to corruption and political accountability (e.g., Avis, Ferraz and Finan 2018; Ferraz and Finan 2008, 2011). For instance, Ferraz and Finan (2011, p. 1285) find that, based on

Figure 4: Evolution over time of CGU's auditing program

(a) Number of municipalities audited by year
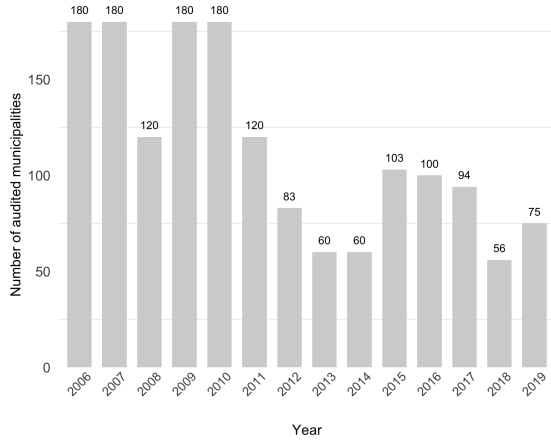
(b) Composition of random and non-random audits by year
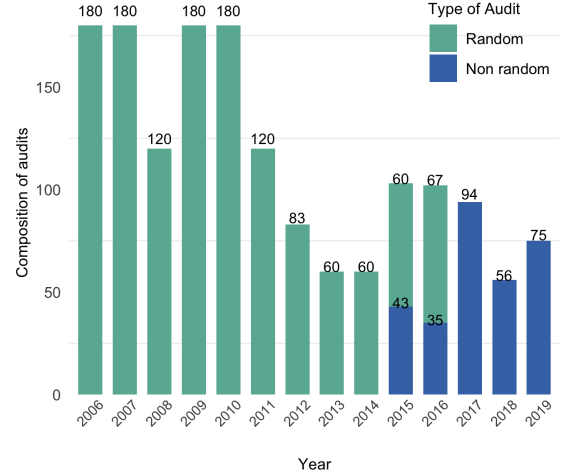
Source: Data provided by the CGU via a freedom of information request.

the audit reports and their manual coding of the findings, 98.6% of the audited municipalities in their sample revealed at least one mismanagement irregularity, and 79% at least one act of corruption (a number that is very similar to that found by Ferraz and Finan (2008), 73%). Audit results are published online for citizen consultation and widely disseminated through the local media and, as expected, tend to be incorporated into the political campaigns of subsequent elections (Ferraz and Finan 2008). Considering the lotteries examined in this paper, and including only random ones, almost the universe of audits include at least one irregularity found by the CGU classified as severe or medium, and 83.4% were classified as severe.[21]

4 Research Design

In this section, I first describe in detail the procedure employed to create a measure of the intensity of candidates' use of anti-corruption appeals. Next, I explain how I operationalize the corruption shock that I examine in this paper. Finally, I provide more information on how I create other variables that are part of the central analysis of the paper.

---

[21]Based on the following own classification: "severe" if *high, very high, severe irregularity*; "medium" if *medium* (i.e., *falha média* and *médio*).

4.1 Measuring Anti-corruption Appeals

To date, there are no existing measures for how "anti-corruption" are candidates running in Brazil's local elections. Therefore, in order to measure this in a systematic and standardized way, I rely on an inductive method, which draws on the manifestos posted by candidates upon registering their candidacy as a self-reported measure of their stance vis-à-vis this issue. Specifically, I relied on machine learning methods to automatize the classification of the universe of paragraphs composing candidates' manifestos, based on whether they were deemed as related to corruption or not. Following that classification, I computed different measures of each candidate's degree of "anti-corruption." Below, I provide an overview of the classification process.[22] First, I downloaded almost fifty thousand manifestos publicly available on the Brazilian Electoral Office's website (the TSE) for the universe of candidates running in the 2012, 2016, and 2020 mayoral elections. I converted these documents into a readable format using OCR. I split candidates' manifestos into paragraphs since this smaller unit of analysis would render more accurate classification results.

Second, after having cleaned out corrupted paragraphs—resulting, for instance, from the OCR or other types of errors,[23] I trained a group of research assistants (RAs) to manually classify a stratified random sample of paragraphs, following a coding procedure.[24] The design of the latter was informed by an exhaustive review of the literature on the politicization of corruption and based on the analysis of a random sample of the manifestos of Brazilian mayoral candidates from 2012 to 2020. The entire manual classification process was blinded, meaning that RAs were never revealed any identifying information about the candidate whose paragraph was being classified. I used this subset of hand-coded paragraphs to obtain two different sets of variables aimed at measuring candidates' reliance on anti-corruption appeals.

First, the hand-coded paragraphs classified as having an anti-corruption narrative served as the basis to create a dictionary of words directly associated with anti-corruption discourses. For example, transparency, management, resources, etc.[25] Next, with this dictionary of words I calculate for each candidate's manifesto a variable corresponding to the total number of words belonging to the anti-corruption dictionary that appear in the candidate's manifesto (*dictionary sum*). From this sum, I compute a variable, which I call *dictionary proportion*, corresponding to the number of times a word belonging to the anti-corruption dictionary

---

[22]In the Appendix, I provide the full details about the coding procedure.

[23]In the code book, I discuss in detail the inclusion rules used in the cleaning process.

[24]Stratifying variables: election year, quintile of manifestos' length, and candidates' party and geographic region of the country.

[25]A word cloud (see Figure 7) and some examples (see Table 18) of paragraphs classified as anti-corruption using this methodology are included in Appendix C.

appears in a paragraph of the candidate's manifesto (i.e., the variable sum of the dictionary) divided by the total number of paragraphs in the manifesto. Next, I trim this variable so that the maximum value is 1.[26]

Second, I also used the subset of hand-coded paragraphs to classify the remaining ones as having or not an anti-corruption discourse, using different classification models and a process of hyperameterization with cross-validation based on the *ROC AUC score* and *accuracy* measures.[27] Specifically, I classified paragraphs using the following algorithms: a logistic regression, Multinomial Bayes Model (MBM), Random Forest Classifier (RFC), Support Vector Machine (SVM) with Sigmoid kernel, Bagging for decision tree classifier, and a Multilayer Perceptron neural network (MLP). Those paragraphs classified by *all* of the different algorithms as anti-corruption were taken to feed a second phase and overcome problems of class imbalance. In a second stage, I fed the classification algorithms with a combination of the latter group of newly and robustly-classified anti-corruption paragraphs and those from the manual classification. Then, I used the best-performing models, i.e., SVM and MLP, to classify the universe of paragraphs as either anti-corruption or not.[28] With all paragraphs' classification, I computed a measure corresponding to the proportion of the manifesto's paragraphs that were classified as referring to corruption.

Once the universe of paragraphs had been classified, these were appended again to the original manifesto to create a series of variables measuring the candidate-level reliance on anti-corruption appeals. An important caveat to this measure is that there could be a gap between the information candidates include in their manifestos and the one they convey in other outlets or venues, such as rallies and TV ads, for instance. To be sure, candidates' oral and written political communication strategies could be worlds apart depending on the audience and purpose. For the purpose of the classification procedure, this means that candidates not classified as anti-corruption as per their manifesto could be classified as so if I were to use other corpora.

---

[26]I create some additional variables: A dummy indicator of whether the number of words belonging to the dictionary is: (i) above the mean across manifestos, (ii) above the median of this variable across manifestos, and (iii) in the 90th percentile of the overall distribution of this variable across manifestos.

[27]This process was done after a twofold depuration. First, I removed from the manifestos "stopwords," which are those deemed as used very frequently that do not provide information for the classification procedure. Second, via the process known as *embedding*, I also cleaned out very common words that, even if not stopwords, may be simple connectors without a relevant meaning. This was done using the following two programs in Python: `Word2Vec` and `TFiD Vectorizer`.

[28]Appendix E describes these algorithms in detail and provides cross-validation results.

4.2 Definitions and Data Sources

Data on the municipal audits was provided by the CGU upon an open data information request. The dataset covers audits conducted between 2006-2019 and includes information on all irregularities found in each inspection order. The data includes the total amount of funds transferred to the municipality, the government sector or program being audited, the severity of the irregularity uncovered a brief description of findings and the date when the audit results were published.

Information on political candidates' characteristics, party affiliation, electoral results, and campaign expenses are publicly made available by the *Tribunal Superior Eleitoral (TSE)*. Information on the characteristics of the municipalities, such as yearly population estimates, municipality GDP, and human development index, comes from the Brazilian Institute of Geography and Statistics (IBGE). To proxy for access to media at the municipal level, I also relied on data provided by the *Atlas da Notícia* project, a journalism initiative from the Institute for the Development of Journalism (Projor) and the Volt Data Lab.[29]

To identify outsider candidates, I use two characteristics: whether the candidate held any political office in the term before and whether the candidate was running with a recently created party in that election.[30] This information comes from the TSE's candidates' registration records, which are publicly available.[31]

(a) **New electoral platform:** corresponds to candidates running with a party that was a newcomer in that election year, i.e., if that was the first election in which that party participated. The criterion considers both local and national elections, which take place in separate years.

(b) **No political experience:** corresponds to candidates with no prior experience in office. To be classified as such, candidates must satisfy two criteria. First, the candidate must not be holding some kind of political or bureaucratic office at the time of registering her candidacy.[32] Second, the candidate cannot have run for office in the previous two

---

[29]More information on the project available at https://www.atlas.jor.br/

[30]In practice, it is possible that the candidates who meet this definition may be façade outsiders, belonging to political dynasties or put forward by traditional candidates. I am processing a database of surnames of the Brazilian bureaucracy and candidates of the last 20 years so that I can purge such cases from the list of outsiders. However, I posit that this should be less of a concern given the case I study. Specifically, I argue that outsiders should be easier to judge given the proximity and richer information environment available to voters in local elections.

[31]Notice that the classification is subject to the data available to construct the variables.

[32]At the time of registering their candidacy with the TSE, candidates must declare their occupation. Candidates are classified as having public office if the latter, for instance, corresponds to deputy, mayor, or senator.

elections.[33] The data to compute these measures is publicly available at the TSE's website.

(c) **Full outsider:** corresponds to candidates fulfilling conditions (a) *and* (b).

To classify candidates according to ideology, I relied on a survey conducted by the Universidade Federal do Paraná (located in Curitiba, Paraná) and the Brazilian Association of Political Science (ABCP), discussed in detail by Bolognesi, Ribeiro and Codato (2021).[34] Based on this measure, a candidate can be considered either extreme-left, left, center-left, center, center-right, right, or extreme-right. For ease of analysis, I collapsed these categories into three: left, right, and center.

4.3 Empirical Strategy

To test the different components of my theory, I pursue a twofold empirical strategy.

First, to assess the underlying assumption that outsider candidates are more prone to rely on valence issues, operationalized in this paper as anti-corruption appeals, I estimate the following candidate-level regression:

$$Y_{i,m,t} = \beta_1 Outsider_{i,m,t} + \beta_2 Ideology_{i,m,t} + \mathbf{X}'_{i,m,t}\psi + \gamma_m + \alpha_t + \varepsilon_{i,m,t} \qquad (1)$$

where $Y_{i,m,t}$ corresponds to one of the four dependent variables described above for candidate i, running in municipality m and in election year t; $Outsider_{i,m,t}$ is an indicator of whether the candidate is classified as an outsider or not, using different conceptualizations of this variable; $Ideology_{i,m,t}$ corresponds to the candidate's ideological classification following Bolognesi, Ribeiro and Codato (2021); $\mathbf{X}'_{i,m,t}$ is a vector of candidate level controls, which includes: age, gender, length of the candidate's manifesto, and level of education; finally, γ_m and α_t correspond to municipal and election year fixed effects, respectively, which account for time-invariant municipal and election specific effects, correspondingly.

The next empirical analysis seeks to identify the effect of valence shocks on the main outcome of interest: outsider candidates' entrance to mayoral elections, which is captured by the share of candidates running as outsiders, based on the different measures I described earlier.

---

[33]In order to compute this measure, I append the list of candidates running in the local elections in Brazil in any municipality from 2004 to 2020. Then, I define a candidate as an outsider if she does not appear in the lists of candidates of the two past elections. Candidates are identified using their unique tax number.

[34]The paper with survey results can be found at https://preprints.scielo.org/index.php/scielo/preprint/view/2552/version/2701.

Based on the theory, valence shocks open windows of opportunity for challengers to capitalize on issues on which they have a comparative advantage. Thus, empirically, this should translate into more outsiders entering races in exposed locations (e.g., municipalities). I estimate this effect by exploiting exogenous variation in the salience of corruption in Brazil induced by the CGU's random audits in Brazilian. In particular, I draw on the fact that, in general, these audits usually encounter corruption irregularities approximately, as discussed in Section 3.3 (see Avis, Ferraz and Finan 2018; Ferraz and Finan 2008). Given this strong "first stage," I use random audits as shocks that increase the salience of corruption. Treatment is then defined as a municipality have been audited in the past four years, which corresponds to mayors' electoral term. As such, I estimate the following model at the municipal level:

$$Outsiders\ Prop_{m,s,t} = \beta_0 + \beta_1 Audit_{m,t-4} + \gamma_m + \alpha_t + \phi_{st} + \sum_{l=1}^{3} \mu_l + \varepsilon_{m,t} \qquad (2)$$

Indexes in (2) correspond outcomes in municipality m located in state (UF) s in election year t. Then, γ_m and α_t are municipal and election years fix effects, respectively; ϕ_{st} are state times election year fixed effects to account for the probability of treatment assignment that varies across states each year, whereby the number of municipalities within state determine the probability that a municipality is audited each year; μ_l are dummy indicators of population strata, which account for the population-based audit eligibility criteria across random audits;[35] and $Audit_{m,t-4}$ is the main treatment of interest, and indicates whether municipality m been audited anytime in the past four years; finally, $Outsiders\_Prop_{m,t}$ corresponds to the share of outsiders who register their candidacies for election year t in municipality m.

Importantly, having been audited proxies for corruption revelation as long as this information is accessible to potential candidates at the time they have to decide whether or not to run for office. Thus, the treatment indicator only takes into account audits whose results were published *before* the deadline for registering candidacies.[36] This condition also ensures that this information is accessible to citizens at the time they are in the voting booth.

---

[35]See Seabra (2018); Litschig and Zamboni (2008).

[36]These dates are: September 26, 2020, August 15, 2016, and July 5, 2012. This information comes a Freedom of Information Request made to the TSE.

4.4 Descriptive Statistics

4.4.1 Brazilian Mayoral Candidates' Anti-corruption Appeals

I begin by describing the dataset resulting from the coding of the universe of manifestos. Table 1 reports the number of manifestos analyzed. I downloaded $49,117$ electoral platforms from the TSE's website, which were split into $16'589,162$ paragraphs.[37] After processing the corpora, the final database to classify contained $6'828,230$ paragraphs, distributed across $48,661$ manifestos.

Table 1: Universe of manifestos analyzed from Brazilian mayoral candidates by year of the election

	2012	2016	2020	Total
Initial manifestos	14,736	16,291	18,055	49,117
Initial paragraphs	1'033,289	6'174,263	9'381,610	16'589,162
Final manifestos	14,551	16,162	17,948	48,661
Final paragraphs	794,453	2'308,567	3'725,210	6'828,230

Note: Author's own calculations.

4.4.2 Outsiders' Anti-Corruption Appeals

I now turn to the description of the main variables used in the analysis of this paper. After cleaning and classifying the universe of paragraphs following the procedure detailed in Section 4.1, I computed a set of measures capturing how much candidates' manifestos used anti-corruption appeals.

Table 2 presents the summary statistics of the main measures used to analyze candidates' leaning on anti-corruption appeals. The variables depicted correspond to the ones discussed in the previous sections. The figures in the table suggest that the set of candidates considered to heavily rely on these appeals varies across measures. First, the average of the variable *dictionary sum*, i.e., sum of times a word belonging to the anti-corruption dictionary appears in the candidate's manifesto, is 39.81 (the SD of this variable is 65.49). On average, these words appear in at least 31% of a candidate's manifesto, but the variance of this variable (*dictionary proportion*) is high, with a standard deviation of 24, suggesting that there are candidates for whom more than half of their manifesto talks about corruption. Finally, 32% of the manifestos are above the mean value of the variable *diccionary* across the sample.

---

[37]Paragraphs were defined as line breaks, identified when converting the documents to text files. For a more detailed discussion about which paragraphs were considered valid, see Appendix C.1.

Overall, the data show that certain candidates use anti-corruption appeals more intensively than others.

The statistics for the measures based on the NLP algorithms are more conservative and yield a lower incidence of anti-corruption appeals: on average, 9.31 paragraphs from a candidate's manifesto were classified as anti-corruption using those methods, representing, on average, 7% of their overall manifesto.

Table 2: Descriptive statistics anti-corruption appeals among Brazil mayoral candidates (proportion)

	Mean	SD	No. of obs
Dictionary sum	39.81	65.49	48,661
Dictionary proportion (trimmed)	0.31	0.24	48,661
Dictionary sum: above mean	0.32	0.46	48,661
Anti-corruption sum (using NLP method)	9.31	18.08	48,661
Anti-corruption proportion (using NLP method)	0.07	0.09	48,661

Notes: Author's own calculations

4.4.3 Outsider Candidates in Brazil's Mayoral elections

Table 3 shows the incidence of outsider candidates in the data across different definitions. Focusing first on the location of candidates with respect to the party system, only 3% of the mayoral candidates in the sample are considered outsiders (i.e., they ran with a new party).[38] Next, turning to candidates' previous experience in office, I use three definitions of outsiders: i) if the candidate has not run for (any) office in the last two elections (prior to the current election year), ii) if the candidate's occupation at the time of registering her candidacy was not related to any publicly elected office or was not a bureaucrat, and iii) the intersection of (i) and (ii), which is the strictest definition of an outsider, and the one I use for the empirical tests. Considering the first measure, only 40 percent of the candidates in the sample are considered outsiders, while this proportion is much higher for the second measure, accounting for 74 percent. Combining both criteria, i.e., those corresponding to the second measure, 35% of the sample are outsiders (third row of Table 3). Combining candidates' location with respect to the party system and prior experience with office, as proposed by Carreras (2012), further reduces the incidence of outsiders in the dataset. Specifically, full outsiders account for only 1%. Given this variation, for robustness, I use different definitions of outsiders in the statistical analysis that follows.[39]

---

[38]Recall that in Brazil, citizens who want to register their candidacy for publicly elected office must be affiliated with a party.

[39]Following the typology proposed by Carreras (2012), 37% of the candidates are either mavericks, amateurs, or full outsiders. *Mavericks* correspond to experienced politicians who run with new parties, *amateurs*

Table 3: Summary statistics: Outsider candidates at the municipality level, according to different measures

	Proportion	SD	No. of obs
Outsider: no party ties	0.03	0.17	52,705
Hasn't ran in the past 2 elections (pol. exp.)	0.40	0.49	52,705
Occupation different from public official	0.74	0.44	52,705
Outsider: no political experience	0.35	0.48	52,705
Outsider type: Full outsider	0.01	0.10	52,705
Maverick, Amateur or Full Outsider	0.37	0.48	52,705

Notes: Author's own calculations. SD is with respect to national averages.

Finally, using data from the 2010 Census, I compare mayoral candidates and Brazil's general population in terms of their sex, age, and education.[40] Tables 10 and 11 in Appendix A present the summary statistics for both populations. There are slightly more outsider female candidates (14.53%) than insider female candidates (12.49%), although the percentage is still much lower than that of the population (51%). Candidates, in general, tend to be older than the general population, but outsiders tend to be younger than insiders (a difference of 1.53 years, p-value< 0.001). This difference is even larger among full outsiders (2.56 years, p-value< 0.001). Candidates from Brazil tend to be more educated than Brazilians in general: the median level is complete higher education, while most Brazilians have no formal education or incomplete primary education.

5 Results

In this section, I provide evidence of the testable implications of my theory. First, I show the results of a test of whether outsiders are indeed more likely to rely on anti-corruption appeals than their establishment counterparts by estimating equation (1). Second, I examine whether shocks increasing the salience of corruption at the municipal level lead to the increased entrance of outsiders on the extensive and intensive margins. Thus, I present the main results of the estimation of equation (2), which provides a causal test for this second observable implication of my theory.

---

correspond to inexperienced politicians running with existing parties, and *full outsiders* are inexperienced candidates running with new parties.

[40]Data collection for the 2022 round is underway.

5.1 Are Outsiders More Likely to Rely on Anti-Corruption Appeals? (H1)

The evidence in Tables 4 and 5 indicates that outsiders are more reliant on anti-corruption appeals than other types of candidates and that this result holds across different measures of the incidence of anti-corruption rhetoric in manifestos and ways of defining outsiders.

Looking at the results depicted in Table 4, depending on the definition used, being an outsider increases between 1.1 and almost 4 percentage points the proportion of paragraphs deemed as anti-corruption (using variable *dictionary proportion*). These findings hold when considering the classification of paragraphs using the NLP methods (Table 5), although the estimated effects are smaller (ranging from 0.4 to 1.2 percentage points).

Notice that, across definitions, outsiders running with center parties, are the ones more intensively relying on anti-corruption appeals: they increase the proportion of paragraphs with anti-corruption words between 2.4 and 9 percentage points, depending on the measure of anti-corruption leaning used. This result is consistent with hypothesis $H1_b$, according to which the ability of outsiders to attract support across the ideological spectrum using valence appeals increases if they adopt more moderate positions, as this allows them to downplay the political dimension. Interestingly, left-wing outsiders are less likely to rely on anti-corruption appeals. A plausible interpretation of the change of sign is related to the fact that the best-known and strongest leftist party in Brazil, the Workers' Party (PT), was in power at the national level between 2003-2016, and then was the main protagonist of Operation Car Wash and the corruption scandals that broke out in 2014 in the country. These factors could have reduced the credibility of the left and, therefore, its ability to mobilize voters on an anti-corruption platform.

Overall, the results presented in this section provide evidence supporting the hypothesis that outsider candidates enjoy a comparative advantage over traditional candidates in mobilizing voters with anti-corruption appeals. Observationally, in the case I study, this translates into outsiders being more likely to include these appeals in their manifestos, for which I find empirical support.

5.2 Do Valence Shocks Allow the Entry of Outsider Candidates? (H2)

The above evidence suggests that outsiders exploit their comparative advantage and are more likely to contest elections with anti-corruption appeals. I now test the second implication derived from my theory: that is, outsiders should be more likely to enter the electoral race

Table 4: The effect of being an outsider candidate on the propensity to rely on anti-corruption appeals

DV: proportion of anti-corruption paragraphs (using dictionary proportion)

Outsider definition:	New platform		No pol. exp.		Full outsider	
Outsider	0.014*	0.005	0.011***	0.014***	0.039***	0.027*
	(0.007)	(0.008)	(0.003)	(0.003)	(0.013)	(0.015)
Center	0.007**	0.006**	0.006**	0.008**	0.007**	0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Outsider × Center		0.090***		-0.005		0.076**
		(0.024)		(0.006)		(0.036)
Left	0.042***	0.042***	0.041***	0.047***	0.042***	0.042***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Outsider × Left		-0.102*		-0.017**		-0.083
		(0.053)		(0.008)		(0.073)
Mean DV	0.315	0.315	0.315	0.315	0.315	0.315
Control Mean DV	0.314	0.314	0.312	0.312	0.314	0.314
R-squared	0.178	0.178	0.178	0.178	0.178	0.178
Sample Size	48612	48612	48612	48612	48612	48612
Number of Municipalities	5568	5568	5568	5568	5568	5568

$^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include election year and municipal level fix effects and the following candidate-level controls: gender, age, and level of education. All models also include the number of paragraphs in the manifesto to normalize the effects by the corpora length. "Control" corresponds to the anti-corruption leaning of the manifestos of the outsiders.

in a place that experiences shocks that increase the salience of corruption. To do so, I turn to random audits of the mayor's account conducted by the Brazilian federal government, which produce exogenous variation in the salience of corruption. Specifically, I test whether municipalities exposed to the shock experienced outsider entry at higher rates than places that were not.

Table **??** presents the results using different definitions of outsiders. It shows that the random exposure of municipalities to shocks that shift the salience of corruption leads to an increased entrance of outsider candidates. Specifically, municipalities audited within the past four years, i.e., most likely exposed to news and allegations related to political corruption, were both more likely to have at least one outsider running for mayor in the elections following right after the audit (extensive margin), and more likely to see a larger proportion of these types of candidates in the ballot (intensive margin). These effects are mostly driven by outsiders coming from new parties, for whom the effect in the extensive margin is 3 percentage points, corresponding to an increase of almost 40% relative to the control mean. When subsetting to mavericks, i.e., candidates with experience but running with new parties, the effect is slightly smaller but still statistically significant (an estimate of 2 pp, 36% from the control mean). I also find an effect for the number of outsiders running:

Table 5: The effect of being an outsider candidate on the propensity to rely on anti-corruption appeals

DV: proportion of anti-corruption paragraphs (using NLP classification algorithm)

Outsider definition:	New platform		No pol. exp.		Full outsider	
Outsider	0.004	0.002	0.004***	0.005***	0.012**	0.011*
	(0.003)	(0.003)	(0.001)	(0.001)	(0.005)	(0.006)
Center	0.003***	0.002**	0.002**	0.002*	0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outsider × Center		0.024***		0.000		0.014
		(0.008)		(0.002)		(0.012)
Left	0.027***	0.027***	0.027***	0.029***	0.027***	0.027***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Outsider × Left		-0.037**		-0.005*		-0.046**
		(0.017)		(0.003)		(0.023)
Mean DV	0.073	0.073	0.073	0.073	0.073	0.073
Control Mean DV	0.073	0.073	0.071	0.071	0.073	0.073
R-squared	0.174	0.174	0.174	0.174	0.174	0.174
Sample Size	48612	48612	48612	48612	48612	48612
Number of Municipalities	5568	5568	5568	5568	5568	5568

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include election year and municipal level fix effects and the following candidate-level controls: gender, age, and level of education. All models also include the number of paragraphs in the manifesto to normalize the effects by the corpora length. "Control" corresponds to the anti-corruption leaning of the manifestos of the outsiders.

the average number increases by 0.012, which corresponds to a substantial increase of 43% from the control mean.

I do not find effects on entrance by outsiders when defined in terms of not having prior experience in office. One possible explanation for this is that, given Brazil's questioned parties, in addition to electoral reforms that have sought to reduce party switching (Novaes 2018), it is possible that running with new parties would be perceived by citizens as a more credible signal of "outsiderness" than running with the traditional tainted ones, even if being amateurs.

I also test whether this salience shock causes greater entrance of outsiders who more heavily center their campaigns on anti-corruption appeals. To test for this, I estimate equation (2) using as the main outcome the proportion of outsiders whose manifestos scored above the mean or median value of variable *dictionary sum* across the sample. Table 13 in the Appendix depicts the results. It shows that the shock increases the share of anti-corruption-leaning outsiders, but only when the latter are defined as "full outsiders."

As a robustness check, I perform the same analysis above but using an alternative measure of outsider, suggested by Hollyer, Klašnja and Titiunik (2021), who conceptualize outsiders in Brazil as young candidates with no previous experience in office, neither as bureaucrats nor

Table 6: The effect of municipal random audits on the entrance of outsider candidates to mayoral races: experience and party platform criteria

Outsider definition:	New party		No pol. exp.		Full outsider	
Dependent Variable:	Outsider = 1	Prop. outsiders	Outsider = 1	Prop. outsiders	Outsider = 1	Prop. outsiders
Audited t-4	0.030**	0.012**	-0.003	-0.005	0.008	0.004
	(0.013)	(0.005)	(0.019)	(0.011)	(0.008)	(0.003)
Control Mean DV	0.076	0.028	0.588	0.287	0.022	0.007
R-squared	0.446	0.448	0.366	0.347	0.360	0.367
Sample Size	16342	16342	16342	16342	16342	16342
Number of Municipalities	5521	5521	5521	5521	5521	5521

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample.

as elected officials. Table 14 in the Appendix presents the results. Although the effects are not statistically significant, probably due to the large divergence in the types of candidates meeting both criteria, the sign of the effect is positive, as expected.

The evidence presented so far corresponds to the effect of random audits on the entry of outsiders (i.e., "reduced form" estimates). As a final test, then, using an instrumental variables approach, I examine the effect of *corruption revealed* on the entry of outsiders.[41] The results, summarized in Table 15 in the Appendix, show that corruption revelations, instrumented by random audits, increase the likelihood of having at least one outsider running for mayor and the proportion of these types of candidates when conceptualized as those who show up with new parties (columns 1 and 2).[42]

5.3   Are Outsiders More Frequently Elected?

One testable implication related to voter response to outsiders' entrance is that outsider candidates running in places experiencing the shock should perform better than those in places without it (H2$_b$). As discussed in the theory section, this is only a secondary implication of the theory, since in the data I can only observe general equilibrium results. In particular, in any election, outsider mayoral candidates not only face relevance and salience constraints, but also compete with the strategies that mainstream candidates may adopt in response

---

[41]This variable is measured as the proportion of audited accounts in which medium or severe corruption has been detected, according to the CGU.

[42]Note that, given the strong (almost deterministic) association between having been audited and finding corruption in the inspected accounts, whereby almost all municipal-level audits yield corruption revelations, the calculated F-statistic is very large (see Table 16).

to their entry, which include both legal ones—such as increased campaign spending—and illegal ones—such as clientelism and other electoral irregularities. This strategic response limits the ability to test the effect of the shock on the electability or electoral performance of outsiders. Given these empirical limitations, I assess this possibility as an extension.

To evaluate whether outsiders' entrance translates into their electoral success, I estimate the following candidate-level model:

$$Y_{i,m,t} = \beta_1 Outsider_{i,m,t} + \beta_2 Audit_{m,t-4} + \beta_3 Outsider_{i,m,t} \times Audit_{m,t-4} + \mathbf{X}'_{i,m,t}\psi + \gamma_m + \alpha_t + \varepsilon_{i,m,t}$$
(3)

where $Y_{i,m,t}$ correspond to either the vote share obtained by candidate i running in municipality m in election year t or whether she was elected; the remaining terms correspond to those defined for equation (1) and (2).

Results.—Before analyzing the results, two observations are worth noting. First, while the effect of running in an audited municipality is well identified, granted the randomization nature of the selection procedure, being an outsider is not. In particular, candidate-level characteristics may be associated with both their likelihood to run as an outsider and to accrue more votes. The vector $\mathbf{X}'_{i,m,t}$ accounts for the available observed characteristics but cannot do so for those that are not unobserved or included in the model. A second warning note comes with the fact that, as shown, outsiders' decision to run is affected by the shock, meaning that $Outsider_{i,m,t} = 1$, included as a regressor in equation (3), is an outcome itself. While this limits the causal interpretation of β_3, this analysis should provide suggestive evidence of whether outsiders in treated municipalities perform better than those running in non-treated ones.

Table 7 presents the estimation of equation (3).[43] Notice, first, that outsiders tend to perform worse across municipalities and elections. This negative effect is attenuated for those who are running in audited municipalities, although the difference is not statistically significant.

6 Discussion

In the previous section, I have shown results consistent with my theory that outsiders are more likely to rely on anti-corruption appeals, which I argue are a case of valence issues, to the extent that they enjoy a comparative advantage when running on these issues vis-à-vis their non-outsider counterparts (see Tables 4 and 5). I have also shown that shocks that

---

[43]I do not run a model with the most restrictive definition of outsider, being *full outsider*, as the variables included do not provide enough variation to estimate effects.

Table 7: The effect of running in an audited municipality on candidates' likelihood of getting elected and vote share by outsider status

	Elected = 1		Vote share	
	New platform	No pol. exp.	New platform	No pol. exp.
Outsider	-0.011	-0.163***	-0.024***	-0.103***
	(0.016)	(0.005)	(0.006)	(0.002)
Audited t-4	-0.004	-0.012	-0.005	-0.009*
	(0.006)	(0.008)	(0.004)	(0.005)
Outsider × Audited	0.010	0.025	0.012	0.012
	(0.045)	(0.020)	(0.018)	(0.009)
Control Mean DV	0.324	0.377	0.340	0.373
R-squared	0.087	0.109	0.291	0.328
Sample Size	48613	48613	47284	47284
Number of Municipalities	5568	5568	5568	5568

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample. *New platform* corresponds to candidates running with a new party. *No pol. exp.* corresponds to candidates with no political experience.

increase the salience of corruption led to an increase in the supply of outsider candidates in Brazil's local mayoral elections (see Table 6).

However, is this effect driven by more outsiders deciding to run when they anticipate that they have a better chance of winning due to the salience shock? Or, do they adapt their campaign strategies in response to the shock, now focusing more on anti-corruption appeals?

To study these questions, I estimate a candidate-level model in which I regress the two main measures of candidates' anti-corruption leaning on an indicator of being outsider, and interact this indicator with another that takes the value of 1 if the candidate is running in a municipality exposed to an audit in the last four years.[44] This model is meant to assess if outsiders' reliance on anti-corruption is contingent on the salience of this valence issue.

The results shown in Tables 8 and 9 (using the dictionary and NLP measures, respectively) indicate that outsider candidates are not more likely to use anti-corruption appeals if faced with a shock. In other words, they do not seem to adjust strategies in response to the shock when entering the electoral race. In contrast, non-outsider candidates do: they resort more to anti-corruption appeals only when there is a shock in the municipality they are running for.

The evidence suggests, on the one hand, that these shocks do not change outsiders' strategies, for whom focusing their campaigns on valence issues is always more strategic than doing so on positional issues (likely to be already owned by existing parties). Yet, corruption

---

[44]I include other covariates discussed in Section 4.3.

salience shocks help them, in turn, to overcome the barriers they face in contesting elections. On the other hand, I find that candidates of established parties do pivot to more heavily campaigning on anti-corruption appeals (a strategy presumably not available for them in the absence of the shock). This finding further supports the interpretation of null effects on outsiders' electoral performance, whereby non-outsiders respond to the entry of outsiders by strategically adjusting their strategies.

Table 8: The effect of being an outsider candidate in a municipality receiving a shock on the propensity to rely on anti-corruption appeals

<div align="center">

DV: proportion of anti-corruption paragraphs (using dictionary)

Outsider definition:	New platform	No pol. exp.	Full outsider
Outsider	0.008	0.015***	0.028*
	(0.008)	(0.003)	(0.016)
Audited t-4	0.013**	0.019***	0.012**
	(0.006)	(0.007)	(0.006)
Outsider × Audited	-0.028	-0.025**	-0.019
	(0.023)	(0.011)	(0.039)
Control Mean DV	0.314	0.312	0.314
R-squared	0.178	0.178	0.178
Sample Size	48613	48613	48613
Number of Municipalities	5568	5568	5568

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

</div>

Note: Robust standard errors clustered at the municipal level. Regressors include municipal level fix effects and the following candidate-level controls: gender, age, and level of education. All models also include the number of paragraphs in the manifesto to normalize the effects by the corpora length.

7   Conclusion

In this paper, I argue that outsiders are better endowed than their non-outsider counterparts to center their electoral campaigns around anti-corruption issues, or more broadly, valence issues. Therefore, I propose that shocks that increase the salience of these issues among the electorate may offer a unique opportunity and help outsiders overcome the barriers to entry they face when attempting to run for office. I test this theory by exploiting exogenous shocks that increase the salience of corruption among certain Brazilian municipalities following random annual audits conducted by the federal government.

First, the empirical evidence corroborates that outsiders—especially those who adopt moderate positions—are more likely to run on valence issues, here operationalized as anti-corruption appeals. I show this using more than forty-nine thousand manifestos of Brazilian candidates running in the 2012, 2016, and 2020 mayoral elections as well as an original

Table 9: The effect of being an outsider candidate in a municipality receiving a shock on the propensity to rely on anti-corruption appeals

DV: proportion of anti-corruption paragraphs using NLP classification algorithm

Outsider definition:	New platform	No pol. exp.	Full outsider
Outsider	0.002	0.005***	0.011*
	(0.003)	(0.001)	(0.006)
Audited t-4	0.004*	0.005**	0.004*
	(0.002)	(0.002)	(0.002)
Outsider × Audited	-0.001	-0.006	-0.007
	(0.009)	(0.005)	(0.015)
Control Mean DV	0.073	0.071	0.073
R-squared	0.174	0.174	0.174
Sample Size	48613	48613	48613
Number of Municipalities	5568	5568	5568

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal level fix effects and the following candidate-level controls: gender, age, and level of education. All models also include the number of paragraphs in the manifesto to normalize the effects by the corpora length.

fine-grained database on candidates' characteristics.

Second, testing the second implication of my theory, I demonstrate that municipalities that have been audited by the CGU within the past four years before the election year are more likely to experience a higher entrance of outsiders. Furthermore, I show that the shock increases the supply of this type of candidate but does not make outsiders more likely to adopt anti-corruption appeals. Ultimately, these results suggest that outsiders do update their decision to enter the electoral race when encountering these shocks, and not their decision to campaign on anti-corruption appeals.

Finally, I find no effect on the electoral prospects of outsiders, suggesting that, in addition to entry, these candidates must overcome other sets of obstacles to win office. In particular, additional analyses show that non-outsiders respond to the anticipated entry of outsiders in response to this shock by incorporating anti-corruption appeals more strongly—and presumably also adapting their campaign strategies in other ways.

By showing that outsiders face barriers to entry that certain shocks help to overcome, this paper has two important, albeit divergent, consequences. First, the literature on political accountability has shown that a key precondition for voters to be willing to hold politicians accountable is the availability of "clean alternatives" (at least those perceived as such, see Agerberg 2020). Works in this strand usually look at vote share as a measure of voter willingness to condemn and punish corrupt politicians.[45] Yet, this corresponds to a general

---

[45]See Grossman and Slough (2021) and Dunning et al. (2019) provide good reviews of this literature. The latter also presents the results of six experiments with a common treatment arm implemented in five

equilibrium effect, which calls for a more careful exploration of both the conditions that facilitate the entry of such alternatives–a gap in the literature that this paper contributes to fill–and the responses of establishment candidates. Additionally, beyond shocks, this paper opens the door to future research on other factors, e.g. institutional arrangements, that may also reduce barriers to entry for these new alternative candidates.

Second, from a less optimistic stance, corruption scandals, which increase the salience of this issue, can further allow the entrance of other types of outsiders associated with poorer democratic records. For instance, a growing number of works have shown the consequences of electing anti-establishment outsiders on the stability of liberal democracy (e.g., Carreras 2014; Hollyer, Klašnja and Titiunik 2021; Levitsky and Cameron 2003; Linz 1990). Moreover, in recent decades, we have seen the breakthrough of candidates who mobilize voters on their discontent with their political system and the political elite. The latter class of outsiders fuels and capitalizes on citizen discontent with the risk of further eroding the institutional checks and balances that presumably constrain their unlimited governing aspirations (Uscinski et al. 2021; Levitsky and Loxton 2013). By shedding light on the conditions that favor outsiders' entrance, this paper thus contributes to our understanding of the roots of democratic backsliding. Moreover, the results also help rationalize recent trends, both in Latin America and elsewhere, of growing citizen anti-establishment sentiment. While corruption is a key driver of this sentiment,[46] the widespread adoption of an anti-corruption appeal— both by outsiders *and* insiders—when faced with shocks that raise the salience of this issue, could explain the general discrediting of the political class and the loss of credibility of this discourse in contexts of widespread corruption.

While the evidence presented in this paper accounts for the entrance of local outsiders, the findings of this paper have implications for our broader understanding of outsider candidates. On the one hand, outsider mayoral candidates are commonplace too. Examples in Latin America include elected mayors in prominent cities (e.g., Claudia López, elected mayor of Bogotá in 2019, Daniel Quintero, elected mayor of Medellín in 2019, Alexis Argüello, elected mayor of Managua in 2008, or Ricardo Belmont, elected mayor of Lima in 1990). In Brazil, the case I investigate in this paper, two of the most important cities of the country, Saõ

---

different countries to test for the effect of information on accountability, which used different proxies of vote choice as the main outcome of interest.

[46]In a project, together with Loreto Cox, we investigate the drivers of anti-establishment citizen sentiment in Colombia in the context of the 2022 presidential elections. To do so, we included an open-ended question where we elicited responses from citizens on the reasons why they feel anger with politicians. The most frequent words in the responses indicate that corruption is one of the most important reasons behind the high anti-establishment sentiment in this country (see word cloud 9 in Appendix I, although more thorough preliminary analyses of these corpora using supervised and semi-supervised NLP methods point in the same direction).

Paulo and Rio de Janeiro, elected outsiders for mayor: Joao Doria and Marcelo Crivella, respectively. Moreover, the local office has also allowed some outsiders as a vehicle to gain visibility and breakthrough in national politics (e.g., 2022 presidential candidate Rodolfo Hernández in Colombia, who had been mayor of Bucaramanga, a major city in the country). Many even make it to become presidents: this has been the case of Mauricio Macri in Argentina (elected president of Argentina in 2015), Andrés Manuel López Obrador (elected president of Mexico in 2018, but who first ran in 2006 when he resigned his position as mayor of Mexico City to run for the presidency). Outside the region, other examples include Mateo Renzi, who had been mayor of Florence and became Italy's prime minister in 2014, and Klaus Iohannis, Romania's current president and previously mayor of the city of Sibiu.

References

Abedi, Amir. 2004. *Anti-political establishment parties: A comparative analysis.* Routledge.

Agerberg, Mattias. 2020. "The lesser evil? Corruption voting and the importance of clean alternatives." *Comparative Political Studies* 53(2):253–287.

Ansolabehere, Stephen and James M Snyder. 2000. "Valence politics and equilibrium in spatial election models." *Public choice* 103(3):327–336.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubín. 2021. "Prior's rule: When do malfeasance revelations help or hurt incumbent parties?" *Journal of the European Economic Association* Forthcoming.

Ashworth, Scott and Ethan Bueno de Mesquita. 2009. "Elections with platform and valence competition." *Games and Economic Behavior* 67(1):191–216.

Avis, Eric, Claudio Ferraz and Frederico Finan. 2018. "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians." *Journal of Political Economy* 126(5):1912–1964.

Bågenholm, Andreas. 2013. "The electoral fate and policy impact of "anti-corruption parties" in Central and Eastern Europe." *Human Affairs* 23(2):174–195.

Bågenholm, Andreas and Nicholas Charron. 2014. "Do politics in Europe benefit from politicising corruption?" *West European Politics* 37(5):903–931.

Bågenholm, Andreas and Nicholas Charron. 2015. Anti-corruption parties and good government. In *Elites, Institutions and the Quality of Government.* Springer pp. 263–282.

Barr, Robert R. 2009. "Populists, outsiders and anti-establishment politics." *Party Politics* 15(1):29–48.

Berliner, Daniel and Aaron Erlich. 2015. "Competing for transparency: political competition and institutional reform in Mexican states." *American Political Science Review* 109(1):110–128.

Binzer Hobolt, Sara and Robert Klemmensen. 2008. "Government responsiveness and political competition in comparative perspective." *Comparative Political Studies* 41(3):309–337.

Blattman, Christopher, Horacio Larreguy, Benjamin Marx and Otis R Reid. 2019. Eat widely, vote wisely? Lessons from a campaign against vote buying in Uganda. Technical report National Bureau of Economic Research.

Bolognesi, Bruno, Ednaldo Ribeiro and Adriano Codato. 2021. "Uma nova classificação ideológica dos partidos políticos brasileiros.".

Carreras, Miguel. 2012. "The Rise of Outsiders in Latin America, 1980–2010." *Comparative Political Studies* 45(12):1451–1482.
URL: *http://dx.doi.org/10.1177/0010414012445753*

Carreras, Miguel. 2014. "Outsiders and Executive-Legislative Conflict in Latin America." *Latin American Politics and Society* 56(3):70–92.

Carreras, Miguel. 2017. "Institutions, governmental performance and the rise of political newcomers." *European Journal of Political Research* 56(2):364–380.

CEP. 2016. "Estudio Nacional de Opinión Pública Número 77, Julio-Agosto 2016." https://www.cepchile.cl/cep/encuestas-cep/encuestas-2010-2019/estudio-nacional-de-opinion-publica-julio-agosto-2016. Centro de Estudios Públicos.

Corrales, Javier. 2008. "Latin America's Neocaudillismo: Ex-Presidents and Newcomers Running for President... and Winning." *Latin American Politics and Society* 50(3):1–35.

Curini, Luigi. 2018. *Corruption, Ideology, and Populism: The Rise of Valence Political Campaigning.* London: Palgrave.

Downs, Anthony. 1957. "An economic theory of political action in a democracy." *Journal of Political Economy* 65(2):135–150.

Droste, Luigi. 2021. "Feeling left behind by political decisionmakers: Anti-establishment sentiment in contemporary democracies." *Politics and Governance* 9(3):288–300.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D Hyde, Craig McIntosh and Gareth Nellis. 2019. *Information, accountability, and cumulative learning: Lessons from Metaketa I.* Cambridge University Press.

Engler, Sarah. 2016. "Corruption and electoral support for new political parties in central and eastern Europe." *West European Politics* 39(2):278–304.

Engler, Sarah. 2020*a*. "Centrist anti-establishment parties and their protest voters: more than a superficial romance?" *European Political Science Review* 12(3):307–325.

Engler, Sarah. 2020*b*. ""Fighting corruption" or "fighting the corrupt elite"? Politicizing corruption within and beyond the populist divide." *Democratization* 27(4):643–661.

Engler, Sarah, Bartek Pytlas and Kevin Deegan-Krause. 2019. "Assessing the diversity of anti-establishment and populist politics in Central and Eastern Europe." *West European Politics* 42(6):1310–1336.

Falleti, Tulia G. 2010. *Decentralization and subnational politics in Latin America.* Cambridge University Press.

Ferraz, Claudio and Frederico Finan. 2008. "Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes." *The Quarterly journal of economics* 123(2):703–745.

Ferraz, Claudio and Frederico Finan. 2011. "Electoral accountability and corruption: Evidence from the audits of local governments." *American Economic Review* 101(4):1274–1311.

Ferreira, Francisco HG, Julian Messina, Jamele Rigolini, Luis-Felipe López-Calva, Maria Ana Lugo, Renos Vakis, Luis Felipe Ló et al. 2012. *Economic mobility and the rise of the Latin American middle class.* World Bank Publications.

Gerbaudo, Paolo. 2014. "Populism 2.0: Social media activism, the generic Internet user and interactive direct democracy." *Social media, politics and the state: Protests, revolutions, riots, crime and policing in the age of Facebook, Twitter and YouTube* pp. 67–87.

Graham, Matthew H and Milan W Svolik. 2020. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *American Political Science Review* 114(2):392–409.

Green, Jane. 2007. "When voters and parties agree: Valence issues and party competition." *Political Studies* 55(3):629–655.

Grossman, Guy and Tara Slough. 2021. "Government Responsiveness in Developing Countries." *Annual Review of Political Science* 25.

Hanley, Seán and Allan Sikk. 2016. "Economy, corruption or floating voters? Explaining the breakthroughs of anti-establishment reform parties in eastern Europe." *Party Politics* 22(4):522–533.

Hollyer, James R, Marko Klašnja and Rocío Titiunik. 2021. "Parties as Disciplinarians: Charisma and Commitment Problems in Programmatic Campaigning." *American Journal of Political Science* .

Ipsos. 2021. "Broken-System Sentiment in 2021: Populism, Anti-elitism and Nativism." Ipsos.
URL: *https://bit.ly/3zURgSo*

Keefer, Philip. 2007. "Clientelism, credibility, and the policy choices of young democracies." *American journal of political science* 51(4):804–821.

Kendall, Chad, Tommaso Nannicini and Francesco Trebbi. 2015. "How do voters respond to information? Evidence from a randomized campaign." *American Economic Review* 105(1):322–53.

Klašnja, Marko, Joshua A. Tucker and Kevin Deegan-Krause. 2016. "Pocketbook vs. Sociotropic Corruption Voting." *British Journal of Political Science* 46(1):67–94.

Levitsky, Steven and James Loxton. 2013. "Populism and competitive authoritarianism in the Andes." *Democratization* 20(1):107–136.

Levitsky, Steven, James Loxton, Brandon Van Dyck and Jorge I Domínguez. 2016. *Challenges of party-building in Latin America.* Cambridge University Press.

Levitsky, Steven and Maxwell A Cameron. 2003. "Democracy without parties? Political parties and regime change in Fujimori's Peru." *Latin American Politics and Society* 45(3):1–33.

Linz, Juan J. 1990. "The perils of presidentialism." *Journal of democracy* 1(1):51–69.

Litschig, Stephan and Yves Zamboni. 2008. "Judicial presence and rent extraction." *GRIPS Discussion Paper 19-20* .

Lupu, Noam, Rachel Beatty Riedl, Jaimie Bleck and Nicolas van de Walle. 2013. "Valence issues in African elections: Navigating uncertainty and the weight of the past." *Comparative Political Studies* 46(11):1394–1421.

Macdonald, Stuart Elaine and George Rabinowitz. 1998. "Solving the paradox of non-convergence: Valence, position, and direction in democratic politics." *Electoral Studies* 17(3):281–300.

Mudde, Cas. 2004. "The populist Zeitgeist." *Government and opposition* 39(4):541–563.

Novaes, Lucas M. 2018. "Disloyal brokers and weak parties." *American Journal of Political Science* 62(1):84–98.

Polk, Jonathan, Jan Rovny, Ryan Bakker, Erica Edwards, Liesbet Hooghe, Seth Jolly, Jelle Koedam, Filip Kostelka, Gary Marks, Gijs Schumacher et al. 2017. "Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data." *Research & Politics* 4(1):2053168016686915.

Pop-Eleches, Grigore. 2010. "Throwing out the Bums: Protest Voting and Unorthodox Parties after Communism." *World Politics* 62(2):221–260.

Reinemann, Carsten, Toril Aalberg and Frank Esser. 2016. Populist political communication: Toward a model of its causes, forms, and effects. In *Populist political communication in Europe.* Routledge pp. 22–36.

Rodrik, Dani. 2020. "Why does globalization fuel populism? Economics, culture, and the rise of right-wing populism." *Annual Review of Economics* 13.

Samuels, David. 2003. *Ambition, federalism, and legislative politics in Brazil.* Cambridge University Press.

Samuels, David and Cesar Zucco. 2014. "The power of partisanship in Brazil: Evidence from survey experiments." *American Journal of Political Science* 58(1):212–225.

Samuels, David and Cesar Zucco. 2016. Party-building in Brazil. In *Challenges of Party-Building in Latin America.* Cambridge University Press pp. 331–355.

Schedler, Andreas. 1996. "Anti-political-establishment parties." *Party politics* 2(3):291–312.

Seabra, Sérgio. 2018. "How auditing can be directed against corruption: the case of Brazil's municipality-facing auditing programme." Building integrity Working Papers Series. Blavatnik School of Government, University of Oxford.

Serra, Gilles. 2018. "The electoral strategies of a populist candidate: Does charisma discourage experience and encourage extremism?" *Journal of Theoretical Politics* 30(1):45–73.

Stokes, Donald. 1992. "Valence politics." *Electoral politics* pp. 141–164.

Uscinski, Joseph E, Adam M Enders, Michelle I Seelig, Casey A Klofstad, John R Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne and Manohar N Murthi. 2021. "American politics in two dimensions: Partisan and ideological identities versus anti-establishment orientations." *American Journal of Political Science* 65(4):877–895.

Weitz-Shapiro, Rebecca. 2014. *Curbing clientelism in Argentina: Politics, poverty, and social policy.* Cambridge University Press.

Appendix

A Comparison of the characteristics of the general population of Brazil with those of outsiders and non-outsiders

Table 10: Socio-demographic characteristics Brazilian population based on 2010 census

Sex	N	Percentage
Male	93,406,990	49%
Female	97,348,809	51%
Age group		
0-4	13,796,158	7%
5-9	14,969,375	8%
10-14	17,166,761	9%
15-19	16,990,872	9%
20-24	17,245,192	9%
25-29	17,104,414	9%
30-34	15,744,512	8%
35-39	13,888,579	7%
40-44	13,009,364	7%
45-49	11,833,352	6%
50-54	10,140,402	5%
55-59	8,276,221	4%
60-64	6,509,120	3%
65-69	4,840,810	3%
70-74	3,741,636	2%
75-79	2,563,447	1%
80-89	2,486,455	1%
90-99	424,893	0%
100 or more	24,236	0%
Education		
No education and elementary incomplete	81,386,577	43%
Elementary school complete and incomplete high school	28,178,794	15%
High school and college incomplete	37,980,515	20%
Higher education complete	13,463,757	7%
Not determined	971,655	1%

Source: IGBE

Table 11: Socio-demographic characteristics mayoral candidates

Election year:	2012				2016				2020				All			
Candidate type:	Outsider		Insider		Outsider		Insider		Outsider		Insider		Outsider		Insider	
Statistic:	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Sex																
Male	5,454	84.77	8,469	88.04	4,986	85.89	9,933	87.60	6,146	85.77	10,664	87.01	16,586	85.47	29,066	87.51
Female	980	15.23	1,150	11.96	819	14.11	1,406	12.40	1,020	14.23	1,592	12.99	2,819	14.53	4,148	12.49
Age group																
20-24	87	1.34	9	0.09	71	1.22	13	0.11	69	0.96	19	0.16	227	1.17	41	0.12
25-29	270	4.15	145	1.51	226	3.89	173	1.53	204	2.85	174	1.42	700	3.59	492	1.48
30-34	509	7.82	480	4.99	462	7.94	561	4.95	464	6.48	623	5.08	1,435	7.37	1,664	5.01
35-39	729	11.20	919	9.55	681	11.71	1,190	10.49	843	11.78	1,265	10.32	2,253	11.57	3,374	10.16
40-44	983	15.10	1,411	14.67	810	13.93	1,610	14.20	1,051	14.68	1,776	14.49	2,844	14.60	4,797	14.44
45-49	1,155	17.75	1,977	20.55	892	15.34	2,013	17.75	1,074	15.01	2,003	16.34	3,121	16.02	5,993	18.04
50-54	1,001	15.38	1,851	19.24	950	16.33	2,143	18.90	1,043	14.57	2,148	17.53	2,994	15.37	6,142	18.49
55-59	785	12.06	1,339	13.92	675	11.61	1,615	14.24	948	13.25	1,867	15.23	2,408	12.36	4,821	14.51
60-64	488	7.50	865	8.99	550	9.46	1,091	9.62	682	9.53	1,202	9.81	1,720	8.83	3,158	9.51
65-69	265	4.07	384	3.99	270	4.64	563	4.97	445	6.22	688	5.61	980	5.03	1,635	4.92
70-74	117	1.80	166	1.73	149	2.56	264	2.33	224	3.13	326	2.66	490	2.52	756	2.28
75-79	32	0.49	50	0.52	51	0.88	77	0.68	81	1.13	126	1.03	164	0.84	253	0.76
80-89	12	0.18	23	0.24	17	0.29	25	0.22	28	0.39	39	0.32	57	0.29	87	0.26
90 or more	75	1.15			12	0.21	1	0.01	1	0.01			88	0.45	1	0.00
Education																
Elementary school complete	1,681	25.83	2,587	26.89	1,444	24.83	2,996	26.42	1,703	23.77	3,029	24.71	4,828	24.77	8,612	25.93
Elementary school incomplete	455	6.99	675	7.02	358	6.16	668	5.89	342	4.77	640	5.22	1,155	5.93	1,983	5.97
High school complete	66	1.01	91	0.95	69	1.19	91	0.80	95	1.33	119	0.97	230	1.18	301	0.91
High school incomplete	205	3.15	270	2.81	159	2.73	311	2.74	188	2.62	293	2.39	552	2.83	874	2.63
Higher education complete	3,181	48.87	4,671	48.56	3,000	51.58	5,893	51.97	4,009	55.94	6,832	55.74	10,190	52.28	17,396	52.38
Higher education incomplete	75	1.15			11	0.19							86	0.44		
Reads and writes	423	6.50	631	6.56	389	6.69	712	6.28	388	5.41	682	5.56	1,200	6.16	2,025	6.10
Undisclosed	423	6.50	694	7.21	386	6.64	668	5.89	441	6.15	661	5.39	1,250	6.41	2,023	6.09
Total	6,509	100.00	9,619	100.00	5,816	100.00	11,339	100.00	7,166	100.00	12,256	100.00	19,491	100.00	33,214	100.00

Source: TSE available at http://www.tse.jus.br/eleicoes/estatisticas/estatisticas-eleitorais and author's own calculations. Outsiders include *mavericks*, *amateurs*, and *full outsiders*, as defined by Carreras (2012).

B Additional analyses

Table 12: The effect of municipal random audits on the entrance of outsider candidates to mayoral races: Carreras (2012) typology

Outsider def:	Maverick		Amateur		Full outsider	
Dependent Variable:	Outsider = 1	Prop. out-siders	Outsider = 1	Prop. out-siders	Outsider = 1	Prop. out-siders
Audited t-4	0.020*	0.009*	-0.011	-0.009	0.008	0.004
	(0.011)	(0.004)	(0.019)	(0.011)	(0.008)	(0.003)
Control Mean DV	0.056	0.021	0.578	0.279	0.022	0.007
R-squared	0.427	0.427	0.364	0.346	0.360	0.367
Sample Size	16342	16342	16342	16342	16342	16342
Number of Municipalities	5521	5521	5521	5521	5521	5521

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample. *Maverick* corresponds to candidates with political experience but running with a new party. *Amateur* corresponds to candidates with no political experience but running with an existing party. *Full outsider* corresponds to candidates with no political experience and running with a new party.

Table 13: The effect of municipal random audits on the entrance of outsider candidates with high levels of intensity of anti-corruption appeals in their manifestos

Outsider def:	New platform		No pol. experience		Full outsider	
DV: *dictionary sum*	Above mean	Above median	Above mean	Above median	Above mean	Above median
Audited t-4	-0.001	0.002	-0.000	-0.008	0.005*	0.001
	(0.006)	(0.002)	(0.001)	(0.008)	(0.003)	(0.002)
Control Mean DV	0.085	0.005	0.002	0.137	0.009	0.003
R-squared	0.412	0.351	0.338	0.405	0.365	0.354
Sample Size	16342	16342	16342	16342	16342	16342
Number of Municipalities	5521	5521	5521	5521	5521	5521

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample. *"Above mean"* corresponds to candidates whose manifesto's number of paragraphs classified as anti-corruption based on the criterion that appears in the anti-corruption dictionary is above the mean number across the sample. *"Above median"* measures the same variable but takes the median value as the reference statistic.

Table 14: The effect of municipal random audits on the entrance of outsider candidates to mayoral races: Hollyer, Klašnja and Titiunik (2021)

Outsider def:	Young and Amateur		Young	
Dependent variable:	Outsider=1	Prop. out-siders	Outsider=1	Prop. out-siders
Audited t-4	0.007	0.002	0.012	0.004
	(0.016)	(0.007)	(0.017)	(0.008)
Control Mean DV	0.202	0.077	0.262	0.104
R-squared	0.408	0.414	0.444	0.456
Sample Size	16342	16342	16342	16342
Number of Municipalities	5521	5521	5521	5521

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample. *Young and Amateur* corresponds to candidates who (i) are 35 years old or younger and (ii) have no political experience. *Young* corresponds to candidates who only meet criterion (i).

Table 15: The effect of municipal corruption revelations on the entrance of outsider candidates to mayoral races: IV estimates

Outsider definition:	New platform		No pol. experience		Full outsider	
Dependent variable:	Outsider=1	Prop. out-siders	Outsider=1	Prop. out-siders	Outsider=1	Prop. out-siders
Corruption Revelation	0.224**	0.091**	-0.022	-0.039	0.057	0.027
	(0.099)	(0.039)	(0.145)	(0.085)	(0.059)	(0.023)
Sample Size	16,342	16,342	16,342	16,342	16,342	16,342

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample. *New platform* corresponds to candidates running with a new party. *No pol. exp.* corresponds to candidates with no political experience. *Full outsider* corresponds to candidates with no political experience and running with a new party.

Table 16: The effect of municipal random audits on corruption revelations: First-stage IV estimates

	Corruption Revelation
Audited t-4	0.133***
	(0.004)
Sample Size	16,342
CD Wald F	16,639
SW S stat.	5.181

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The table reports the *Stock-Wright* LM S-statistic for weak-instrument-robust inference, and *Cragg-Donald Wald* F-statistic for weak identification test.

Table 17: The effect of running in an audited municipality on candidates' likelihood of getting elected, vote share and total number of votes by outsider status and type of candidate

	Elected = 1			Vote share			Total votes (standardized)		
	New platform	No pol. exp.	Full outsider	New platform	No pol. exp.	Full outsider	New platform	No pol. exp.	Full outsider
Outsider	-0.108***	-0.246***	-0.189***	-0.090***	-0.152***	-0.139***	-0.360***	-0.459***	-0.546***
	(0.02)	(0.01)	(0.03)	(0.01)	(0.00)	(0.01)	(0.03)	(0.02)	(0.05)
Reelected	0.122***	0.081***	0.123***	0.089***	0.068***	0.088***	0.249***	0.194***	0.247***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
Challenger	-0.128***	-0.119***	-0.126***	-0.088***	-0.075***	-0.088***	-0.278***	-0.229***	-0.279***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
Audited t-4	-0.021	-0.040	-0.022	0.009	0.008	0.008	0.005	0.026	0.005
	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.04)	(0.05)	(0.04)
Outsider × Audited	0.028	0.054	-0.025	0.001	-0.004	-0.008	0.123	-0.045	-0.059
	(0.05)	(0.03)	(0.08)	(0.02)	(0.01)	(0.04)	(0.10)	(0.06)	(0.13)
Reelected × Audited	0.013	0.026	0.013	-0.009	-0.007	-0.008	0.050	0.038	0.052
	0.04	0.04	0.04	0.01	0.01	0.01	0.09	0.09	0.09
Challenger × Audited	0.024	0.036	0.027	-0.015	-0.019*	-0.013	-0.022	-0.037	-0.009
	0.03	0.03	0.03	0.01	0.01	0.01	0.05	0.05	0.05
Control Mean DV	0.325	0.378	0.327	0.340	0.374	0.342	0.006	0.078	0.007
R-squared	0.107	0.121	0.107	0.334	0.353	0.334	0.382	0.390	0.383
Sample Size	48,239	48,239	48,239	46,918	46,918	46,918	45,670	45,670	45,670
Number of Municipalities	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568	5,568

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors clustered at the municipal level. Regressors include municipal, election year, and combinations of state crossed with election years fix effects. Municipalities not eligible to be audited, based on the population and capital city criteria, are excluded from the sample. *New platform* corresponds to candidates running with a new party. *No pol. exp.* corresponds to candidates with no political experience. *Full outsider* corresponds to candidates with no political experience and running with a new party.

49

C   Coding procedure used in the pre-processing of candidates' manifestos

In this section, I describe in detail the pre-processing of candidates' manifestos in preparation for their subsequent classification in terms of their anti-corruption leaning. The processing of the universe of manifestos of Brazilian mayoral candidates was designed and coordinated by Natalia Garbiras-Díaz, and implemented with the research assistance of Javier Herrera.

C.1   Valid paragraphs

The following filters were applied to the universe of paragraphs to be considered valid and thus be entered into the classification procedure.

1. First, I computed the total length of the paragraph in terms of its number of characters; and those with a length of less than five characters were eliminated.

2. Second, I carried out a *tokenization* process, where each paragraph was separated into words for better analysis. Taking the words, I eliminated "stop words" with numerical values and special characters.

3. Third, I counted the unique words within the paragraph and those composed by only five or fewer different words were eliminated, thus avoiding undetected text patterns.

4. Finally, I eliminated paragraphs with atypical sizes. To do so, for each year, I calculated the distribution of paragraphs' length and then trimmed paragraphs at the top 1%.

Overall, following this four-step filtering process, I only kept paragraphs with at least five unique words—excluding *stop words* and with words of at least three characters.

Figure 5: Candidates' manifestos paragraphs examples

```
PLANO DE GOVERNO DO CANDIDATO A PREFEITO DE
JABOATÃO DOS GUARARAPES PELO PMN PAULO SERGIO
RIBEIRO VAREJÃO

Para alcançar um desenvolvimento pleno, Jaboatão dos Guararapes, por
setor, tem que ter:

EDUCAÇÃO - Uma Cidade Universitária, com todas as faculdades
tradicionais, como Direito, Medicina, Engenharia, Economia, e pelo menos
mais duas Escolas Técnicas. Afora isso, tem que ter, em cada distrito, escolas
de porte e estrutura como o GEO, que é privada, para ensino integral e
prática de desportos vários, como atletismo, natação, futebol, basket, futsal,
judô ou jiu jitsu, para a disciplina, dentre outros. E além desses
equipamentos de ensino, espaço para danças diversas, incluindo-se aí o
ballet, as artes, e também conhecimentos da cultura. Tudo isso aliado a uma
política salarial de valorização do professor, e também a um processo
constante de capacitação.

DESENVOLVIMENTO ECONOMICO — Distrito Industrial, em área
destinada a este fim, com lotes industriais para serem dados em comodato
para novas industrias que queiram se implantar neste município, e adoção de
lei de incentivos fiscais, observada a Lei de Responsabilidade Fiscal. Este
distrito irá gerar emprego e a renda, e estimulará outras industrias e
empresas de serviços menores em seu entorno, contribuindo para o
desenvolvimento municipal.
```

Note: Document 8 of year 2012. Based on the processing rules, in this example, there would be three valid paragraphs.

C.2 A stratified sample of paragraphs to code manually

I randomly sampled a selection of the paragraphs to be then hand-coded by the group of research in terms of whether they refer or not to an anti-corruption or anti-establishment appeals. To ensure coding paragraphs from candidates with a wide variety of profiles (i.e., sampling manifestos), I selected the random sample of paragraphs to hand-code, stratifying on variables deemed to be prognostic of different styles and incidence of anti-corruption rhetoric across mayoral candidates. In this section, I describe how I conducted this procedure.

After the process of selecting valid paragraphs, I built a database grouping, in each election year, the manifestos for each candidate. The stratification was made according to the following criteria: party classification in the political spectrum (right, left, center-right, center-left, center), geographical region, and quantiles of the distribution of the manifesto's length.

In the sampling process, I attributed a probability of 0.35 to selecting a manifesto from the two highest quintiles of the length distribution, and a probability of 0.1 to the manifestos in the three lowest quintiles. Given the filtering process described above, the larger manifestos were more likely to have more informative content for the investigation of political rhetoric,

51

such that higher relative weights were attributed to larger manifestos.

In each election year, a random sample of 3% of candidates was drawn from the universe of filtered paragraphs. The final database included only the manifestos from the selected candidates. Again, to favor higher informational content in the selection, a higher relative weight was attributed to longer paragraphs. I attributed a 0.4 probability of selection to the two highest quartiles, and a 0.1 probability to the paragraphs belonging to the three lowest quartiles.

C.3 Manual classification

The dictionary of anti-corruption related words was created in the following process. First, a group of six research assistants (RA) were hired to manually classify paragraphs from the manifestos of mayor candidates running in the 2012, 2016 and 2020 municipal elections in Brazil. The RAs were trained and instructed on the meaning of an anti-corruption rhetoric.[47] The manual coding was initially coordinated in a "pilot" training session, in which a sample of paragraphs extracted from the manifestos was coded by more than one person. This initial sample was excluded from the subsequent analysis.

The coding was anonymized throughout the whole process, that is, the RAs were never shown any information identifying the candidate to which the manifesto belonged. Furthermore, the RAs worked independently the whole time.

To ensure the validity of the hand-coding procedure, I selected a "redundant" sample of paragraphs to be coded by more than one RA, usually two. This verification set allowed me to assess the reliability of the coding procedure as well as the understanding of the concepts across RAs.

The consistency of the procedure is depicted in Figure 6, which indicates a high intercoder reliability for the classification of paragraphs included in this "redundant sample." Specifically, it shows that there was a perfect match in 895 out of the 945 paragraphs that were coded by more than one RA (corresponding to 94%).

Then, the RAs manually classified the random and stratified sample extracted from the manifestos by the RAs. With this procedure, I created an anti-corruption dictionary, containing words and expressions associated to paragraphs that were classified as containing anti-corruption stances by the RAs (see Figure 7 and Table 18). Afterwards, this dictionary was used to process the information on the full database with the manifesto documents. Three different variables were created: dictionary, dictionary dummy and a dictionary proportion. They indicate, respectively: sum of the number of times that one or more anti-

---

[47]The training and classification also included anti-establishment appeals, which are not analyzed here.

Figure 6: Classification of Redundant Sample
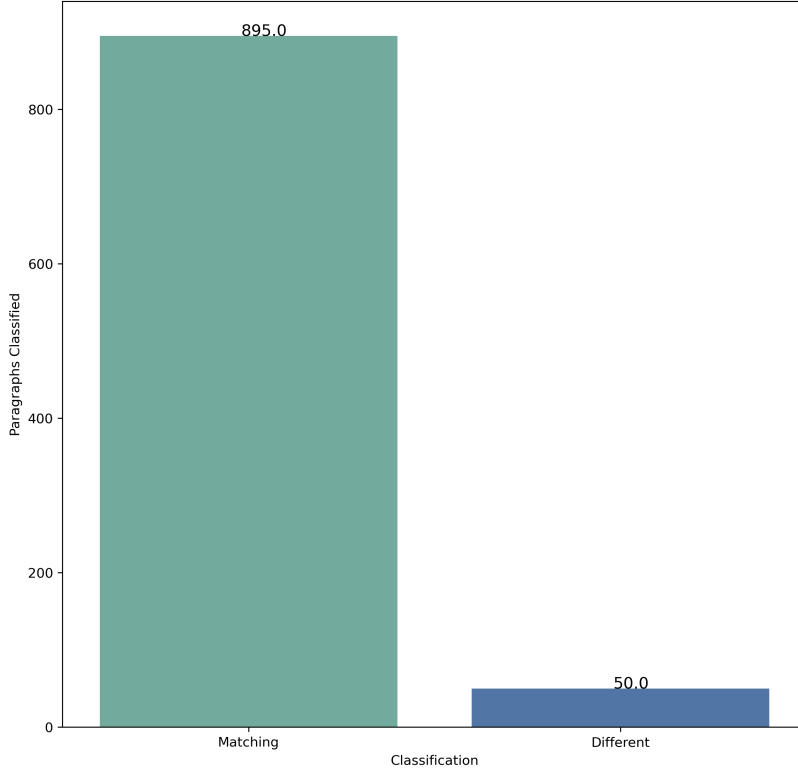
corruption words occur in the manifesto, at the paragraph level; manifestos that contained at least one of the anti-corruption entries in the dictionary; and the number of anti-corruption words identified over the number of total paragraphs in the manifesto.

D   Coding procedure and codebook for the manual classification of paragraphs

This codebook has been prepared by Natalia Garbiras-Díaz and Renata Avila. This is the original manual RAs were given during the training sessions.

D.1   Some preliminary concepts

The anti-establishment rhetoric is implemented through a challenge to established politics or to the "elites." It is a challenge to the class of individuals yielding power, an "us versus them" approach, opposed to the class of individuals wielding power (Barr 2009). Anti-establishment opposition focuses on attacks not merely on incumbent governments but on "established politics" or all parties (Schedler 1996; Engler, Pytlas and Deegan-Krause 2019).

Figure 7: Anti-corruption dictionary word cloud (manual classification)

It is therefore important to note that anti-establishment rhetoric is usually different from specific anti-incumbent claims.

Furthermore, anti-establishment rhetoric may or may not be accompanied by other elements that comprise the definition of populism, as in Mudde (2004), i.e., people-centred statements and invocations of general will. Parties that rely on anti-establishment discourse can be found across the ideological spectrum, including the center, and are not necessarily linked to other elements of populist discourse (Engler, Pytlas and Deegan-Krause 2019).

The anti-corruption rhetoric emphasizes the need to eradicate corruption from the political system. Anti-corruption stances may be used alongside an anti-establishment approach, as parties attacking the political establishment usually politicize corruption, stating that the political elite is dishonest and only-driven by self-interest. However, anti-corruption rhetoric can also feature prominently as part of a "good governance framework," without attacks to the political elite. Therefore, candidates may also approach the fight against corruption

Table 18: Examples of anti-corruption paragraphs (manual classification)

Portuguese	English
- "Reforçar os elos entre governo e sociedade radicalizando a democracia no planejamento eficiência transparência e controle da administração municipal governando com austeridade."	- "Strengthen the links between government and society radicalizing democracy in planning, efficiency, transparency, and control of the municipal administration by governing with austerity."
- "Garantir a participação popular na formulação e acompanhamento das políticas públicas, praticar a ética e combater a corrupção por meio de auditorias estratégicas, controle interno e monitoramento dos serviços públicos inclusive terceirizados de forma transparente."	- "Ensure popular participation in the formulation and monitoring of public policies, practice ethics, and fight corruption through strategic audits, internal control, and monitoring of public services including outsourced services in a transparent manner."
- "Publicar nos meios de comunicação o balanço geral do município bem como os relatórios exigidos pela Lei de Responsabilidade Fiscal LRF em linguagem objetiva para que o cidadão compreenda comfacilidade. Alimentar constantemente o Portal da Transparência com as informações de cada secretaria. Promover o Planejamento de Gestão Participativa."	- "Publish in the media the general balance sheet of the municipality as well as the reports required by the LRF Fiscal Responsibility Law in objective language so that citizens can easily understand. Constantly feed the Transparency Portal with information from each secretariat. Promote Participatory Management Planning."

as means of achieving economic prosperity, promoting values such as state efficiency and transparency (Engler 2020b).

As explained above, political science research shows that differences between anti- establishment and anti-incumbent rhetoric are salient in many contexts. However, given the choice of the method to analyze candidate rhetoric, i.e., through their manifestos, some considerations must be made. The manifestos (*programas* de governo in Portuguese) are presented to the Superior Electoral Court as official campaign documents and published on the Court's website alongside any criminal records that the candidate may have. It is reasonable to expect that, in this type of document, candidates might have incentives to tone down attacks on broadly defined "elites" or even on specific groups, while nevertheless relying (more heavily) on this type of language in other contexts of campaigning (e.g., public rallies or media advertisements). This would be consistent with research showing that social media is one of the key channels of political communication for anti-establishment parties (Reinemann, Aalberg and Esser 2016; Gerbaudo 2014).

D.2 Variables

In order to identify the different types of rhetoric in the manifestos, the following variables will be coded by the RA:

- **screening**: equal to 1 if the paragraph shown is minimally cohesive and translates a

complete idea, and 0 otherwise.

- **anti_corruption**: equal to 1 if the anti-corruption approach, and 0 otherwise. Words in Portuguese frequently associated to this stance are "transparência", "eficiência," "probidade," "governança," as well as vows to "acabar com a corrupção" (erradicate corruption). Initiatives such as open government data, monitoring of government purchases and evaluation of programs are examples of anti-corruption oriented reforms. In Portuguese, these may appear, for example, as "transparência digital," "prestação de contas," and mentions to the "Lei de Acesso à Informação."

- **anti_incumbent**: equal to 1 if the paragraph adopts an anti-corruption stance, 0 otherwise. In government proposals, challengers who adopt anti-incumbent discourse will most often bring forward shortcomings of the current administration, which translates to "atual gestão" in Portuguese. The language will typically highlight shortcomings of the current administration or deterioration brought about in the city, contrasting such failures with the challenger's own projects for the next administration. Candidates may also appeal to the population directly by relying often on people-centric statements such as "nosso povo," "nossa cidade," while calling citizens to enact change and vote the incumbent out of office.

- **anti_establishment**: the variable should be equal to 1 if the paragraph adopts an anti-establishment approach, 0 otherwise. You should identify whether the candidate adopts a more general anti-elite or anti-political-establishment approach, beyond an attack on the current administration. To illustrate, the candidate may denounce policies that benefit only particular groups, or claim that the population is suffering and the city perishing while those holding political power only attend to their own interests. Examples of excerpts in Portuguese are "distribuição de cargos públicos em benefício pessoal," "afastamento da política dos interesses do povo," "ausência de compromisso dos que estão no poder." Note that this rhetoric may be characterized by a more narrative approach.
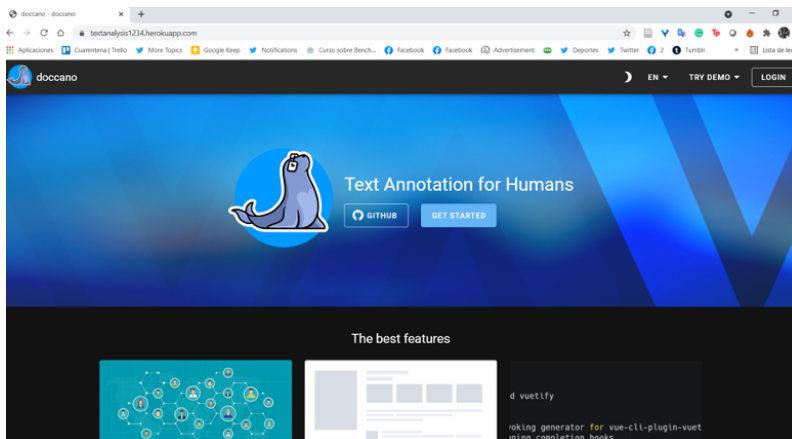
Please read the instructions carefully before beginning.

D.3 Coding Instructions

These instructions below were presented to the research assistants. It was elaborated by Natalia Garbiras-Díaz, Renata Avila and Javier Herrera.

How to access the platform?

1. Link to Doccano: https://textanalysis1234.herokuapp.com/

2. Click on Get Started, and insert your username and password:

3. To access the classification interface, click on "Prueba_1":


4. To start:

5. To access the classification instructions, click on the following:

How does it work?

In this page, you will find paragraphs that have been randomly assigned to you for classification. The goal is to determine whether the content of paragraphs has an anti-corruption rhetoric, an anti-establishment rhetoric, both of them or none. In practical terms, for those paragraphs that you have classified as having some or both approaches, you should underline the words that you led you to classify it in that particular manner.
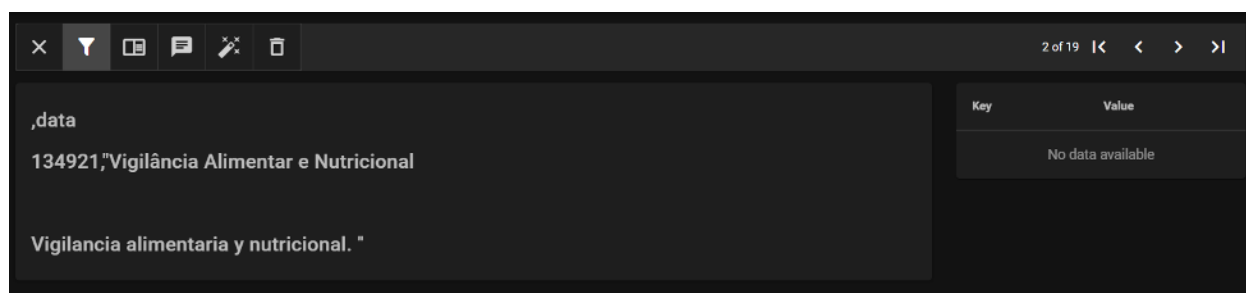
The paragraph will be shown in Portuguese and in Spanish, and you must underline the words in **PORTUGUESE** that made you classify it in such a way.

Please note that the Spanish translation is made automatically via Google Translate. If you think another translation or further research is necessary in order to fully understand the meaning of the paragraph, you can do so.

Filtering question

You should only classify the paragraphs that convey a complete and coherent idea. Therefore, when starting, you must indicate whether the paragraph has a complete meaning or not, in other words, whether it expresses an idea in itself. In this way:

- If the paragraph does **NOT** express a coherent meaning, you may underline any of the words with the label "Sin Sentido", and continue.[48]

- If the paragraph **DOES HAVE** a complete idea, then proceed to classify it in the following categories:

  - **Anti-corruption**: if the paragraph expresses ideas related or alluding to the fighting against corruption. For example: "transparency", "efficiency", or expressions related to, for instance, transparency in the use of public resources or criticism to the inappropriate use of political power.
  - **Anti-incumbent**: if the paragraph contains direct or indirect criticism to the current administration (incumbent).
  - **Anti-establishment**: rhetoric against the political elites that goes beyond specific criticism to the current administration.
  - **Ninguno**: if the paragraph conveys none of the three ideas above.

Note: There are paragraphs that may have more than one type of rhetoric, for example: anti-corruption and anti-establishment, anti-corruption and anti-incumbent, anti-incumbent and anti-establishment, or the three of them.

Repeat the previous steps each time a new paragraph is shown. You can pause or exit the platform at any time.

There are two additional documents you can use in the coding process. The first is a "language directory," where you can register commonly observed language patterns for both kinds of rhetoric, as you advance in the classification of the paragraphs. Since this is a shared directory, doing so will facilitate future work. You can access the document here.

---

[48]The group of RAs who went through the classification process were all Spanish-native speakers.

Please use the second document to relate any technical problems that you experience with the platform, as well as any issue or general feedback regarding the classification of paragraphs. You can access it here. Random audits will be occasionally conducted on classified paragraph to detect any major problems. In the second section of this same document, you can enter any substantive questions you come across when classifying paragraphs (e.g., specific paragraphs for which it may be hard to determine the rhetoric employed). Furthermore, if you have doubts on the meaning of Portuguese words and the language in general, you can contact Renata Avila at avilarenatac@gmail.com

How to classify?

1- When reading a paragraph, underline the word that you wish to select, and the platform will automatically show you a list of the different labels. Select the appropriate one in each case. See the image below:
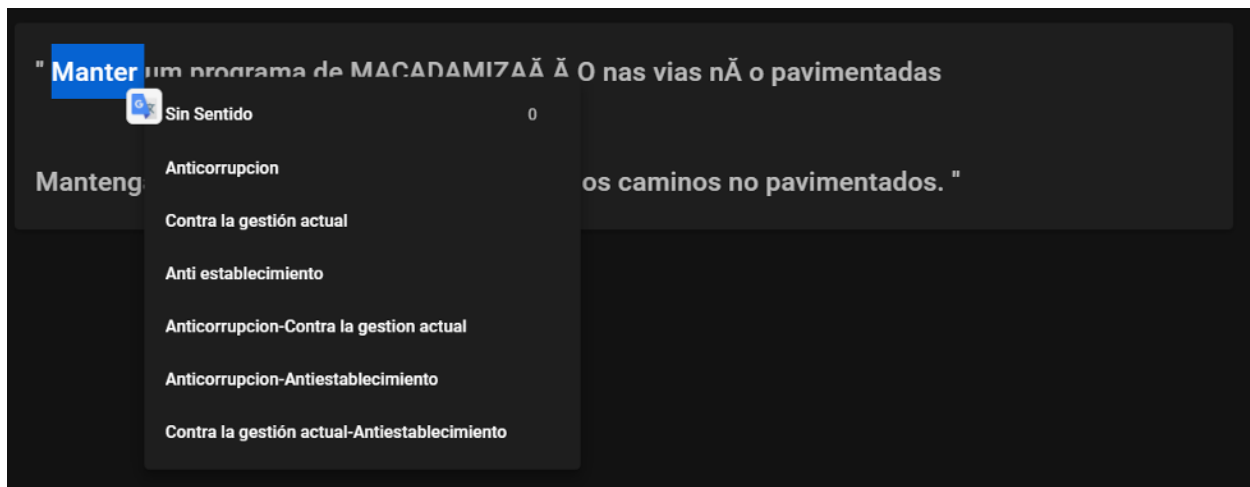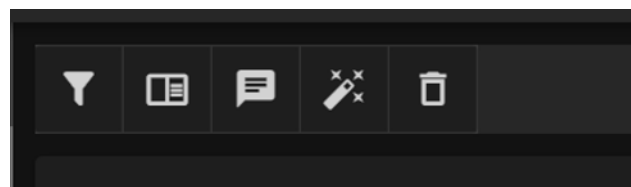

2 - You must select **ALL** words in the paragraph that correspond to each type of rhetoric that you want to classify.
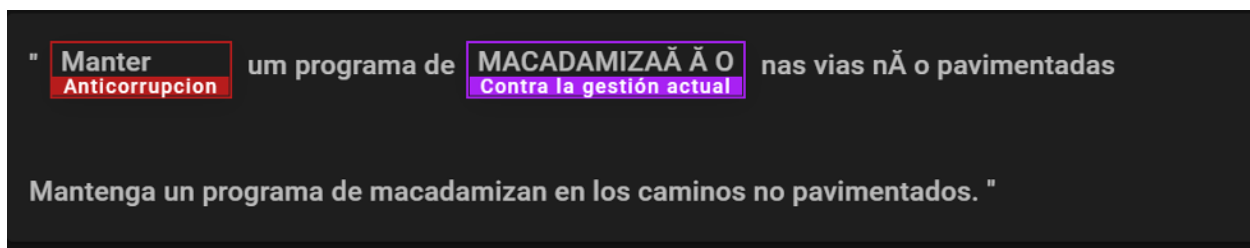
3 - Once you have finished, click the "next" arrow, as shown in the image below, in the top-right corner. In the toolbar, in the filter button, as shown in the top-left corner of the image below, you can revise the paragraphs that have not yet been classified.
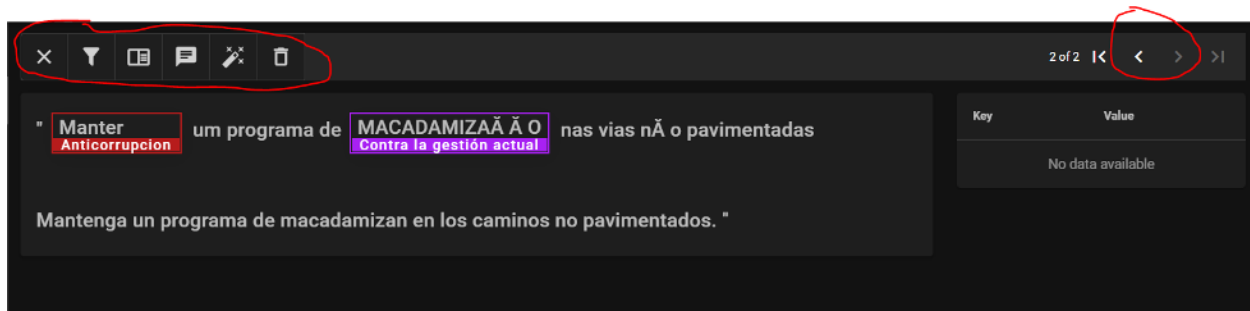
Toolbar

The first button corresponds to the "filter" option, where you can select paragraphs that have been already classified or not. The second one corresponds to the "instructions," which you will have read in the beginning of the exercise. The following button contains an option to make "comments," and the next one is "automatic selection," but you do not have access to this latter option. The final button can be used to "clean labels," if you want to delete the current selections.

E   Supervised classification

E.1   Description of the machine-learning algorithms used to classify paragraphs

NLP algorithm	Description
Logistic Regression	Is a supervised classification algorithm that creates a regression model to predict the probability that a given dataset belongs to a certain category. It does so by determining the best predicted weights (i.e. the estimation coefficients) through a maximum-likelihood estimation.
Multinomial Bayes Model	Is a machine learning classification method based on the Bayes' theorem. This algorithm assigns documents to different groups by determining the probability that a text belongs to the same category of other documents conditional on having the same topic/subject.
Random Forest Classifier	Is a supervised classification algorithm that is constituted by a large number of decision trees. Each tree returns a prediction set and the most frequent/voted set becomes the model's prediction. All trees are built on different samples and are trained using the "bagging" method.
Support Vector Machine with Sigmoid Kernel	Is a machine learning classification method that works by assigning a dataset to a high-dimensional plane where data points can be classified, even when they cannot be separated linearly. This algorithm transforms the data in a way that a hyperplane (line that separates categories) can be drawn between classes. Afterwards, it uses the transformed data to predict the category to which a new item should belong.

Table 19 – continued from previous page

NLP algorithm	Description
Bagging for Decision Tree Classifier	Similar to the Random Forest Classifier (which is an improvement of this method), this classification algorithm trains different decision trees using subsets of the original data in order to combine each prediction and obtain a more accurate result.
Multilayer Perceptron Neural Network	Is a neural network algorithm that can be used as a classification model. This method takes inputs, combines them in a weighted sum (called activation function), and feeds the next layers with the results of this computation. Afterwards, it employs a learning mechanism called backpropagation to adjust weights and minimize the stochastic gradient descent (function that minimizes the distance between misclassified points and the decision boundary). Finally, after all gradients have converged, the algorithm classifies each item based on the results obtained in the output layer.

E.2 Cross validation

Table 20: Second-stage metrics' scores after cross-validation

NLP algorithm	Metric	
	ROC AUC	Accuracy
Logistic regression	0.93825	0.96178
Naive Bayes - multinomial	0.86020	0.95086
Naive Bayes - binomial	0.96879	0.95008
Random forest classifier	0.88415	0.95554
SVM with sigmoid Kernel	0.96442	0.96568
Bagging for decision tree	0.94701	0.93058
Multilayer Perceptron	0.91517	0.95398

F   Details on Brazil's local electoral rules

Candidates are legally obliged to disclose their campaign funding and spending information to the Tribunal Superior Eleitoral (TSE). Every candidate must open a bank account specifically for campaign donations. In the 2012-2020 period, there was some difference in campaigning funding rules in place. Starting in 2015, the Supreme Court introduced a ban on corporate donations to political candidates. Before that year, corporations were entitled to donate up to 2% of their gross annual revenues to political candidates or parties. From this point onward, candidates could only obtain campaign money from their political parties, citizens, or self-financing.

Political parties finance candidates through the *Party Fund*, which is constituted by public resources, donations, and electoral fines, among other sources. The creation of this fund was established in the 1998 Brazilian Constitution. Furthermore, in 2017, to compensate for the extinction of corporate donations, the *Fundo Especial de Financiamento de Campanha (FEFC)*, popularly known as the Electoral Fund, was also created (Law n. 13.487/17). In addition, Law n. 13.488/17 also established caps on campaign spending, which vary according to the position for which each candidate is running.

It is also worth noting that Brazilian legislation introduced gender quotas for proportionate elections in 1995, and the rules currently in place stipulate that, for every party, 30% of candidacies for proportionate elections must be female (Law n. 9504/1997). In 2018,

the Superior Court also determined that 30% of resources from the party's fund should be destined for the support of female candidacies. However, vice-candidacies are also considered in the scope of this rule.

G   Details on CGU audits and lotteries

From 2003 to 2015, the *Controladoria-Geral da União* (CGU) conducted 40 rounds of audits based on a randomized process. Municipalities below a population threshold, usually $500,000$ inhabitants,[49] excluding capitals, were eligible to the draft. In each lottery, a predetermined number of municipalities from each state was drafted.

Since the CGU audits program was expanded in 2015 and renamed *Programa de Fiscalização em Entes Federativos* (FEF), different methodologies of selecting municipalities were adopted. We henceforth refer to each edition of the *FEF* program as *cycles*, following the nomenclature used by the CGU. In cycle 1, the selection of municipalities was based on the highest scores in a 'vulnerability matrix', in which scores were assigned to the municipalities based on previous data to determine those that were potentially at-risk for corruption. In cycle 2, the audits were conducted in all state capitals. There was no randomization in these cycles.

In cycle 3, municipalities were selected via random lottery. A total of 70 municipalities were chosen in a universe of 1,520 eligible ones. Random numbers were assigned to all eligible municipalities, according to the number of municipalities that are to be audited in each state.

Cycles 4, 5 and 6 were also based on the "vulnerability" scores, with no random selection.

Finally, a seventh cycle was initiated in 2020 and 60 municipalities were selected by random lottery. However, due to the Covid-19 pandemic, the audits were delayed and the results were not available as of September 2021.

G.1   Audits and election years

Below, I specify the overlap between **random** CGU audits and years preceding the relevant municipal elections.

- 2012: Lotteries 32 and 33. Draws were made at the end of 2010, results published in January 2011. Lotteries 34 and 35 were conducted in 2011 and results were published in March and April 2012 (07/03/12 and 24/04/12).

---

[49]In this link it is possible to inquire the legislation delimiting selection criteria for each draft.

- 2016: Lotteries 40, 39. The Cycle 1 of the FEF, also published in 2015, was not randomized. Cycle 3 of the FEF, indeed random, happened during the election year, and results were published only in October or December 2016.

- 2020: No audit results available as of September 2021, only list of eligible municipalities.

G.2 Lottery selection

In each of the relevant lotteries, 60 municipalities were selected among those with 500,000 or fewer inhabitants, excluding state capitals. Different areas of government and public transfers were inspected in each one of the lotteries, as described below. In each draft, municipalities that were inspected in any of the previous three lotteries were not subject to selection in the current lottery.

G.2.1 Lottery 32

In municipalities with a population above 20,000, the audits will inspect public administration activities in one of 5 following groups: Commerce and Services, Agriculture and Culture; Agrarian Organization, Energy and Envinronmental; Public security, Industry and Science and Technology; Habitation, Sanitation and Urbanism; Communications, Social Security, Employment and Leisure

For those municipalities with up to 100,000 inhabitants, the three following areas were added to the inspection: Social Assistance, Education and Health. For municipalities with over 100,000 inhabitants, one of the three areas were added to the inspection: Social Assistance, Education and Health.

G.2.2 Lottery 33

The same rules of Lottery 32 were followed. As well as the municipalities that were previously audited in any of the lotteries 30, 31 and 32, other exclusions to the universe of eligible municipalities were made on account of the calamities caused by heavy rainfall in the states of Alagoas and Pernambuco.

The municipalities that were excluded and respective states were: Murici (AL), Quebrângulo (AL), Santana do Mundaú (AL), São José da Laje (AL), União dos Palmares (AL), Água Preta (PE), Barra de Guabiraba (PE), Barreiros (PE), Catende (PE), Correntes (PE), Côrtes (PE), Jaqueira (PE), Maraial (PE), Palmares (PE), Primavera (PE), São Benedito do Sul (PE), Vitória de Santo Antão (PE).

G.2.3   Lottery 34

The universe of eligible municipalities was divided in two groups:

- *Faixa A*: For municipalities with at most 50,000 inhabitants, funds transferred by the following national ministries were audited: Education, Health, Social Development and Hunger Eradication.

- *Faixa B*: For municipalities with a population between 50,000 and 500,000,000, funds transferred by the following national ministries were audited: (1) Health, (2) Social Development and Hunger Eradication

G.2.4   Lottery 35

The universe of eligible municipalities was again divided in two groups:

- *Faixa A*: For municipalities with at most 50,000 inhabitants, funds transferred by the following national ministries were audited: Education, Health, Social Development and Hunger Eradication.

- *Faixa B*: For municipalities with a population between 50,000 and 500,000,000, funds transferred by one of the following national ministries were audited: (1) Health, or (2) Education

G.2.5   Lottery 39

In this lottery, two municipalities were to be selected among the states of Rondônia, Acre and Roraima. The state to which the first selected municipality belonged was excluded from the draft of the second municipality. Furthermore, municipalities from the states of Amazonas and Amapá were subject to a grace period, following Lottery 38.

The universe of eligible municipalities was divided in two groups:

- *Faixa A*: For municipalities with at most 50,000 inhabitants, funds transferred by the following national ministries were audited: Education, Health, Social Development and Hunger Eradication.

- *Faixa B*: For municipalities with a population between 50,000 and 500,000,000, funds transferred by the following national ministries were audited: (1) Health, (2) Social Development and Hunger Eradication

G.2.6 Lottery 40

In this lottery, only municipalities with at most 100,000 inhabitants were eligible. No particular area of government was ex-ante selected for inspection, and funds and programs audited were to be defined by the CGU. Municipalities in the states of Acre and Rondônia were also excluded no from this draft, following Lottery 38.

G.3 Classification of audit findings

The data provided by the CGU includes the registering of all irregularities found in each inspection. Each inspection order is associated to a variable that classifies its finding according to the severity of the irregularity uncovered. Upon request, the CGU provided a summary of the meaning of each category:

- **Severe**: missing funds or public property; illegal, illegitimate or anti-economic administration activities; accountability failure or omission, including withholding of information critical to the audit inspection; infraction of legal or administrative rules, with potentially severe losses to the Treasury or grave deviations from public administration principles.

- **Medium**: illegal, illegitimate or anti-economic administration activity, or legal / administrative infraction not included the hypotheses above.

- **Formal**: irregularities that do not result in damages to the Treasury, legal or administrative infractions, or deviations from public administration principles

- **Information**: no irregularities found. Standard administration activities or positive feedback regarding the quality or implementation of programs and operations.

Additional categories appear in the dataset after the adoption of a new Technical Manual by the CGU in 2017, whereupon the audit findings were classified according to their impact on the goals of the program or activity targeted by the audit inspections, such as: cost, deadline, quality, delivery of service or public goods. The following levels were established:

- **Low**: the execution of programs or operations was found to be substandard, with a low impact on quality, cost or timeline.

- **Médio**: the execution of programs or operations was discrepant with the established criteria or delay in activities was found, with significant but reversible impact on the goals

- **Alto**: the execution of program or operations was discrepant with the established criteria, or delay in activities was found, causing significant and difficult to reverse impacts on the goals of the program

- **Muito Alto**: execution of programs or operations or paralization of activities, with severe or irreversible damage to the goals of the program

G.4 Balance audited and non-audited municipalities

Figure 8: Covariate Balance
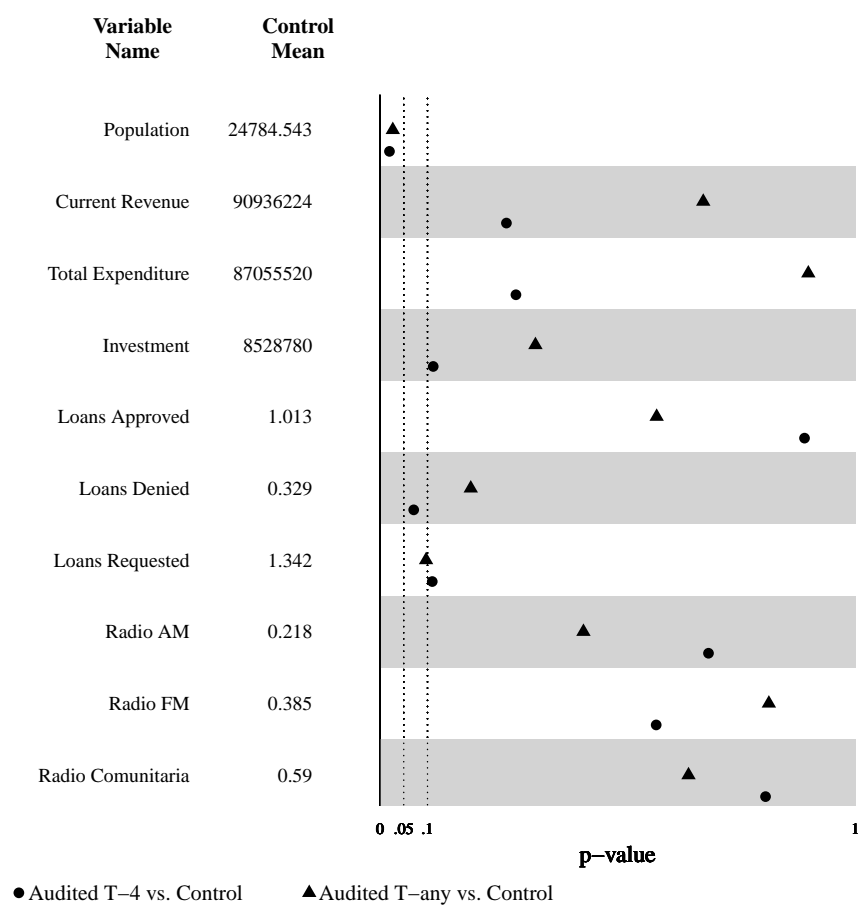
Note: p-values correspond to t-tests comparing audited and non-audited municipalities.

H Code book of variables used for the empirical analysis

H.1 Election and political variables

- **office**: candidate held some kind of political office at the time of the election

- **office_or_public**: candidate either held some political office or was employed in public service.

- **classification**: political parties broadly classified as left, center-left, center, center-right, and right, according to the party's self-declared legend.

- **classification_3**: political party classification based on political expert study, conducted by the UFPR and the ABCP. Based on this measure, a candidate was classified as either extreme-left, left, center-left, center, center-right, right, or extreme left. For the ease of the analysis, the categories were collapsed into three levels: left, right, and center. The category *right* includes extreme-right and right parties; *center* includes those classified as either center-right, center or center-left; and the category *left* includes parties classified as left or extreme-left.

- **new party election**: candidate was running with a party that was a newcomer in that election year, i.e if that was the first election in which that party participated. The criterium considers both local and national elections, which take place in separate years.

- **outsider**: factor variable, assuming values 0, 1 or 2, if, respectively, the candidate fulfills none, one or both of the outsider characteristics, i.e, not holding political office and running with a new party in the respective election

- **outsider_aggregate**: equal to 1 if the outsider factor variable is equal to either 1 or 2, indicating that the candidate fulfills at least one of the outsider characteristics

- **gender**: factor variable that indicates whether the candidate is male or female. The original variable re-leveled such that 1 indicates female.

- **ds_genero**: description of gender factor variable, values are "male" or "female"

- **education**: factor variable indicating the (highest) level of education attained by the candidate. The levels, from 2-8, indicate, in ascending order: *alphabetized, incomplete*

basic education, basic education, incomplete high school, high school, incomplete higher education, higher education.

- **ds_grau_instrucao**: contains the description of the education factor variable given above

- **sg_partido**: code/initials of the political party with which the candidate was running.

- **sg_partido**: name of the political party with which the candidate was running.

- **dt_limite**: date established by the TSE as the deadline for registering candidacies.

- **diccionario**: sum of the number of times that one or more that words belonging to the anti-corruption dictionary appear in the manifesto, at the paragraph level. It identifies the words and counts whether the word. It identifies the words and counts whether the word appears at least once. Then, the number of paragraphs in which the anti-corruption words were located is summed up.

- **diccionario_dummy**: indicator variable that is equal to one if the manifesto of the candidate contains at least one of the words belonging to the anti-corruption dictionary.

- **diccionario_proporcion**: number of anti-corruption words identified over the total number of paragraphs in the manifesto.

- **anticorrupcion**: number of paragraphs in the candidate's manifesto that were classified as anti-corruption according to the NLP models.

- **anticorrupcion_dummy**: indicator variable that equals 1 if at least one of the paragraphs in the candidate's manifesto were classified as anti-corruption, according to the NLP models.

- **anticorrupcion_proporcion**: number of paragraphs classified as anti-corruption, according to the NLP models, over the total number of paragraphs in the manifesto.

H.2 New parties by election

New party variable takes the value of 1 if the candidate's party was created 3 years or less prior to election, where the creation date is taken from the TSE: https://docs.google.com/spreadsheets/d/1t3A2CHRpvKiDlgaLppc58aBoZmlFqimqAPLXQ0Y5EuY/edit#gid=0

- **PSD** and **PATRIOTA**: new in 2012

- **SOLIDAREIDADE** and **PROS**: created in Sep 2013, borderline new in 2016 (federal election in 2014, but first municipal election is 2016).

- **NOVO** and **REDE**: created in 2015, new in 2016.

- **UP**: created in 2019, new in 2020.

H.3 Municipality-level controls

- **id_municipio**: unique municipality identifier (7-digit code) used by the IBGE

- **population**: population count at the municipal level

- **gdp**: municipal-level gross domestic product. The last observation available is for 2018.

- **gdppc**: GDP per capita at the municipality level.

- **idhm**: Human Development Index at the municipality level, calculated by the United Nations' UNDP and based on life expectancy, education and income measures extracted from the 2010 Census.

- **idhm_l, idhm_e, idhm_r**: longevity, education and income components of the IDH, respectively. The geometric mean of the three measures is the IDHM of the municipality.

- **qtd_veiculos**: number of media outlets in the municipality in year 2020. The indicator is taken from the *Atlas de Notícias* project.

- **veiculos_por_100k_hab**: number of media outlets per 100km, in 2020. The indicator is taken from the *Atlas de Notícias* project.

- **ano_populacao**: year of population count used to calculate the *veiculos_por_100k_hab* measure. The value is 2020 for all observations in the sample.

- **id_municipio_tse**: unique municipality identifier in the TSE datasets

- **capital_uf**: indicator variable that equals 1 if the municipality is the capital of the federative unit (state or federal district) where it is located.

- **nome**: name of the municipality

- **sigla_uf**: two-letter code of federative unit (state or federal district)

H.4 Audits

The variables related to audit findings were provided by the CGU, at the inspection order level.

- **lottery**: number or code indicating the lottery number or cycle number, assigned by the CGU. In the aggregate data provided by the CGU, the codes for the random lotteries are the pure numbers (until 40), and *V03*. Codes *V01*, *V02*, *V04*, *V05* and *V06* correspond to non-random audits. Results for the seventh cycle, *V07*, conducted in 2020 and which was randomized, are not yet available in the data.

- **ano_lottery**: the year in which the random lottery was conducted. This is the year variable that is used in the graph that shows the number of audits per year.

- **date**: date of publication of the audit results by the CGU.

- **ano**: the year when the audit results were made publicly available. Corresponds to the year in the date above.

- **Tipo_constatacao**: classification of audit findings provided by the CGU. Possible categories are: high ('alto'), very high ('muito alto'), medium ('médio'), severe irregularity ('falha grave'), medium irregularity ('falha média), low ('baixo'), formal irregularity ('falha formal), and information ('informação'). The details on each classification were provided by the CGU upon request, and are explained in the Appendix.

- **Programa**: specific government program whose federal transfers were subject to inspection

Further variables were created at the municipality level with information on 'corruption', based on the CGU's classification on findings, which considers the severity of each irregularity found in the inspection order.

- **corruption_any**: equals to 1 if any irregularities classified as "medium" or "severe" were found by the CGU audits in that municipality

- **corrup_any_severe**: indicator variable that equals 1 if any severe irregularity was found by the CGU audits in that municipality.

- **corruption_any_broad**: indicator variable that equals 1 if any kind of irregularity was found by the CGU audit in that municipality.

- **id_municipio**: unique municipality identifier (7-digit code) used by the IBGE

- **ano_match**: identifier year variable, used as a match in the merging process between the municipal audit data and the election data at the candidate level. Corresponds to the year **after** the audit happened, **or the same year** as long as the audit results were published before the limit date for registering candidacies. Note further that the merging process only considers only **random** audits.

H.5 Other variables in the fully merged dataset:

- **diff_days**: time difference, in days, between the date in which the audit results were published and the limit date for registering political candidates in the municipal election in that year. The variable is equal to zero for municipalities that were not matched in the audit data (either not audited or not audited in the year before/of the municipal election)

- **population_500**: indicator variable that equals 1 if the municipality has **at most** 500,000 inhabitants.

I Drivers of anti-establishment sentiments: Evidence from Colombia

Figure 9: Word cloud from elicited answers on the drivers of anti-establishment sentiments

Note: This word cloud draws on a study conducted in the context of Colombia's 2022 presidential election. Answers come from a follow-up survey of an online panel of 2,000 individuals. Respondents were asked the following open-ended question: "When thinking about politicians, many people say they feel angry, feel rage, or resentment. According to your experience in Colombia, mention at least three things that make you angry or generate anger or resentment from these politicians. We are very interested in your views on this. Please take at least 30 seconds to answer this question without rushing."

J Further information about the database on the universe of presidential candidates running in Latin America from 1980 until 2020

I compiled a database with a comprehensive classification of the electoral platforms—i.e., the vehicle used to register the candidacy (e.g., a party, a movement, a party coalition, among others)—of the universe of presidential candidates running in Latin America since the third wave of democratization. To do so, I designed a coding procedure informed by the existing literature on party systems in Latin America and expanding existing classifications of electoral platforms (e.g., Levitsky et al. 2016; Carreras 2012). For instance, I created the category of *marginal outsiders* to account for cases of candidates who run with existing marginal parties and are, thus, not classified as outsiders according to Carreras (2012) but should be, following Barr (2009), to the extent that they share the barriers and features of those outsiders running with new parties.[50] To keep the exercise within feasibility, and seeking to only consider relevant candidates who had a possibility of winning the election, I restricted the sample to those candidates meeting two criteria: i) that were on the list of countries constituting the largest democracies in Latin America,[51] and, ii) who obtained at least 5% of the votes in the election.[52] Then, a group of research assistants classified the electoral platforms of this final list of candidates following the coding procedure (available in the ancillary coding procedure for "*A database of the universe of presidential candidates running in Latin America between 1980 and 2020*").

---

[50]Party classification follows Levitsky et al. (2016), whereby a party is marginal if it has obtained between 1 and 10% of the votes in five or more consecutive national legislative elections. For the full definitions, see the ancillary coding procedure for "*A database of the universe of presidential candidates running in Latin America between 1980 and 2020*").

[51]These countries are: Bolivia, Uruguay, Peru, Mexico, Chile, Brazil, Argentina, Venezuela, and Colombia.

[52]For countries with runoff elections, I apply this criterion to the vote obtained by the candidate in the first round. This strategy follows other classifications (see, for instance, Carreras (2012)). Notice that others have used higher thresholds, which risk excluding relevant cases to consider (e.g., Corrales (2008), who uses a 10% threshold).