Department of Economics- FEA/USP

Play to Your Strengths: Incumbency Effect and Issue Emphasis in Brazilian Local Elections

LEILA PEREIRA

WORKING PAPER SERIES Nº 2026-13

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2026-13

Play to Your Strengths: Incumbency Effect and Issue Emphasis in Brazilian Local Elections

Leila Pereira (leilapereira@usp.br)

Abstract:

Incumbents' reputations acquired during their time in office influence voters' choices. Some studies suggest that, when designing their campaign messages, incumbents may benefit from playing to their strengths and emphasizing fewer issues than challengers. In this paper, I provide causal empirical evidence on this behavior using data from Brazilian mayoral elections and a regression discontinuity design. The findings indicate that *marginal incumbents*—those elected by a small vote margin—produce more concentrated campaign manifestos than challengers who narrowly lost. Moreover, incumbents belonging to cohesive and programmatic parties tend to emphasize party-specific issues more strongly, suggesting that reputation plays an important role in issue selection. To measure issue emphasis and manifesto concentration, I leverage a unique database of more than 28,000 political manifestos and apply a Latent Dirichlet Allocation (LDA) algorithm to estimate a vector of topics—or the issues emphasized by each candidate. Using this information, I calculate Herfindahl-Hirschman indices (HHI) to assess the degree of issue concentration in campaign manifestos.

Keywords: Elections, Electoral Competition, Issue Selection, Incumbency, Manifestos.

JEL Codes: D72, D83, P00

Play to Your Strengths: Incumbency Effect and Issue Emphasis in Brazilian Local Elections

Leila Pereira *

April 16, 2026

Abstract

Incumbents' reputations acquired during their time in office influence voters' choices. Some studies suggest that, when designing their campaign messages, incumbents may benefit from playing to their strengths and emphasizing fewer issues than challengers. In this paper, I provide causal empirical evidence on this behavior using data from Brazilian mayoral elections and a regression discontinuity design. The findings indicate that *marginal incumbents*—those elected by a small vote margin—produce more concentrated campaign manifestos than challengers who narrowly lost. Moreover, incumbents belonging to cohesive and programmatic parties tend to emphasize party-specific issues more strongly, suggesting that reputation plays an important role in issue selection. To measure issue emphasis and manifesto concentration, I leverage a unique database of more than 28,000 political manifestos and apply a Latent Dirichlet Allocation (LDA) algorithm to estimate a vector of topics—or the issues emphasized by each candidate. Using this information, I calculate Herfindahl–Hirschman indices (HHI) to assess the degree of issue concentration in campaign manifestos.

JEL Codes: D72, P00

Keywords: Elections, Electoral Competition, Issue Selection, Incumbency, Manifestos, Brazilian politics.

*University of São Paulo (FEA-USP), Address: R. Avenida Professor Luciano Gualberto, 908, São Paulo, 05508-010, SP, Brazil. Email: leilapereira@usp.br. I thank Rodrigo Soares, Sergio Firpo, Raphael Corbi, Lucas Novaes, Fernanda Estevan, Natália Bueno, Rafael Pucci and all seminar and conference participants at Insper, at the 4th Aix-Marseille Aix-Marseille School of Economics (AMSE) summer school, and at LACEA-LAMES 2023 for their extremely valuable comments and suggestions. I also thank the São Paulo State Research Foundation (FAPESP) - scholarship process 2018/16476-8 - and CAPES/PSDE - scholarship process 88881.362146/2019-01 - for their financial support.

1 Introduction

In canonical models of electoral behavior, incumbents' actions and policies during their time in office are central to voters' choices. Rational voters evaluate incumbents retrospectively, using their past performance to inform electoral decisions (Downs, 1957). In these models, observed outcomes provide sufficient information about future performance, so that voters rely primarily on incumbents' records rather than on campaign promises when forming expectations about future utility.

In practice, however, voters face uncertainty and imperfect information, which limit their ability to accurately assess incumbents' past performance. Information about policy outcomes may be noisy, multidimensional, or difficult to attribute, making retrospective evaluation costly or imprecise. In such environments, campaign communication becomes consequential. Party manifestos, for instance, can reduce uncertainty about the likelihood that a candidate will implement a given policy after the election (Bernhardt and Ingberman, 1985). Candidates can also design their campaign messages to clarify or persuade voters about which issues are most relevant in a given election (Robertson, 1976; Schattschneider, 1975; Budge and Farlie, 1983; Petrocik, 1996; Kaplan et al., 2006; Vavreck, 2009; Amorós and Puy, 2013; Dolezal et al., 2014; Aragonès et al., 2015; Dragu and Fan, 2016; Denter, 2020). Moreover, depending on electoral competitiveness, candidates may moderate their positions to appeal to a broader electorate (Henderson, 2013; Le Pennec, 2023). As a result, even incumbents—despite having a public record—may benefit from strategically emphasizing certain issues over others in their campaigns.

But how do incumbents and challengers differ in their strategic selection of campaign messages? For Bernhardt and Ingberman (1985), incumbents' first-term reputation limits their policy positions during elections because shifting to a different position raises voter uncertainty about the policies they will pursue if re-elected. This idea that reputation limits incumbents' campaigns aligns with the findings of Greene (2020), who shows that incumbents are electorally rewarded for focusing their manifestos on fewer issues. Importantly, Greene (2020) studies the electoral consequences of manifesto concentration, rather than whether incumbents strategically choose to concentrate their messages in the first place. Vavreck (2009) also supports this argument by showing that U.S. presidential incumbents with negative economic records who focus on this in their campaigns are less likely to be re-elected. Druckman et al. (2020) note that incumbents improve their chances by emphasizing their positive experiences and actions while in office. Similarly, this idea aligns with the theoretical results of Colomer and Llavador (2012), which show that a successful issue selection strategy includes both discussing salient issues and specializing in issues that give candidates a competitive advantage. Therefore, incumbents

should avoid discussing poorly evaluated policies from their first term.

These findings, taken together, suggest that, due to reputation constraints, incumbents should choose to play to their strengths and emphasize fewer issues in their campaign messages compared to challengers. Despite this evidence, the literature has not yet established whether incumbency causally affects how parties structure their campaign messages or whether incumbents actually narrow their manifestos relative to challengers. Furthermore, the literature has not addressed whether these reputation-driven incentives affect incumbents' campaign message choices in local, low-stakes settings.

In this paper, I take a step forward by shifting the focus from electoral outcomes to campaign behavior. Rather than asking whether manifesto concentration benefits incumbents electorally, as in Greene (2020), I study whether incumbency itself causally affects the concentration of campaign manifestos. To do so, I use a unique dataset comprising over 28,000 political manifestos from Brazilian mayoral elections in 2012 and 2016. In Brazil, mayoral candidates are legally required to submit a campaign manifesto as part of their candidacy registration prior to the election, and these documents are collected and made publicly available by the Superior Electoral Court. The manifestos are short texts outlining candidates' policy priorities and provide a unique source of information on campaign messaging.[1] To estimate the issues emphasized by each candidate, I employ a Latent Dirichlet Allocation (LDA) model and derive a vector giving the probability that a manifesto's content is devoted to different topics (Blei et al., 2003). Using this information, I calculate Herfindahl–Hirschman indices (HHI) for each manifesto to evaluate their relative issue concentration.

I rely on a regression discontinuity (RD) design to compare manifesto issue concentration between parties that barely won the previous election (i.e., the incumbents) and parties that barely lost (i.e., the challengers). In close elections, parties are considered to be almost randomly assigned to winning or losing and should not differ significantly in any other dimension. Hence, the estimates can be interpreted as the causal effect of winning office on manifesto issue concentration in the following election. In this sense, my design allows me to test a key behavioral implication of Greene (2020)'s argument, namely, whether incumbents narrow the scope of their campaign messages.

A caveat of this empirical strategy is that, because Brazilian rerunning rates are not as high as in other countries, a sample selection problem could bias the incumbency effect estimates (Magalhaes, 2015). To alleviate this concern, I show that the results are robust to three alternative samples that address the selection issue. The first sample conditions on at least one of the two

[1] These documents are only available from 2012 onwards. Section 3 details their structure and contents.

parties in a municipality — the incumbent or the challenger—rerunning in two consecutive elections. In the second sample, I condition on observing the winning party in two consecutive elections in a given municipality, using the best-ranked non-incumbent party in $t+1$ as the comparison group when the original challenger does not rerun. In the third and most restrictive sample, I keep only parties running in municipalities where both the winner and the runner-up contested two successive elections.

Results confirm the hypothesis drawn from the literature and indicate that, among parties involved in closely contested elections, incumbents' manifestos are, on average, approximately 17% more concentrated than challengers'. This pattern is robust to alternative specifications and also holds when exploiting within-party variation in manifesto concentration over time using a difference-in-discontinuities strategy. Moreover, the higher manifesto issue concentration among incumbents appears to be driven, in part, by a rhetorical topic containing words related to continuity. One possible interpretation is that, while this emphasis on continuity does not necessarily imply that incumbents discuss fewer policy issues, it may reflect a more constrained scope of campaign messaging—framing their agenda around the continuation of existing policies rather than the introduction of entirely new proposals. While the analysis does not allow me to directly test this mechanism, the pattern is consistent with the findings of Druckman et al. (2020), which suggest that incumbents tend to emphasize their experience by making more campaign statements about their actions while in office.

To disentangle whether parties play to their strengths when selecting which issues to emphasize, I compute the heterogeneity of incumbency effects across the four major parties in Brazilian local elections. The premise is that, if reputation and comparative advantages play a role in issue selection, then more programmatic parties, which historically advocate for specific policies, should face stronger constraints when crafting their campaign messages. In Brazil, most elections revolve around personalistic links between candidates and constituents, and parties generally lack programmatic platforms (Mainwaring, 1993; Banck, 1998; Kitschelt et al., 2010). One exception, however, is the Workers' Party (PT), which is known to be a more programmatic and disciplined party committed to improving social policies (Samuels and Zucco, 2014; Klašnja and Titiunik, 2017). In this exercise, consistent with theoretical predictions, I find that PT incumbents place greater emphasis on social issues.

This paper highlights an aspect often overlooked by spatial models of electoral competition following a Downsian approach (Downs, 1957).[2] Specifically, it underscores that incumbents and challengers have incentives to craft divergent campaign messages. This notion appears in

[2]Excellent surveys of the literature on spatial models of electoral competition are provided by Osborne (1995) and Persson and Tabellini (2002).

Downs (1957), Samuelson (1984), and Bernhardt and Ingberman (1985), where the authors argue that an incumbent's reputation in their first term may influence candidates' strategies during elections. However, empirical evidence demonstrating that incumbents and challengers indeed produce distinct campaign messages remains scarce.

In addition to engaging in a broad dialogue with Downsian spatial models, this paper speaks more directly to the literature on issue selection in electoral competition (Robertson, 1976; Budge and Farlie, 1983; Petrocik, 1996; Schattschneider, 1975; Kaplan et al., 2006; Vavreck, 2009; Green-Pedersen and Mortensen, 2010; Amorós and Puy, 2013; Dolezal et al., 2014; Aragonès et al., 2015; Dragu and Fan, 2016; Denter, 2020). This body of work explores how candidates strategically emphasize certain issues during elections in order to attract constituents. When choosing which issues to address in their manifestos, candidates might be attempting to shape the election agenda. They may also face constraints due to their party and individual reputations.

In this paper, I focus primarily on the reputational aspects of issue selection strategies, aligning with the work of Vavreck (2009), Colomer and Llavador (2012), Greene (2020), and Druckman et al. (2020). Taken together, these studies suggest that incumbents who concentrate their campaign messages on their positive track records are more likely to win elections. I contribute to this literature by providing causal evidence that, among parties involved in closely contested elections, incumbents focus on fewer issues in their manifestos than their opponents.

From an empirical standpoint, this paper also contributes as a first attempt to identify the causal effects of incumbency on manifesto issue concentration using within-country data and without relying on human-coded text data. This approach contrasts with prior empirical work that either relies on cross-country correlations (Greene, 2020) or constructs text measures through manual inspection (Druckman et al., 2009; Green-Pedersen and Mortensen, 2010; Greene, 2020; Druckman et al., 2020).

New statistical text analysis methods improve on manual approaches in two key ways. First, manually coded text data are typically available only for limited or selective samples, which constrains much of the existing literature to descriptive or correlational analyses. By using unsupervised topic modeling, this paper substantially expands the feasible scale of analysis while improving consistency and replicability, as the classification of text does not depend on coder-specific judgments or predefined coding rules. Second, access to text-derived measures for nearly the entire universe of mayoral elections enables this paper to integrate text data into a regression discontinuity design, allowing it to causally assess whether incumbency increases manifesto concentration in close elections.

Within this broader methodological contribution, this paper applies new statistical text anal-

ysis methods to construct a unique large-scale dataset on the issues discussed by Brazilian mayoral candidates in the 2012 and 2016 elections (Blei et al., 2003; Grimmer and Stewart, 2013; Gentzkow et al., 2019a). In that sense, this paper connects with efforts such as Le Pennec (2023), which build datasets on individual politicians' campaigns at the local level. While text data have been used extensively in various contexts within economics and the social sciences, the focus has primarily been on Europe and the United States (Grimmer, 2010; Mueller and Rauh, 2018; Hansen et al., 2018; Gentzkow et al., 2019b; Le Pennec, 2023). So far, Brazilian manifesto data have received limited attention. Studies such as Salles (2019) and Salles and Guarnieri (2019) explore only subsets of the data, while others, such as Novaes (2023), use them primarily for robustness exercises.

The remainder of the paper is structured as follows. Section 2 conceptually discusses the incentives that lead incumbents to produce campaign manifestos focused on fewer issues than challengers. Section 3 presents the unique dataset on mayoral political manifestos and defines the measure of manifesto issue concentration. Section 4 describes the empirical strategy. Section 5 presents evidence that, in closely contested elections, incumbents produce more concentrated manifestos. Section 6 investigates the mechanisms underlying this result, presenting evidence that incumbents play to their strengths when choosing which issues to emphasize. Section 7 concludes.

2 Incumbency and Manifesto Issue Concentration

Incumbency provides voters with additional information about the governing candidate (Bevia and Llavador, 2009). It unveils aspects like the candidate's type or quality (Ashworth et al., 2019; Caselli et al., 2014), as well as their preferred policy position (Samuelson, 1984; Bernhardt and Ingberman, 1985).

This additional information—not available for challengers—can either benefit or hinder the incumbent's electoral prospects. For instance, the model in Bevia and Llavador (2009) shows that high-quality candidates benefit from incumbency revealing their type, whereas low-quality candidates would be better off if their type were not revealed. Even though the incumbent's record may create comparative advantages or disadvantages during the race, incumbents and challengers can respond to this information by crafting different campaign messages. As Vavreck (2009) points out, *"even in the face of strong structural conditions like a declining economy and a lengthy war, what candidates say and do in their campaigns can affect outcomes."* (p.xxi)

It is well established that candidates compete and strategize over issue selection in their campaign messages to sway the election agenda and improve their electoral chances. So far, there has been an extensive debate on different issue selection strategies, the possible equilibrium outcomes, and the mechanisms that explain these equilibria (Spiliotes and Vavreck, 2002; Sigelman and Buell, 2004; Kaplan et al., 2006; Amorós and Puy, 2013; Aragonès et al., 2015; Dragu and Fan, 2016; Denter, 2020).

However, fewer studies have examined how issue selection strategies change in the presence of incumbency (Bernhardt and Ingberman, 1985; Vavreck, 2009; Colomer and Llavador, 2012; Greene, 2020; Druckman et al., 2020). In this literature, incumbency restricts the set of issues and positions that candidates can adopt, stimulating specialization around specific topics. The central insight is that incumbency constrains campaign strategies by tying candidates to their past records in office. As a result, incumbents cannot freely choose which issues to emphasize and instead may benefit from concentrating their campaigns on issues that highlight their strengths or positively evaluated performance, while avoiding issues associated with poorly evaluated policies.

In the paragraphs below, I detail the mechanisms highlighted in each of these studies and how they lead to the prediction that incumbents strategically focus on particular issues.

Focusing on U.S. presidential elections, Vavreck (2009) shows that incumbents' campaign strategies are shaped by their economic record. When economic performance is favorable, incumbents benefit from *clarifying campaigns* that emphasize the economy, whereas challengers attempt to shift attention to alternative issues on which their positions are better aligned with voters' preferences. When economic performance is poor, these roles reverse. More generally, the study highlights that the incumbent's economic record determines the best strategies for both the incumbent and the challenger.

Druckman et al. (2020) examine the mechanisms behind incumbency advantages in U.S. congressional elections. Through a content analysis of candidates' websites and an experiment, the authors conclude that incumbents benefit when their campaign strategies emphasize aspects related to their current positions, such as their established relationships with voters and the successful actions they have implemented in their constituencies. Hence, even though, according to Druckman et al. (2020), incumbents' campaigns should not be about specific policy issues, they should focus on their positive records.

Colomer and Llavador (2012) develop an agenda-setting model in which candidates' equilibrium issue choices depend on issue salience and the degree of voter consensus over policy alternatives. The model predicts that candidates optimally emphasize issues on which their pre-

ferred policies are expected to receive stronger voter support, even when those issues are not the most salient. Empirical illustrations from U.S. presidential elections support this logic, showing that candidates prioritize issues that confer a comparative electoral advantage.

Furthering the research on issue selection in electoral campaigns, Greene (2020) argues that the scope of issues addressed in party manifestos affects voter support differently for incumbents and challengers. Incumbents benefit from focusing on fewer issues, as their recent policy activities and established reputations limit the effectiveness of a broad issue scope. In contrast, challengers, who lack recent governing records, gain from addressing a broader range of issues in order to appeal to a wider coalition of voters. Using manifesto data from OECD countries over several decades, Greene (2020) documents a negative correlation between issue diversity and incumbents' vote shares, suggesting that incumbents are electorally rewarded for more focused campaign manifestos on issues where they have a comparative advantage.

Building on these results, this paper shifts the focus from electoral outcomes to campaign behavior. Rather than examining whether issue concentration affects vote shares, I study whether incumbency causally affects how parties structure their campaign messages. The first testable implication I examine is whether incumbents, due to reputational constraints, emphasize fewer issues in their manifestos than challengers. The second is whether parties strategically concentrate their campaign messages on issues where they have a comparative advantage. The following section describes the data used to test these hypotheses.

3 Brazilian Local Elections and Manifesto Data

Brazil's political and administrative organization comprises the Union (federal government), 26 states, 5,569 municipalities, and the Federal District (Brasília). These levels of government are all autonomous and share responsibility for providing essential public services. In this system, mayors play a crucial role, which entails, in addition to legislating on local matters, organizing and providing transportation, education, and health services (Novaes and Schiumerini, 2021).[3] This raises the importance of local electoral competition, as voters are likely to be concerned about mayors' performance when deciding for whom to vote.

Municipal elections are held every four years and determine mayors and the local legislative body. These local contests are not synchronized with federal elections, which take place with a two-year gap and in which voters choose the president, governors, and their respective legislative

[3] Article 30 of the Brazilian Federal Constitution of 1988 describes municipalities' responsibilities.

bodies. The majority of mayoral elections (98%) occur under a single-ballot plurality system.[4] Mayoral candidates must be Brazilian citizens who are at least 21 years old and reside in the municipality in which they run. They must also be affiliated with a political party and may not serve more than two consecutive terms. Even though there are more than 30 parties in Brazil, the median number of candidates contesting a local election is two.

Since 2009, Law No. 12034/09 requires candidates to include their campaign manifestos in the candidacy registration applications, which are submitted approximately two months before election day. Moreover, the Superior Electoral Court (TSE) provides this information jointly with data on the candidates' characteristics, the election results, the campaign contributions and expenditures, and the electorate characteristics.[5] These manifestos are brief documents outlining each candidate's policy priorities. Appendix A.1 provides examples of selected manifestos, while Table 1, presented later in this section, summarizes descriptive statistics from the manifesto dataset. Although constituents do not usually access these documents directly, media outlets use them as sources of information on candidates' political priorities (Sarmento and Vidigal, 2018; Globo, 2020; Souza, 2020).

In this paper, I rely on manifestos from the 2012 and 2016 mayoral elections to construct measures of issue emphasis and manifesto concentration during electoral campaigns. Overall, out of 14,933 and 15,758 eligible candidates in 2012 and 2016, I successfully recover manifesto texts from 12,699 (85%) and 15,549 (98.7%) candidates, respectively. Although most documents are easily readable using available text extraction tools, some consist of image files (i.e., in JPEG or non-searchable PDF format). In these cases, I employ an Optical Character Recognition (OCR) method to convert images into machine-encoded text.[6]

The key empirical challenge of this paper is to construct measures of issue emphasis and issue concentration for more than 28,000 manifestos. For instance, I need to identify whether a candidate emphasizes social or economic issues and whether incumbents concentrate more on fewer issues than their opponents. To overcome this challenge, I employ an unsupervised learning method from computational linguistics known as Latent Dirichlet Allocation (LDA), originally developed by Blei et al. (2003), which has been increasingly used in the social sciences

[4] In municipalities with more than 200,000 eligible voters, elections are held under a dual-ballot (or runoff) system. These municipalities are excluded from this analysis.

[5] Manifestos for 2012 and 2016 were obtained by web scraping the following website: bit.ly/41gU3BW. More recently, TSE made these documents available through its data repository at bit.ly/3Un9038.

[6] To extract text from PDF files, I use the package pdftools (bit.ly/3UqWijR). Approximately 18% of the manifestos were converted using OCR. The OCR methods used were Tesseract (bit.ly/3Umqn47) and the Google Vision API (bit.ly/40ULDjR). The chosen text output for each document was the one produced by the method that yielded the lowest proportion of misspelled words. The proportion of misspelled words was computed using the hunspell package in R (bit.ly/43mgk2P).

and economics (Grimmer and Stewart, 2013; Wilkerson and Casas, 2017; Benoit, 2020).

LDA is a Bayesian probabilistic model. It assumes that a collection of documents is generated by two latent structures: the topic—a vector of word probabilities—and the document, which is itself a vector of topic probabilities. The LDA model assumes that these probability vectors follow Dirichlet distributions (hence the name Latent Dirichlet Allocation). Therefore, topics provide information on the different issues discussed in an election. At the same time, LDA reveals the likelihood that each manifesto discusses a given topic, which is a direct measure of issue emphasis. For instance, the words 'doctors', 'hospital', and 'exams' are more likely to appear in manifestos emphasizing health issues.[7]

Following the literature on text as data, I perform a series of text preprocessing procedures before estimating the LDA model to reduce dimensionality and retain only the elements essential to the analysis (Grimmer and Stewart, 2013; Taddy, 2013; Gentzkow et al., 2019a). I remove punctuation, numbers, and non-word text elements. I also remove stop words such as articles and prepositions and discard words that are too frequent or too rare using term frequency–inverse document frequency (tf-idf). Moreover, the LDA model is conditioned on three exogenous parameters: the number of topics and two Dirichlet priors for the vectors of topic probabilities within documents and word probabilities within topics. To define the values of these parameters, I follow Griffiths and Steyvers (2004) and Hansen et al. (2018). The first paper establishes fixed priors for the Dirichlet parameters to generate more distinguishable topics. The second argues that choosing a larger number of topics may lead to better statistical goodness of fit, while fewer topics improve interpretability. In this paper, I use a baseline model with 40 topics and show that the overall conclusions are robust to an alternative specification with 100 topics.[8]

Figure 1 summarizes the LDA output considering 40 topics.[9] It presents, for the 2016 election, both the average probability that a manifesto discusses each topic (i.e., the size of each rectangle) as well as the five most frequent words within each topic. Following common practice in the literature, I classify and label topics subjectively based on their most frequent and highest-probability words, using the dominant terms within each topic to guide interpretation (Hansen et al., 2018; Mueller and Rauh, 2018). Although the within-topic word probability vector spans all unique terms in the corpus, only a small subset of words has relatively high probability within each topic. Moreover, many topics combine words related to specific policy

[7]For brevity, I do not formally describe the model here. See Blei et al. (2003) for a detailed description of the LDA method. Estimations were conducted using the R package topicmodels by Grün and Hornik (2011).

[8]Appendix A provides a detailed description of the text preprocessing procedures and the choice of parameters for the LDA model. Section 5.2 presents the sensitivity of the results to a model with 100 topics.

[9]The original Portuguese version of Figure 1 is displayed in Appendix Figure B.4. Table B.1 reports the ordered topics, their top five terms, assigned themes, and average topic probabilities used to construct Figure 1.

issues, as illustrated in Figure 1. For example, topic 4 is classified as *culture* because its most likely words include *music*, *theater*, and *festival*. Other topics, such as topic 2, are not linked to a specific policy issue but instead capture features of candidates' discourse, with *continuity* and *continue* among the most probable words. Finally, some topics cannot be assigned to specific issue groups and are shown in the light gray block.

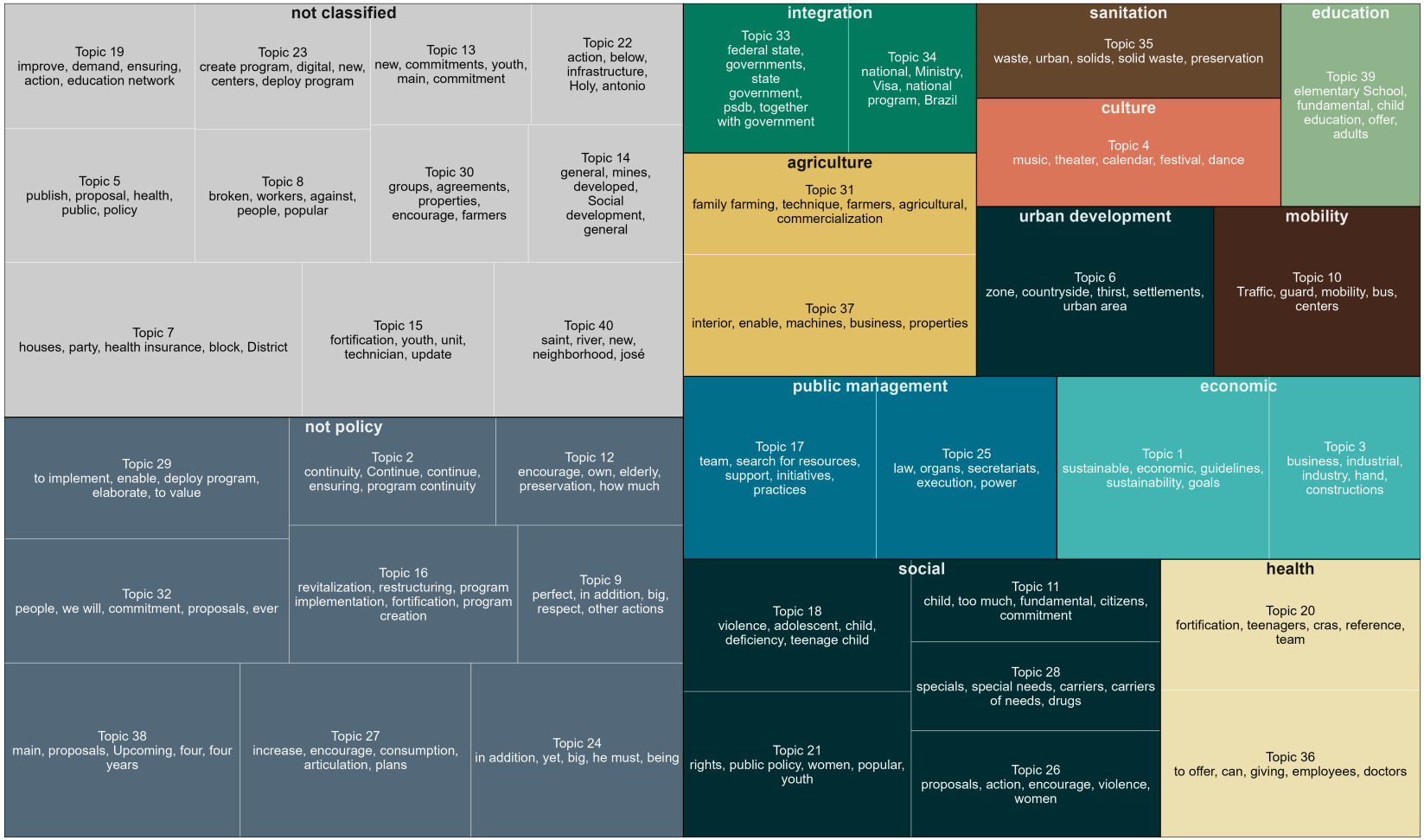


Figure 1: Topic model output - top five most frequent words in each of the 40 topics

Notes: The size of each block represents the average probability of a topic appearing in a manifesto in 2016. Colors indicate a subjective classification of topics into 'policy issues', 'not policy issues', or 'unclassified'.

To validate the topics obtained, Figure 2 shows that manifesto contents follow expected geographical patterns and emphasize regionally specific issues.[10] The left panel features a word cloud of the 50 most frequent words in topic 10 (*urban mobility*) and a correlation plot linking urban population share to the likelihood of discussing urban mobility. As expected, candidates in more urbanized areas—especially those with over 75% urban population—emphasize this issue more in their manifestos. The right panel shows a negative correlation between the probability of discussing topic 6 (*urban development*) and the municipality's Human Development Index (HDI), with candidates in less developed areas more likely to address issues related to poor urban infrastructure.

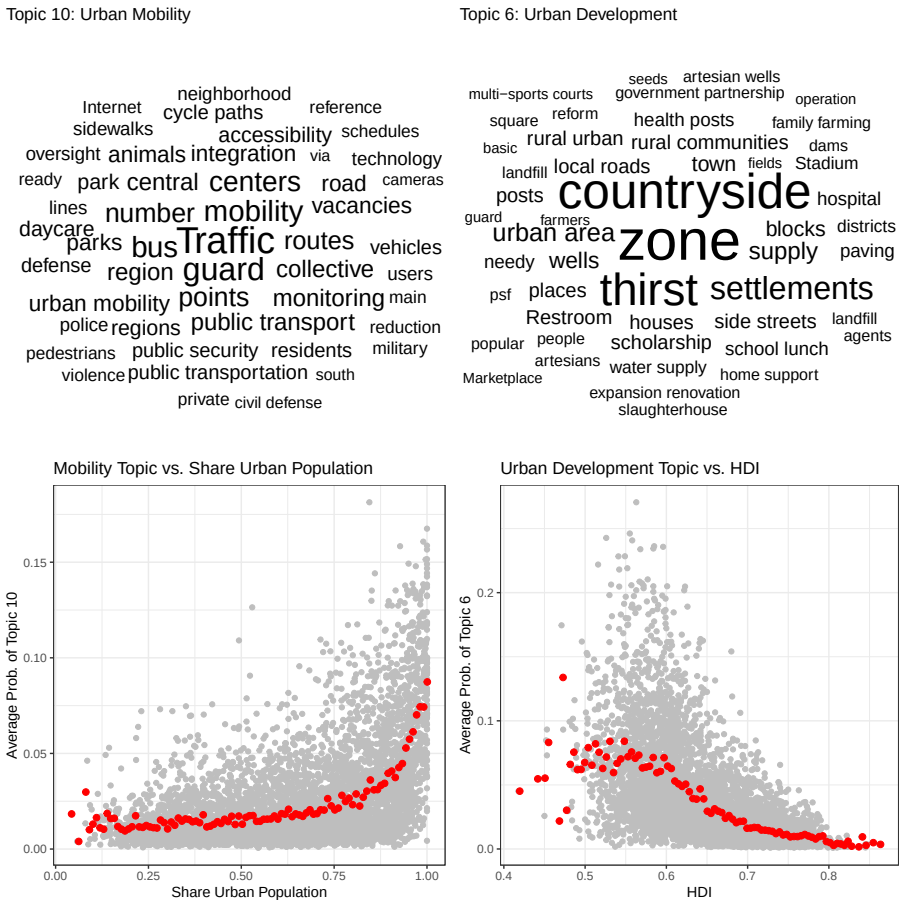


Figure 2: Validation - emphasis on mobility and urban development issues follows expected geographical patterns

Notes: Manifestos in more urban municipalities are more likely to discuss urban mobility, while those in areas with lower HDI place greater emphasis on urban development. Grey dots represent municipal average topic probabilities; red dots represent binned averages (100 bins).

[10]The original Portuguese version of Figure 2 is displayed in Appendix Figure B.5.

Figure 3 presents an additional example of meaningful geographic clustering of topics. The word clouds indicate that topics 31 and 37 are related to agriculture.[11] The lower-left plot displays a positive correlation between the number of agricultural establishments per thousand inhabitants in a municipality (according to the 2017 agricultural census) and the likelihood of discussing agriculture in manifestos. The lower-right plot indicates a negative correlation between the urban population share and the probability of discussing agriculture-related topics.

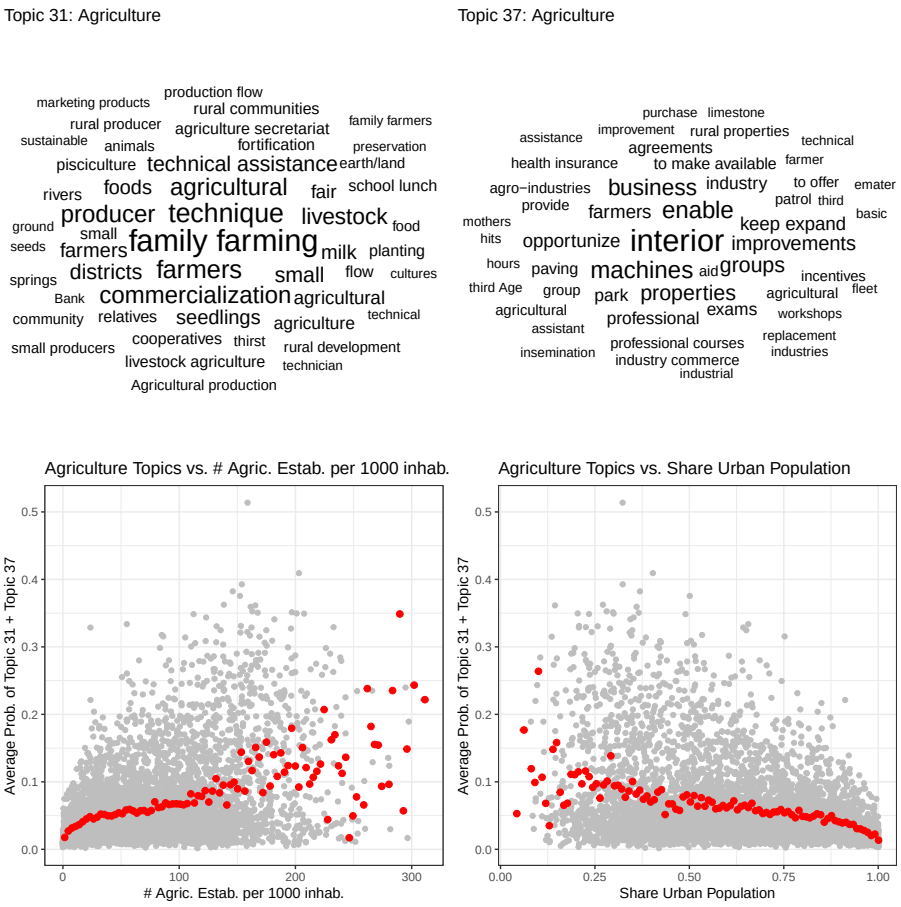


Figure 3: Validation - emphasis on agricultural issues follows expected geographical patterns

Notes: Manifestos in municipalities with more agricultural establishments per 1,000 inhabitants are more likely to discuss agriculture, while the likelihood of discussing agriculture decreases as the urban population share increases. Grey dots represent municipal average topic probabilities; red dots represent binned averages (100 bins).

As a second type of validation of the topics obtained, I present evidence that changes in the average manifesto topic distribution in the 2016 election, relative to the 2012 election, follow a logical pattern. For instance, in the 2016 election, mayoral candidates were more likely to

[11] The original Portuguese version of Figure 3 is displayed in Appendix Figure B.6.

discuss topics related to *sustainable development* (topic 1) and *public management* (topic 17) than in the 2012 election. Figure 4 summarizes these findings.[12]

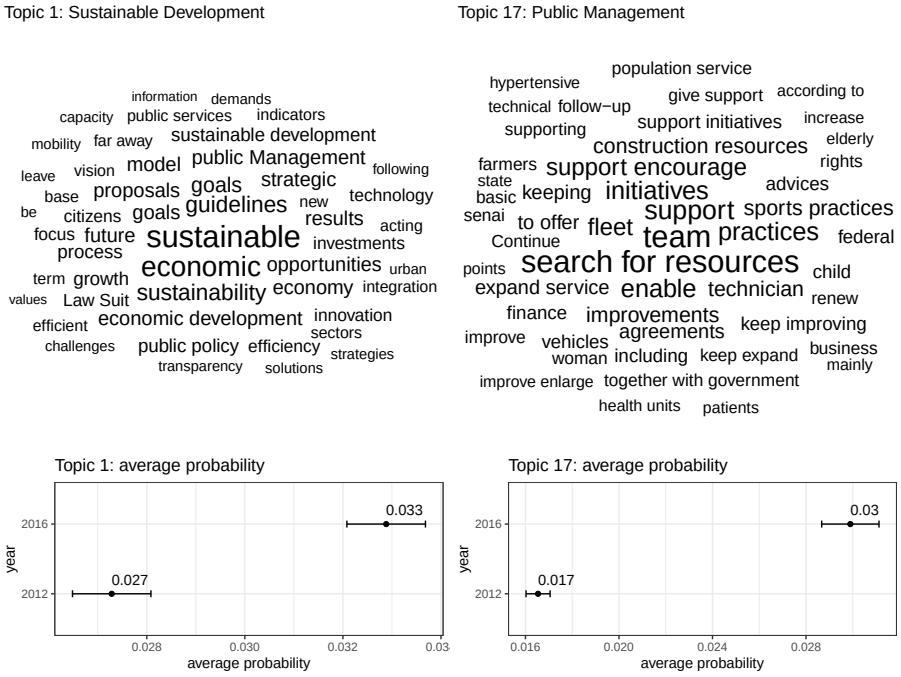


Figure 4: Validation - emphasis on sustainable development and public management issues follows expected temporal patterns

Notes: Sustainable development and public management topics are more likely to be discussed in the 2016 election relative to the 2012 election. In the bottom plots, dots represent the average probability of discussing each topic by year. Lines indicate the corresponding 95% confidence intervals.

This change in manifesto topic distribution appears to be a logical response by candidates to the deterioration of the Brazilian economy (GDP fell by more than 3% per year in 2015 and 2016) and the political instability following the impeachment of President Dilma Rousseff in August 2016. Candidates may have felt compelled to discuss ways to recover from the economic crisis and to improve the municipal fiscal and administrative situation.

Finally, Figure 5 shows that the average probability that a manifesto discusses topics related to essential needs does not change across elections.[13] The topics of *sanitation* (topic 35) and *education* (topic 39) reflect issues that persist in Brazilian municipalities over time. Hence, mayoral candidates in the 2012 election should be equally likely to discuss these issues as candidates in the 2016 election.

[12]The original Portuguese version of Figure 4 is displayed in Appendix Figure B.7.

[13]The original Portuguese version of Figure 5 is displayed in Appendix Figure B.8.

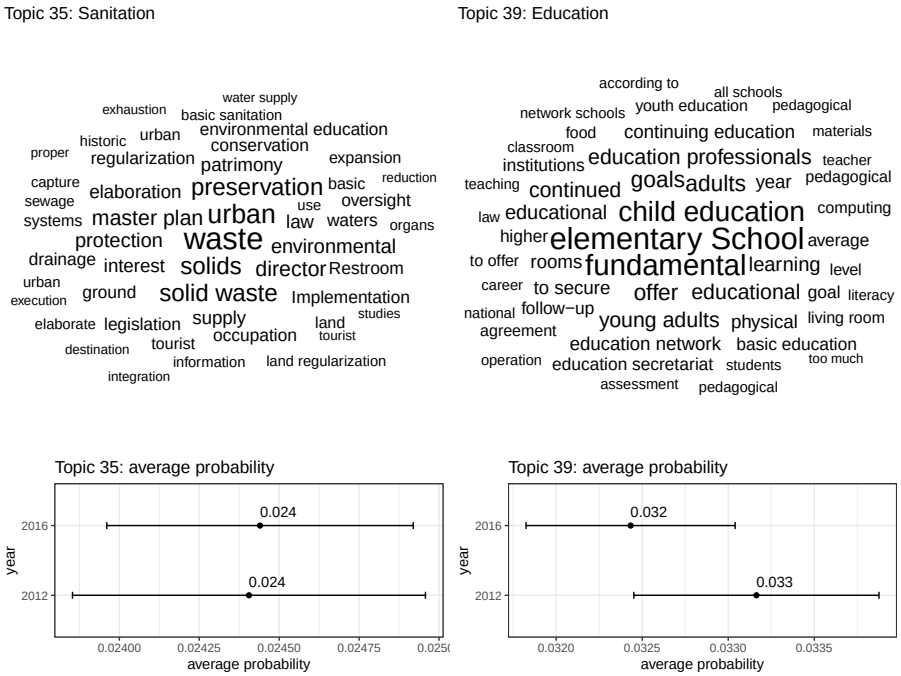


Figure 5: Validation - emphasis on essential needs issues follows expected temporal patterns

Notes: The probability of discussing topics that reflect essential needs, such as sanitation and education, remains relatively stable over time. In the bottom plots, dots represent the average probability of discussing each topic by year. Lines indicate the corresponding 95% confidence intervals.

The validation exercises above show that the LDA model successfully recovers relevant topics and provides meaningful information about the probability that a manifesto emphasizes different policy issues. Relying on this vector of topic probabilities, I construct a measure of manifesto issue concentration. More specifically, I calculate a Herfindahl–Hirschman Index (HHI) for each manifesto based on its vector of topic probabilities.[14] More formally, let θ_{dt} be a vector containing the probabilities that a manifesto d in election t discusses each of the K different topics estimated using the LDA model. In the baseline case, as I estimate a model with 40 topics, the length of the vector θ_{dt} is 1×40. The following expression defines the HHI measure.

$$HHI_{dt} = \sum_{k=1}^{K} \theta_{kdt}^2 \tag{1}$$

Hence, the HHI helps identify whether a manifesto is more focused on particular topics, although it does not indicate precisely which topics it emphasizes. In Section 5.2, I verify that the results are robust to alternative measures of manifesto concentration.

Table 1 provides descriptive statistics for the manifesto dataset analyzed in this study. It reports medians, means, and standard deviations for variables related to manifesto size and structure, the HHI indicator, and the estimated average probabilities of discussing each of the issues described in Figure 1. After text preprocessing, manifestos contain, on average, approximately 1,300 words. About 70% of the text consists of unique terms. Moreover, politicians tend to use bullet points and short, simple sentences, with the average sentence containing 18 words (roughly two lines). Issue concentration, as measured by the HHI, averages around 0.12.[15]

[14] The HHI measure is frequently applied in industrial organization research to evaluate market power. It is usually calculated using firms' market shares (Rhoades, 1993).

[15] Overall, a comparison of incumbent and challenger manifestos shows no systematic differences in these characteristics. The table is available upon request.

Table 1: Descriptive Statistics of Political Manifestos, 2012-2016

	Median	Mean	Std Dev.
N. Words	930	1277.17	1323.11
Prop. Un. Words	0.61	0.62	0.12
N. Words by Phrase	17.09	18.08	7.43
HHI	0.07	0.12	0.14
Agriculture	0.04	0.06	0.07
Continuity	0.01	0.02	0.05
Culture	0.02	0.03	0.03
Economic	0.04	0.05	0.05
Education	0.01	0.02	0.03
Health	0.04	0.06	0.05
Integration	0.03	0.04	0.04
Mobility	0.01	0.02	0.03
Not Classified	0.23	0.25	0.13
Not Policy	0.18	0.21	0.14
Public Management	0.03	0.05	0.07
Sanitation	0.01	0.02	0.03
Social	0.08	0.12	0.13
Urban Development	0.02	0.04	0.05

Notes: Table includes all candidates manifestos for which topics were successfully recovered; Excluding municipalities with more than 200 thous. voters (runoff elections); N.Words: word count of each manifesto; Prop. N. Words: proportion of unique words excluding repetitions; N.Words by Phrase: N.Words/(End Point Count + Semi Colon Count);

4 Research Design

This paper studies whether holding the mayor's office affects the decision to produce a manifesto focused on fewer issues. However, this effect cannot be measured without bias by simply comparing the average manifesto issue concentration between incumbents and challengers, since assignment to treatment (i.e., being the incumbent) is not random. For instance, a candidate's previous experience and unobservable abilities may simultaneously be correlated with both her decision to focus her manifesto on certain issues and her probability of becoming the incumbent.

To avoid this problem, I rely on a regression discontinuity design (RDD) (Lee, 2008; Lee and Lemieux, 2010; Calonico et al., 2015; Cattaneo et al., 2019). In this design, units with scores exceeding a threshold are assigned to treatment. Under certain assumptions, units marginally below or above the threshold can be considered almost randomly assigned to treatment and control groups, allowing for the estimation of the causal effect of treatment on an outcome of interest.

In this application, I exploit the discontinuity in the election outcome generated by the vote margin (the running variable). The main unit of analysis is the party (i) in municipality (m) in

a given election year (t). The treatment and control groups comprise parties that respectively barely won or lost election t in municipality m. The outcome of interest is the party's manifesto issue concentration (HHI) in the following election, $t+1$. The RDD estimates identify the local average treatment effect of incumbency for parties involved in closely contested elections. Throughout the remainder of the paper, references to the "incumbency effect" should be understood in this local sense.

For a party to be treated in $t+1$ (i.e., to be the incumbent in the $t+1$ election), in addition to being elected in t (i.e., having a positive vote margin in the t election), it must also run in two consecutive elections. As Magalhaes (2015) and Klašnja and Titiunik (2017) point out, in Brazilian local elections, this restriction might lead to selection bias when estimating the incumbency effect, since many candidates do not run in two consecutive elections, and the election outcome in t may differentially influence the decision to run again in $t+1$ for winners and runners-up, depending on characteristics such as ability, ambition, or outside career opportunities.

To illustrate the selection problem, following Magalhaes (2015), consider a mayoral election in which two candidates have similar electoral strength but differ in underlying political ability or ambition. A high-ability runner-up who loses the election in t by a narrow margin may interpret the close defeat as a signal of viability and choose to run again in $t+1$ or pursue higher office. In contrast, a low-ability runner-up facing higher costs of rerunning may exit electoral politics altogether. Among incumbents, however, rerunning is typically less costly and more likely regardless of ability. As a result, conditioning on parties or candidates that contest two consecutive elections induces a selection issue: among rerunners, runner-ups tend to be positively selected on ability relative to incumbents.

This selection is problematic for estimations that condition on rerunning, because differences in outcomes in the $t+1$ election, such as manifesto concentration, may reflect underlying candidate quality rather than the causal effect of incumbency. For instance, if high-ability runners-up are more likely to produce focused manifestos, then the estimates in this paper would understate the true effect.

The papers above are interested in estimating the effect of incumbency on vote share or the probability of getting elected. Both address the selection problem by estimating an unconditional version of the incumbency effect, where the outcome variable in $t+1$ (i.e., vote share or the election outcome) is replaced by zero when the candidate or party does not run in the following election. I cannot fully implement a similar solution in this paper, since I do not observe manifestos for candidates who do not run in the election. This information is crucial, especially for the treated units, because I expect that incumbents should be the ones concentrating their

campaign manifestos on fewer issues.

To overcome this potential selection issue, I show that the results are consistent and robust when estimating effects for alternative samples. First, I estimate the most straightforward case, conditioning on the party rerunning, as described more formally in the equation below:

$$TE = \lim_{c \downarrow 0} \mathbb{E}[Y_{i,t+1} | V_{i,t} = c, R_{i,t+1} = 1] - \lim_{c \uparrow 0} \mathbb{E}[Y_{i,t+1} | V_{i,t} = c, R_{i,t+1} = 1] \quad (2)$$

Where $Y_{i,t+1}$ is the HHI measure of party i's manifesto issue concentration in the $t+1$ election (either 2012 or 2016), $V_{i,t}$ corresponds to the vote margin of party i in election t (i.e., the running variable), $R_{i,t+1} = 1$ indicates that party i runs in both elections t and $t+1$, and c is the cutoff point above which the party is elected and becomes treated. Following the guiding hypothesis, I expect the treatment effect to be positive, as incumbents should present more focused manifestos than their opponents.

Second, I show that the results for the incumbency effect hold under a weaker sample restriction, where estimates are conditional only on the winner contesting two consecutive elections. In this case, I use the manifesto concentration of the best-ranked party (different from the incumbent) in municipality m and election $t+1$ as a proxy for the challenger that did not run in the following election. Third, I provide evidence that the estimates are robust to a more restrictive case, conditional on observing in $t+1$ the manifestos of both the winner and the runner-up parties in municipality m and election t.

To estimate the incumbency effect on manifesto issue concentration in $t+1$, I use local first- or second-order polynomials estimated on each side of the cutoff, weighted by triangular kernels based on the distance between the cutoff and each observation's score. As in most RDD applications (Cattaneo et al., 2019), I show that the results are robust to different bandwidth values (h) around the cutoff. To select these values, I first use the algorithm provided by Calonico et al. (2015) to obtain the optimal bandwidth ($\hat{h}$) for each outcome variable.[16] Then, following the RDD literature, I verify the robustness of the results to half and twice the value of the optimal bandwidth.

The primary analysis is conducted at the party-municipality level for three main reasons. First, candidate-level data are more limited because, as noted earlier, re-election rates are low in Brazil. Second, while candidates face a two-term limit, parties do not. Therefore, comparing elections in which an incumbent party fields a new candidate to those in which both the party and the candidate are incumbents can provide valuable insights into the relationship between

[16] For implementation, I use the `rdrobust` software available in R at https://cran.r-project.org/web/packages/rdrobust/rdrobust.pdf.

reputation and manifesto issue concentration. Third, party switching is common in Brazilian politics—34.6% of candidates who ran in both the 2012 and 2016 elections changed parties in 2016. Candidates who switch parties may face different incentives when deciding which issues to emphasize. Nonetheless, Section 5.2 shows that the results are consistent even when estimated at the candidate-municipality level.

To track parties over time, I use their registration numbers, ensuring continuity even if party names change. No parties were dissolved during the analysis period. However, five new parties were established: the Social Democratic Party (PSD) in 2011, Solidarity (Solidariedade) in 2013, and the New Party (Novo), Sustainability Network (Rede Sustentabilidade), and Brazilian Women's Party (PMB) in 2015. Of these, PSD was the most relevant, with its candidates ranking among the top two in 16.6% of municipalities in the 2012 election. In contrast, the other four parties collectively had top-two candidates in only 3.1% of municipalities in the 2016 election. Since these new parties did not clearly emerge from specific predecessors, I treat them as entirely new entities. For example, if a municipality elected PT in 2008 but PSD in 2012, I classify this as the incumbent party not being re-elected in 2012.[17]

Finally, the sample analyzed includes only parties in municipalities with single-ballot elections (i.e., cities with fewer than 200,000 voters), since electoral competition is likely to differ in dual-ballot elections. Moreover, the main results are based on an HHI measure of manifesto issue concentration computed using the LDA model with 40 topics. In Section 5.2, I present a series of exercises demonstrating the robustness of the results to alternative specifications, including a candidate-level analysis, an LDA model with 100 topics, and alternative measures of manifesto issue concentration.

5 Results and Robustness

5.1 Main Results

In this section, I estimate the causal effect of incumbency on parties' manifesto issue concentration. Table 2 reports the results for the different alternatives for dealing with the selection issue described in the previous section. In each panel, I report the bias-corrected estimator and standard errors following (Calonico et al., 2020). Panel A shows the results conditional on at least one of the two parties in a municipality-the incumbent or the challenger-rerunning in the subsequent election. Panel B reports estimates conditional only on the winner contesting two

[17] Main results are robust to excluding municipalities where a new party was running.

consecutive elections, using the best-ranked non-incumbent party in $t+1$ as the comparison group when the original challenger does not rerun. The most restrictive specification is presented in Panel C, where estimates are conditional on observing, in $t+1$, the manifestos of both the winner and the runner-up parties in t in the same municipality.

Sample sizes vary across panels due to different rerunning requirements. Panel A includes 7,441 elections, with 5,532 incumbent and 3,877 challenger observations, for a total of 9,404 candidate observations. Panel B includes 5,532 elections—one for each incumbent rerunning—and 5,201 challenger observations. The difference of 331 challenger observations arises because, in some municipalities, no manifesto information is available for any challenger in election $t+1$ to proxy the runner-up from election t. The original challenger from election t reruns in approximately 70% of cases; in the remaining 30%, the challenger's HHI is proxied using the best-ranked non-incumbent party in $t+1$. Panel C includes 1,968 elections in which both the winner and the runner-up from election t are observed again in $t+1$.

The columns in Table 2 explore the robustness of the results to different specification choices. Column (1) shows the mean difference between incumbents and challengers without restricting the sample to close elections. Column (2) presents the main estimates based on the MSE-optimal bandwidth ($\hat{h}$) proposed by Calonico et al. (2015) for a local linear specification. Columns (3) and (4) report the sensitivity of the estimates to bandwidths equal to half and twice the MSE-optimal bandwidth. Column (5) adds municipal characteristics as controls.[18] Columns (6) and (7) present estimates based on a quadratic specification.

[18] Municipal controls include per capita GDP, population, the shares of agriculture, industry, and services in GDP, and the number of apt candidates in each election.

Table 2: RD Effect of Winning at t on the Manifesto's HHI at t+1 (Party)

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: At Least One Party Reruns (Incumbent or Challenger) in $t+1$							
Control Mean	0.1121	0.1122	0.1087	0.1122	0.1117	0.1129	0.1132
Estimate	0.0046	0.0173**	0.0258*	0.017**	0.0182**	0.0226**	0.0215**
Std.Err.	(0.0037)	(0.0074)	(0.0137)	(0.0073)	(0.0078)	(0.0099)	(0.0096)
BW	1	0.1647	0.0823	0.3294	0.1469	0.2011	0.2098
P-Value	0.219	0.0198	0.0606	0.0193	0.0196	0.022	0.0252
N.Obs	9409	6563	4031	8507	6116	7274	7378
Panel B: Incumbent Reruns, Challenger Proxied when Missing in $t+1$							
Control Mean	0.1134	0.1129	0.1085	0.114	0.1127	0.1129	0.113
Estimate	0.0027	0.0187**	0.0211	0.0166**	0.0196**	0.0249**	0.0246***
Std.Err.	(0.0034)	(0.0078)	(0.0146)	(0.0073)	(0.0079)	(0.0097)	(0.0095)
BW	1	0.1403	0.0702	0.2807	0.1326	0.189	0.1922
P-Value	0.4389	0.016	0.1484	0.0237	0.0134	0.0102	0.0096
N.Obs	10733	6454	3721	9166	6206	7731	7808
Panel C: Both Winner and Runner-Up Rerun in $t+1$							
Control Mean	0.1114	0.1094	0.1017	0.1108	0.1093	0.1096	0.1095
Estimate	-1e-04	0.0278**	0.0492**	0.0219*	0.027**	0.0483***	0.0483***
Std.Err.	(0.0057)	(0.0128)	(0.0238)	(0.0121)	(0.0126)	(0.0164)	(0.0163)
BW	1	0.1246	0.0623	0.2492	0.1271	0.1495	0.1499
P-Value	0.9876	0.0295	0.0384	0.0691	0.0319	0.0033	0.0031
N.Obs	3936	2392	1424	3390	2428	2694	2694

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates exclude cities with runoff voting. The outcome variable is the manifesto's Herfindahl-Hirschman Index (HHI) of topic concentration in t+1 (2012 or 2016). The running variable is the vote margin in t (2008 or 2012). Table reports bias-corrected estimates and standard errors following (Calonico et al., 2020). Panel A conditions on at least one of the two parties (incumbent or challenger) rerunning in $t+1$. Panel B conditions only on the incumbent rerunning, using the best-ranked non-incumbent party in $t+1$ as the comparison group when the original challenger does not rerun. Panel C conditions on observing both the winner and the runner-up from election t in $t+1$ within the same municipality.

The results reveal that when the incumbent party barely wins the election in t and runs again in the next election, it produces a manifesto that is 17% more focused than when it barely loses. For the linear specification in Panel B, the average HHI of an incumbent party's manifesto is approximately 0.019 higher than the control mean of 0.11. Reassuringly, the estimates are statistically significant at the 95% level and consistent across alternative specifications and the different sample selection approaches presented in Panels A to C.

To better understand the magnitude of this effect, it is useful to consider an example of what an increase of this size in the HHI index could imply for the structure of a manifesto. Because the HHI is computed as the sum of squared topic probabilities, it places a greater weight on the most salient issues. For example, moving from a manifesto in which the two most prominent issues together account for 40% of total attention (20% each) to one in which a single issue receives 31% of the content, while the remaining attention is distributed evenly across other topics, implies an increase in concentration of approximately 0.018. This change is very close to the estimated effect of approximately 0.019 reported above. Hence, an increase of this mag-

nitude is consistent with a meaningful reallocation of manifesto content toward the top issues. Such a shift reflects a visibly more focused manifesto and is consistent with incumbent parties strategically narrowing their issue emphasis in subsequent elections.

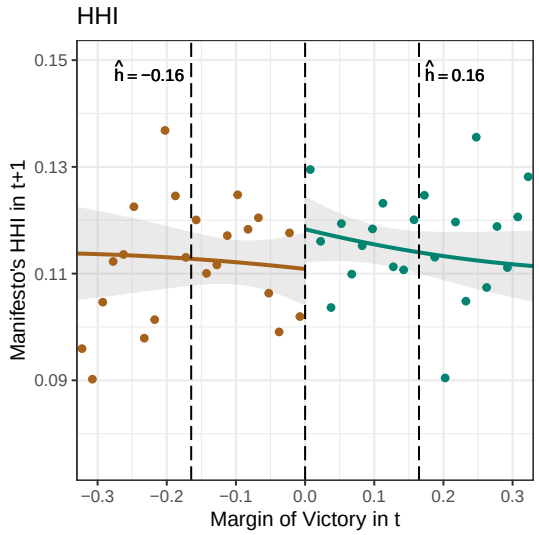
Figure 6 illustrates the effect. The dots represent the average manifesto HHI in $t+1$ for local binned means (bin width 0.015) of the party's vote margin in t. The solid lines show predicted values from local second-order polynomials fitted on each side of the cutoff at zero, and the grey shaded areas denote 95% confidence intervals.

Despite the estimated effects being strong and robust, I must still verify that there are no threats to their validity. Following the literature, I present two validity checks of the RD design Cattaneo et al. (2019). I begin by examining whether predetermined covariates are balanced around the cutoff c. Then, I assess whether the density of the vote margin is discontinuous at the cutoff.

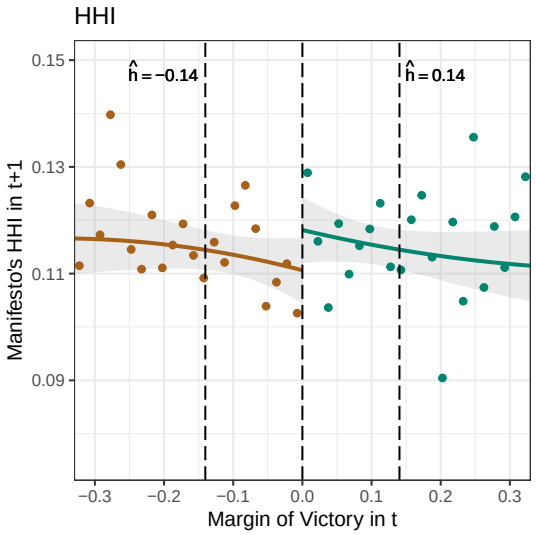
To verify covariate balance, I run an RD specification similar to the one presented above, using covariates that should not be affected by the election outcome at t as dependent variables. This test aims to ensure that treated and control units are comparable at baseline. Table 3 presents the results for the Panel A sample, which is conditional on at least one of the two parties in a municipality—the incumbent or the challenger—rerunning in the subsequent election. The estimated treatment effects for the different predetermined covariates are reported in each row. The results show no evidence of discontinuities in covariates at the cutoff. This pattern also holds for the samples in Panels B and C; the corresponding results are reported in Tables C.2 and C.3 in the Appendix.

In addition, to address the concern that incumbency may affect campaign resources in the subsequent election and thereby confound the estimated effects on manifesto concentration, I examine whether party campaign expenditures in $t+1$ exhibit a discontinuity at the electoral cutoff. The results, also reported in Table 3 and in Appendix Tables C.2 and C.3, show no evidence of discontinuous jumps in campaign spending in $t+1$ at the threshold.

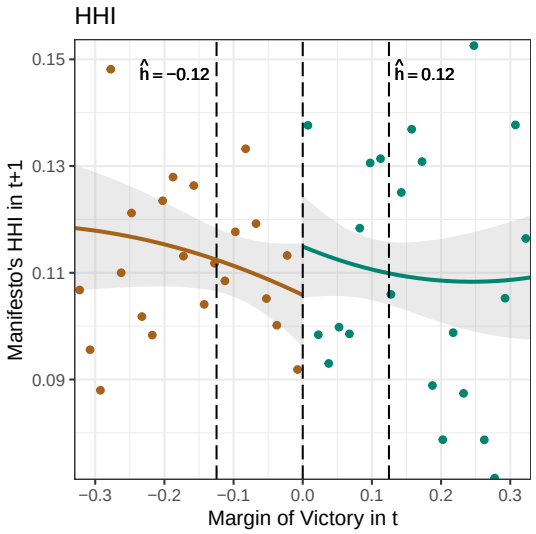
Figure C.9 graphically illustrates this result. It also shows that, while campaign expenditures increase smoothly with electoral performance, incumbent parties elected by small vote margins in t do not experience a discrete increase in campaign resources in $t+1$ relative to parties that barely lost. This pattern, together with the declining trend in manifesto concentration shown in Figure 6, suggests that for parties winning by large margins—where campaign spending is substantially higher—there may be some substitution between campaign resources and manifesto issue concentration. By contrast, such substitution does not appear to operate in closely contested elections around the cutoff.



(a) At Least One Party Reruns (Incumbent or Challenger) in $t+1$



(b) Incumbent Reruns, Challenger Proxied when Missing in $t+1$



(c) Both Winner and Runner-Up Rerun in municipality m in $t+1$

Figure 6: RD Effect of Winning at t on the Party Manifesto's HHI at $t+1$

Notes: The dots represent the average manifesto HHI in $t+1$ for local binned means (bin width 0.015) of the party's vote margin in t. The solid lines show predicted values from local second-order polynomials fitted on each side of the cutoff at zero, and the grey shaded areas denote 95% confidence intervals. Vertical dashed lines indicate the estimated bandwidth $\hat{h}$ and the cutoff.

Away from the discontinuity, the decline in manifesto concentration may reflect differences in electoral incentives. In less competitive elections, incumbents with larger margins may rely more on campaign resources and broader outreach strategies, reducing the need to focus their

Table 3: Predetermined Covariate Smoothness in t
Panel A: At Least One Party Reruns (Incumbent or Challenger) in $t+1$

Dep. Var.	Mean Control	Estimate	St. Err.	P-Val	BW	N
GDP per capita ('000 BRL)	12.676	0.185	0.670	0.782	0.135	5,811
Population ('000)	20.052	-1.495	1.707	0.381	0.148	6,162
Agric. Share GDP	0.213	0.002	0.009	0.823	0.162	6,497
Indus. Share GDP	0.129	0.004	0.009	0.699	0.108	4,956
Serv. Share GDP	0.285	-0.005	0.006	0.464	0.136	5,850
Number of Apt Candidates	2.718	-0.021	0.060	0.728	0.143	6,040
Small Party	0.058	0.006	0.015	0.691	0.141	5,988
MDB Wins	0.220	-0.012	0.025	0.632	0.154	6,245
PSDB Wins	0.151	0.005	0.020	0.802	0.192	7,046
PP Wins	0.135	-0.013	0.021	0.547	0.137	5,823
PT Wins	0.093	-0.010	0.019	0.613	0.128	5,590
PSB Wins	0.036	-0.00001	0.011	1.000	0.159	6,368
Age	48.049	-0.302	0.627	0.630	0.134	5,794
Gender: Female	0.108	0.012	0.019	0.545	0.155	6,338
Educational Attainment: High School or Higher	0.797	0.011	0.024	0.637	0.148	6,162
Campaign Expenditure in t ('000 BRL)	84.359	-7.251	9.969	0.467	0.110	4,993
Campaign Expenditure in t+1 ('000 BRL)	94.294	-3.760	10.018	0.707	0.148	6,160
Delta Expenditure ('000 BRL)	10.936	-1.077	10.039	0.915	0.111	5,073
HHI in t (no observations for 2008)	0.128	-0.015	0.014	0.274	0.121	2,391
CK5 in t (no observations for 2008)	0.520	-0.002	0.014	0.911	0.128	2,507

Notes: Table shows estimates of separate RD treat. effects for each predetermined covariate; Excl. municipalities with runoff voting; BW: vote margin in t; Estimates obtained using local linear regressions and triangular kernel weights; Estimates and 95% confidence intervals are robust-bias corrected following (Calonico et al., 2020); Candidates were not required to publish political manifestos in 2008. Therefore, this year is excluded from the calculation of RD effects for predetermined outcomes (HHI and CK5). Delta Expenditure is equal to Campaign Expenditure in t+1 - Campaign Expenditure in t.

manifestos on a narrower set of issues. In addition, candidates who win by large margins may already be better known to voters, so that valence may play a relatively larger role. Conversely, in close races, parties may rely more on focused campaign messages, leading to higher issue concentration.

A remaining limitation is that the data do not allow for a direct test of whether incumbency affects access to media following the election. Nevertheless, within the RD setting, the absence of discontinuities in observable campaign resources in $t+1$ points against mechanisms operating primarily through differential access to funds or to the media, and is more consistent with interpretations in which the observed increase in issue concentration reflects incentives associated with incumbency status itself.

I proceed by examining if the density of the vote margin is discontinuous at the cutoff. A discontinuity in the running variable would indicate that units select themselves into treatment, ultimately compromising the RD estimates. Figure C.10 in Appendix C shows the density tests following Cattaneo et al. (2019). As mentioned in the previous section, conditioning on parties' decisions to rerun in the subsequent election may introduce a selection problem. This issue arises in Panel A of Table 2, which conditions on at least one of the two parties in a municipality—the

incumbent or the challenger—rerunning in the next election. In this case, for a 90% confidence interval, one fails to reject the null hypothesis of no discontinuity in the density of treated and control groups at the cutoff. In contrast, I find no evidence of discontinuities when conditioning solely on the incumbent party rerunning in the subsequent election (Panel B of Table 2) or when conditioning on observing both the winner and the runner-up parties from the same municipality rerunning in the subsequent election (Panel C of Table 2).

To further strengthen the results, I show that the findings also hold when focusing on within-party variation over time, rather than relying exclusively on cross-sectional comparisons. Specifically, I find that incumbent parties elected by a narrow vote margin in t concentrate their manifestos more in $t+1$, even when comparing the same party across elections. To this end, I implement a difference-in-discontinuities design in which the dependent variable is the change in manifesto concentration, measured as the difference in HHI between consecutive elections ($\Delta HHI = HHI_{i,t+1} - HHI_{i,t}$), rather than the level of HHI in $t+1$. By exploiting within-party changes over time, this approach differences out unobservable party characteristics that are fixed across elections. The results are reported in Table 4.

Table 4: RD Effect of Winning at t on the Manifesto's Delta HHI (Party)

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control Mean	-0.0221	-0.0221	-0.028	-0.025	-0.0232	-0.0268	-0.0266
Estimate	0.0233**	0.0413**	0.0011	0.0374**	0.0421**	0.0456**	0.0461**
Std.Err.	(0.0095)	(0.019)	(0.0198)	(0.018)	(0.0185)	(0.0211)	(0.0205)
BW	1	0.0848	0.0424	0.1696	0.0826	0.1281	0.1261
P-Value	0.0142	0.0299	0.957	0.0376	0.0231	0.0304	0.0249
N.Obs	1598	732	402	1172	718	978	966

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates exclude cities with runoff voting. Conditioning on both the winner or runner-up parties in 2012 running again in 2016. The outcome variable is the manifesto's Delta Herfindahl Hirshman Index (HHI) of topic concentration (2016-2012). The running variable is the vote margin in 2012. Table reports bias-corrected estimates and standard errors following (Calonico et al., 2020).

Across most specifications, the estimated effects are positive and statistically significant. These results indicate that incumbent parties that won the 2012 elections by a small vote margin increase the concentration of their manifestos in the subsequent election relative to parties that barely lost in 2012 and ran again in 2016. Two points are worth noting. First, because manifesto data are available only for the 2012 and 2016 elections, it is not possible to directly test whether the parallel trends assumption holds. Second, the mean change in HHI among control parties is negative, suggesting that part of the estimated relative increase in concentration for winners reflects a decrease in concentration among narrowly losing parties. This pattern is nonetheless consistent with the theoretical framework in (Greene, 2020), which argues that challengers face

incentives to adopt less concentrated manifestos.

5.2 Robustness Exercises

In this section, I investigate whether the estimated effects are robust to a series of additional tests. From this point onward, all results are based on the sample in Panel B in Table 2. Using this sample ensures that there is no discontinuity in the density of treated and control groups. Moreover, compared to Panel C in Table 2, this sample provides greater power to estimate coefficients and heterogeneous effects that I will present in the following section.

***Different specifications*:** Figure 7 presents a more thorough analysis of how different bandwidth choices, kernels, and polynomial orders affect the main results. Overall, the estimates are very robust to alternative kernels, polynomial orders, and bandwidths.

***Candidate-level analysis*:** Results do not change significantly when the unit of analysis is the candidate-municipality level instead of the party-municipality level. Table 5 shows the corresponding estimates at the candidate-municipality level. Although the estimates are less statistically significant in some specifications due to the more restrictive sample, their magnitude is similar to the results presented in Table 2.

Table 5: RD Effect of Winning at t on the Manifesto's HHI at t+1 (Candidate)

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control Mean	0.1146	0.1166	0.112	0.1151	0.1172	0.1146	0.1145
Estimate	0.0031	0.0214**	0.0185	0.0181**	0.0236**	0.0264**	0.0262**
Std.Err.	(0.004)	(0.0094)	(0.0186)	(0.0089)	(0.0098)	(0.0111)	(0.011)
BW	1	0.1235	0.0618	0.2471	0.1118	0.1887	0.1904
P-Value	0.4419	0.0231	0.3199	0.0414	0.016	0.018	0.0166
N.Obs	8590	5109	2879	7327	4733	6540	6561

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates exclude cities with runoff voting. Conditioning on municipalities where the winning candidate in t also run in t+1. The outcome variable is the manifesto's Herfindahl Hirshman Index (HHI) of topic concentration in t+1 (2012 or 2016). Herfindahl Hirshman Index (HHI) of topic concentration built using outputs of LDA model with 40 topics. The running variable is the vote margin in t (2008 or 2012). Table reports bias-corrected estimates and standard errors following (Calonico et al., 2020).

***LDA model with 100 topics*:** To address the concern that alternative LDA specifications may yield different estimates, Table 6 shows that the results are similar when a topic model with 100 topics is employed.

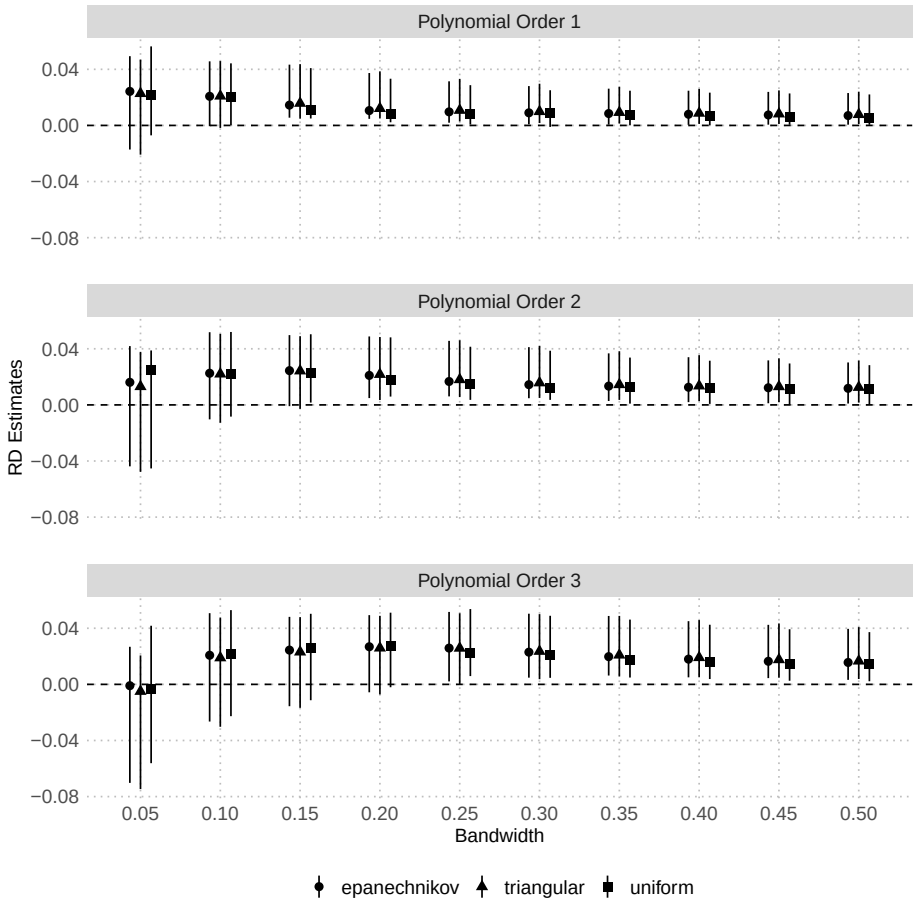


Figure 7: RD Estimates of Winning at t on Manifesto's HHI at $t+1$: Robustness to Bandwidth, Kernel, and Polynomial Order (Panel B)

Notes: The Figure plots RD estimates together with 95% robust-bias corrected confidence intervals following Calonico et al. (2020) for different bandwidths, kernels and polynomial orders.

Table 6: RD Effect of Winning at t on the Manifesto's HHI at t+1 (Party)

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control Mean	0.0898	0.0884	0.085	0.0904	0.089	0.0893	0.089
Estimate	0.0051	0.0178**	0.0186	0.0153**	0.0189**	0.0236**	0.0234**
Std.Err.	(0.0035)	(0.0076)	(0.0145)	(0.0071)	(0.0078)	(0.0096)	(0.0094)
BW	1	0.1485	0.0743	0.2971	0.1381	0.1933	0.197
P-Value	0.1403	0.0192	0.2011	0.0323	0.015	0.0141	0.0126
N.Obs	10733	6676	3930	9307	6379	7842	7933

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates exclude cities with runoff voting. Conditioning on municipalities where the winning party in t also run in t+1. The outcome variable is the manifesto's Herfindahl Hirshman Index (HHI) of topic concentration in t+1 (2012 or 2016). Herfindahl Hirshman Index (HHI) of topic concentration built using outputs of LDA model with 100 topics. The running variable is the vote margin in t (2008 or 2012). Table reports bias-corrected estimates and standard errors following (Calonico et al., 2020).

Alternative Measures of Issue Concentration: I proceed by investigating if the results are robust to the following alternative measures of issue concentration: (i) the maximum topic probability within each manifesto, (ii) the cumulative probability of the top five topics (CK5), and (iii) one minus Shannon's entropy.

The first two measures are more straightforward to interpret, as they are constructed directly from topic probabilities. To compute the third concentration measure, I rely on a transformation of Shannon's entropy index, a measure widely used in physics, economics, and political communication (Shannon, 1948; Jennings et al., 2011; Greene, 2016, 2020). Shannon's entropy is defined as follows:

$$\text{Shannon's } H_i = -\sum_{k=1}^{K} \theta_{i,k} \ln(\theta_{i,k}) \tag{3}$$

Where $\theta_{i,k}$ denotes the share of party i's manifesto devoted to issue k (i.e., the topic probability of issue k), with $k = 1, \ldots, K$. Since the topic model used in the analysis includes 40 topics, $K = 40$.

This measure captures how dispersed a manifesto is relative to a uniform distribution of attention across issues. For example, consider a manifesto that is highly concentrated, such that one topic appears with probability 0.999, while the remaining probability mass (0.001) is evenly distributed across the other topics. In this case, Shannon's entropy is close to zero (e.g., Shannon's $H_i \approx 0.01$). More generally, higher concentration is associated with lower values of Shannon's entropy. Conversely, if a manifesto allocates attention uniformly across all topics, so that each topic appears with probability $1/40$, then Shannon's entropy reaches its maximum value: Shannon's $H_i = -(40 \times 1/40 \times \ln(1/40)) \approx 3.69$.

Because all outcome variables in the analysis are intended to measure concentration, I further transform this variable to facilitate interpretation. First, I normalize Shannon's entropy by dividing it by $\ln(K)$, so that the perfectly uniform distribution case is mapped to one.[19] The normalized entropy measure is thus defined as Shannon's $H_i^{\text{norm}} = \frac{\text{Shannon's } H_i}{\ln(K)}$.

Second, to obtain a concentration index that ranges between zero and one and increases with concentration, I define:

$$C_i = 1 - \text{Shannon's } H_i^{\text{norm}}.$$

[19] For a perfectly uniform distribution of attention across the K issues, $\theta_{i,k} = 1/K$ for all k. Substituting into Shannon's entropy definition yields Shannon's $H_i = -\sum_{k=1}^{K}(1/K)\ln(1/K) = -K(1/K)(-\ln K) = \ln(K)$, which is the theoretical maximum of Shannon's entropy.

Under this normalization, C_i lies in the unit interval: a perfectly uniform distribution of attention across issues implies $C_i = 0$, while a manifesto entirely focused on a single issue implies $C_i = 1$. This transformation preserves the desirable properties of Shannon's entropy while yielding a measure whose interpretation aligns closely with the concept of issue concentration used throughout the analysis.

Table 7 reports estimates using the three alternative measures of manifesto issue concentration discussed above. Across all three outcomes, the estimated effects are positive, indicating that parties that won elections by small vote margins at t tend to produce more concentrated manifestos in $t+1$ than parties that barely lost. For the maximum topic probability and the cumulative probability of the top five topics (CK5), although point estimates are consistently positive, they are generally imprecise and rarely statistically significant. This pattern is expected, as both measures focus on a limited subset of the manifesto topic distribution and are therefore noisier.

In contrast, the results are stronger and more precisely estimated when concentration is measured using one minus normalized Shannon's entropy, which incorporates the full topic distribution. For this measure, the RD estimates are statistically significant and similar across most specifications. Overall, these findings suggest that the results are robust to considering alternative measures of issue concentration.

Table 7: RD Effect of Winning at t on Alternative Measures of Manifesto's Issue Concentration at t+1

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable: Maximum topic probability							
Control Mean	0.2251	0.2243	0.2219	0.2264	0.2237	0.2247	0.2247
Estimate	0.0012	0.0131	0.0165	0.0121	0.015	0.0205*	0.0195*
Std.Err.	(0.0044)	(0.009)	(0.0162)	(0.0085)	(0.0094)	(0.0117)	(0.0114)
BW	1	0.1706	0.0853	0.3412	0.1546	0.2039	0.2121
P-Value	0.7888	0.1471	0.3083	0.1526	0.1077	0.0801	0.0866
N.Obs	10733	7284	4393	9636	6847	8079	8204
Dependent Variable: Sum of top 5 topic probabilities (CK5)							
Control Mean	0.5108	0.5086	0.5059	0.5111	0.5089	0.5097	0.51
Estimate	0.0019	0.0123	0.0173	0.0103	0.0126	0.0134	0.0131
Std.Err.	(0.0036)	(0.0078)	(0.0136)	(0.0072)	(0.0078)	(0.0089)	(0.0087)
BW	1	0.148	0.074	0.2961	0.143	0.2293	0.2363
P-Value	0.5983	0.1122	0.2042	0.1539	0.1059	0.1315	0.133
N.Obs	10733	6658	3914	9305	6506	8500	8594
Dependent Variable: 1 minus normalized Shannon entropy							
Control Mean	0.2015	0.2002	0.1956	0.202	0.2005	0.2007	0.2006
Estimate	0.002	0.0169**	0.0214	0.0146*	0.0176**	0.0222**	0.0218**
Std.Err.	(0.0036)	(0.008)	(0.0147)	(0.0075)	(0.0081)	(0.0099)	(0.0097)
BW	1	0.1438	0.0719	0.2877	0.137	0.1943	0.1995
P-Value	0.5895	0.0351	0.1447	0.0527	0.0309	0.0247	0.0241
N.Obs	10733	6537	3822	9241	6339	7876	8002

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates exclude cities with runoff voting and are based on the sample used in panel B of Table 2. The outcome variables are: (i) Maximum topic probability; (ii) Sum of top 5 topic probabilities (CK5); and (iii) 1 minus normalized Shannon's entropy. All of these variables are measured for each manifesto in t+1 (2012 or 2016). The running variable is the vote margin in t (2008 or 2012). Table report bias-corrected estimates and standard errors following (Calonico et al., 2020).

6 Investigating the Mechanisms Behind Manifesto Concentration

Results presented in Section 5 show that, in close elections, incumbent parties produce more concentrated campaign manifestos than challengers. In this section, I investigate the mechanisms behind this result. In particular, I examine whether incumbent parties' reputations built during the first term in office help explain the observed increase in manifesto concentration, and how these reputational constraints shape which issues incumbent parties decide to focus on.

6.1 First-Term Reputation and Manifesto Concentration

I start by analyzing whether the greater party-level manifesto concentration documented in Section 5 is associated with incumbents' first-term reputation. To do so, following an approach similar to that proposed by Klašnja and Titiunik (2017), I exploit heterogeneity across two distinct party-level subsamples: the Open Seat Sample in $t+1$ and the Incumbent Sample in $t+1$.

Because candidates are subject to a two-term limit, while parties are not, when a party i elects a candidate for a second term in election t, it must run with a different candidate in $t+1$. This case corresponds to what Klašnja and Titiunik (2017) refer to as the Open Seat Sample in $t+1$. In contrast, the Incumbent Sample comprises parties that elected a candidate for a first term in t. In this case, the same candidate runs again in the election $t+1$. If the experience of governing during the first term——and the reputation built during that period——matters for issue selection, one should expect stronger effects precisely for the Incumbent Sample in $t+1$.

Table 8 reports the results of this heterogeneity analysis. It mirrors Table 2, with columns exploring the robustness of results to alternative specification choices. Panels A and B now present estimates for the Incumbent and Open Seat samples, respectively. The estimated coefficients are generally positive and similar to those presented in Table 2, but remain statistically significant only for the Incumbent Sample. This is consistent with the interpretation that first-term reputation plays an important role in driving incumbents' decisions to concentrate their manifestos. Nevertheless, because the Open Seat Sample is substantially smaller and standard errors are correspondingly larger, the absence of statistical significance in Panel B should be interpreted with caution. In particular, the estimates do not allow us to rule out a positive effect for the Open Seat Sample.

Table 8: RD Effect of Winning at t on the Manifesto's HHI at t+1 for Selected Samples

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Incumbent Sample in $t+1$							
Control Mean	0.1119	0.112	0.1123	0.1125	0.1122	0.112	0.112
Estimate	0.0034	0.0144*	0.0244*	0.0143*	0.0175**	0.0259**	0.0257**
Std.Err.	(0.004)	(0.0076)	(0.0143)	(0.0074)	(0.0085)	(0.0111)	(0.0109)
BW	1	0.1805	0.0903	0.3611	0.145	0.1788	0.1825
P-Value	0.3892	0.0587	0.0894	0.0536	0.0391	0.0197	0.0184
N.Obs	8254	6188	3817	7713	5396	6146	6214
Panel B: Open Seat Sample in $t+1$							
Control Mean	0.1184	0.116	0.1118	0.1205	0.116	0.1212	0.1212
Estimate	-0.0038	0.0157	0.0107	0.0157	0.0152	0.0171	0.0165
Std.Err.	(0.0072)	(0.0134)	(0.0247)	(0.0127)	(0.0133)	(0.0162)	(0.016)
BW	1	0.2316	0.1158	0.4633	0.229	0.3354	0.3336
P-Value	0.5999	0.2428	0.6632	0.2176	0.2529	0.2924	0.3007
N.Obs	2479	1625	973	2200	1617	1976	1972

Notes: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. All estimates exclude cities with runoff voting. The outcome variable is the manifesto's hhi in t+1 (2012 or 2016). The running variable is the vote margin in t (2008 or 2012). Table report bias-corrected estimates and standard errors following (Calonico et al., 2020).

To provide more robust evidence on this mechanism, Table 9 focuses on within-party changes in issue concentration. Using ΔHHI as the outcome, as in Table 4, I find that among the Incumbent Sample, parties that barely win in t significantly increase manifesto concentration in $t+1$ relative to their own previous manifesto. In contrast, point estimates for parties in the Open Seat Sample are uniformly negative and statistically indistinguishable from zero. Because this specification differences out time-invariant party characteristics and long-run ideological positioning, it provides a second source of evidence that the reputation acquired during the first term in office drives incumbent parties' strategic choice to concentrate their manifestos. In this sense, the within-party results complement the cross-sectional party-level estimates in Section 5 and in Table 8 by showing further evidence supporting that the incumbency effect on manifesto concentration is strongest when the party runs with the same candidate, and therefore reflects the joint reputation of the party and the governing candidate formed during the first term.

Table 9: RD Effect of Winning at t on the Manifesto's Delta HHI for Selected Samples

Function	Mean Diff.	Linear				Quadratic	
	All	$\hat{h}$	$\hat{h}/2$	$2\hat{h}$	Controls	$\hat{h}$	Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Incumbent Sample in $t+1$							
Control Mean	-0.0226	-0.0214	-0.0295	-0.0246	-0.0191	-0.0219	-0.0219
Estimate	0.0198**	0.0501***	0.0244	0.0457**	0.0549***	0.0538**	0.0557**
Std.Err.	(0.01)	(0.019)	(0.0186)	(0.0185)	(0.0195)	(0.0225)	(0.0219)
BW	1	0.1024	0.0512	0.2049	0.0854	0.1219	0.122
P-Value	0.0475	0.0083	0.1909	0.0135	0.0049	0.0167	0.0108
N.Obs	1282	698	418	1058	628	786	784
Panel B: Open Seat Sample in $t+1$							
Control Mean	-0.0203	-0.0387	-0.0208	-0.0405	-0.0394	-0.0461	-0.0451
Estimate	0.0383	-0.0527	-0.0313	-0.0569	-0.0526	-0.0538	-0.0588
Std.Err.	(0.026)	(0.0406)	(0.0867)	(0.0378)	(0.0401)	(0.0452)	(0.0432)
BW	1	0.1122	0.0562	0.2247	0.111	0.2023	0.2099
P-Value	0.1404	0.194	0.7184	0.1322	0.1893	0.2343	0.1733
N.Obs	316	146	76	232	144	220	222

Notes: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. All estimates exclude cities with runoff voting. The outcome variable is the manifesto's Delta HHI (2016 - 2012). The running variable is the vote margin in 2012. Table reports bias-corrected estimates and standard errors following (Calonico et al., 2020).

6.2 Issue Selection by Incumbents

Having shown that incumbents' increased manifesto concentration is driven by first-term reputation, this subsection examines how these reputational constraints shape issue selection in campaign manifestos. In particular, I analyze whether incumbents systematically emphasize specific policy issues or rhetorical themes, and whether issue selection reflects parties' comparative advantages and reputational strengths.

To this end, I compute the effect of a party winning at t on the likelihood of discussing specific issues at $t+1$ for the entire sample and for the four largest parties. Figure 8 plots separate RD effects for the policy issues presented and described in Figure 1. It reports bias-corrected point estimates and 95% confidence intervals following (Calonico et al., 2020). As before, the estimates are based on local linear regressions with the MSE-optimal bandwidth and triangular kernel weights.

The results show no systematic increase in emphasis on specific policy areas such as agriculture, education, or the economy. Instead, in close elections, incumbents appear to place greater emphasis on a more rhetorical issue centered around continuity. The most frequent terms in this issue (originally topic 2 in the LDA model) are associated with maintaining ongoing actions (e.g., the verbs *continue*, *keep*, and *ensure*).[20]

[20]Figure D.11 in the Appendix shows a word cloud with the top 50 words in topic 2.

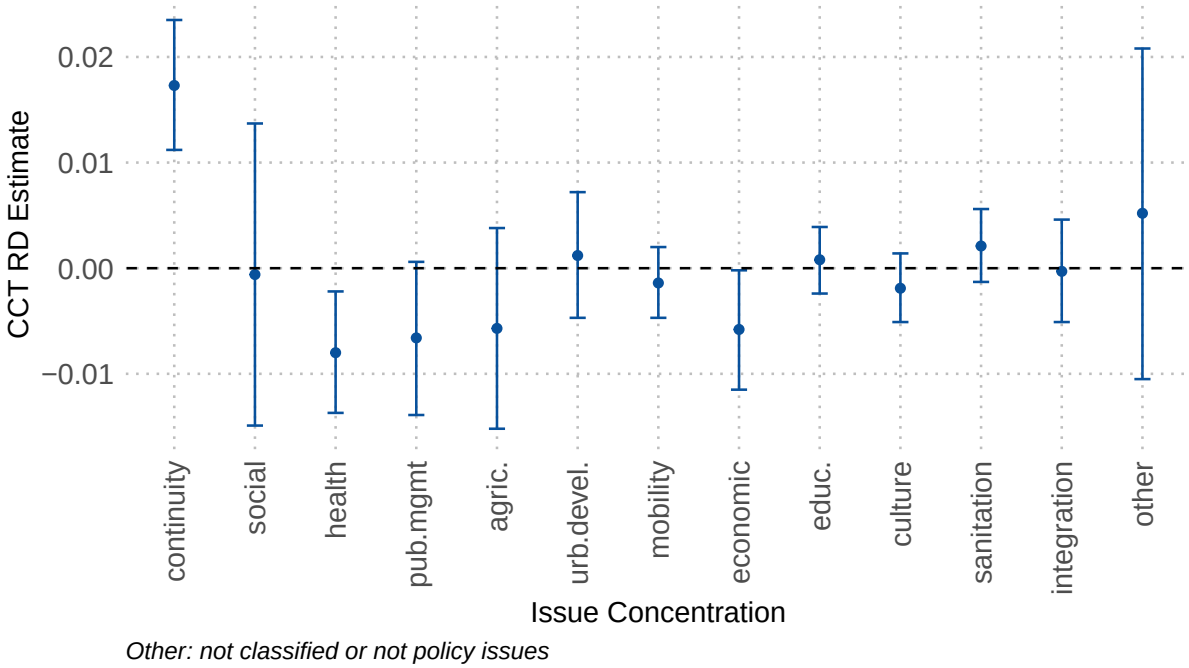


Figure 8: RD Effect of Winning at t on HHI and Issue Emphasis in $t+1$

Notes: The Figure plots separate RD effects for the probability of discussing the different policy issues presented and described in Figure 1. Effects were estimated using local linear regressions and triangular kernel weights. Estimates and 95% confidence intervals are robust-bias corrected following (Calonico et al., 2020).

This pattern does not necessarily follow mechanically from the finding that incumbents produce more concentrated manifestos. A priori, other plausible mechanisms for increased concentration could be at work. For instance, if incumbents responded primarily to voters' issue salience, one might expect greater emphasis on highly salient policy areas such as health.[21] Alternatively, following Vavreck (2009), unfavorable economic conditions could lead incumbents to de-emphasize economic issues relative to challengers. I do not find strong evidence of these patterns in the data. At the same time, while terms related to continuity may be more likely to appear in incumbents' manifestos, this does not follow from the empirical design or from the construction of the topic measure; rather, it emerges from the empirical analysis.

One possible interpretation for the greater use of the continuity topic among incumbents is that they emphasize different substantive policy issues depending on their specific first-term performance and reputational strengths at the municipal level. While the analysis does not allow me to directly verify this mechanism, the pattern is consistent with such an interpretation. In par-

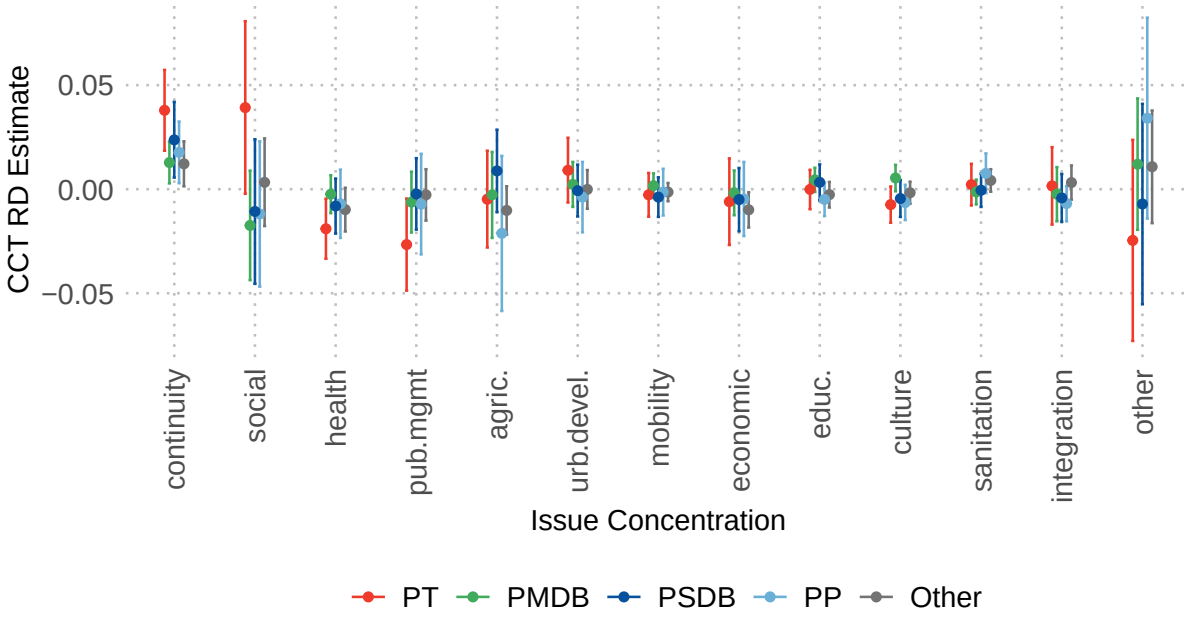
[21] Based on a Datafolha public opinion survey, health is consistently cited as one of the main concerns among Brazilians. More information is available at https://g1.globo.com/politica/noticia/2025/12/13/analise-preocupacoes-datafolha.ghtml.

ticular, heterogeneous issue emphasis may not appear systematically in the aggregate analysis, whereas a common rhetorical strategy centered on continuity may. In this sense, the results are suggestive of reputational constraints stemming from governing experience: rather than shifting attention toward new or universally salient policy issues, incumbents may tend to highlight what they have already done while in office. Consistent with Druckman et al. (2020), incumbents appear more likely to reinforce their experience and past performance than to diversify their campaign discourse.

Next, I estimate the heterogeneous incumbency effect separately for the four largest parties in Brazilian local elections. In this analysis, I run RD estimates using only municipalities that elected a candidate from one of these parties in period t, ensuring that the incumbent in $t+1$ is always from the same party. Consequently, the estimates assess whether incumbents from a given party concentrate their manifestos more than challengers, without imposing restrictions on the opponent's party (i.e., the party that lost in t).

The largest parties were selected based on the number of mayoral candidates elected between 2008 and 2016 and are the Brazilian Democratic Movement (MDB), the Brazilian Social Democracy Party (PSDB), the Progressive Party (PP), and the Workers' Party (PT). Among these parties, PT is a relatively more programmatic, cohesive, and high-discipline party committed to improving social policies (Samuels and Zucco, 2014; Klašnja and Titiunik, 2017). Therefore, compared to the other parties, PT should be more able to discipline its candidates and limit manifesto content to social issues. In contrast, candidates from the other parties are less likely to be subject to similar constraints and should be freer to emphasize issues in which they have a comparative advantage.

The results are displayed in Figure 9. Similar to Figure 8, it plots separate RD effects for the probability of emphasizing different policy issues. However, I now estimate party-specific coefficients. Most parties still do not systematically emphasize any particular policy issue. However, PT incumbents place greater emphasis on social issues, suggesting that party reputation influences the manifesto issue selection.



Other: not classified or not policy issues

Figure 9: RD Effect of Winning at t on HHI and Issue Emphasis in $t+1$ for selected parties

Notes: The Figure plots separate RD effects for selected parties and the probability of discussing the different policy issues presented and described in Figure B.4. Effects were estimated using local linear regressions and triangular kernel weights. Estimates and 95% confidence intervals are robust-bias corrected following (Calonico et al., 2020).

7 Conclusion

In this paper, I empirically investigate the effect of incumbency on manifesto issue concentration. Focusing on parties involved in closely contested elections, I test the hypothesis that incumbents' manifestos are more concentrated than those of challengers. The findings indicate that incumbent parties elected by small vote margins produce campaign manifestos concentrated on fewer issues and tend to play to their strengths by specializing in themes over which they hold a comparative advantage.

Using a regression discontinuity design and data from Brazilian municipal elections, I show that incumbents' manifestos are, on average, approximately 17% more concentrated than those of challengers. The average HHI of an incumbent party's manifesto is 0.019 higher than the control mean of 0.11. This effect is statistically significant and robust to alternative functional forms, bandwidth choices, and sample selections. Additional robustness exercises confirm that the results hold under a candidate-level analysis, an LDA model with 100 topics, and alternative measures of manifesto issue concentration.

The findings also hold when focusing on within-party variation over time, rather than rely-

ing exclusively on cross-sectional comparisons. Specifically, even when comparing the same party across elections, incumbent parties elected by a small vote margin in t concentrate their manifestos more in $t+1$.

Furthermore, the results suggest that parties play to their strengths when selecting which issues to emphasize. First, I compare the estimates for a sample in which both the incumbent party and the candidate elected in t remain the same in $t+1$ (i.e., the Incumbent Sample in $t+1$) with a sample in which, due to two-term limits, the candidate running for a party in $t+1$ changes (i.e., the Open Seat Sample in $t+1$). I show that first-term performance matters for issue selection, as the estimates are stronger for the Incumbent Sample in $t+1$. In other words, when both the candidate's and the party's reputations from the previous mandate are at stake, incumbents produce more focused manifestos.

Second, I compute the effect of a party winning at t on the likelihood of discussing specific issues at $t+1$. In this exercise, I estimate the coefficients for the full sample and for each of the four largest parties in terms of the number of elected candidates in municipal elections between 2008 and 2016. Results indicate that, when analyzed jointly, incumbent parties place relatively greater emphasis on a rhetorical issue associated with "continuity" (topic 2). One possible interpretation, consistent with Druckman et al. (2020), is that incumbents frame their campaign messages around the continuation of actions and policies implemented during their first term. While the analysis does not allow me to directly verify this mechanism, the pattern is suggestive of such a strategy. At the same time, because the performance of each party is likely to differ across policy areas, the coefficients for the probability of discussing specific policy issues are not systematically different from zero.

The party-specific estimates reveal that Workers' Party (PT) incumbents place greater emphasis on social issues than incumbents from other parties. This suggests that, because PT is a relatively more programmatic, cohesive, and disciplined party, it can constrain its candidates to emphasize in their campaign messages the issues in which the party holds a comparative advantage.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used Grammarly and ChatGPT services in order to improve language and readability. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

P. Amorós and M. S. Puy. Issue convergence or issue divergence in a political campaign? *Public Choice*, 155:355–371, 2013. doi: 10.1007/sl1127-011-9865-0. URL https://www.jstor.org/stable/42003105.

E. Aragonès, M. Castanheira, and M. Giani. Electoral competition through issue selection. *American Journal of Political Science*, 59:71–90, 1 2015. ISSN 15405907. doi: 10.1111/ajps.12120.

S. Ashworth, E. B. D. Mesquita, and A. Friedenberg. Incumbency and information. Working Paper, 2019.

G. A. Banck. Personalism in the brazilian body politic: Political rallies and public ceremonies in the era of mass democracy. *European Review of Latin American and Caribbean Studies*, pages 25–43, 1998.

K. Benoit. Text as data: An introduction. In L. Curini and R. Franzese, editors, *The SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE, 2020.

M. D. Bernhardt and D. E. Ingberman. Candidate reputations and the 'incumbency effect'. *Journal of Public Economics*, 27:47–67, 1985.

C. Bevia and H. Llavador. The informational value of incumbency. *Journal of Public Economic Theory*, 11:773–796, 2009. doi: https://doi.org/10.1111/j.1467-9779.2009.01429.x.

D. M. Blei and J. D. Lafferty. Topic models. In A. N. Srivastava and M. Sahami, editors, *Text Mining: Classification, Clustering, and Applications*. Chapman and Hall/CRC, 1st edition edition, 2009.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

I. Budge and D. Farlie. *Explaining and predicting elections: Issue effects and party strategies in twenty-three democracies*. George Allen Unwin, 1983.

S. Calonico, M. D. Cattaneo, and R. Titiunik. Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110:1753–1769, 10 2015. ISSN 1537274X. doi: 10.1080/01621459.2015.1017578.

S. Calonico, M. D. Cattaneo, and M. H. Farrell. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23:192–210, 2020. ISSN 1368423X. doi: https://doi.org/10.1093/ectj/utz022.

F. Caselli, T. Cunningham, M. Morelli, and I. M. de Barreda. The incumbency effects of signalling. *Economica*, 81:397–418, 2014. ISSN 14680335. doi: 10.1111/ecca.12060.

M. D. Cattaneo, N. Idrobo, and R. Titiunik. *A Practical Introduction to Regression Discontinuity Designs*. Cambridge University Press, 11 2019. ISBN 9781108684606. doi: 10.1017/9781108684606.

J. M. Colomer and H. Llavador. An agenda-setting model of electoral competition. *SERIEs*, 3: 73–93, 2012. ISSN 18694195. doi: 10.1007/s13209-011-0056-5.

M. da Silva Conrado, A. D. Felippo, T. A. S. Pardo, and S. O. Rezende. A survey of automatic term extraction for brazilian portuguese. *Journal of the Brazilian Computer Society*, 20, 5 2014. ISSN 16784804. doi: 10.1186/1678-4804-20-12.

D. Demszky, N. Garg, R. Voigt, J. Zou, M. Gentzkow, J. Shapiro, and D. Jurafsky. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 4 2019. doi: doi.org/10.48550/arXiv.1904.01596. URL http://arxiv.org/abs/1904.01596.

P. Denter. Campaign contests. *European Economic Review*, 127, 8 2020. ISSN 00142921. doi: 10.1016/j.euroecorev.2020.103459.

M. Dolezal, L. Ennser-Jedenastik, W. C. Müller, and A. K. Winkler. How parties compete for votes: A test of saliency theory. *European Journal of Political Research*, 53:57–76, 2 2014. ISSN 03044130. doi: 10.1111/1475-6765.12017.

A. Downs. *An Economic Theory of Democracy*. Harper Row, 1957.

T. Dragu and X. Fan. An agenda-setting theory of electoral competition. *Journal of Politics*, 78:1170–1183, 10 2016. ISSN 14682508. doi: 10.1086/686310.

J. N. Druckman, M. J. Kifer, and M. Parkin. Campaign communications in u.s. congressional elections. *American Political Science Review*, 103:343–366, 8 2009. ISSN 00030554. doi: 10.1017/S0003055409990037.

J. N. Druckman, M. J. Kifer, and M. Parkin. Campaign rhetoric and the incumbency advantage. *American Politics Research*, 48:22–43, 1 2020. ISSN 15523373. doi: 10.1177/1532673X18822314.

N. Egami, C. J. Fong, J. Grimmer, M. E. Roberts, and B. M. Stewart. How to make causal inferences using texts. *Sci. Adv*, 8:2652, 2022. doi: 10.1126/sciadv.abg2652. URL https://www.science.org/doi/10.1126/sciadv.abg2652.

F. N. Flores and V. P. Moreira. Assessing the impact of stemming accuracy on information retrieval – a multilingual perspective. *Information Processing and Management*, 52:840–854, 9 2016. ISSN 03064573. doi: 10.1016/j.ipm.2016.03.004.

M. Gentzkow, B. Kelly, and M. Taddy. Text as data. *Journal of Economic Literature*, 57:535–574, 2019a. ISSN 00220515. doi: 10.1257/jel.20181020.

M. Gentzkow, J. M. Shapiro, and M. Taddy. Measuring group differences in high□dimensional choices: Method and application to congressional speech. *Econometrica*, 87:1307–1340, 2019b. ISSN 0012-9682. doi: 10.3982/ecta16566.

Globo. Eleições 2020: O que dizem os programas de governo de capitão nelson e dimas gadelha, candidatos à prefeitura de são gonçalo, no rj. *G1 RJ*, 11 2020. URL http://glo.bo/3VFJrMY.

C. Green-Pedersen and P. B. Mortensen. Who sets the agenda and who responds to it in the danish parliament? a new model of issue competition and agenda-setting. *European Journal of Political Research*, 49:257–281, 3 2010. ISSN 03044130. doi: 10.1111/j.1475-6765.2009.01897.x.

Z. Greene. Competing on the issues: How experience in government and economic conditions influence the scope of parties' policy messages. *Party Politics*, 22:809–822, 11 2016. ISSN 14603683. doi: 10.1177/1354068814567026.

Z. Greene. Being heard above the noise: The role of incumbent issue diversity in election campaigns. *Political Behavior*, 42:487–507, 6 2020. ISSN 15736687. doi: 10.1007/s11109-018-9504-2.

T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl_1):5228–5235, 2004. doi: 10.1073/pnas.0307752101. URL https://www.pnas.org/doi/abs/10.1073/pnas.0307752101.

J. Grimmer. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18:1–35, 2010. ISSN 14764989. doi: 10.1093/pan/mpp034.

J. Grimmer and B. M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21:267–297, 2013. ISSN 14764989. doi: 10.1093/pan/mps028.

B. Grün and K. Hornik. topicmodels: An r package for fitting topic models. *JSS Journal of Statistical Software*, 40:1–30, 2011. doi: https://doi.org/10.18637/jss.v040.i13. URL http://www.jstatsoft.org/.

S. Hansen, M. McMahon, and A. Prat. Transparency and deliberation within the fomc: A computational linguistics approach. *Quarterly Journal of Economics*, 133:801–870, 5 2018. ISSN 15314650. doi: 10.1093/qje/qjx045.

J. A. Henderson. Downs' revenge: Elections, responsibility and the rise of congressional polarization. 2013. URL https://escholarship.org/uc/item/2457g1p0.

W. Jennings, S. Bevan, A. Timmermans, G. Breeman, S. Brouard, L. Chaqués-Bonafont, C. Green-Pedersen, P. John, P. B. Mortensen, and A. M. Palau. Effects of the core functions of government on the diversity of executive agendas. *Comparative Political Studies*, 44 (8):1001–1030, 2011.

J. S. Justeson and S. M. Katz. Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering*, 1:9–27, 1995. ISSN 14698110. doi: 10.1017/S1351324900000048.

N. Kaplan, D. K. Park, and T. N. Ridout. Dialogue in american political campaigns? an examination of issue convergence in candidate television advertising. *American Journal of Political Science*, 50:724–736, 2006. URL http://www.jstor.org/stable/3694245.

H. Kitschelt, K. A. Hawkins, J. P. Luna, G. Rosas, and E. J. Zechmeister. *Latin American Party Systems*. Cambridge University Press, 2010.

M. Klašnja and R. Titiunik. The incumbency curse: Weak parties, term limits, and unfulfilled accountability. *American Political Science Review*, 111:129–148, 2 2017. ISSN 15375943. doi: 10.1017/S0003055416000575.

C. Le Pennec. Strategic campaign communication: Evidence from 30,000 candidate manifestos. *The Economic Journal*, 134:785–810, 2023. doi: https://doi.org/10.1093/ej/uead082.

D. S. Lee. Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142:675–697, 2 2008. ISSN 03044076. doi: 10.1016/j.jeconom.2007.05.004.

D. S. Lee and T. Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48:281–355, 6 2010. ISSN 00220515. doi: 10.1257/jel.48.2.281.

L. D. Magalhaes. Incumbency effects in a comparative perspective: Evidence from brazilian mayoral elections. *Political Analysis*, 23:113–126, 2015. ISSN 14764989. doi: 10.1093/pan/mpu012.

S. Mainwaring. Presidentialism, multipartism, and democracy: the difficult combination. *Comparative Political Studies*, 26:198–228, 1993. ISSN 15523829. doi: 10.1177/0010414093026002003.

A. Mendes and S. Antunes. Collocations in portuguese: A corpus-based approach to lexical patterns. *Collocations Cross-Linguistically. Corpora, Dictionaries and Language Teaching*, pages 141–166, 2016.

H. Mueller and C. Rauh. Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, 112:358–375, 5 2018. ISSN 15375943. doi: 10.1017/S0003055417000570.

L. M. Novaes. The violence of law-and-order politics: The case of law enforcement candidates in brazil. *American Political Science Review*, 118:1–20, 2 2023. ISSN 15375943. doi: 10.1017/S0003055423000540.

L. M. Novaes and L. Schiumerini. Commodity shocks and incumbency effects. *British Journal of Political Science*, 52:1–20, 10 2021. ISSN 14692112. doi: 10.1017/S0007123421000478.

M. J. Osborne. Spatial models of political competition under plurality rule: A survey of some explanations of the number of candidates and the positions they take. *Canadian Journal of Economics*, 28:261–301, 1995. doi: https://doi.org/10.2307/136033.

T. Persson and G. Tabellini. *Political economics: explaining economic policy*. MIT press, 2002.

J. R. Petrocik. Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science*, 40:825–850, 1996. URL http://www.jstor.org/stable/2111797.

S. A. Rhoades. The herfindahl-hirschman index. *Fed. Res. Bull.*, 79, 1993.

D. Robertson. *A theory of party competition*. Wiley, 1976.

R. G. Rossi. Classificação automática de textos por meio de aprendizado de máquina baseado em redes. Doctoral Dissertation. Universidade de São Paulo, 2016.

N. Salles. Programs and parties: Rethinking electoral competition through analysis of brazilian 'grotões'. *Brazilian Political Science Review*, page 13, 2019. doi: 10.1590/1981-3821201900020003/For. URL http://dx.doi.org/10.1590/1981-3821201900020003/.

N. Salles and F. Guarnieri. Electoral strategy in brazilian municipalities: Programmatic component and partisan alignment. *Revista de Sociologia e Politica*, 27, 2019. ISSN 16789873. doi: 10.1590/1678-987319277201.

D. Samuels and C. Zucco. The power of partisanship in brazil: Evidence from survey experiments. *American Journal of Political Science*, 58:212–225, 2014. doi: https://doi.org/10.1111/ajps.12050.

L. Samuelson. Electoral equilibria with restricted strategies. *Source: Public Choice*, 43:307–327, 1984.

G. Sarmento and L. Vidigal. Veja as palavras mais citadas nos programas de governo dos 13 candidatos à presidência. *G1*, 9 2018. URL http://glo.bo/3Rr5Lr7.

E. E. Schattschneider. *The semisovereign people: A realist's view of democracy in America*. Wadsworth Publishing Company, 1975.

C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

L. Sigelman and E. H. Buell. Avoidance or engagement? issue convergence in u.s. presidential campaigns, 1960-2000. *American Journal of Political Science*, 48:650–661, 2004. URL http://www.jstor.org/stable/1519925.

J. Souza. Eleições 2020: plano de governo dos candidatos à prefeitura de feira de santana. *G1 BA*, 11 2020. URL http://glo.bo/3XjINpC.

C. J. Spiliotes and L. Vavreck. Campaign advertising: Partisan convergence or divergence? *The Journal of Politics*, 64:249–261, 2002.

M. Taddy. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108:755–770, 2013. ISSN 01621459. doi: 10.1080/01621459.2012.734168.

K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 252–259, 2003.

L. Vavreck. *The Message Matters: The Economy and Presidential Campaigns*. Princeton University Press, 2009.

J. Wilkerson and A. Casas. Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, pages 529–544, 2017. doi: 10.1146/annurev-polisci-052615. URL https://doi.org/10.1146/annurev-polisci-052615-.

A Appendix - Dataset

A.1 Example of Manifestos

This appendix presents examples of mayoral candidate manifestos from states across Brazil's five regions. Due to space constraints, only the first four pages of each manifesto are displayed.

Overall, the manifestos are relatively simple documents. They typically include a brief overview of the candidate's priorities and are organized into sections addressing various policy areas (e.g., health, education, culture, etc.). These sections often consist of bullet-point lists outlining specific proposals or goals.

The length of the manifestos varies considerably. For example, the documents shown in Figures A.1 and A.5 are only one page long, while those in Figures A.2, A.3, and A.4 are significantly longer, with 14, 19, and 13 pages, respectively.

PMN – Município de Tabaporã-MT

Candidato a Prefeito: Tonhão – 33

Candidato a Vice Prefeito: João do Caminhão – 33

Plano de governo

Propostas:

Saúde

Providenciar e aparelhar o hospital, diminuindo assim a necessidade de ter que transferir doentes para hospitais de outros municípios. Providenciar também, centro de UTI e incubadoras para recém-natos.

Educação

Investir nos profissionais de educação, providenciar ações educativas que possam vir a diminuir a incidência de marginalidade. Investir também nas demais áreas da cultura do município.

Agricultura

Trabalhar e beneficiar os pequenos e médios agricultores, investindo em cursos e palestras, em tecnologias e meios de produzir e comercializar seus produtos. Criar e implantar granjas de frango e porcos, investir nas áreas de laticínio e demais produtos agrícolas.

Criar pequenas e médias empresas, colaborando assim para a diminuição do desemprego, tornando menos provável a migração para outras cidades a procura de emprego, permitindo que a cidade evolua.

Meio ambiente

Trabalhar na conservação de matas e nascentes conforme as regras e normas da Sema.

Tabaporã-MT; 11 de agosto de 2016.

Figure A.1: Manifesto of Mayoral Candidate in Municipality of Tabaporã - MT (2016) - Mid-West Region

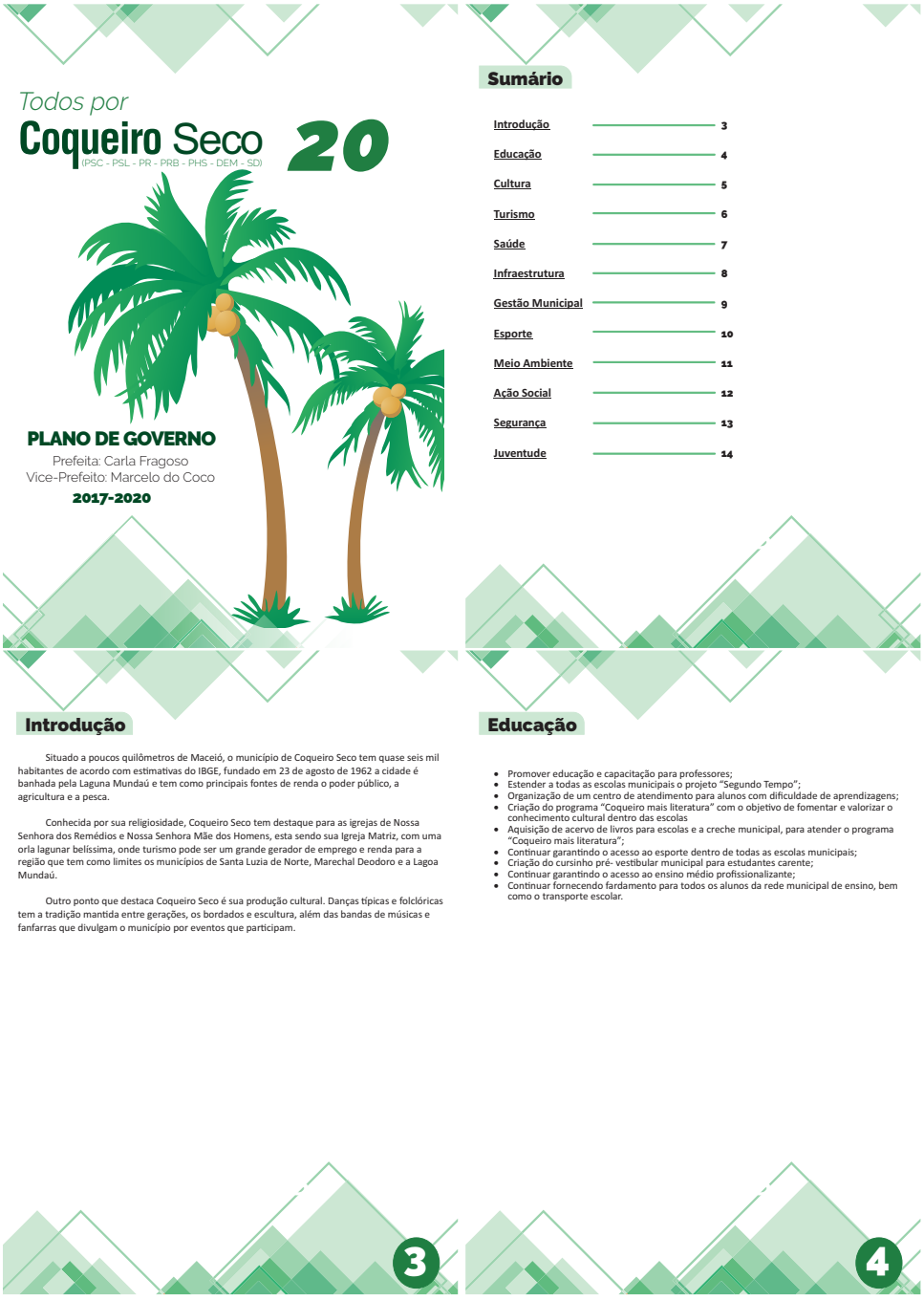


Figure A.2: Manifesto of Mayoral Candidate in Municipality of Coqueiro Seco - AL (2016) - Northeast Region

INDICES

VILSON 15

1

AVEIRO

Histórico: Tem a sua origem numa aldeia de índios Mundurucus (tapajós-tapera) que foi elevada a freguesia de Nossa Senhora da Conceição de Aveiro em 1781. Essa aldeia obteve a denominação portuguesa de lugar de Aveiro, por ato do governador e capitão-general, José de Nápoles Tello de Menezes, em 23 de agosto de 1781, que nomeou, na mesma ocasião o morador Francisco Alves Nobre para administrá-la. Constatou-se, no registro oficial a existência antes de 1781, da freguesia de Nossa Senhora da Conceição do Aveiro, do que se concluiu, portanto, que o ato de criação desse lugar foi somente uma confirmação, pois o local já era conhecido como Aveiro. (Fonte: IBGE) Aveiro foi elevado a município em 1883 e compreende atualmente os distritos de Aveiro, Brasília Legal, Pinhel e Fordlandia.

POPULAÇÃO:15 767 hab. *IBGE/2010*

IDH – M: 0,541 BAIXO PNUD/2010

PIB: R$ 59.110,804 MIL*IBGE/2010*

PIB per CAPITA: R$ 3.717,80*IBGE/2010*

AREA: 17.074.290 KM²

DENSIDADE: 0,92 HAB/KM²

CLIMA: EQUATORIAL

2

PRINCÍPIOS DO GOVERNO

VILSON 15

Governo da Verdade

Só a transparência vai garantir uma atuação eficiente.

Governo do Acolhimento

Todos tem direito a cidadania e a felicidade.

Governo da Coragem

É preciso determinação e firmeza para mudar hábitos e processos.

Governo da Diversidade

Assim é a alma de Aveiro.

Governo da Competência

Um futuro melhor se faz com o conhecimento, com o aproveitamento do potencial humano.

3

OBJETIVOS DO GOVERNO

VILSON 15

- ✓ CONSTRUIR UMA AVEIRO MAIS DIGNA PARA NOSSOS MORADORES;
- ✓ ENFRENTAR AS DESIGUALDADES SOCIOESPECIALS;
- ✓ REALIZAR UMA AÇÃO PARTICIPATIVA, COM AÇÕES SETORIAIS.
- ✓ DAR MAIOR EFICIÊNCIA À ADMINISTRAÇÃO PUBLICA MUNICIPAL.

4

Figure A.3: Manifesto of Mayoral Candidate in Municipality of Aveiro - PA (2016) - North Region

PLANO DE GOVERNO

TRANSPARÊNCIA E GESTÃO EFICIENTE

PSDB SOCIAL DEMOCRACIA

COLIGAÇÃO AVANÇA PARNAÍBA

GESTÃO 2017-2020

Apresentação:

O Plano de Governo da Coligação Avança Parnaíba foi elaborado a partir das principais demandas da cidade de Santana de Parnaíba, considerando sua capacidade de realização e a continuidade no desenvolvimento econômico e social do município frente aos anseios de sua população.

As propostas a seguir tem a função de nortear a gestão de governo no período de janeiro de 2017 a dezembro de 2020, de modo a garantir a execução e entrega dos projetos em conformidade com os recursos financeiros, materiais e humanos que cada um envolve.

Nosso Plano de Governo tem como premissa a gestão transparente e eficiente dos recursos públicos, elementos fundamentais para governar em favor da população, principalmente porque a cidade vive seu melhor momento e tem sido avaliada por especialistas como uma das melhores cidades do País para investimentos e para se viver, além de ter alcançado hoje a condição de cidade mais segura da região metropolitana do estado de São Paulo e uma das que melhor administra o dinheiro público no Brasil.

Com os expressivos avanços conquistados nos últimos anos, vimos que o planejamento vai nos permitir continuar avançando no desenvolvimento e preparação para o crescimento sustentável, justo e viável do município nos próximos quatro anos, construindo uma das melhores cidades do Brasil em qualidade de vida e crescimento planejado.

Propostas:

NA EDUCAÇÃO

Educação de qualidade e um modelo para o País é a nossa meta. Continuaremos dedicando todos os esforços para garantir o acesso à educação para todas as crianças de Santana de Parnaíba, com o compromisso de zerar a fila de "creche" na cidade, aperfeiçoando o atendimento no ensino fundamental e demais modalidades, como a educação especial e a educação de jovens e adultos. Vamos elevar ainda mais o padrão de qualidade educacional no município por meio de investimentos em manutenção e ampliação de unidades escolares existentes e a construção de novas unidades; aperfeiçoamento e qualificação do corpo docente; atualização permanente de material didático; uniformização dos estudantes; alimentação escolar de qualidade; transporte escolar; valorização dos profissionais da educação e implementação de novas ferramentas e metodologias de estudo, incluindo recursos digitais e idiomas.

A seguir, as principais propostas de atuação na área da educação:

- *Implantação do Programa Jovem Médico, que consiste em oferecer cursos preparatórios aos estudantes do ensino médio municipal para ingressarem nos cursos de medicina de faculdades públicas.*
- *Construção de nove novos colégios de educação infantil na cidade, instalados nos bairros do Refúgio dos Bandeirantes, São Pedro, Chácara Solar, Ingaí, Jaguari, Chácara das Garças, Crystal e Jardim Itapuã.*
- *Implantação da Escola Pública de Idiomas, que vai oferecer cursos de idiomas (inglês, espanhol e mandarim) para os estudantes com melhores desempenhos na rede pública de ensino.*
- *Implantação do Projeto Escola.com, que vai disponibilizar uso de "tablets" na grade de ensino público para desenvolvimento do conhecimento e uso de recursos tecnológicos no aprendizado dos estudantes e suporte dos professores.*
- *Revisão do Plano de Carreira e Estatuto do Magistério.*
- *Inclusão de professores para atendimento de crianças com deficiência na educação infantil, garantindo o direito e acessibilidade ao estudo para todos.*
- *Implantação do Centro de Atenção à Pessoa Humana, local dedicado e adaptado para desenvolvimento educacional de crianças e adolescentes com deficiência.*
- *Ampliação do programa de formação continuada para todos os profissionais da educação, em parceria com instituições de ensino superior da região.*

- *Criação da política de incentivo ao estudo, com a realização de olimpíada do conhecimento, bolsas de estudos no exterior e bolsas para o ensino superior.*
- *Construção de 02 Centros Educacionais Integrados - CEIs (Creche, Colégio Infantil até o ensino Médio, com piscina semiolímpica, ginásio de esportes e centro cultural) na região Central e na fazendinha, com atendimento para todo o município.*
- *Implantação da plataforma Google de aprendizado à distância para todos os alunos da rede pública.*
- *Inclusão de aulas de criação de games na grade escolar*

NA SAÚDE

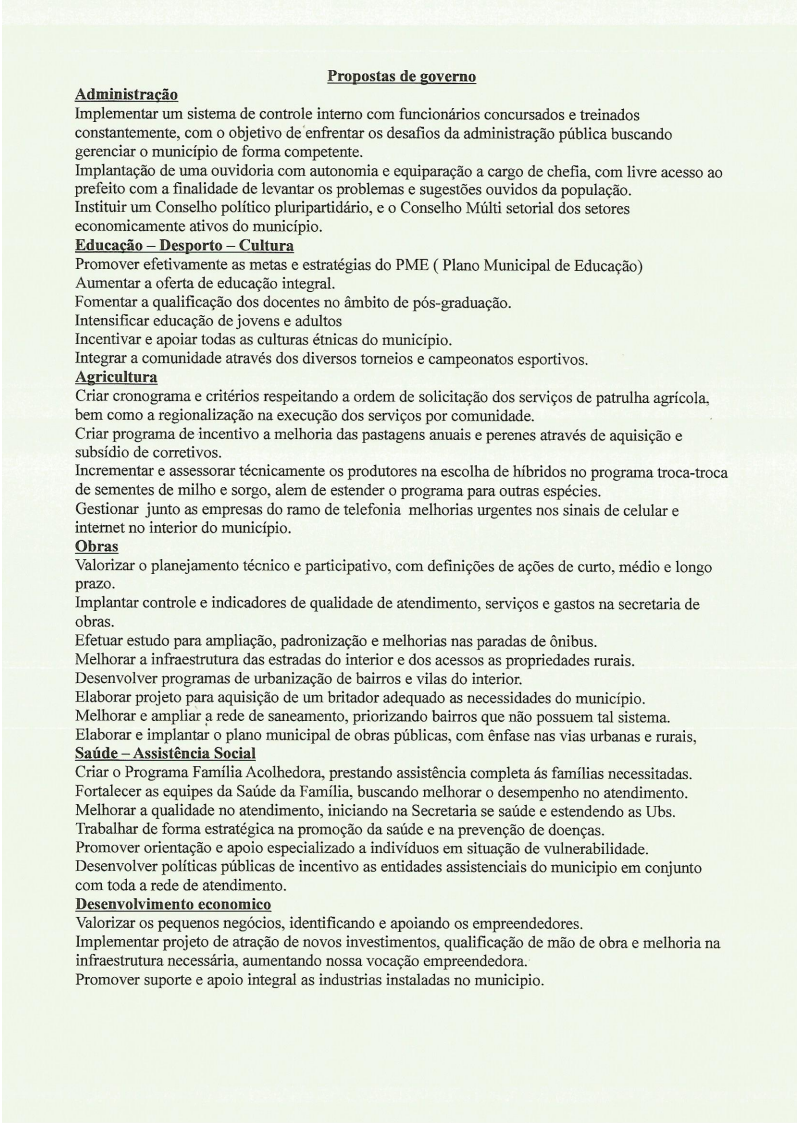
A melhoria contínua dos serviços de saúde é prioridade em nosso plano de governo. Entendemos que os investimentos nesta área passam obrigatoriamente pela humanização do atendimento, em todas as áreas. Além disso, é necessário promover a prevenção e recuperação da saúde por meio de programas dirigidos às crianças, mulheres, homens e pessoas na melhor idade, de forma acolhedora e integral do cidadão, observando os objetivos de Desenvolvimento do Milênio da ONU.

Vamos promover ainda mais avanços na área da saúde, com novas unidades de saúde, melhoria no sistema de atendimento e serviços e integração da comunidade com as iniciativas voltadas a qualidade de vida e saúde.

Nossas principais propostas nesta área são:

- *Entrega da maternidade parnaibana.*
- *Construção do Pronto Socorro Infantil.*
- *Implantação de seis novas Unidades Básicas de Saúde – UBS, instaladas nos bairros Refúgio dos Bandeirantes, Ingaí, Chácara Solar, Suru, Chácara das Garças e Sítio Morro.*
- *Implantação da Clínica Municipal de Exames e Diagnósticos - CMED, voltada para atender de forma rápida as demandas de exames de imagens e laboratoriais dos munícipes.*
- *Ampliação do Programa Saúde da Família, para toda a cidade.*
- *Implantação do sistema de agendamento de consultas e exames Online.*
- *Implantação da Farmácia de referência 24h.*

Figure A.4: Manifesto of Mayoral Candidate in Municipality of Santana do Parnaíba - SP (2016) - Southeast Region



Propostas de governo

Administração

Implementar um sistema de controle interno com funcionários concursados e treinados constantemente, com o objetivo de enfrentar os desafios da administração pública buscando gerenciar o município de forma competente.
Implantação de uma ouvidoria com autonomia e equiparação a cargo de chefia, com livre acesso ao prefeito com a finalidade de levantar os problemas e sugestões ouvidos da população.
Instituir um Conselho político pluripartidário, e o Conselho Múlti setorial dos setores economicamente ativos do município.

Educação – Desporto – Cultura

Promover efetivamente as metas e estratégias do PME (Plano Municipal de Educação)
Aumentar a oferta de educação integral.
Fomentar a qualificação dos docentes no âmbito de pós-graduação.
Intensificar educação de jovens e adultos
Incentivar e apoiar todas as culturas étnicas do município.
Integrar a comunidade através dos diversos torneios e campeonatos esportivos.

Agricultura

Criar cronograma e critérios respeitando a ordem de solicitação dos serviços de patrulha agrícola, bem como a regionalização na execução dos serviços por comunidade.
Criar programa de incentivo a melhoria das pastagens anuais e perenes através de aquisição e subsídio de corretivos.
Incrementar e assessorar técnicamente os produtores na escolha de híbridos no programa troca-troca de sementes de milho e sorgo, alem de estender o programa para outras espécies.
Gestionar junto as empresas do ramo de telefonia melhorias urgentes nos sinais de celular e internet no interior do município.

Obras

Valorizar o planejamento técnico e participativo, com definições de ações de curto, médio e longo prazo.
Implantar controle e indicadores de qualidade de atendimento, serviços e gastos na secretaria de obras.
Efetuar estudo para ampliação, padronização e melhorias nas paradas de ônibus.
Melhorar a infraestrutura das estradas do interior e dos acessos as propriedades rurais.
Desenvolver programas de urbanização de bairros e vilas do interior.
Elaborar projeto para aquisição de um britador adequado as necessidades do município.
Melhorar e ampliar a rede de saneamento, priorizando bairros que não possuem tal sistema.
Elaborar e implantar o plano municipal de obras públicas, com ênfase nas vias urbanas e rurais,

Saúde – Assistência Social

Criar o Programa Família Acolhedora, prestando assistência completa ás famílias necessitadas.
Fortalecer as equipes da Saúde da Família, buscando melhorar o desempenho no atendimento.
Melhorar a qualidade no atendimento, iniciando na Secretaria se saúde e estendendo as Ubs.
Trabalhar de forma estratégica na promoção da saúde e na prevenção de doenças.
Promover orientação e apoio especializado a indivíduos em situação de vulnerabilidade.
Desenvolver políticas públicas de incentivo as entidades assistenciais do municipio em conjunto com toda a rede de atendimento.

Desenvolvimento economico

Valorizar os pequenos negócios, identificando e apoiando os empreendedores.
Implementar projeto de atração de novos investimentos, qualificação de mão de obra e melhoria na infraestrutura necessária, aumentando nossa vocação empreendedora.
Promover suporte e apoio integral as industrias instaladas no municipio.

Figure A.5: Manifesto of Mayoral Candidate in Municipality of Getúlio Vargas - RS (2016) - South Region

A.2 Text Preprocessing Procedures

The literature on text as data emphasizes that an important step before applying statistical text analysis methods, such as Latent Dirichlet Allocation, is the definition of the vocabulary, or text cleaning (Grimmer and Stewart, 2013; Gentzkow et al., 2019a; Egami et al., 2022). The key idea of this procedure is to reduce text dimensionality by performing a sequence of steps that define which elements are essential to the analysis and remove those that are not. The steps most commonly used in the literature are: (i) eliminating raw text elements other than words, such as punctuation, numbers, and web page addresses; (ii) removing stop words; (iii) stemming words; (iv) identifying collocations; and (v) removing words that are either too frequent or too rare using tf-idf (*term frequency–inverse document frequency*). One important caveat is that, although these steps help reduce data dimensionality and yield more interpretable results, they require careful decisions about which vocabulary to retain in each application (Gentzkow et al., 2019a). I present the decisions made for each cleaning step in more detail below.

Following the literature, I start with step (i) and remove punctuation, numbers, and web page addresses. I also convert uppercase letters to lowercase and remove stop words, as recommended in step (ii). Stop words are commonly used words such as articles, prepositions, and some very common verbs. In English, *the* and *a* are examples of stop words, while in Portuguese, *a*, *é*, and *uma* are examples of stop words.[22] Since these elements do not intrinsically convey meaning, they are not essential for uncovering the main subjects of political manifestos. For instance, the verb *é* (equivalent to the verb *to be* in English) does not carry meaning by itself. Nevertheless, its use is widespread.

Step (iii) consists of stemming words, replacing terms in the text with their root (stem). For example, the words *fisher* and *fishing* are reduced to *fish* after the stemming procedure. There are many available algorithms for stemming text. I use the RSLP (an acronym for Portuguese Language Suffix Remover) algorithm.[23] Flores and Moreira (2016) evaluates different algorithms for stemming words in Portuguese and identifies RSLP as the best stemmer in terms of the Error Rate Relative to Truncation. It also produces the lowest Underestimating Index (a measure that captures the number of times a stemmer fails to remove a suffix).

One potential issue with stemming the text corpus is that the algorithm sometimes mistakenly removes more stem parts than it should. Hence, words with different meanings may produce the

[22] To implement this step, I use the list of Portuguese stop words available in the R package `tm` and add some additional words that are likely to be present in all documents (e.g., *município* (municipality), *municipais* (municipalities), *plano governo* (government plan), *proposta governo* (government proposal), and *programa governo* (government program)).

[23] This algorithm is used to stem words in Portuguese texts and is available in the R package `ptstem`.

same stem (e.g., *adoçante* (sweetener) and *adoção* (to adopt) are both reduced to the same stem *adoç*). Because of this issue, I decided not to perform the stemming procedure in the baseline dataset used in this paper.

Step (iv) identifies collocations. To define a collocation according to the Natural Language Processing (NLP) literature, one starts by defining an n-gram as *n* words that appear side by side: a uni-gram consists of a single term, a bi-gram is a combination of two terms, and so on. Hence, an n-gram is based on word proximity rather than meaning. A collocation is an expression that conveys a specific meaning and is produced by multiple words that co-occur in a sentence more frequently than would be expected by chance. Therefore, a collocation can be understood as an n-gram that carries meaning. For instance, the bi-gram *urban infrastructure* is a collocation formed by two tokens, *urban* and *infrastructure*. Each of these tokens has meaning on its own, but when combined, they produce a distinct expression of interest. Hence, the vocabulary I consider contains both uni-grams and bi-grams.[24]

The algorithm I use to compute these collocations consists of first forming all possible uni-grams and bi-grams from the text and then removing all expressions that are too infrequent. To illustrate this step, consider the following phrase: *Increase life expectancy by providing better primary health care services*. The procedure for computing bi-grams combines every pair of tokens that appear side by side. Hence, the phrase above contains ten uni-grams (*increase*, *life*, etc.) and nine bi-grams (*increase life*, *life expectancy*, *expectancy by*, etc.). This example makes clear that, although some bi-grams constitute actual collocations (e.g., *life expectancy*), others do not (e.g., *expectancy by*). Moreover, the number of meaningless bi-grams (i.e., not actual collocations) increases significantly with the size and number of documents analyzed.

After performing this step, the Document Term Matrix (DTM)[25] has 7.1M unique uni-grams and bi-grams, which is far more than the estimated number of words in the Portuguese language (381,000 words according to the *Portuguese Orthographic Vocabulary of the Brazilian Academy of Letters*).[26] However, after removing very infrequent terms (i.e., those that appear in 5% or fewer documents), the resulting DTM has 4,086 unique terms.

There are two caveats to this step. First, in addition to removing meaningless bi-grams (and retaining actual collocations), this step also helps identify and remove misspelled uni-grams and

[24] I follow Demszky et al. (2019) and consider only uni-grams and bi-grams. Other authors also include tri-grams, but for simplicity, I do not use them. Moreover, I use the R package `tm` to compute these collocations.

[25] A Document Term Matrix contains the number of times that each word (whether a uni-gram or bi-gram) appears in each document. Hence, rows indicate documents, and columns indicate terms. The DTM is usually a sparse matrix because each document contains only a small fraction of the set of words in the entire text collection (corpus).

[26] Information available at: http://www.academia.org.br/nossa-lingua/busca-no-vocabulario?sid=19.

other very infrequent terms. Second, although some authors, such as Hansen et al. (2018), adopt a methodology based on English language structure (morphosyntactic information) to identify collocations,[27] I follow a simpler approach and identify collocations based on the frequency of pairs of words that appear side by side. I make this decision because the available tools for part-of-speech tagging and collocation identification are not yet as well developed for Portuguese as they are for English. Moreover, according to Rossi (2016), part-of-speech tagging techniques, besides being computationally intensive, make the performance of text classification dependent on the quality of these techniques.[28]

The last procedure (step v) is used by most authors applying statistical methods for text analysis (Blei and Lafferty, 2009). It consists of pruning vocabulary by ranking words by tf-idf (*term frequency-inverse document frequency*) and removing the words at the bottom of this rank. The tf-idf weight is defined as:

$$\textit{tf-idf} = (1 + \ln(n_w)) \times \ln\left(\frac{D}{D_w}\right) \tag{A.1}$$

Where n_w represents how often the term w appears in the whole corpus; D is the number of documents, and D_w is the number of documents containing the term w. From A.1, one notes that, on the one hand, if all documents contain a specific term (i.e., $D = D_w$), the tf-idf weight goes to zero. On the other hand, if a term is infrequent (i.e., n_w is low), its tf-idf will also be low. Hence, the tf-idf weight punishes both very common and infrequent terms (Gentzkow et al., 2019a). Figure A.1 below plots the words ranked by tf-idf. The tf-idf weights, depicted in Figure A.1, increase sharply for words ranked above the 95% quantile. At the same time, these weights decrease strongly for words ranked below the 5% quantile.

To better understand which words fall into these groups, Figure A.2 displays the 50 most frequent words ranked below the 5% quantile or above the 95% quantile. On the left-hand side, the word cloud shows terms that appear in most documents and consequently have low tf-idf values. For instance, the words *saúde* (health), *educação* (education), *programa* (program), and *cidade* (city) are expected to appear in most political manifestos because they either reflect major issues discussed by mayoral candidates (such as *health* and *education*) or more generally indicate that the document constitutes a campaign manifesto within a particular city. Hence,

[27] Hansen et al. (2018)'s approach to identifying collocations consists first of assigning each term a grammatical category using a part-of-speech tagger described by Toutanova et al. (2003). Then, it computes the frequencies of the identified part-of-speech patterns to infer which are likely to correspond to collocations, following Justeson and Katz (1995).

[28] Some references for working with texts in Portuguese include da Silva Conrado et al. (2014), Mendes and Antunes (2016), and `https://www.sketchengine.eu/corpus-brasileiro/`.

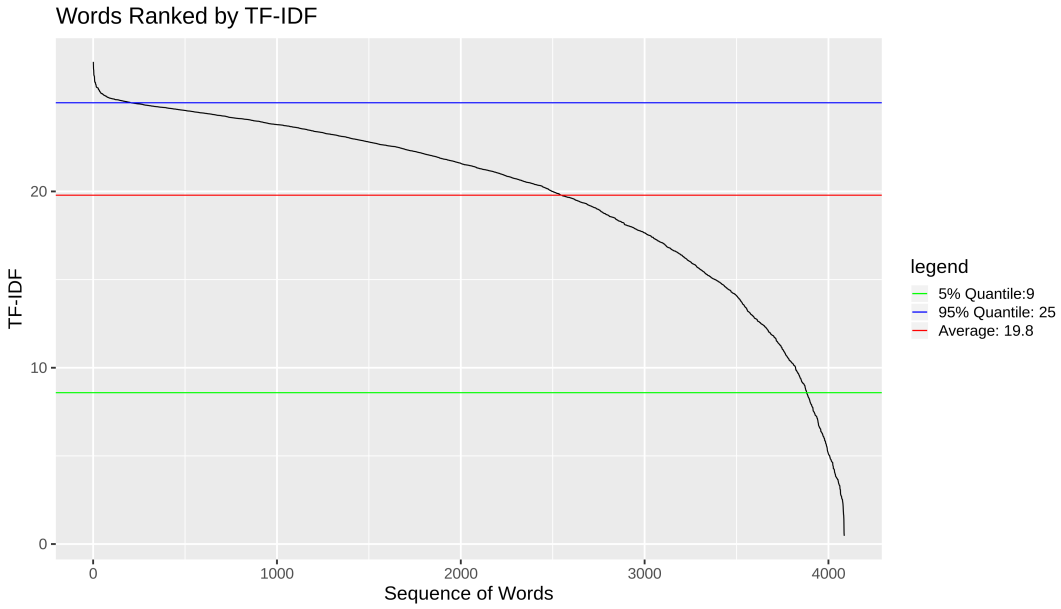


Figure A.1: Ranking words by term frequency-inverse document frequency

these words do not contribute significantly to identifying different topics or comparing manifesto issue emphasis and concentration. For this reason, following Blei and Lafferty (2009) and Hansen et al. (2018), I remove all words ranked below the 5% quantile of tf-idf (left-hand plot of Figure A.2).

The right-hand side plot of Figure A.2 shows words ranked above the 95% quantile by tf-idf weights. One characteristic of this group is that it is composed of common city names or proper names (e.g., *souza* and *carlos*), party acronyms (e.g., *pps*, *psc*, and *prb*), and other words that do not carry significant meaning (e.g., *ações propostas*, *eixo*). These words have high tf-idf values because, although they are not very frequent in the full set of documents (i.e., low n_w), they are frequent in the few documents in which they appear, which raises the idf component of the weight. Therefore, since these words do not help identify clear thematic topics, I remove them from the analysis. An alternative procedure would be to remove these types of words using dictionaries of party names and acronyms, city names, and proper names, similar to the removal of stop words. Nevertheless, this process is computationally intensive due to the size of the required dictionaries and is also subject to potential errors in selecting which words to include in those dictionaries.

After all those preprocessing steps, the final DTM contains 3,676 unique terms and is used as the baseline dataset for recovering the main topics discussed by mayoral political manifestos.

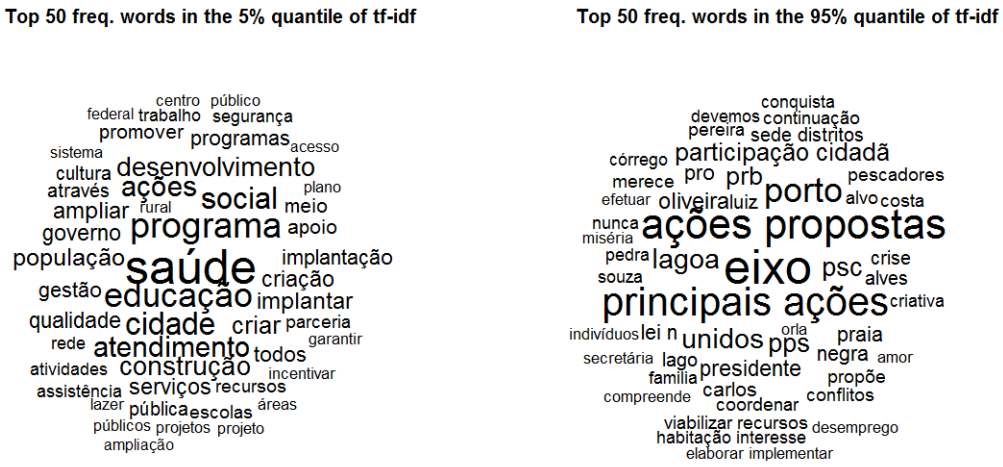


Figure A.2: Top 50 most frequent words in the 5% and 95% quantiles of tf-idf weights

A.3 Selection of LDA parameters

The LDA model is conditioned on three exogenous parameters: the number of topics K and two Dirichlet hyperparameters, α and η. α refers to the Dirichlet prior for the distribution of topics within each document (usually denoted by θ). η refers to the Dirichlet prior for the distribution of words within topics (usually denoted by β). To define the values of these parameters, I follow Griffiths and Steyvers (2004), who consider the following fixed priors for the Dirichlet hyperparameters: $\alpha = 50/K$ and $\eta = 0.1$. Due to the properties of the Dirichlet distribution, assigning low values to the hyperparameter η (i.e., $0 < \eta < 1$) helps generate topic distributions that concentrate on a small number of prominent words and are therefore more distinguishable.

Hansen et al. (2018) argue that the choice of the number of topics (K) is a persistent problem in topic modeling because, on the one hand, a higher number of topics leads to better statistical goodness of fit, while, on the other hand, a smaller number of topics leads to greater interpretability. Following these authors, I choose, as a baseline, to work with a model with 40 topics in order to produce more interpretable results. Nevertheless, I show that the overall conclusions are robust to an alternative model specification with 100 topics.

Figure A.3 provides evidence on the trade-off between statistical goodness of fit and interpretability described by Hansen et al. (2018). It shows the results of a 5-fold cross-validation exercise of the LDA model, varying K and measuring how well each model predicts out of sample. This exercise is similar to the one performed by Blei et al. (2003). The measure used to

compare the out-of-sample predictive power of each model is called perplexity and is presented below:

$$Perplexity(TestSet) = \exp\left\{ -\frac{\sum_{d=1}^{D} \log(p(w_d))}{\sum_{d=1}^{D} N_d} \right\} \tag{A.2}$$

Where $\sum_{d=1}^{D} \log(p(w_d))$ is the likelihood of all the words in the testing set estimated by the model run in the training set; N_d is the number of terms in each document d in the testing set. The expression above indicates that the better the model out-of-sample prediction is, the higher the likelihood in the numerator. Hence, the perplexity decreases when this likelihood increases. In the exercise, I compute the perplexity measure on the testing set for each of the five training sets and a sequence of different values of K.

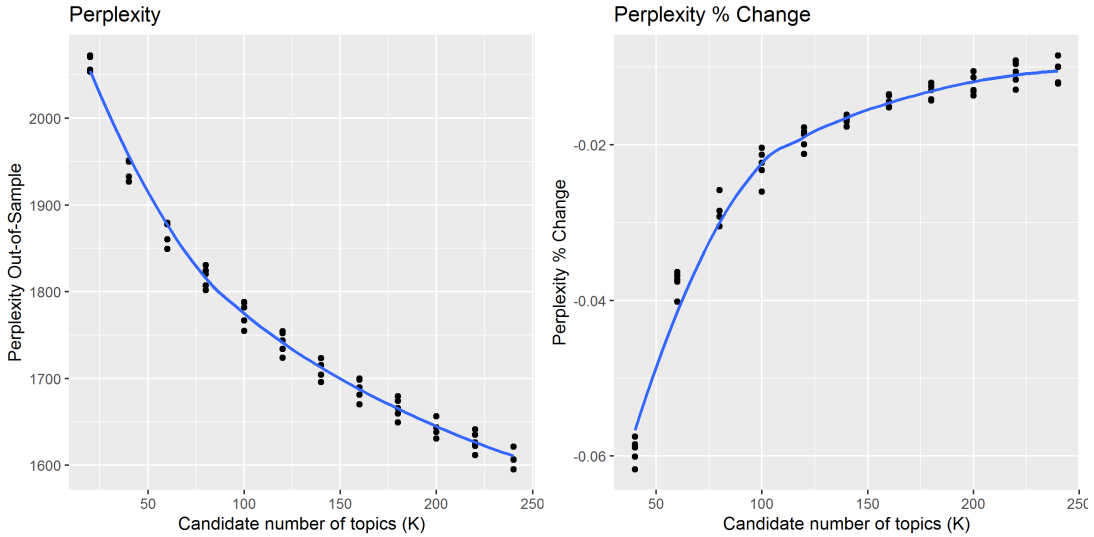


Figure A.3: 5-fold cross-validation exercise of LDA model for alternative values of K

Notes: each dot corresponds to a model trained with 1/5 of the sample and a perplexity measure computed using the other 4/5 of the sample.

The left-hand plot in Figure A.3 shows that perplexity decreases as the number of topics increases: goodness of fit improves for models with higher K. Moreover, the marginal improvement from increasing the number of topics is greater for models with up to 100 topics. As the right-hand plot indicates, when the number of topics increases from 20 to 40, perplexity decreases by around 6%. In contrast, when the number of topics increases from 100 to 120, perplexity decreases by only 2%. Hence, the marginal benefit (in terms of predictive power) of increasing the number of topics becomes smaller after reaching 100 topics. Therefore, using a

baseline model with 40 topics and a model with 100 topics for robustness appears to be a reasonable choice. Although models with more than 100 topics yield better out-of-sample predictions, their marginal improvement does not seem large enough to offset the increased difficulty in interpretation.

B Appendix - LDA Outcomes

B.1 Portuguese Figures (Original Files)

This appendix shows the original versions of Figures 1 to 5. The Google Cloud Translation API and the `googleLanguageR` package were used to translate these figures into English.

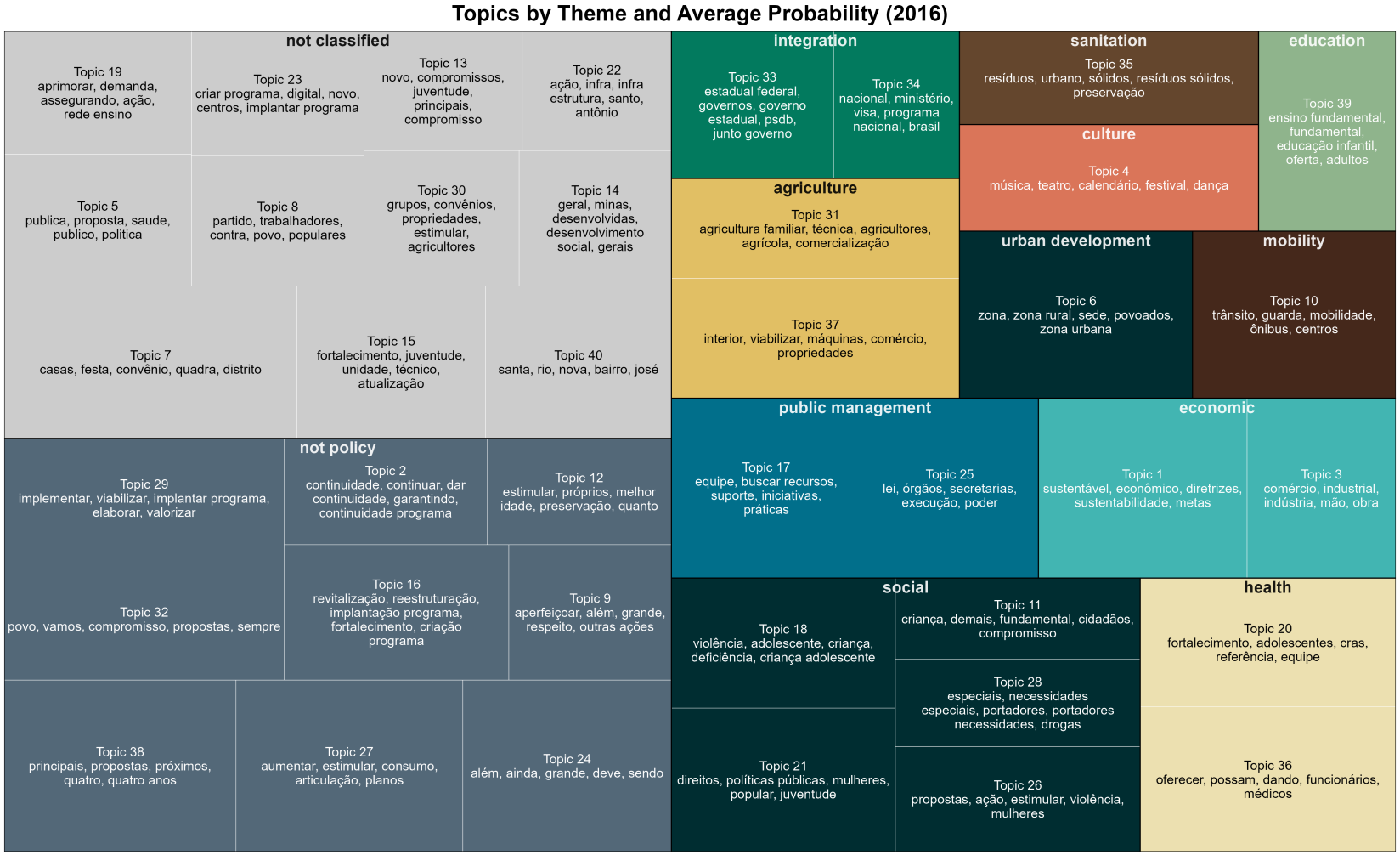


Figure B.4: Topic model output - top 5 most frequent words in each one of the 40 topics

Notes: The blocks' size represents the average probability of that topic appearing in one manifesto in 2016. The colors indicate a subjective classification of topics into 'policy issues', 'not policy issues', or 'unclassified'.

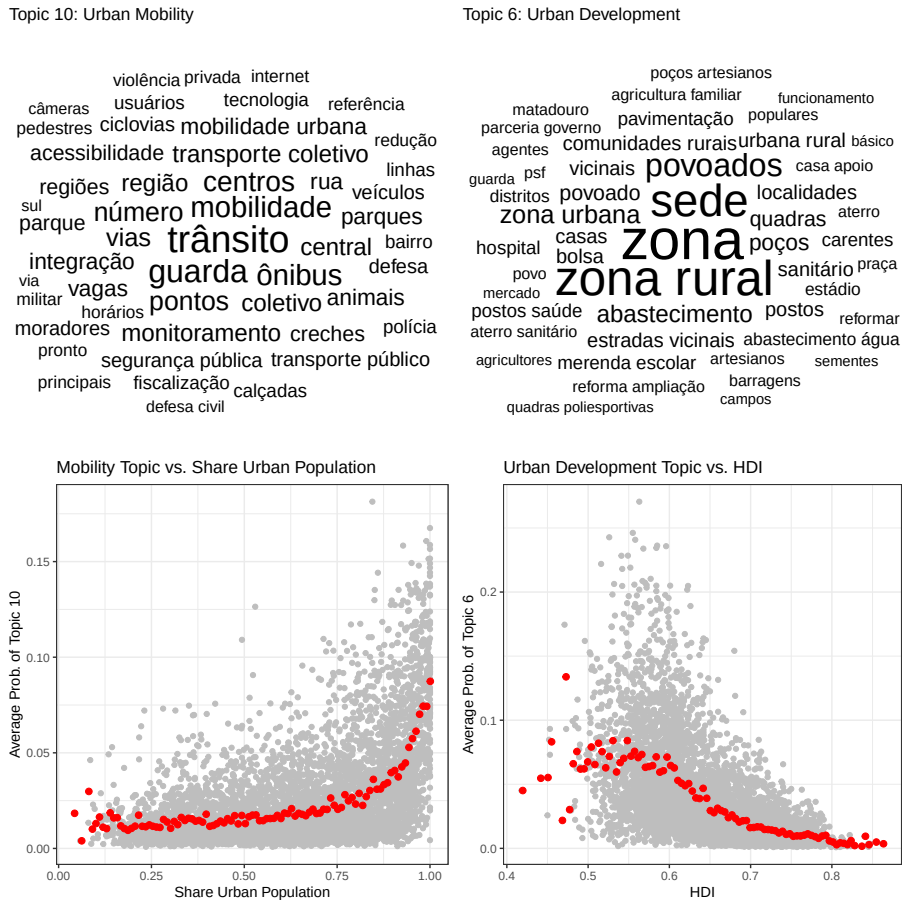


Figure B.5: Validation - emphasis on mobility and urban development issues follows an expected geographical pattern

Notes: Manifestos in more urban municipalities tend to discuss urban mobility more. Those in areas with low HDI focus more on urban development. Grey dots represent municipal average topic probabilities; red dots are binned averages (100 bins).

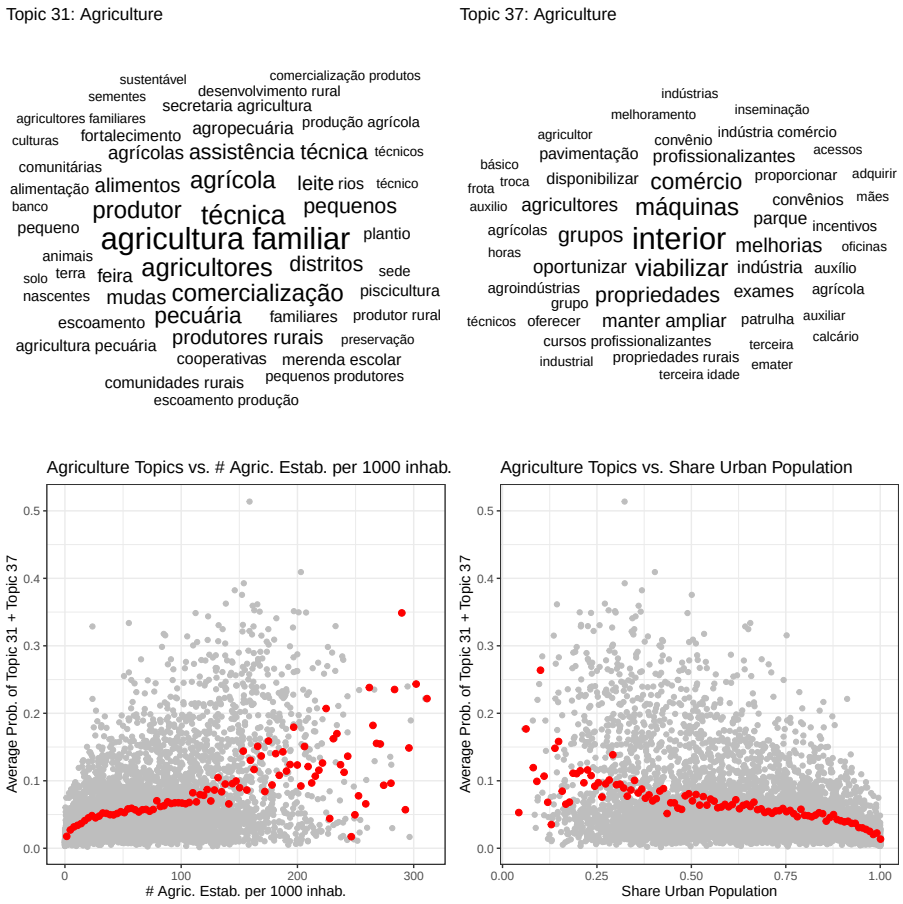


Figure B.6: Validation - emphasis on agriculture issues follows an expected geographical pattern

Notes: Manifestos in municipalities with more agricultural establishments per 1000 inhabitants are more likely to discuss agriculture. Conversely, the likelihood of discussing agriculture decreases as the urban population share increases. Grey dots show municipal average topic probabilities; red dots show binned averages (100 bins).

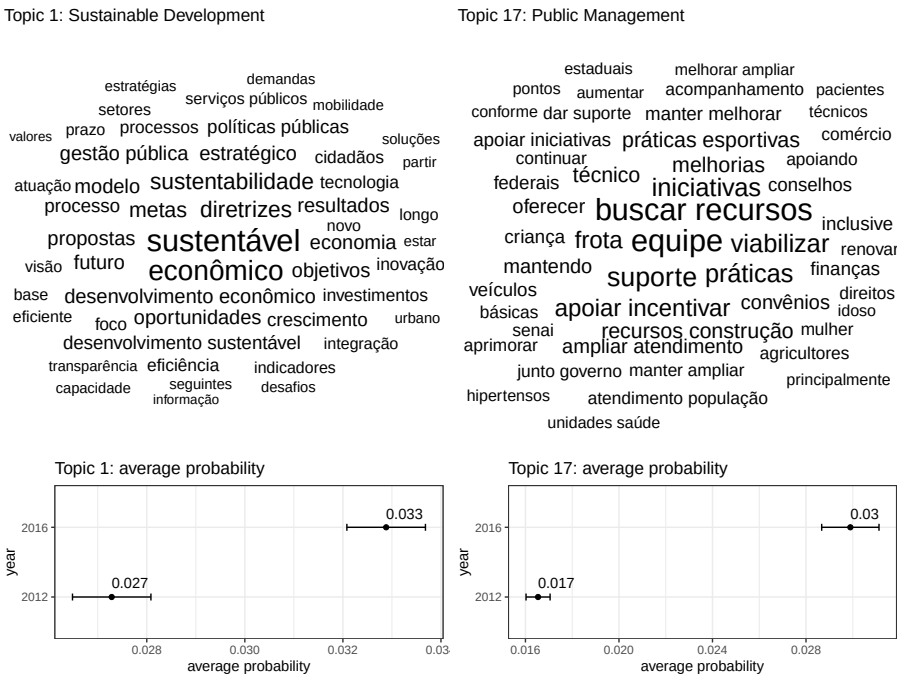


Figure B.7: Validation - emphasis on sustainable development and public management issues follows an expected temporal pattern

Notes: Sustainable development and public management topics are more likely to be discussed in the 2016 election, compared to the 2012 election. In the bottom plots, dots indicate the average probability of discussing that topic each year. Lines indicate the respective 95% intervals.

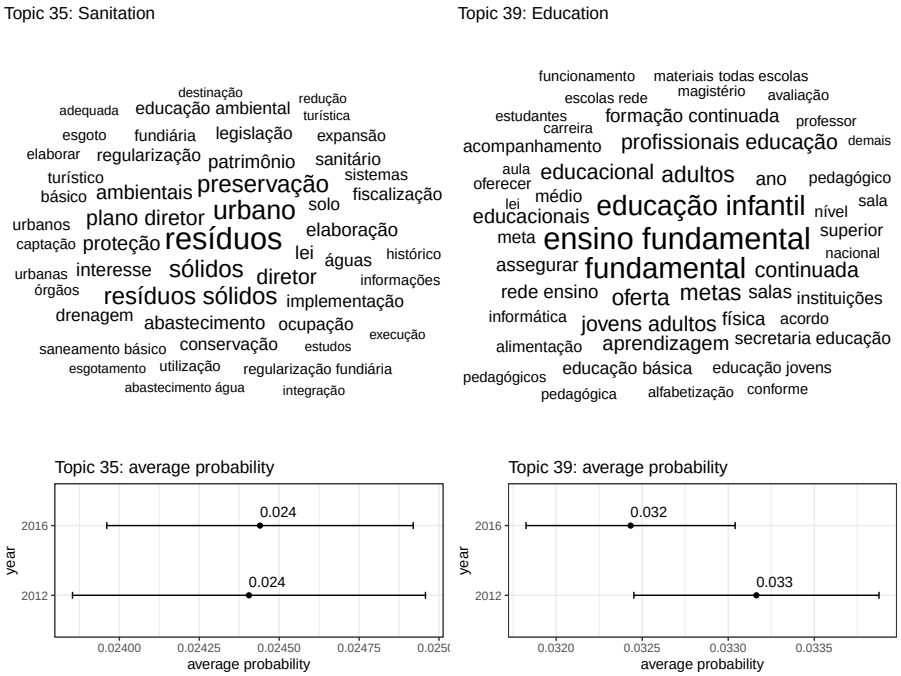


Figure B.8: Validation - emphasis on essential needs issues follows an expected temporal pattern

Notes: Probability of discussing topics that reflect essential needs, such as sanitation and education, is more stable over time. In the bottom plots, dots indicate the average probability of discussing that topic each year. Lines indicate the respective 95% intervals.

B.2 Table Describing Topics and Probabilities for 2016 Average Manifesto

Table B.1: Topics, Themes, and Average Topic Weights (2016)

Topic	**Top-5 terms**	**Theme**	**Mean θ**
Topic 1	sustainable, economic, guidelines, sustainability, goals	*economic*	0.033
Topic 2	continuity, Continue, continue, ensuring, program continuity	*not policy*	0.019
Topic 3	business, industrial, industry, hand, constructions	*economic*	0.023
Topic 4	music, theater, calendar, festival, dance	*culture*	0.028
Topic 5	publish, proposal, health, public, policy	*not classified*	0.022
Topic 6	zone, countryside, thirst, settlements, urban area	*urban development*	0.034
Topic 7	houses, party, health insurance, block, District	*not classified*	0.039
Topic 8	broken, workers, against, people, popular	*not classified*	0.020
Topic 9	perfect, in addition, big, respect, other actions	*not policy*	0.019
Topic 10	Traffic, guard, mobility, bus, centers	*mobility*	0.030
Topic 11	child, too much, fundamental, citizens, commitment	*social*	0.017
Topic 12	encourage, own, elderly, preservation, how much	*not policy*	0.017
Topic 13	new, commitments, youth, main, commitment	*not classified*	0.017
Topic 14	general, mines, developed, Social development, general	*not classified*	0.018
Topic 15	fortification, youth, unit, technician, update	*not classified*	0.025
Topic 16	revitalization, restructuring, program implementation, fortification, program creation	*not policy*	0.027

Table B.1: Topics, Themes, and Average Topic Weights (2016) *(continued)*

Topic	**Top-5 terms**	**Theme**	**Mean θ**
Topic 17	team, search for resources, support, initiatives, practices	*public management*	0.030
Topic 18	violence, adolescent, child, deficiency, teenage child	*social*	0.025
Topic 19	improve, demand, ensuring, action, education network	*not classified*	0.020
Topic 20	fortification, teenagers, cras, reference, team	*health*	0.029
Topic 21	rights, public policy, women, popular, youth	*social*	0.028
Topic 22	action, below, infrastructure, Holy, antonio	*not classified*	0.016
Topic 23	create program, digital, new, centers, deploy program	*not classified*	0.019
Topic 24	in addition, yet, big, he must, being	*not policy*	0.031
Topic 25	law, organs, secretariats, execution, power	*public management*	0.028
Topic 26	proposals, action, encourage, violence, women	*social*	0.023
Topic 27	increase, encourage, consumption, articulation, plans	*not policy*	0.034
Topic 28	specials, special needs, carriers, carriers of needs, drugs	*social*	0.019
Topic 29	to implement, enable, deploy program, elaborate, to value	*not policy*	0.029
Topic 30	groups, agreements, properties, encourage, farmers	*not classified*	0.018
Topic 31	family farming, technique, farmers, agricultural, commercialization	*agriculture*	0.025
Topic 32	people, we will, commitment, proposals, ever	*not policy*	0.030

Table B.1: Topics, Themes, and Average Topic Weights (2016) *(continued)*

Topic	**Top-5 terms**	**Theme**	**Mean θ**
Topic 33	federal state, governments, state government, psdb, together with government	*integration*	0.021
Topic 34	national, Ministry, Visa, national program, Brazil	*integration*	0.016
Topic 35	waste, urban, solids, solid waste, preservation	*sanitation*	0.024
Topic 36	to offer, can, giving, employees, doctors	*health*	0.032
Topic 37	interior, enable, machines, business, properties	*agriculture*	0.030
Topic 38	main, proposals, Upcoming, four, four years	*not policy*	0.035
Topic 39	elementary School, fundamental, child education, offer, adults	*education*	0.024
Topic 40	saint, river, new, neighborhood, josé	*not classified*	0.025

C Appendix - Validation of the RD Design

This section validates the RD design. First, I show that covariates are continuous at the cutoff regardless of the selected sample in Table 2. Tables C.2 and C.3 repeat the exercise in Table 3 in the main text for the samples in Panels B and C. Each row presents the estimated treatment effects for different predetermined covariates. The first part of the table shows the effect of the incumbent barely winning in t on municipality characteristics in t, including per capita GDP, the share of agriculture in GDP, and the number of candidates running in the mayoral election. The table also examines the incumbency effect on candidate characteristics in t, such as age, gender, educational attainment, and campaign expenditure. Moreover, the table shows that among parties involved in closely contested elections in t, incumbents do not spend more on their campaigns than challengers in the subsequent election ($t+1$). The last two rows assess the effect on predetermined values of the two manifesto issue concentration measures used as outcomes (HHI and CK5). As expected in a valid RD Design, the RD effect is indistinguishable from zero in all cases.

Table C.2: Predetermined Covariate Smoothness in t
Panel B: Incumbent Reruns, Challenger Proxied in $t+1$

Dep. Var.	Mean Control	Estimate	St. Err.	P-Val	BW	N
GDP per capita ('000 BRL)	12.711	0.058	0.689	0.933	0.125	5,924
Population ('000)	19.158	-0.463	1.672	0.782	0.151	6,755
Agric. Share GDP	0.214	0.0004	0.008	0.961	0.195	7,891
Indus. Share GDP	0.131	0.001	0.009	0.877	0.127	6,015
Serv. Share GDP	0.283	-0.001	0.006	0.903	0.167	7,196
Number of Apt Candidates	2.718	-0.013	0.057	0.824	0.163	7,086
Small Party	0.084	-0.018	0.016	0.258	0.162	7,065
MDB Wins	0.211	-0.003	0.022	0.885	0.183	7,557
PSDB Wins	0.158	-0.002	0.020	0.923	0.175	7,359
PP Wins	0.123	0.002	0.020	0.922	0.154	6,764
PT Wins	0.083	-0.001	0.018	0.937	0.130	6,065
PSB Wins	0.034	0.003	0.011	0.772	0.140	6,397
Age	48.687	-0.869	0.600	0.148	0.142	6,497
Gender: Female	0.119	0.001	0.016	0.939	0.209	8,148
Educational Attainment: High School or Higher	0.787	0.025	0.022	0.272	0.156	6,886
Campaign Expenditure in t ('000 BRL)	82.105	1.953	7.778	0.802	0.169	7,262
Campaign Expenditure in t+1 ('000 BRL)	83.007	7.296	8.293	0.379	0.155	6,864
Delta Expenditure ('000 BRL)	1.695	5.835	7.571	0.441	0.154	6,833
HHI in t (no observations for 2008)	0.129	-0.016	0.011	0.167	0.176	3,307
CK5 in t (no observations for 2008)	0.530	-0.015	0.012	0.227	0.160	3,103

Notes: Table shows estimates of separate RD treat. effects for each predetermined covariate; Excl. municipalities with runoff voting; BW: vote margin in t; Estimates obtained using local linear regressions and triangular kernel weights; Estimates and 95% confidence intervals are robust-bias corrected following (Calonico et al., 2020); Candidates were not required to publish political manifestos in 2008. Therefore, this year is excluded from the calculation of RD effects for predetermined outcomes (HHI and CK5). Delta Expenditure is equal to Campaign Expenditure in t+1 - Campaign Expenditure in t.

Figure C.9 provides a graphical illustration of the RD for a subset of the covariates reported

Table C.3: Predetermined Covariate Smoothness in t
Panel C: Both Winner and Runner-Up Rerun in $t+1$

Dep. Var.	Mean Control	Estimate	St. Err.	P-Val	BW	N
GDP per capita ('000 BRL)	12.900	-0.008	1.006	0.994	0.134	2,526
Population ('000)	20.600	-0.091	2.767	0.974	0.144	2,638
Agric. Share GDP	0.209	-0.002	0.016	0.917	0.131	2,492
Indus. Share GDP	0.133	0.002	0.012	0.883	0.151	2,724
Serv. Share GDP	0.292	0.0003	0.009	0.972	0.166	2,872
Number of Apt Candidates	2.704	0.012	0.088	0.894	0.151	2,718
Small Party	0.049	0.022	0.023	0.330	0.141	2,606
MDB Wins	0.232	-0.002	0.037	0.957	0.154	2,726
PSDB Wins	0.163	-0.001	0.033	0.965	0.155	2,740
PP Wins	0.145	-0.001	0.035	0.966	0.139	2,578
PT Wins	0.077	-0.001	0.028	0.975	0.111	2,200
PSB Wins	0.040	0.008	0.017	0.653	0.176	2,936
Age	48.589	-1.028	0.930	0.269	0.130	2,468
Gender: Female	0.110	0.014	0.029	0.613	0.154	2,742
Educational Attainment: High School or Higher	0.814	-0.017	0.033	0.608	0.159	2,808
Campaign Expenditure in t ('000 BRL)	80.793	-2.311	14.815	0.876	0.110	2,184
Campaign Expenditure in t+1 ('000 BRL)	89.917	5.568	14.923	0.709	0.169	2,896
Delta Expenditure ('000 BRL)	8.984	4.838	16.094	0.764	0.120	2,324
HHI in t (no observations for 2008)	0.116	-0.028	0.017	0.097	0.101	926
CK5 in t (no observations for 2008)	0.507	-0.008	0.018	0.639	0.113	997

Notes: Table shows estimates of separate RD treat. effects for each predetermined covariate; Excl. municipalities with runoff voting; BW: vote margin in t; Estimates obtained using local linear regressions and triangular kernel weights; Estimates and 95% confidence intervals are robust-bias corrected following (Calonico et al., 2020); Candidates were not required to publish political manifestos in 2008. Therefore, this year is excluded from the calculation of RD effects for predetermined outcomes (HHI and CK5). Delta Expenditure is equal to Campaign Expenditure in t+1 - Campaign Expenditure in t.

in Table C.2. The plots corroborate the tabular results, showing no evidence of discontinuities at the cutoff for municipal characteristics (e.g., per capita GDP, number of candidates), candidate characteristics (e.g., age, education, campaign expenditure), or the predetermined values of the outcome variable (HHI).

Second, I examine whether the density of the vote margin is discontinuous around the cutoff. Figure C.10 shows a graphical representation of the continuity test for the vote margin at the cutoff. It presents locally estimated densities following Cattaneo et al. (2019), with shaded 95% confidence intervals. Panels (a) to (c) report results for increasingly restrictive samples. Panel (a) conditions on at least one of the two parties in a municipality—the incumbent or the challenger—rerunning in the subsequent election. Panel (b) conditions only on the incumbent rerunning, using the best-ranked non-incumbent party as the comparison group when the original challenger does not rerun. Panel (c) conditions on observing both the winner and the runner-up parties from the same municipality rerunning in the subsequent election.

In Panel (a), the density estimate indicates a slight jump in the vote margin, though not significant at the 95% level. More precisely, the discontinuity test has a p-value of 0.074, meaning that for a 95% significance, one fails to reject the null hypothesis of no discontinuity in the den-

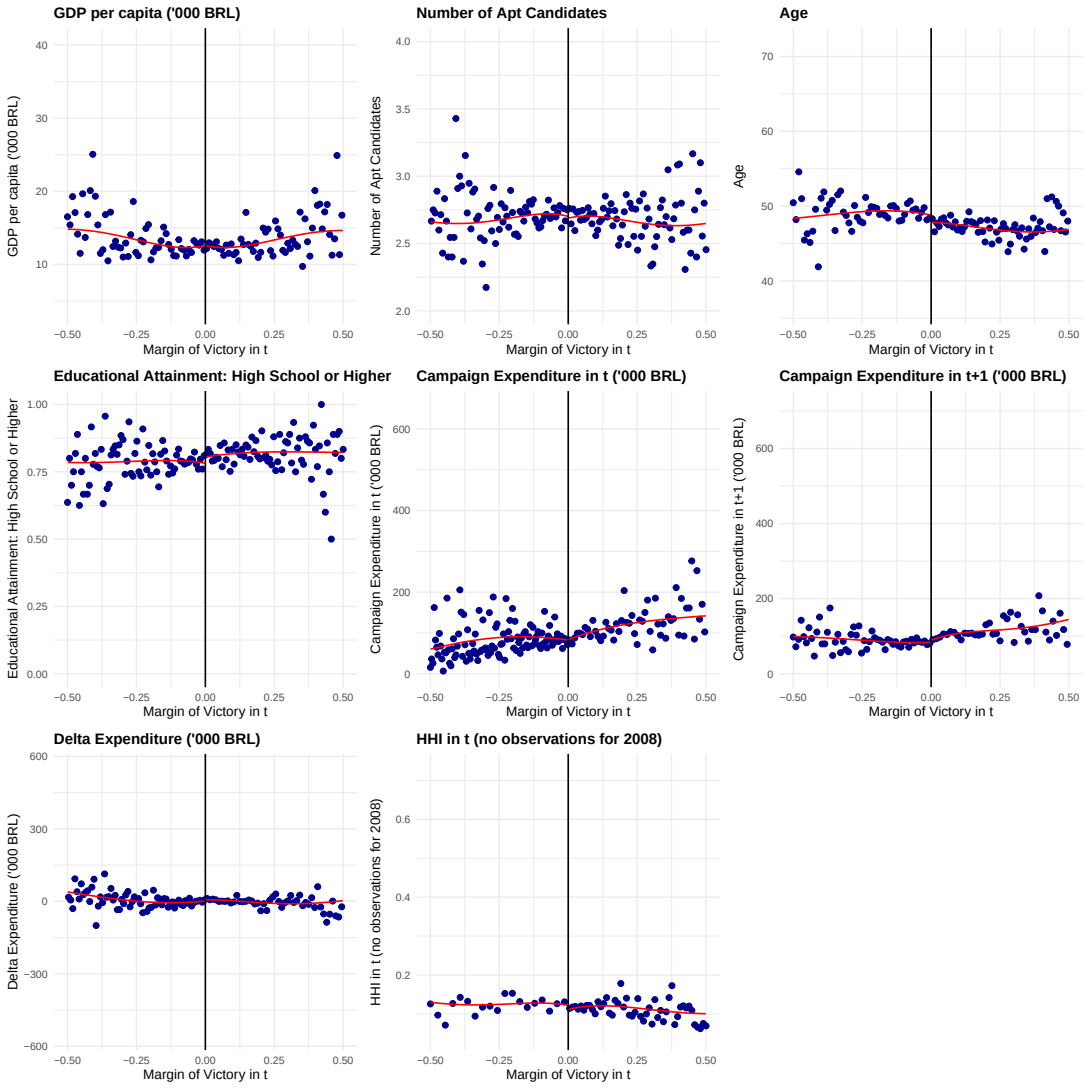
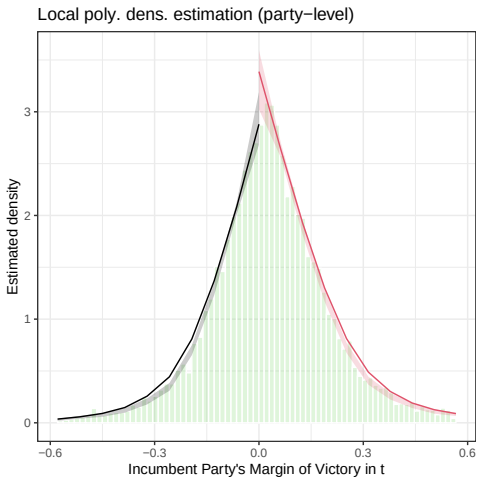


Figure C.9: Covariate Smoothness for Selected Variables

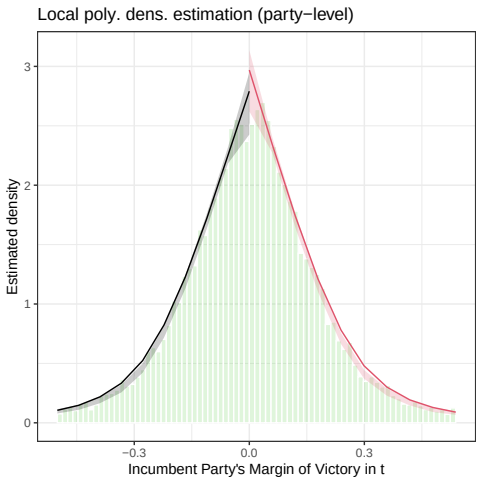
Notes: The Figure shows plots of RD regressions for selected covariates presented in Table C.2. It uses the preferred sample of Panel B (Incumbent Reruns, Challenger Proxied when Missing in $t+1$). The horizontal axis shows the margin of victory in t.

sity of treated and control groups at the cutoff. In Panels (b) and (c), the p-values are 0.334 and 0.999, respectively, suggesting no evidence of discontinuity in the density of observations around the cutoff.

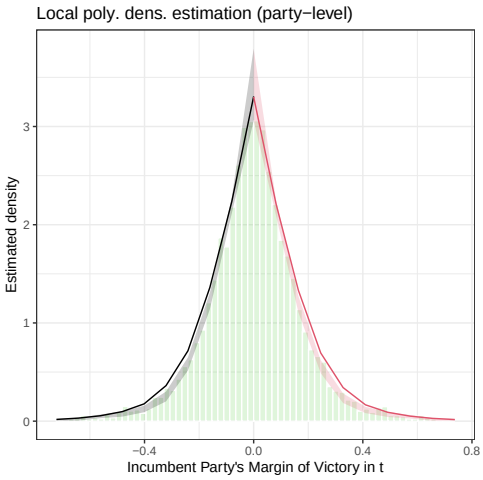
Figure C.10: Test of Density Discontinuity at the Cutoff



(a) At Least One Party Reruns (Incumbent or Challenger) in $t+1$



(b) Incumbent Reruns, Challenger Proxied when Missing in $t+1$



(c) Both Winner and Runner-Up Rerun in municipality m in $t+1$

Notes: The figure presents local polynomial density estimates according to Cattaneo et al. (2019). The p-value for the discontinuity test in Panels (a) to (c) is 0.074, 0.334, 0.999, respectively. Hence, the test only fails to reject H0 of no discontinuity at the cutoff in panel (a).

D Appendix - Continuity Issue

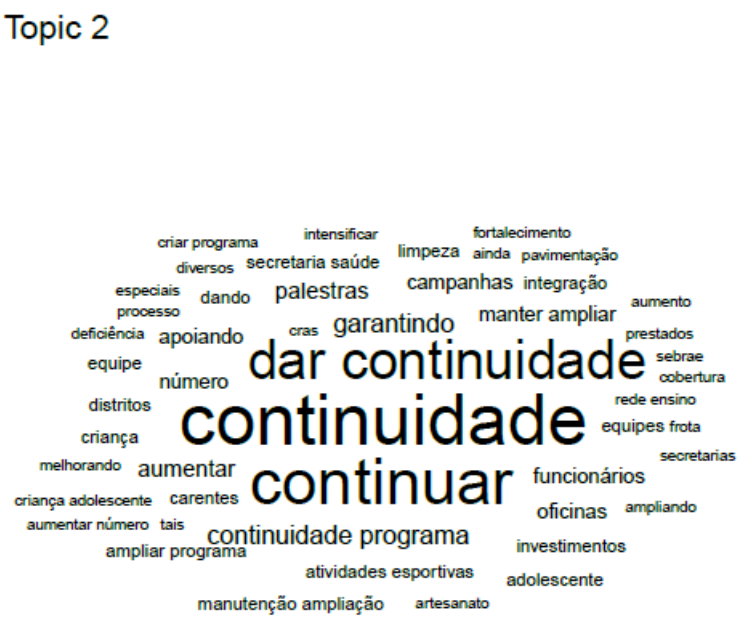


Figure D.11: Top 50 most frequent words in topic 2 (continuity)