When Women Run Against Men: Evidence from Political Platforms

Marli Fernandes*

Nova School of Business and Economics

JMP
Click here for the most recent version
November 28, 2023

Abstract

Do female and male candidates differ in their political campaigns? Do they adjust to their opponent´s gender? Using individual political platforms from legislative elections in France, I combine computational text analysis with a regression discontinuity design setup in the two-round French legislative elections to understand political platform content differences between women and men. I find that women present themselves closer to the party line and give more salience to topics such as security and foreign policy. This is a strategic response, as I causally determine that when women run against a man, as opposed to a woman, they give more prevalence to male-stereotypical topics. However, once elected, women address issues like health and education compared to male colleagues. In contrast, when male politicians run against women, they adapt their platforms more marginally.

Keywords: Elections, gender, text as data

JEL classification: D72, J16, P0

1 Introduction

Citizen-candidate models demonstrate that politicians not only care about being elected, but also about implementing their preferred policies (Alesina (1988), Osborne and Slivinski (1996), Besley and Coate (1997)). The citizen-candidate model points to the relevance of identity in policy-making. Female representatives tend

*PhD student at Nova School of Business and Economics. E-mail: marliagfernandes@gmail.com. Orcid id: 0000-0001-8120-5654. Address: Rua da Holanda no 1, 2775-405, Carcavelos, Portugal. I am grateful to my supervisor José Tavares for his support. I am grateful for the comments made by Alex Armand, Pedro Martins, Victoire Gerard, and seminar participants at Nova SBE and the Lisbon Micro Group. I thank Odile Gaultier-Voituriez in providing support to access the CEVIPOF 2017 manifestos. I acknowledge financial support from FCT - Fundação para a Ciência e Tecnologia - PD/BD/135298/2017 and COVID/BD/152031/2021. All errors are my own. The author does not have any conflict of interest to disclose.

to address specific issues that, by nature or due to traditional gender roles, mainly affect women (Chattopadhyay and Duflo (2004), Hessami and da Fonseca (2020)). In contrast to the citizen-candidate model, the median voter theorem (Hotelling (1929), Downs (1957)) predicts that politicians converge on policy to cater to the median voter preferences and secure being elected. According to this theorem, gender is irrelevant in political campaigns.

Being elected is incredibly challenging for women since research has shown that they face voter bias (De Paola, Scoppa, and Lombardo (2010), Le Barbanchon and Sauvagnat (2022), Eyméoud and Vertier (2023)). If women believe that the voters discriminate against their characteristics, they might adopt non-discriminated traits to maximize their chances of being elected. This last situation implies that female politicians cannot signal to voters their true type, in contrast to the predictions of the citizen-candidate model, where female politicians focus on their preferred policies, conveying correct information during campaigns to voters.

Do women focus on their preferred policy topics during political campaigns, or do they prefer to adapt their discourse to voters' preferences to maximize the opportunity of being elected? Testing this hypothesis is methodologically challenging because a female presence in a race is endogenous to electoral districts' characteristics. In addition, data on political discourse during campaigns is often lacking. In most countries, there is no record of campaigns run by individual politicians, except France. Because the majority of the data only covers election winners in their political office, research is mostly concentrated on how women perform once elected.

To understand whether women signal their type during elections, I study gender differences in political campaigns during legislative elections in France. The same pool of voters is called to participate in France's first and second rounds. With an average of nine candidates in the first round and two in the second, politicians face the challenge of appealing to a broader electorate. To overcome endogeneity issues, I exploit two-stage elections to estimate the impact of gender in campaign strategies. I use a regression discontinuity design to isolate quasi-random variation in the candidates' gender. The research design exploits that in a two-round system, where a female politician, instead of a man, barely makes it to the final round, and those who just miss the qualification threshold, the electoral district and opponents' characteristics are arguably comparable in observed and unobserved characteristics.

I use a comprehensive candidates' political platforms data set that provides the record of the campaign messages sent by every competing politician during the legislative elections in France. Individual candidates issue their own campaign platforms, which the State prints and sends to all registered voters a few days before the election. Politicians use these platforms to inform their constituents about themselves, their program and appeal to vote.

I use various methods of computational text analysis to construct several text-based outcomes that permit to test the following hypothesis: (i) differences in tone/sentiments, (ii) differences in the ideology, (ii) differences in the originality or personalization of the campaign, (iii) differences in the prevalence of topics.

I first document causally, to my knowledge for the first time, that there are differences between women and men when they campaign. I find that women convey more to the party line. The second interesting finding is that women campaign more at the right than at the left and discourse more on security & foreign policy than males. These two last characteristics are traditionally stereotyped as male traits. Furthermore, I causally show that a higher prevalence of security & foreign policy topics is an adaptive behavior; when women marginally compete with a man instead of a woman, they campaign by a large magnitude more on this topic. In other words, expecting voters' bias, women adopt a stereotyped male trait when they compete against men.

Another potential explanation for gender differences in political discourse is a gender gap in campaign financing. If women face donor or party discrimination and collect less funds, they might have a weaker campaign advisors' team. I prove that women do not receive less funds from donors or their party; however, surprisingly, when they compete with a man instead of a woman, they use less of their personal funds. This result reinforces the idea that when women compete against men, they feel less confident about their success.

According to the median voter theorem, in a representative democracy, politicians will converge to the viewpoint of the median voter. Candidates adjust their platform to their opponent's platform, which leads them to converge to the center (Di Tella, Kotty, Le Pennec, and Pons (2023)). In more detail, the results might reflect the predictions of the median voter theorem and not a consequence of women expecting voters' bias. However, I find very little evidence that males adopt female-stereotyped traits. When a woman is marginally elected to the second round of the election, male politicians change their political platforms to a small extent, and this change depends on the male's ideology. Left-wing politicians write more, while right-wing politicians do not change their campaign communication.

Last, focusing on mixed-gender races, I compare barely elected female candidates with males elected by a small margin. I causally prove that women are as active as men in legislative debates, a participation that requires charismatic and good rhetorical abilities. Moreover, in legislative debates, women speak more about health & education than men by a significant margin. This last result has several implications. First, it reinforces the idea that women strategically adapt stereotyped male traits during their campaigns to increase their chances of winning. Second, it shows that issues favored by women do not get more attention in political campaigns by the most legitimate politicians. Third, it proves voters cannot infer women's correct type during campaigns.

The paper is organized as follows. In Section 2 I overview the related literature. In Section 3 I describe the institutional context. Section 4 presents the data, Section 5 the methodology, Section 6 the results and Section 7 concludes.

2 Literature review

The results of the paper contribute to several strands of the literature. First, the paper is related to the impact of politicians' identity. The Median Voter Theorem (Downs (1957)) assumes the parties' unique objective is winning elections. Consequently, if two parties have the same information about voters' preferences, they will converge to the viewpoint of the median voter, ignoring politicians' identity. Le Pennec (2023) and Di Tella et al. (2023) empirically demonstrate the convergence mechanism underlying the median voter theorem. However, candidates only converge to a certain extent. In citizen-candidate models, parties do not only care about winning elections but also about implementing their preferred policies (Alesina (1988), Osborne and Slivinski (1996), Besley and Coate (1997)). The findings of this paper extend this literature in understanding the influence of alternative dimensions of identity, in this case, gender. It contributes to clarifying whether any of the models can predict politicians' behavior in identity situations.

Second, the paper also contributes to the extensive literature on gender and competition. For a review, see Niederle and Vesterlund (2011). A series of laboratory studies documents that, conditional on performance, women are often more reluctant to compete than men (e.g., Niederle and Vesterlund (2007), Markowsky and Beblo (2022)). This pattern has been confirmed in the case of elections (Kanthak and Woon (2015), Barber, Butler, Preece, et al. (2016)). These differences might be a reflection of social learning (Booth and Nolen (2012)) or culture (Gneezy, Leonard, and List (2009)) rather than inherent gender traits. Research has also shown that these differences depend on their own gender and on the gender of people with whom they interact. In a two-person bargaining game, competition and retaliation are higher when the bargaining partners have the same gender (Sutter, Bosman, Kocher, and van Winden (2009)). Finding opposing results, Datta Gupta, Poulsen, and Villeval (2013) conclude that individuals compete less with same-sex opponents in tournaments. Indeed, literature on how individuals compete, considering the opponents' gender, is limited and needs further exploitation. This paper contributes to this literature by studying, to my knowledge, for the first time, how politicians adapt their races when they compete with the opposing sex.

The third strand relates to the large literature on gender gaps in politics. Research has been documenting the causes of gender differences in political success. Potential determinants include a lack of political ambition (Fox and Lawless (2004), Fox and Lawless (2014)), voter bias (Le Barbanchon and Sauvagnat (2022), Eyméoud and Vertier (2023)) or party bias (Esteve-Volart and Bagues (2012), Gonzalez-Eiras and Sanz (2021), Lippmann (2021)). I contribute to this literature by studying gender differences in political campaigns.

3 Institutional context

3.1 French parliamentary elections

The paper focuses on parliamentary elections. The parliamentary elections elect all the members of the National Assembly, the lower house of the Parliament.

These elections are held under direct universal suffrage with a two-round plurality voting rule. Candidates can directly win in the first round if they obtain a number of votes greater than 50 percent of the votes and 25 percent of the registered citizens. In most elections, no candidate is elected in the first round, and a second round is held one week later. The second round is decided by simple plurality: the candidate with the largest vote share wins the election.

Candidates that obtain at least 12.5 percent of the vote share of the eligible voters are qualified for the second round. However, if only one of the candidates (or none) meets the threshold, the two candidates with the largest vote share can proceed to the second round.

Selection of political candidates It is possible to run for one of the 577 French constituencies without being affiliated to a party, as long as the aspiring MP meets all the necessary conditions to be able to enter the race, in particular being at least 18 years old, having the right to vote and not be ineligible (because of a court decision or a function incompatible with the mandate of an MP, such as being mayor). However, most candidates run under a party label, making them much more visible to voters.

The selection of political candidates for each electoral constituency varies from party to party. For a description of how each party selects its candidates, see Murray (2010). For example, in 2022, the LR used a ”National Investiture Commission", composed of around sixty members, that nominated the candidates for several constituencies (Loaec (2022)). In contrast, the Ecologist party used an internal commission, the "Permanent Electoral Commission", which designated the first candidates in connection with local proposals; still, the choices needed to be validated by a vote of party members.

According to the 2000 gender parity law, parties must present an equal fraction of male and female candidates across the electoral districts. If the difference between female and male candidates exceeds 4% (48% females and 52% males, or the reverse), non-compliance with the gender parity rules results in a financial penalty. The financial penalty is computed as follows: "public funding provided to political parties based on the number of votes they receive in the first round of elections is reduced by a percentage equivalent to one-half of the difference between the total number of candidates of each sex, out of the total number of candidates" (Le Barbanchon and Sauvagnat (2022)). Notice that in France, voters vote for two candidates: the leading candidate, if elected, will become an MP and the substitute. The last one replaces the top candidate only if he leaves office; potential reasons are being nominated to a function incompatible with the mandate of an MP (e.g., minister, secretary of state, elected as a mayor) or death/health issues. In most cases, the substitute never becomes an MP during the legislative term. After the approval of the law, in the 2002 legislative elections, women were 38.8% of the leading candidates, compared to 23% in the 1997 legislative elections.

Ideological classification I use the official party labels provided by the Ministry of the Interior to classify candidates (including independent ones) and following Jolly et al. (2022) I classify candidates into six partisan families: far-left, left, liberal, right, far-right, and other. The last category refers to politicians who do not fall into any of these traditional ideological categories or do not classify themselves into any ideology.

In the rest of the paper, I refer to political orientation as the broader categories "left" (far-left and left) and "right" (liberal, right and far-right), unless specified otherwise. Essentially, I classify "left" parties as the ones that score less than 5 in the left-right dimension of Jolly et al. (2022) and "right" as the ones that score more than 6.[1] Parties classification is available in the Appendix Section A.

[1] According to Jolly et al. (2022), all liberal parties in the sample have a left-right score between 6 and 7, in a spectrum between 0 and 10.

3.2 Political platforms

During the legislative campaign, individual candidates can emit one political platform (*trans.* profession de foi) before each election round.[2] The appendix provides one example in the Figure B.1. What is the traditional content of a political platform? The manifesto permits a candidate to present his program and ideas that he plans to commit to when elected. Candidates might run a campaign by focusing on national policies and/or addressing local issues since they represent an electoral district. They can also run a more partisan program or personalize their campaign, focus on preferred policy topics, or, as expected, appeal to vote and criticize the opposition.

Candidates are responsible for printing these platforms, and the state can reimburse their costs if they gather at least 5% of the votes in one of the rounds (Electoral law, articles R39 and L216). An official local propaganda committee is responsible for mailing the manifestos to voters at least four days before the first round and three days before the second round (if it happens).[3]

According to the Ipsos - CEVIPOF 2022 Presidential electoral survey, the primary sources of candidates' information are: 38% of the electors follow the television, 15% the internet (many candidates share their platforms online), 13% newspapers, 12% the manifestos received in the mailbox, 6% use other sources around them, and 16% do not inform. Relatively to the 1988, 1993, and 1997 manifestos, the internet was yet to be widely available and used. Furthermore, per election, there are 577 constituencies and an average of 4079 candidates, making it impossible for television to give coverage to all candidates. Consequently, it is likely that platforms are more important in legislative elections than in presidential elections.

4 Data

4.1 Electoral data

Each dataset records the number of registered voters, abstentions, cast votes, valid and invalid votes, and the votes for each candidate in each electoral district. The electoral data for French elections comes from the Ministry of Interior.

[2]Throughout the paper, I use the words "political platform" and "manifesto" interchangeably; in this paper, they are synonymous.

[3]These platforms are only mailed if they obey these rules: they must have a maximum size of 210x297 millimeters, weigh between 60 and 80 grams per square meter (Electoral law, article R29), and cannot combine the three colors of the French flag, except if they are part of the party's emblem (Le Pennec (2023)).

4.2 Political platforms

Candidate manifestos for the 1988 and 1993 elections were digitized by the Archelec project (Gaultier-Voituriez (2016)).[4] Until 1993, the CEVIPOF collected manifestos each election with the government's support. Each departmental administration mailed the manifestos distributed in their district. Unfortunately, this practice finished in 1993, so manifestos between 2002 and 2012 are not available.

Platforms for the 1997 elections were digitized from the National Archives.

For 2017 they come from several sources: during the campaign, the Ministry of Interior shares the manifestos submitted by the candidates on their website[5] and they were web-scraped by a non-profit organization calleg RegardsCitoyens[6]. In additon, I manually digitized missing manifestos at CEVIPOF (Sciences Po), the National Assembly website also shares the manifestos of all elected MPs[7], some missing manifestos were also found on several local news or candidates websites. For 2022, platforms come from the Ministry of Interior website and several local news and candidates websites.

The dataset comprises 27934 political platforms, 22487 for the first round and 5447 for the second round.

I pre-process platforms' content by removing capitalization, punctuation, stop words, and special characters. I tokenize documents at the single-word level and lemmatize each word using Spacy's French model.

4.3 Legislative work

I web-scraped the speeches from the Assemblée Nationale website covering the 1998-2022 period, and transform them into a novel dataset.[8][9] I restrict my analysis to elected politicians, excluding presidents and vice-presidents of the Parliament. I eliminate procedural words in parliamentary speech because they appear frequently and their use is unlikely to be informative about group differences (Gentzkow, Shapiro, and Taddy (2019)).[10] I also remove speeches with less than

[4] They are available at `https://archive.org/details/archiveselectoralesducevipof/`

[5] They are available at: `https://programme-candidats.interieur.gouv.fr/`

[6] They are available at: `https://github.com/regardscitoyens/professions-foi-candidats/tree/master/documents/LG17`

[7] They are available at: `https://www.assemblee-nationale.fr/dyn/decouvrir-l-assemblee/histoire/barodet2/recueil-des-professions-de-foi-de-la-15eme-legislature`

[8] They are available at: https://www.assemblee-nationale.fr/dyn/15/comptes-rendus/seance

[9] I do not include the years 2001 and the end of the 11th legislature (2002) because the website during this period is designed differently and is difficult to web-scrape.

[10] I obtain the list of procedural phrases from the following websites: https://www2.assemblee-nationale.fr/decouvrir-l-assemblee/folder/lexique and https://www.assemblee-

five words; these speeches tend to be minor reactions to an opponent's speech and are not informative in terms of group or topic. I aggregate speeches so that a document captures all speeches by a given speaker in one agenda of a plenary session; I remove aggregate speeches with less than 15 words; concise speeches are not ideal for detecting group differences or topics. The dataset includes a total of 155,207 documents.

Second, I web-scrape the National Assembly website to obtain the content of the written questions. These questions are directed to a minister to express the MP concern on a specific topic.[11] The dataset includes a total of 590,185 questions.

For further details on these two data sources, see the Appendix E.1 and E.2.

4.4 Campaign contributions

Data between 1993 and 2017 on campaign expenditures and contributions is from Bekkouche, Cagé, and Dewitte (2022). For 2022, I use data from the National Commission on Campaign Accounts and Political Financing (CNCCFP). In French parliamentary elections, candidates who receive at least 1% of votes in the first round must submit their campaign accounts to the French CNCCFP. This commission was created in 1990, so data before that date is unavailable.

For each year, electoral district, and candidate, I observe for the whole campaign the total amount spent by the candidate, the total amount of contributions he received, and the amount of each type of contribution: contributions received from the candidate's political party, donations, and personal funds.

5 Methodology

5.1 Measuring tone

I compute a quantitative measure of the tone using a dictionary-based approach. The tone of each platform is computed as follows:

$$Tone_i = \frac{W_i^P - W_i^N}{W_i^T} \tag{1}$$

where $Tone_i$ is the tone of manifesto i, W_i^P is the number of positive words within manifesto i, W_i^N is the number of negative words within manifesto i, W_i^T is the

nationale.fr/connaissance/lexique.asp .

[11] They are available at: https://www2.assemblee-nationale.fr/recherche/questions

number of total words within manifesto i.[12]

I use the LSDFr dictionary (Duval and Pétry (2016)) to perform the tone assessment, which political communication researchers created. The authors also show that their dictionary over-performs the LIWC (Piolat, Booth, Chung, Davids, and Pennebaker (2011)).

In the Appendix, I demonstrate that radical parties have more negative political platforms as expected (Table C1) and show the most positive and negative manifestos (Figures C.1 and C.2).

5.2 Left-right dimension

I project each document onto a left-right dimension. Following Gentzkow et al. (2019), I adopt a semi-supervised machine learning approach to project all political platforms onto an ideological scale, considering the known party affiliation of politicians and the ideological leaning of these parties from left to right. In more detail, for each election year, I restrict the vocabulary to words used by at least 0.5 and 50% of the platforms, and I start by building a matrix with the frequency of each word in a document. Then, I follow the multinomial inverse regression approach proposed by Taddy (2013) and Taddy (2015), and use a penalized estimator to control for finite-sample bias as recommended by Gentzkow et al. (2019).[13] Cagé, Le Pennec, and Mougin (n.d.) adopt the same approach to project all manifestos onto an ideological scale and give them a left-right partisan score. After obtaining a score for each word, I calculate the average ideological score of each document. As a result, a document with a negative (positive) score is associated with left (right) ideology, meaning that it includes words that are associated with left (right) manifestos and are less often included in right (left) manifestos. On the other hand, if a manifesto has a score close to zero, it means that it tends to use neutral words (words that tend to be used by left and right politicians) or polarising words from both sides. In the Appendix Section F, I provide further details on the methodology, figures with the kernel density of partisan scores, and the more left and right words by year.

[12] A disadvantage of the dictionary method is that it requires to be reliably specified. To overcome this issue, Gennaro and Ash (2022) use word embeddings in a sample of six million speeches. However, word embeddings are not adequate for the case of small samples, as it is the case of this paper.

[13] I use the R package dmr (Taddy and Taddy (2022)).

5.3 Topic classification

In this part, I restrict the vocabulary to words used by at least 1% and 50% of the platforms for the whole sample; in the case of the legislative debates, the minimum number is 0.5% since the number of observations is considerably larger.

I measure the prevalence of specific policy topics in campaign manifestos. The topics of the platforms, legislative debates or written questions are not explicitly stated, and therefore does not exist a training data set containing pre-defined categories. Absent this information, I rely on topic modeling techniques to retrieve the topics and construct the outcomes of interest, specifically seeded Latent Dirichlet Allocation (seeded LDA).

LDA (Blei, Ng, and Jordan (2003)) is a generative probabilistic model based on the assumption that each document is a mixture of topics and that latent topics generate the words observed in the document of a corpus. LDA is an unsupervised method, while seeded LDA (Lu, Ott, Cardie, and Tsou (2011), Watanabe and Baturo (2023)) is a semi-supervised machine learning technique. Seeded LDA extracts these topics based on a prior 'seed' of selected words that capture the topic of interest. Watanabe and Baturo (2023) show that this method improves the inconsistency of topics that LDA generally produces.

The central tuning parameter of a LDA model is the number of topics K to be estimated. If K is too small, documents about different topics will be lumped together in the same estimated topic. If K is too large, documents that belong to the same topic are split. I calculate topic coherence in a simple LDA to estimate the number of topics for platforms. The highest value is at 8. I define the following eight topics: economy & employment, environment; health & education; security, justice & foreign policy; local politics, national politics, and the remaining two are other.

For legislative debates, the coherence score advises to use around 12 topics. Note that the number of documents is much higher than in the case of platforms because the number of observations is much higher. Nonetheless, to ensure comparability with the platforms, I regrouped the topics. At the beginning of each legislative session, the President of the Assembly announces the works of the session; this means that each session tends to be about a specific topic. I classify a document as about a specific topic if the highest value refers to that topic.

In the Appendix section B, section G and section H I provide further details on the method, the seed words and the top words for each topic in the platforms, legislative debates and written questions.

5.4 Text similarity

I construct a measure of textual similarity between documents. I create two measures of similarity. Politicians might opt to follow the party line or personalize their campaign. The first creates a similarity measure between a document and an average of all documents from the same party. The second one measures a similarity between a document and an average of all documents from the women from the same party. In a few cases, less than two women from the same party run in the second round; in those situations, I compute the average between women from the same political orientation. I scale this measure to have a standard deviation of one. In these estimates, I build the similarity measure considering manifestos from the respective round, since politicians tend to change their speech between rounds (Le Pennec (2023)).

For any given document vectors x_i and x_j, the cosine similarity is the normalized dot product between the vectors:

$$cosine(x_i, x_j) = \frac{x_i \cdot x_j}{||x_i||||x_j||} \tag{2}$$

5.5 Descriptive statistics

Table 1 presents descriptive statistics relative to gender differences in political platforms. I confirm statistically significant differences at 1% in all the characteristics analyzed in both the first and second rounds (except tone in the second round). I corroborate a vote gender gap in both first and second-round elections; on average, women receive less votes by 6.19 percentage points than males in the second round. Regarding the political platform characteristics, women are, on average, 0.19 standard deviations more at the right than males. Women substantially write more about health & education and security & foreign policy than males, by 0.82 and 0.63 standard deviations. In contrast, they give less salience to national politics by 0.63 standard deviations.

Nonetheless, it is premature to refer to gender differences in political discourse, given that the female presence may be endogenous to perceived voter bias, party bias, or male and female candidates' characteristics. In the following subsection, I explain better the seriousness of endogeneity and my strategy to solve this issue.

In the Appendix Table C3, I also present descriptive statistics about differences between elected and non-elected politicians in the second round by gender. The differences are smaller than the gender differences; furthermore, the sign of the gender differences does not depend, on average, if the focus is on elected or non-elected politicians.

Table 1: Differences between female and male 2nd round political candidates (independently of adversary)

	1st Round			2nd Round		
	Female	Male	Difference	Female	Male	Different
% votes	4.10	7.37	-3.27***	23.19	29.38	-6.19***
Number words	0.14	0.06	0.08***	0.05	-0.43	0.48***
Tone	-0.28	-0.00	-0.28***	0.29	0.33	-0.05
Left right	-0.11	-0.08	-0.03**	0.52	0.33	0.19***
Similar to party	0.23	-0.09	0.32***	0.12	-0.03	0.15***
Similar to women	0.74	0.64	0.10***	0.20	-0.05	0.26***
Economy & employment	0.31	0.03	0.28***	-0.32	-0.48	0.16***
Environment	0.18	0.01	0.17***	-0.10	-0.27	0.17***
Health & education	0.06	-0.10	0.16***	0.89	0.07	0.82***
Security & foreign policy	0.13	-0.08	0.21***	0.59	-0.04	0.63***
Local politics	0.14	-0.02	0.16***	0.13	-0.15	0.28***
National politics	-0.24	-0.07	-0.17***	0.00	0.63	-0.63***
Observations	9177	21911	31088	1275	4721	5996

5.6 Empirical strategy

Women are under-represented in politics. Female under-representation can be due to anti-female party bias: party leaders may field fewer female candidates or place them in unfavorable spots on party lists (Esteve-Volart and Bagues (2012), Lippmann (2021)); or as an alternative hypothesis, even if party leaders are willing to promote female candidates, they may anticipate anti-female voter bias and be less willing to field female candidates in competitive races (Le Barbanchon and Sauvagnat (2022)). Women are less likely to run for office, less likely to believe they are qualified to seek office (Fox and Lawless (2011)), less likely to receive encouragement to run for office (Fox and Lawless (2004), Fox and Lawless (2014)), which might culminate with lower political ambition (Pate and Fox (2018)) and a higher aversion to competition (Preece and Stoddard (2015), Kanthak and Woon (2015)). Other potential explanations for women's under-representation are gender inequalities in campaign finance (Barber et al. (2016)), household income and bread-winning responsibilities (Bernhard, Shames, and Teele (2021)). Previous exposure to female leadership positions leads to changes in voter attitudes, and women more likely to stand for and win elected positions (Beaman, Chattopadhyay, Duflo, Pande, and Topalova (2009), Bhavnani (2009), Baskaran and Hessami (2018)), in addition previous female success might affect the entry of new women into politics (Bhalotra, Clots-Figueras, and Iyer (2018)). All these factors demonstrate that female presence in a race is endogenous, and they also might affect the type of women present in the race and how opponents respond to their presence. A quasi-experiment is necessary to differentiate female and male politicians during campaigns causally or understand how politicians change their campaigns considering their opponents' gender.

To answer these questions, I use a regression discontinuity design and compare districts where a woman barely qualified for the second round with districts where a woman did not qualify by a small margin. At the cutoff, the female presence is orthogonal to voters' and electoral district characteristics.

Throughout the paper, I use several regression discontinuity designs. This section explains the main sharp and fuzzy regression discontinuity design (RDD). For the secondary sharp RDDs that I employ, I will explain the running variable at the beginning of each results section.

5.6.1 Sharp regression discontinuity design

To estimate differences between female and male political platforms in the second round, I use a sharp regression discontinuity design and estimate the following equation:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 R_i + \beta_3 R_i T_i + \beta_4 X_i + \mu_i \tag{3}$$

The treatment variable T is a dummy equal to one if it is a woman and 0 if it is a male in the second round. The running variable R is the qualifying margin of the candidate in the first round. Remember from Section 3.1, in races where only one or no candidate obtained 12.5% of the votes, only the two most voted candidates are eligible for the second round. In this situation, the running variable must be the difference between the most-voted woman and the second-most-voted man. The unit of observation is the candidate, and there is one observation per electoral district.

Identification assumption The validity of the RDD relies on the assumption that first-round candidates of a particular type (e.g., males) do not systematically sort on the right of the qualification threshold. Such manipulation is unlikely since France is a democracy, and international observers qualify the elections as pluralistic, competitive, and respectful of fundamental rights.[14] In addition, manipulation is difficult because it requires predicting the outcome of the first election stage with great accuracy. I test the assumption's validity using the McCrary (2008) test and check if there is a jump in the density of the running variable at the threshold. As Figure C.3 in the Appendix demonstrates, there is no jump at the margin. I also confirm the results with the Cattaneo, Jansson, and Ma (2018) test.

[14] *France. Presidential Election 10 and 24 April 2022. ODIHR Election Assessment Mission Final Report* (2022).

Another implication of the identifying assumption is that districts' characteristics are continuous at the threshold. I run balancing tests for: votes, turnout, number of candidates, number of enrolled voters, victory margin, number of candidates per ideology, and the sum of votes for left and right-wing candidates in the first round. Tables C4 and C5 in the Appendix show the sharp RDD estimates of the effect of having a woman in the second round in alternative to a man on first-round electoral district characteristics. Of the 12 regressions, the coefficient on the treatment variable is not statistically significant in any of them. The pre-treatment characteristics of the districts are balanced.

For completeness, I also run balancing tests for the dependent variables. Of the eleven regressions, five are statistically significant at 10% level (Tables C8 and C9 in the Appendix). The statistically significant regressions are: tone, similar to party, health & education, local and national politics. These results start to provide evidence of gender differences in political campaigns at the gender level. Finding a balance in these characteristics is not relevant in this case since what is necessary to ensure is that the electoral district and voters' characteristics are orthogonal at the cutoff but not the gender differences in campaign discourse. I include in my regressions controls for first-round campaign discourse characteristics.

The analysis can also be affected by endogenous sample selection. A potential concern is if a political platform is more observed for one gender than the other or when a woman is present (or not) in the second round. Column 1 of Table C4 shows that this is not the case, there is no significant jump in the probability of having a first-round manifesto available at the qualification threshold.

A further concern is the confounding effects due to the ideology of candidates. Male and female candidates may systematically differ in their ideology. For instance, female candidates may be mostly from left-wing parties. Hence, the results obtained from estimating Equation (1) may be due to the ideology rather than the gender of the candidate. However, I do not find significant differences in the ideology of female and male candidates around the cutoff (see Table C6 in the Appendix). Di Tella et al. (2023) demonstrate empirically that candidates strategically adjust their platform to get closer to their opponent. It could be that the results are confounded because the opponents of female candidates are mostly from a specific ideology. However, I also do not find significant differences in the ideology of female and male opponents around the cutoff (Table C7 in the Appendix).

5.6.2 Fuzzy regression discontinuity design

I also try to understand if male candidates adapt their political platform to the presence of a female on the race. The running variable in this case is the margin of qualification of the woman. It has two cutoffs since it is possible to qualify for the second round in two different ways. First, all candidates that obtain at least 12.5% of the registered votes qualify for the second round; in this case, in races where at least two male candidates obtained 12.5%, I estimate the margin of a third candidate (woman) obtaining the required threshold. Second, if only one candidate (or none) obtained 12.5%, the two most-voted candidates are eligible for the second round; in this situation, the margin is the difference between the most-voted woman and the second most-voted man. When more than two candidates are selected for the second round, candidates can decide to drop from the race, making the regression discontinuity design fuzzy.

I pull both sets of races to estimate the overall impact of a man running against a woman. The estimated equation is similar to equation (2) but T_i is instrumented with D_i as shown in the following first-stage equation:

$$T_i = \alpha_0 + \alpha_1 D_i + \alpha_2 R_i + \alpha_3 R_i D_i + \alpha_4 X_i + \epsilon_i \tag{4}$$

I call compliers the districts in which a woman qualifies ($D = 1$) and runs in the second round ($T = 1$).

In both specifications (sharp and fuzzy RDD), I follow Calonico, Cattaneo, and Titiunik (2014) and use a non-parametric approach, fitting a local linear regression on each side of the threshold within an optimal bandwidth selected by the MSERD procedure. I use the mean squared error optimal bandwidths selection procedure proposed by Calonico et al. (2014) in its covariate-adjusted version (Calonico, Cattaneo, Farrell, and Titiunik (2019)). This procedure is data-driven, implying that bandwidth size varies with the outcome under consideration.[15] Indeed, X_i is a vector of first-round independent variables.[16]. These controls are not necessary for identification but improve efficiency.

Identifying assumption and robustness. The identifying assumption is that the win margin of qualification cannot be precisely manipulated. To assess its plausibility, I test for the continuity of the density of the running variable. I test the assumption's validity using the McCrary (2008) test and check if there is a jump in the density of the running variable at the threshold. As advised by Cattaneo, Titiunik, Vazquez-Bare, and Keele (2016) in the context of multi-cutoffs, I explore the density of observations around each cutoff. Regarding the margin between the most voted woman and the 12.5%, Figure C.4 in the Appendix demonstrates no jump in any cutoffs. I also confirm the results with the Cattaneo et al. (2018) test.

[15] I use the Stata package rdrobust (Calonico, Cattaneo, Farrell, and Titiunik (2017)).

[16] Control variables are presented in Tables C4, C5, C8 and C9

Relatively to the other cutoff, the results were referred to in the previous subsection (see Figure C.3).

Second, Figure C.5 in the Appendix plots the treatment against the running variable and demonstrates that first stage is very strong.

Third, the main implication of the identifying assumption is that districts' characteristics are continuous at the threshold. I run balancing tests in the same spirit as the sharp RDD. The Appendix Tables C10 through C14 show the results. Of the 29 regressions, I only find statistical significance for two variables. I find that treated males have an ideological score more at the right (variable "Left right" in Table C12, statistically significant at 1%) and have a lower chance of being associated with a left ideology (variable "Left" in Table C14, statistically significant at 5%), which is concerning. Reassuringly, I will run regressions with and without controlling for the "Left right" covariate.

6 Results

6.1 Baseline results: Marginally eligible women versus marginally eligible men

6.1.1 Political platforms: Positioning and topic content

I use my regression discontinuity design strategy to test whether there are gender differences in political platforms, as stated in Section 5.6.1. To understand if women campaign differently from men, I restrict my analysis to races where only two candidates are permitted to pass to the second round, and the margin is the difference between the share of votes between the most voted woman and the second most voted man; in this case, women are the treatment group and men the counterfactual.

Table 2 presents results for the main platforms' characteristics. I test if women diverge ideologically from men in a left-right dimension. Column 3 shows that women are more at the right by 0.270 standard deviations than men.

An important finding is that women personalize their platforms less by 0.462 standard deviations, an estimate that is significant at 1% level (column 6). Several potential interpretations might be given to this result. Personalizing a campaign shows effort and rhetorical skills. On the other hand, it can also mean that women feel a higher need to convey to the party line and prefer to give more emphasis to partisan policy issues rather than the candidate's attachment to the district or personal achievements. Unsurprisingly, I find that women have a speech more similar to the women from the same party by 0.711 standard deviations (column 7) compared to men, a value that is higher than similar to the party, showing that part of the effect comes from the women.

Table 2: Differences between female and male 2nd round political candidates (independently of adversary)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Woman	0.024	-0.193	0.270*	0.544**	0.711***
	(0.126)	(0.146)	(0.155)	(0.218)	(0.244)
Observations	788	788	788	784	743
Eff. number of obs	261	300	243	209	192
Robust p-value	0.783	0.178	0.094	0.014	0.004
Polyn. order	1	1	1	1	1
Bandwidth	3.899	4.458	3.545	3.054	3.083
Outcome mean	-0.158	0.042	0.160	-0.172	-0.207

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All the dependent variables are standardized. In column 1, the outcome is the number of words used in the platform. For column 2, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 3 is explained in 5.2. The methodology to obtain columns 4 and 5 is explained in 5.4.

Furthermore, I also explore whether women write more or in a more positive or negative tone (columns 1 and 2), but I do not find any evidence.

In Table 3 I study differences in topics covered. I conclude that women work more on security & foreign policy by 0.501 standard deviations, topics that traditionally are associated with men. These results are not necessarily surprising. Previous experiments have shown that female candidates can successfully reverse gender stereotypes by portraying themselves as possessing stereotypical masculine traits (Huddy and Terkildsen (1993), Bauer (2017)).[17]

[17] Herrnson, Lay, and Stokes (2003) refer several situations in the U.S. where women emphasized issues traditionally thought of as best handled by men to downplay differences perceived by voters between male and female candidates. "In her race for governor of California in 1990, Dianne Feinstein emphasized her support for the death penalty, and as a vice presidential candidate in 1984, Geraldine Ferraro accentuated her tough stance on crime. Women campaigning for the U.S. Senate

Table 3: Differences between female and male 2nd round political candidates (independently of adversary)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Woman	0.048	-0.047	-0.167	0.501*	-0.323*	0.111
	(0.136)	(0.049)	(0.161)	(0.284)	(0.178)	(0.186)
Observations	788	788	788	788	788	788
Eff. number of obs	285	234	315	203	254	261
Robust p-value	0.632	0.343	0.267	0.128	0.098	0.579
Polyn. order	1	1	1	1	1	1
Bandwidth	4.201	3.423	4.945	2.938	3.770	3.892
Outcome mean	-0.141	-0.167	-0.081	0.122	0.385	0.308

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

The results of this section are robust to a larger bandwidth and are present in the Appendix Tables D15 and D16; the magnitudes are higher in that case.

Heterogeneity I exploit heterogeneity concerning district characteristics and political ideology to detect whether the results are consistent with the conjecture that there are gender differences in political discourse

Several studies have been showing that exposure to female representation can have a powerful effect on how voters perceive women and how women see themselves and their role in society (see, e.g., Beaman et al. (2009), Beaman, Duflo, Pande, and Topalova (2012), Bhavnani (2009), Baskaran and Hessami (2018)). I explore how results change when I restrict my analysis to districts that never elected a woman, at least until that election.[18] Table 4 Panel A displays the platform's main characteristics results. Results are slightly smaller in magnitudes compared to Table 2. Table 5 Panel A presents the results about the topics covered in the platforms. Women give more salience to security & foreign policy than males by 0.606 standard deviations, statistically significant at 1%, a magnitude more prominent than the one found in Table 3. This higher coverage for security & foreign policy issues is substituted by giving a lower salience to local issues by 0.463 standard deviations. Therefore, these results seem to be driven by women believing that

between 1982 and 1986 were more likely than men to appear in their own ads and to dress formally in their commercials to convince voters of their legitimacy and professionalism."

[18] I do not present results for places that previously elected a woman since the number of observations on the margin is lower than 50 observations.

voters prefer male-stereotyped characteristics in their electoral districts.

Table 4: Differences between female and male 2nd round political candidates (independently of adversary)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Panel A: Districts that never elected a woman					
Woman	-0.003	-0.179	0.225	0.473**	0.614**
	(0.127)	(0.163)	(0.152)	(0.230)	(0.262)
Observations	556	556	556	553	519
Eff. number of obs	236	247	215	178	167
Robust p-value	0.995	0.217	0.160	0.069	0.027
Polyn. order	1	1	1	1	1
Bandwidth	5.103	5.479	4.257	3.511	3.549
Outcome mean	-0.300	0.061	0.174	-0.166	-0.209
Panel B: Left candidates					
Woman	0.066	0.254	0.382**	0.456**	0.435**
	(0.186)	(0.208)	(0.180)	(0.216)	(0.208)
Observations	235	235	235	232	219
Eff. number of obs	91	77	112	84	76
Robust p-value	0.899	0.203	0.051	0.026	0.031
Polyn. order	1	1	1	1	1
Bandwidth	3.458	2.833	4.569	3.187	3.066
Outcome mean	-0.272	0.084	0.168	-0.167	-0.213
Panel C: Right candidates					
Woman	0.205	-0.400**	0.143	1.114***	1.228***
	(0.158)	(0.177)	(0.136)	(0.308)	(0.323)
Observations	541	541	541	541	524
Eff. number of obs	154	177	181	106	98
Robust p-value	0.242	0.028	0.355	0.000	0.000
Polyn. order	1	1	1	1	1
Bandwidth	3.858	4.440	4.504	2.796	2.664
Outcome mean	-0.319	0.042	0.170	-0.187	-0.243

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All the dependent variables are standardized.

Table 5: Differences between female and male 2nd round political candidates (independently of adversary)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
	Panel A: Districts that never elected a woman					
Woman	0.055	-0.095*	-0.123	0.606***	-0.463***	0.155
	(0.153)	(0.049)	(0.215)	(0.215)	(0.168)	(0.191)
Observations	556	556	556	556	556	556
Eff. number of obs	228	153	215	215	209	226
Robust p-value	0.603	0.069	0.473	0.018	0.009	0.335
Polyn. order	1	1	1	1	1	1
Bandwidth	4.837	3.026	4.252	4.267	4.100	4.649
Outcome mean	-0.131	-0.158	-0.062	0.096	0.365	0.314
	Panel B: Left candidates					
Woman	-0.110	0.111*	-0.115	0.237***	0.059	0.168
	(0.149)	(0.058)	(0.076)	(0.071)	(0.215)	(0.189)
Observations	235	235	235	235	235	235
Eff. number of obs	84	92	88	92	77	112
Robust p-value	0.470	0.054	0.222	0.003	0.411	0.479
Polyn. order	1	1	1	1	1	1
Bandwidth	3.072	3.541	3.325	3.695	2.727	4.551
Outcome mean	-0.123	-0.169	-0.054	0.108	0.478	0.310
	Panel C: Right candidates					
Woman	0.377*	-0.022	-0.228	0.831*	-0.282	-0.284
	(0.224)	(0.043)	(0.286)	(0.494)	(0.329)	(0.282)
Observations	541	541	541	541	541	541
Eff. number of obs	96	132	181	108	105	116
Robust p-value	0.101	0.694	0.391	0.178	0.530	0.357
Polyn. order	1	1	1	1	1	1
Bandwidth	2.400	3.324	4.523	2.854	2.741	2.990
Outcome mean	-0.104	-0.165	-0.071	0.082	0.470	0.287

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

Considering that women are more likely to vote for left-wing parties (Edlund and Pande (2002), Abendschön and Steinmetz (2014), Gethin, Martínez-Toledano, and Piketty (2022)) and parties respond to their voters' preferences (Griffin and Newman (2005)), there could exist heterogeneity at the party level. In addition, studying heterogeneity at the ideology level permits assessing the findings' external validity and whether the effects are specific to certain parties within the French elections. I present results for left and right-wing politicians in Panel B and C, respectively. Both left and right-wing female politicians follow more the party line than males, although the magnitudes are larger for right-wing women. Left-wing women are 0.382 standard deviations more at the right than left-wing males, and hence, the results in Table 3 are driven by left-wing women. In addition, right-wing women have a more negative discourse by 0.400 standard deviations (significant at 5%), a male-stereotyped characteristic. Finally, Table 5 shows that both left and right-wing women give more salience to security issues & foreign policy than their male colleagues of the same ideology.

6.1.2 Can campaign financing explain gender differences in platforms?

If campaign funds are essential for a politician's success, then large differences in the amount of money that male and female candidates raise might impact how women campaign in their race. A lower amount of funding potentially translates into less funding to hire a team of advisors, implying that this team is smaller and/or of a lower quality. Consequently, women might receive lower-quality advice on the topics of the campaign, what words to choose, and how to personalize their message. For example, women could personalize their platforms less because they have less money to hire campaign advisors who can advise them on what words to choose. Moreover, given that I also have data on funding from the party, it also permits me to understand if there is party bias against women, and part of the results can be explained by less/more support from the party. Previous research has causally shown that there is a gender gap in campaign fundraising for US state legislators (Barber et al. (2016)) and Brazilian mayoral candidates (Brollo and Troiano (2016), Ferraz, Nogueira, and Tavares (2022)).

I test whether there are gender differences in campaign financing in France that potentially can explain gender differences in political platforms. I estimate these differences by employing a sharp regression discontinuity design similar to the one explained in Section 5.6.1. As shown in Table 6, I find no significant gender difference in campaign expenditures or contributions. Moreover, I do not find any evidence of differences between left-wing women and men and right-wing women and men (Table D17 in the Appendix). I also confirm the robustness of my results to a polynomial of order 2 (Table D18).

Table 6: Gender differences in campaign financing

	(1) Total expenditures	(2) Total revenues	(3) Party contribution	(4) Private donations	(5) Personal contribution
Woman	-0.048	-0.025	-0.006	0.064	-0.048
	(0.193)	(0.211)	(0.110)	(0.203)	(0.090)
Observations	1135	1135	1132	1132	1132
Eff. number of obs	475	466	353	382	459
Robust p-value	0.985	0.907	0.788	0.600	0.767
Polyn. order	1	1	1	1	1
Bandwidth	4.746	4.539	3.241	3.639	4.447
Outcome mean	1.179	1.267	0.194	0.346	0.602

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. Each outcome uses the number of registered voters as the denominator.

6.1.3 Electoral outcomes

Campaign information affects voters' beliefs about the candidates, and voters are responsive to information and the different strategies of persuasive communication conveyed by politicians (Kendall, Nannicini, and Trebbi (2015), Cruz, Keefer, Labonne, and Trebbi (2018)). For the sake of completeness, I also explore whether there are gender differences in two electoral outcomes: vote share and election probability. Notice that electoral outcomes are an output, a result of how voters evaluate politicians' campaigns, their policy proposals, and the ideas they convey. At the same time, electoral outcomes also reflect other voters' preferences, such as voters' bias against women. This paper's quasi-experimental design does not permit disentangling the two effects. I could implement a Gelbach (2016) decomposition to measure how gender differences in the platform contribute to the gender vote gap. However, the Gelbach (2016) departs from an OLS regression and not the RDD; moreover, voters might react differently to the same information provided by the two genders.

In Table 7, I present the results. For columns 1-3, the dependent variable is the share of votes received, and for columns 4-6, the dependent variable is a dummy equal to 1 if the candidate was elected, 0 if not. Women do not significantly receive fewer votes than males (column 1), except right-wing women, who receive less 1.150 percentage points, statistically significant at 5%. Women have a 0.201 lower probability of being elected, but right-wing women primarily drive this result. These findings align with the conclusions of Eyméoud and Vertier (2023) for French local elections.

Table 7: Impact of a marginal presence of a woman on male candidates in the 2nd round

	Votes			Win election		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Left	Right	All	Left	Right
Woman	-0.096	0.641	-1.150**	-0.201**	-0.096	-0.235**
	(0.704)	(0.774)	(0.558)	(0.093)	(0.111)	(0.119)
Observations	835	240	582	835	240	582
Eff. number of obs	264	88	181	230	106	132
Robust p-value	0.879	0.395	0.030	0.048	0.473	0.074
Polyn. order	1	1	1	1	1	1
Bandwidth	3.629	3.246	4.015	3.085	4.198	3.051
Outcome mean	20.514	20.498	20.730	0.231	0.242	0.231

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. For columns 4-6, the dependent variable is a dummy equal to 1 if the candidate won the election, 0 if not. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

6.2 Do women adapt to the gender of their opponent?

6.2.1 Platform content: Positioning and topic content

The gender gap in political discourse shown in Tables 2 and 3 might be related to female politicians often running against men. In this paper section, I try to understand how women adapt their discourse when they run against a man instead of a woman. To explore this question, I restrict my sample to races where at least one woman reached the second round. In cases where only one candidate (or none) obtained the 12.5%, the margin is the difference between the most voted man and the second most voted woman; the treatment group is women who compete against a man, and the counterfactual is women who compete against another woman. In other words, I focus my analysis on the most voted woman in the first round and try to understand how this woman in the second round adapts her platform when she runs with a man who barely passed to the second round (instead of a woman) in comparison with races where another woman barely passed to the second round instead of a man.

The validity of the RDD relies on the key assumption that first-round candidates of a particular type (e.g., female candidates) do not systematically sort on the right of the qualification threshold. I implement the tests proposed by McCrary (2008) and Cattaneo et al. (2018) and verify that there is no discontinuity in the density of the running variable at the threshold (Appendix Figure D.6).

The main implication of the identifying assumption is that electoral districts' characteristics are continuous at the threshold. I run balancing tests for first-round election characteristics. Tables D19 to D24 in the Appendix show the results. Considering 34 balance tests, only three covariates are statistically significant. The number of far-right candidates is statistically significant at 5%. Controlling or not controlling for this variable does not significantly change the results. The variables "Similar to party" and "Similar to woman" are statistically significant at 1%, which is worrisome. Controlling for these variables only affects the results of the referred variables.[19]

Table 8 presents the results of political platforms' general characteristics. Table 9 refers to the topics covered. The variable "Similar to party" is not statistically significant; this means that women do not have a lower or higher tendency to personalize their campaign when they run against men. On the other hand, women do not write more (column 1 Table 8) when they compete against a man, nor change the sentiment of their language (column 2 Table 8), nor change their ideological score (column 3).

Conversely, when women run against a man, they talk more about security & foreign policy (column 4 Table 9) by 0.775 standard deviations, a result statistically significant at 1%. At the same time, they do not change the coverage of other topics. Hence, women strategically adapt to the gender of their candidate; expecting female characteristics to be penalized by voters, they opt to focus more in their campaigns on stereotyped male topics. It needs to be highlighted that treated and counterfactual women are not necessarily the same as in Section 6.1 since, in that case, the treated women were the ones who barely passed the first round. This reinforces the external validity of the paper. In the Appendix Tables D25 and D26, I test the robustness of the results to a larger bandwidth.

[19] Results without controlling for the statistically significant covariates are available upon request.

Table 8: Impact of a marginal presence of a man on female candidates in the 2nd round

	(1) No. words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Treatment	0.118	-0.180	0.259	0.349	0.208
	(0.180)	(0.143)	(0.162)	(0.216)	(0.199)
Observations	465	465	465	461	453
Eff. number of obs	139	147	176	158	176
Robust p-value	0.532	0.173	0.097	0.110	0.298
Polyn. order	1	1	1	1	1
Bandwidth	3.060	3.262	3.893	3.441	3.957
Outcome mean	0.165	0.282	0.466	0.192	0.273

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman competing in the second round with a man instead of a woman. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All the dependent variables are standardized. In column 1, the outcome is the number of words used in the platform. For column 2, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 3 is explained in 5.2. The methodology to obtain columns 4 and 5 is explained in 5.4.

Table 9: Impact of a marginal presence of a man on female candidates in the 2nd round

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Treatment	0.059	-0.010	0.445	0.775***	-0.379*	-0.167
	(0.120)	(0.038)	(0.291)	(0.287)	(0.214)	(0.195)
Observations	465	465	465	465	465	465
Eff. number of obs	131	145	152	131	135	147
Robust p-value	0.421	0.837	0.157	0.029	0.122	0.329
Polyn. order	1	1	1	1	1	1
Bandwidth	2.926	3.214	3.329	2.914	2.979	3.252
Outcome mean	-0.310	-0.103	1.264	0.696	0.486	-0.290

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman competing in the second round with a man instead of a woman. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

6.2.2 Campaign financing

Just like women adapt their political platforms to the gender of their opponent, changes in campaign financing might happen. For example, parties might provide less funding to women when they compete against men because they expect them to have lower chances of winning and prefer to focus on safer races. Another hypothesis is that donors discriminate against women, and when they can choose between the two genders, they prefer to donate to men. These situations might impact the content of political platforms. For instance, women might adapt their campaign discourse after noticing that their male opponent receives more private donations and locals prefer a male MP.

I explore whether the gender of a woman's opponent impacts their funding. Results are displayed in Table 10. I find that if a woman competes against a man, in alternative to a woman, they spend on average 1.208 euros less per voter (column 1), and their total revenues are 1.717 euros lower per voter (column 2). Fewer party contributions or private donations cannot explain the difference in total revenues, meaning that parties do not support less women when they compete against a man or donors discriminate against them. On the contrary, it is explained by women using less 0.158 euros per voter of their personal contribution (column 5, statistically significant at 1%). Considering that an electoral district, on average, has around 83 322 voters, an envelope calculation estimates a reduction in personal contributions by 13 165 euros. This result is aligned with the previous finding of women changing the salience of their topics in their campaigns; women fearing voter bias change the coverage of the topics in the campaign and invest less of their own money.

6.3 Do men adapt to the presence of a woman in the race?

Politicians' discourse and policy proposals follow voters' preferences in the median voter theorem. Di Tella et al. (2023) demonstrate empirically that convergence to the center results from candidates adjusting to the rival they will face in the second round. In Section 6.2.1, I find that when women run against a man instead of a woman, they focus on male-stereotyped topics significantly. Suppose the median voter theorem can predict politicians' performance in identity situations, when competing with an opponent of the opposing sex. In that case, one should expect male politicians to focus more on female-stereotyped topics, such as the environment, health & education.

In this section of the paper, I try to understand how male candidates adapt to the presence of a woman who was barely eligible to participate in the second

Table 10: Female campaign financing when a woman competes with a man instead of a woman

	(1) Total expenditures	(2) Total revenues	(3) Party contribution	(4) Private donations	(5) Personal contribution
Treatment	-1.208*	-1.717*	-0.003	-0.176	-0.158***
	(0.628)	(0.882)	(0.020)	(0.176)	(0.060)
Observations	554	554	554	554	554
Eff. number of obs	133	137	122	126	156
Robust p-value	0.109	0.088	0.471	0.325	0.002
Polyn. order	1	1	1	1	1
Bandwidth	2.578	2.676	2.419	2.478	3.034
Outcome mean	0.238	0.251	0.017	0.064	0.072

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. Each outcome uses the number of registered voters as the denominator.

round. I perform a fuzzy regression discontinuity design and use the two cutoffs that permit eligibility. The methodology is explained in Section 5.6.2.

Table 11 presents the results relative to the main platform's characteristics. and Table 12 relative to the topics. Panel A shows the results for all candidates. In general, I do not find that male politicians adapt to the presence of a woman since none of the platforms' outcomes are statistically significant. I explore whether these results are significantly masked by heterogeneity. In Panel B, I show the results for left-wing politicians and find that when male politicians compete with a woman in the race, they write more by 1.085 standard deviations, a finding statistically significant at 1%. Compared to races without a woman, they give more salience to economy & employment by 0.662 standard deviations and less to security & foreign policy by 0.411 standard deviations, statistically significant at 5 and 1%, respectively. I do not find that left-wing male politicians adapt in other forms. In addition, I also do not find that right-wing male politicians adapt to the presence of a woman in the race, given that none of the outcomes are statistically significant.

Table 11: Impact of a marginal presence of a woman on male candidates in the 2nd round

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
	Panel A: All candidates				
Treatment	0.016	-0.181	-0.113	0.200	0.174
	(0.129)	(0.151)	(0.090)	(0.137)	(0.150)
Observations	2495	2495	2495	2429	2252
Eff. number of obs	614	485	559	597	508
Robust p-value	0.967	0.223	0.252	0.252	0.400
Polyn. order	1	1	1	1	1
Bandwidth	4.485	3.624	4.119	4.422	4.155
Outcome mean	-0.180	0.216	0.306	-0.070	0.696
	Panel B: Left candidates				
Treatment	1.085***	-0.381	-0.236	-0.392	-0.618
	(0.394)	(0.335)	(0.195)	(0.410)	(0.376)
Observations	923	923	923	918	748
Eff. number of obs	110	94	107	78	62
Robust p-value	0.004	0.280	0.156	0.210	0.060
Polyn. order	1	1	1	1	1
Bandwidth	2.707	2.415	2.676	2.027	2.119
Outcome mean	-0.099	0.251	0.300	-0.091	0.704
	Panel B: Right candidates				
Treatment	-0.182	-0.140	-0.136	0.207	0.277
	(0.163)	(0.155)	(0.089)	(0.147)	(0.178)
Observations	1561	1561	1561	1501	1504
Eff. number of obs	365	437	420	398	323
Robust p-value	0.223	0.458	0.144	0.193	0.120
Polyn. order	1	1	1	1	1
Bandwidth	3.854	4.591	4.371	4.237	3.410
Outcome mean	-0.169	0.247	0.317	-0.075	0.702

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. The dependent variables of columns 2-8 are standardized. In column 1, the outcome is the number of votes obtained divided by the number of voters enrolled in the district. In column 2, the outcome is the number of words used in the manifesto. For columns 3 and 4, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 5 is explained in 5.2. The methodology to obtain columns 6-9 is explained in 5.4.

Table 12: Impact of a marginal presence of a woman on male candidates in the 2nd round

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
			Panel A: All candidates			
Treatment	0.077	-0.031	-0.080	0.080	-0.027	0.065
	(0.103)	(0.026)	(0.110)	(0.133)	(0.121)	(0.163)
Observations	2495	2495	2495	2495	2495	2495
Eff. number of obs	586	487	683	765	621	581
Robust p-value	0.385	0.169	0.534	0.649	0.987	0.884
Polyn. order	1	1	1	1	1	1
Bandwidth	4.292	3.683	5.046	5.671	4.597	4.262
Outcome mean	-0.347	-0.210	0.554	0.064	0.204	0.243
			Panel B: Left candidates			
Treatment	0.662**	-0.071	-0.147*	-0.411***	0.096	0.104
	(0.337)	(0.089)	(0.089)	(0.104)	(0.244)	(0.315)
Observations	923	923	923	923	923	923
Eff. number of obs	86	87	134	76	99	120
Robust p-value	0.040	0.323	0.129	0.000	0.508	0.898
Polyn. order	1	1	1	1	1	1
Bandwidth	2.204	2.237	3.688	1.959	2.551	3.140
Outcome mean	-0.348	-0.198	0.589	-0.013	0.287	0.252
			Panel C: Right candidates			
Treatment	-0.024	-0.037	-0.112	0.089	0.011	0.118
	(0.109)	(0.025)	(0.159)	(0.173)	(0.139)	(0.208)
Observations	1561	1561	1561	1561	1561	1561
Eff. number of obs	402	300	420	535	460	344
Robust p-value	0.947	0.127	0.559	0.695	0.792	0.676
Polyn. order	1	1	1	1	1	1
Bandwidth	4.199	3.068	4.360	5.677	4.891	3.567
Outcome mean	-0.345	-0.203	0.570	0.064	0.183	0.237

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

All the referred results are robust to several specifications. Given the unbalance in the first round in the variable "Left right", I compare the results controlling for the variable (as presented here) and without controlling (Appendix Table D27), and do not find significant differences. Second, I test the robustness of the results to a polynomial of order 2 (Appendix Tables D28 and D29). Importantly, I also restrict my analysis to races with only two candidates in the second round. Results are similar for the majority of the outcomes, with some exceptions. The magnitudes of the variables economy & employment and security & foreign policy are considerably smaller; plus, right-wing politicians move to the left by 0.183 standard deviations, statistically significant at 10%.

In summary, the evidence that female politicians adapt to the gender of their opponent is much stronger than males adapting to the presence of a woman on the race.

6.4 After elections: Do elected females debate differently from elected males?

In this paper section, I compare women's behavior during campaigns with their behavior once elected. I adopt a standard sharp regression discontinuity design to estimate the causal impact of gender on participation in legislative debates and the topics focused on. Focusing on mixed-gender elections, I use the female margin of victory as a forcing variable in the sharp RD design. I compute the female margin of victory as the difference in the vote share of the female and the male candidates relative to the share of votes obtained by both. This method has been widely used in previous research (e.g., Bhalotra et al. (2018), Casarico, Lattanzio, and Profeta (2022), Chauvin and Tricaud (2023)).

As referred to for the other RDDs, the identification assumption is that all candidate characteristics change continuously around the threshold and, therefore, that the only discrete change occurring at this threshold is the shift in the gender of the MP. Sorting of candidates across the discontinuity threatens the validity of this assumption if it occurs precisely at the cutoff. To bring empirical support for the identification assumption, I check if there is a jump in the density of the running variable at the threshold, using the McCrary (2008) and Cattaneo et al. (2018) manipulation tests. I do not find any evidence of manipulation at the threshold (Figure E.7 in the Appendix). I also conducted 12 balance tests to bring empirical support to the identifying assumption that districts' characteristics are continuous at the threshold (Appendix Tables E32 and E33). Two are statistically significant at 5%: number of candidates and number of left candidates. Controlling or not controlling for these covariates does not substantially change the results.

Table 13 presents the results. In Panel A, I show results for all politicians; in Panel B and C, I include results for left and right-wing politicians, respectively, to confirm that a specific type of politician does not drive the results. First of all, women participate in legislative debates as much as men, an activity that favors charismatic politicians and requires good rhetorical skills. Right-wing women participate less in legislative debates by 0.189 standard deviations, but this result is not statistically significant if I use a larger bandwidth.

Table 13: Differences between female and male MPs during parliamentary work - legislative debates

	(1) Debates	(2) Economy & employment	(3) Environment	(4) Health & education	(5) Security & foreign policy	(6) Local politics
			Panel A: All politicians			
Woman	-0.063	0.007	-0.007	0.720***	-0.205	-0.100
	(0.124)	(0.163)	(0.154)	(0.189)	(0.159)	(0.169)
Observations	1319	1319	1319	1319	1319	1319
Eff. number of obs	548	648	816	478	679	560
Robust p-value	0.660	0.919	0.985	0.000	0.183	0.580
Polyn. order	1	1	1	1	1	1
Bandwidth	5.019	6.020	7.915	4.265	6.434	5.138
Outcome mean	0.028	0.004	-0.021	0.061	-0.016	-0.050
			Panel B: Left politicians			
Woman	0.051	-0.047	0.136	1.015***	-0.331	-0.477**
	(0.229)	(0.304)	(0.232)	(0.336)	(0.282)	(0.230)
Observations	501	501	501	501	501	501
Eff. number of obs	209	272	232	151	221	242
Robust p-value	0.817	0.764	0.492	0.003	0.249	0.074
Polyn. order	1	1	1	1	1	1
Bandwidth	5.993	8.316	6.804	4.062	6.553	7.140
Outcome mean	0.028	0.004	-0.021	0.061	-0.016	-0.050
			Panel C: Right politicians			
Woman	-0.189*	-0.061	-0.057	0.659**	-0.087	0.123
	(0.109)	(0.195)	(0.226)	(0.256)	(0.259)	(0.240)
Observations	794	794	794	794	794	794
Eff. number of obs	311	423	426	350	356	308
Robust p-value	0.133	0.979	0.776	0.024	0.712	0.720
Polyn. order	1	1	1	1	1	1
Bandwidth	4.291	5.891	5.958	4.947	5.009	4.244
Outcome mean	0.028	0.004	-0.021	0.061	-0.016	-0.050

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman being elected in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

In conformity to the findings of the literature (Hessami and da Fonseca (2020), Lippmann (2021)), I show that female politicians significantly work more in topics of health & education (column 4), by 0.720 standard deviations, a finding significant at 1%. I test the robustness of these results in two ways. First, to a polynomial of order two (Appendix Table E34). Second, I explore whether this pattern is only observed for legislative debates. Written questions to the government are the most constituency-focused activity a MP can engage in. In the Appendix Table E35, I also show that women work more on health & education than males by a significant magnitude, 0.537 standard deviations and statistically significant at 1%.

In addition, left-wing women work less on local issues (column 6) by 0.477 standard deviations. Relatively to other topics, no statistically significant differences exist between female and male MPs.

These results contrast with the findings during elections. Therefore, during campaigns, women strategically do not work on their preferred topics but prefer to focus on topics that traditionally are associated with men to increase their chances of winning. One can also argue that women in the elections signal their true type, meaning they prefer to work more on security & foreign policy. However, once elected, their peers from their party pressure them to work on health & education because voters find them more legitimate to work on these topics. Such a hypothesis seems unlikely since it goes against significant literature. Several papers demonstrate that local female leaders and voters who are more independent in defining their policies than MPs prefer to invest in public goods, particularly in the areas of health and education (Chattopadhyay and Duflo (2004), Holman (2014), Funk and Gathmann (2015)).

7 Conclusion

Several prestigious professions are male-dominated environments. In politics, women face more difficulties in succeeding. In this paper, I provide the first causal evidence of gender differences in political campaigns. I explore the case of French legislative elections for which individual political platforms are available from 1981 to 1997 and 2017-22 for both first and second-round races.

In order to isolate the causal impact of gender on campaigns, I implement a regression discontinuity design and compare female candidates who narrowly were eligible for the second round against a male candidate. I causally show that women write a platform more similar to their party and personalize less their political campaigns. Notably, female candidates advertise more on security & foreign policy than men, left-wing women are ideologically at the right of left-wing men, and right-wing women write in a more negative tone, characteristics that are stereotyped as male traits. I prove that given that there are no gender differences in campaign financing, they cannot explain differences in political platforms.

I explore whether women consider the gender of their opponents' when writing their platforms. I compare women who competed against a barely eligible man with women who competed against a woman who barely passed to the second round instead of a man. When women run against a man, they do not have a higher chance of personalizing their platform, but they significantly increase the coverage of security & foreign policy topics. Women strategically adapt to their opponents and adopt stereotyped male traits to increase their chances of winning. Moreover, when women compete against a man instead of a woman, they use fewer personal resources to fund their campaigns, reinforcing the idea that women feel less confident in competing against men.

On the contrary, evidence that male politicians adapt to the presence of a woman in the second round is less clear and heterogeneous. I conduct a regression discontinuity design and compare races where a woman was barely eligible for the second round with races where she was not present. Only left-wing politicians seem to adapt by writing more, and the topics that they prefer to focus on depend on the number of candidates in the race. Right-wing politicians do not change their platforms.

Once elected, women participate in debates as much as men but focus more on health & education topics by a large magnitude. Again, I prove that focusing on stereotyped male topics during campaigns is a strategic behavior to prevent voter discrimination. Furthermore, women participate in legislative debates as much as men, a task that requires good rhetorical abilities.

The paper's research design does not permit us to understand if women's campaign strategy pays off or if, alternatively, implementing a different campaign would permit electing more women to the office. Future research should try to understand how gender differences in campaigns explain the gender vote gap. This research should also pay special attention to understanding if voters react differently to similar messages conveyed by women instead of men. Understanding how voters react to gender in campaign information and how this reflects in votes is crucial to advising future female politicians and helping them be elected.

References

Abendschön, S., & Steinmetz, S. (2014). The gender gap in voting revisited: Women's party preferences in a european context. *Social Politics*, *21*(2), 315–344.

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

Ash, E., Krümmel, J., & Slapin, J. B. (2023). Better to be jeered than ignored? gender and reactions during parliamentary debates. *Center for Law & Economics Working Paper Series*, *2*.

Barber, M., Butler, D. M., Preece, J., et al. (2016). Gender inequalities in campaign finance. *Quarterly Journal of Political Science*, *11*(2), 219–48.

Baskaran, T., & Hessami, Z. (2018). Does the election of a female leader clear the way for more women in politics? *American Economic Journal: Economic Policy*, *10*(3), 95–121.

Bauer, N. M. (2017). The effects of counterstereotypic gender strategies on candidate evaluations. *Political Psychology*, *38*(2), 279–295.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, *124*(4), 1497–1540.

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *science*, *335*(6068), 582–586.

Bekkouche, Y., Cagé, J., & Dewitte, E. (2022). The heterogeneous price of a vote: Evidence from multiparty systems, 1993–2017. *Journal of Public Economics*, *206*, 104559.

Bernhard, R., Shames, S., & Teele, D. L. (2021). To emerge? breadwinning, motherhood, and women's decisions to run for office. *American Political Science Review*, *115*(2), 379–394.

Besley, T., & Coate, S. (1997). An economic model of representative democracy. *The quarterly journal of economics*, *112*(1), 85–114.

Bhalotra, S., Clots-Figueras, I., & Iyer, L. (2018). Pathbreakers? women's electoral success and future political participation. *The Economic Journal*, *128*(613), 1844–1878.

Bhavnani, R. R. (2009). Do electoral quotas work after they are withdrawn? evidence from a natural experiment in india. *American Political Science Review*, *103*(1), 23–35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of* *machine Learning research*, *3*(Jan), 993–1022.

Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, *81*(2), 542–555.

Brollo, F., & Troiano, U. (2016). What happens when a woman wins an election? evidence from close races in brazil. *Journal of Development Economics*, *122*, 28–45.

Cagé, J., Le Pennec, C., & Mougin, E. (n.d.). Firm donations and political rhetoric: Evidence from a national ban. *American Economic Journal: Economic Policy*.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, *101*(3), 442–451.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Casarico, A., Lattanzio, S., & Profeta, P. (2022). Women and local public finance. *European Journal of Political Economy*, *72*, 102096.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), 234–261.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., & Keele, L. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics*, *78*(4), 1229–1248.

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica*, *72*(5), 1409–1443.

Chauvin, J. P., & Tricaud, C. (2023). *Gender and electoral incentives: Evidence from crisis response*. Centre for Economic Policy Research.

Cruz, C., Keefer, P., Labonne, J., & Trebbi, F. (2018). *Making policies matter: Voter responses to campaign promises* (Tech. Rep.). National Bureau of Economic Research.

Curini, L., & Vignoli, V. (2021). Committed moderates and uncommitted extremists: Ideological leaning and parties' narratives on military interventions in italy. *Foreign Policy Analysis*, *17*(3), orab016.

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, *51*(1), 816–835.

De Paola, M., Scoppa, V., & Lombardo, R. (2010). Can gender quotas break down negative stereotypes? evidence from changes in electoral rules. *Journal of Public Economics*, *94*(5-6), 344–353.

Di Tella, R., Kotty, R., Le Pennec, C., & Pons, V. (2023). Keep your enemies closer: Strategic platform adjustments during u.s. and french elections. *NBER Working Paper*(31503).

Djourelova, M. (2023). Persuasion through slanted language: Evidence from the media coverage of immigration. *American Economic Review*, *113*(3), 800–835.

Djourelova, M., Durante, R., & Martin, G. (2021). The impact of online competition on local newspapers: Evidence from the introduction of craigslist.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, *65*(2), 135–150.

Duval, D., & Pétry, F. (2016). L'analyse automatisée du ton médiatique: construction et utilisation de la version française du lexicoder sentiment dictionary. *Canadian Journal of Political Science/Revue canadienne de science politique*, *49*(2), 197–220.

Edlund, L., & Pande, R. (2002). Why have women become left-wing? the political gender gap and the decline in marriage. *The Quarterly Journal of Economics*, *117*(3), 917–961.

Esteve-Volart, B., & Bagues, M. (2012). Are women pawns in the political game? evidence from elections to the spanish senate. *Journal of public Economics*, *96*(3-4), 387–399.

Eyméoud, J.-B., & Vertier, P. (2023). Gender biases: evidence from a natural experiment in french local elections. *Economic Policy*, *38*(113), 3–56.

Ferraz, C., Nogueira, C., & Tavares, J. (2022). Does gender-targeted electoral funding work? evidence from brazil.

Fox, R. L., & Lawless, J. L. (2004). Entering the arena? gender and the decision to run for office. *American Journal of Political Science*, *48*(2), 264–280.

Fox, R. L., & Lawless, J. L. (2011). Gendered perceptions and political candidacies: A central barrier to women's equality in electoral politics. *American Journal of Political Science*, *55*(1), 59–73.

Fox, R. L., & Lawless, J. L. (2014). Uncovering the origins of the gender gap in political ambition. *American Political Science Review*, *108*(3), 499–519.

Fraccaroli, N., & Pizzigolotto, A. (2021). Credit shocks and populism.

France. presidential election 10 and 24 april 2022. odihr election assessment mission final report. (2022).

Funk, P., & Gathmann, C. (2015). Gender gaps in policy making: Evidence from direct democracy in switzerland. *Economic Policy*, *30*(81), 141–181.

Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics*, *5*, 529–542.

Gaultier-Voituriez, O. (2016). Archelec, les archives électorales de la v e république, du papier au numérique. *Histoire@ Politique*, *3*, 213–220.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics*, *34*(2), 509–543.

Gennaro, G., & Ash, E. (2022). Emotion and reason in political language. *The Economic Journal*, *132*(643), 1037–1059.

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

Gethin, A., Martínez-Toledano, C., & Piketty, T. (2022). Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020. *The Quarterly Journal of Economics*, *137*(1), 1–48.

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, *77*(5), 1637–1664.

Gonzalez-Eiras, M., & Sanz, C. (2021). Women's representation in politics: The effect of electoral systems. *Journal of Public Economics*, *198*, 104399.

Granzier, R., Pons, V., & Tricaud, C. (2023). Coordination and bandwagon effects: How past rankings shape the behavior of voters and candidates. *American Economic Journal: Applied Economics*, *15*(4), 177–217.

Griffin, J. D., & Newman, B. (2005). Are voters better represented? *The Journal of Politics*, *67*(4), 1206–1227.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the fomc: A computational linguistics approach. *The Quarterly Journal of Economics*, *133*(2), 801–870.

Herrnson, P. S., Lay, J. C., & Stokes, A. K. (2003). Women running "as women": Candidate gender, campaign issues, and voter-targeting strategies. *The Journal of Politics*, *65*(1), 244–255.

Hessami, Z., & da Fonseca, M. L. (2020). Female political representation and substantive effects on policies: A literature review. *European Journal of Political Economy*, *63*, 101896.

Holman, M. R. (2014). Sex and the city: Female leaders and spending on social welfare programs in us municipalities. *Journal of Urban Affairs*, *36*(4), 701–715.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*(153), 41-57.

Huddy, L., & Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American journal of political science*, 119–147.

Jolly, S., Bakker, R., Hooghe, L., Marks, G., Polk, J., Rovny, J., ... Vachudova, M. A.

(2022). Chapel hill expert survey trend file, 1999–2019. *Electoral studies*, *75*, 102420.

Kanthak, K., & Woon, J. (2015). Women don't run? election aversion and candidate entry. *American journal of political science*, *59*(3), 595–612.

Kendall, C., Nannicini, T., & Trebbi, F. (2015). How do voters respond to information? evidence from a randomized campaign. *American Economic Review*, *105*(1), 322–353.

Le Barbanchon, T., & Sauvagnat, J. (2022). Electoral competition, voter bias, and women in politics. *Journal of the European Economic Association*, *20*(1), 352–394.

Le Pennec, C. (2023). Strategic campaign communication: Evidence from 30,000 candidate manifestos. *The Economic Journal*, uead082.

Lippmann, Q. (2021). Are gender quotas on candidates bound to be ineffective? *Journal of Economic Behavior & Organization*, *191*, 661–678.

Loaec, M. (2022). Législatives: comment les partis désignent-ils leurs candidats? *TF1 Info*. Retrieved 2022-10-04, from `https://www.tf1info.fr/politique/elections-legislatives-2022-comment-les-partis-lfi-ps-lr-larem-designent-ils-leurs-candidats-2214390.html`

Lu, B., Ott, M., Cardie, C., & Tsou, B. K. (2011). Multi-aspect sentiment analysis with topic models. In *2011 ieee 11th international conference on data mining workshops* (pp. 81–88).

Markowsky, E., & Beblo, M. (2022). When do we observe a gender gap in competition entry? a meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, *198*, 139–163.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, *142*(2), 698–714.

Murray, R. (2010). *Parties, gender quotas and candidate selection in france*. Springer.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, *122*(3), 1067–1101.

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, *3*(1), 601–630.

Osborne, M. J., & Slivinski, A. (1996). A model of political competition with citizen-candidates. *The Quarterly Journal of Economics*, *111*(1), 65–96.

Pate, J., & Fox, R. (2018). Getting past the gender gap in political ambition. *Journal of Economic Behavior & Organization*, *156*, 166–183.

Piolat, A., Booth, R. J., Chung, C. K., Davids, M., & Pennebaker, J. W. (2011). La version française du dictionnaire pour le liwc: modalités de construction et exemples d'utilisation. *Psychologie française*, *56*(3), 145–159.

Preece, J., & Stoddard, O. (2015). Why women don't run: Experimental evidence on gender differences in political competition aversion. *Journal of Economic Behavior & Organization*, *117*, 296–308.

Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining—beware the same sex! *Experimental Economics*, *12*, 318–331.

Taddy, M. (2013). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, *108*(503), 755–770.

Taddy, M. (2015). Distributed multinomial regression. *The Annals of Applied Statistics*, *9*(3), 1394–1414.

Taddy, M., & Taddy, M. M. (2022). Package 'distrom'.

Watanabe, K., & Baturo, A. (2023). Seeded sequential lda: A semi-supervised algorithm for topic-specific analysis of sentences. *Social Science Computer Review*, 08944393231178605.

Watanabe, K., Xuan-Hieu, P., & Watanabe, M. K. (2023). Package 'seededlda'.

Weigel, J. L. (2020). The participation dividend of taxation: How citizens in congo engage more with the state when it tries to tax them. *The Quarterly Journal of Economics*, *135*(4), 1849–1903.

A Ideological classification

I allocate candidates into seven political orientations (far-left, left, liberal, right, far-right and other). I use the party classifications from the Chapel Hill Expert Survey (CHES) and when missing from the ParlGov and Granzier, Pons, and Tricaud (2023). I consider the party positioning on a scale between 0 (left) and 1 (right) and family classification.

1981 Parliamentary Elections	
Political label	**Political orientation**
Parti Communiste Français	Far-left
Divers Droite	Right
Divers Gauche	Left
Ecologistes	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
Indépendants	Other
Non Classés	Other
Rassemblement pour la République	Right
Socialistes	Left
Union pour la Démocratie Française	Right

1988 Parliamentary Elections	
Political label	**Political orientation**
Parti Communiste Français	Far-left
Divers Droite	Right
Ecologistes	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
Front National	Far-right
Majorité Présidentielle	Left
Radicaux de Gauche	Left
Régionalistes	Other
Rassemblement pour la République	Right
Socialistes	Left
Union pour la Démocratie Française	Right

1993 Parliamentary Elections	
Political label	**Political orientation**
Parti Communiste Français	Far-left
Divers	Other
Divers Droite	Right
Extrême Droite	Far-right
Extrême Gauche	Far-left
Front National	Far-right
Génération Ecologie	Left
Majorité Présidentielle	Left
Radicaux de Gauche	Left
Régionalistes	Other
Rassemblement pour la République	Right
Parti Socialiste	Left
Union pour la Démocratie Française	Right
Les Verts	Left

1997 Parliamentary Elections	
Political label	**Political orientation**
Parti Communiste Français	Far-left
Divers	Other
Divers Droite	Right
Divers Gauche	Left
Ecologistes	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
Front National	Far-right
Mouvement Pour la France	Far-right
Parti Radical Socialiste	Left
Rassemblement pour la République	Right
Socialistes	Left
Union pour la Démocratie Française	Right

2002 Parliamentary Elections	
Political label	**Political orientation**
Communistes	Far-left
Chasse, Pêche, Nature et Traditions	Right
Divers	Other
Démocratie Libérale	Right
Divers Droite	Right
Divers Gauche	Left
Ecologistes	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
Front National	Far-right
Ligue Communiste Révolutionnaire	Far-left
Lutte Ouvrière	Far-left
Mouvement des Citoyens	Left
Mouvement National Républicain	Far-right
Mouvement pour la France	Right
Pôle Républicain	Left
Radicaux de Gauche	Left
Régionalistes	Other
Rassemblement pour la France	Right
Socialistes	Left
Union pour la Démocratie Française	Liberal
Union pour un Mouvement Populaire	Right
Les Verts	Left

2007 Parliamentary Elections	
Political label	**Political orientation**
Communistes	Far-left
Chasse, Pêche, Nature et Traditions	Right
Divers	Other
Divers Droite	Right
Divers Gauche	Left
Ecologistes	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
Front National	Far-right
Majorité Présidentielle	Right
Mouvement pour la France	Right
Radicaux de Gauche	Left
Régionalistes	Other
Rassemblement pour la France	Right
Socialistes	Left
Union pour la Démocratie Française - Mouvement Démocrate	Liberal
Union pour un Mouvement Populaire	Right
Les Verts	Left

2012 Parliamentary Elections	
Political label	**Political orientation**
Alliance Centriste	Liberal
Autres	Other
Centre pour la France	Liberal
Communistes	Far-left
Divers Droite	Right
Divers Gauche	Left
Ecologistes	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
Front de Gauche	Far-left
Front National	Far-right
Nouveau Centre	Liberal
Parti Radical	Right
Radicaux de Gauche	Left
Régionalistes	Other
Socialistes	Left
Union pour un Mouvement Populaire	Right
Europe Ecologie - Les Verts	Left

2017 Parliamentary Elections	
Political label	**Political orientation**
Parti Communiste Français	Far-left
Debout la France	Far-right
Divers	Other
Divers Droite	Right
Divers Gauche	Left
Europe Écologie les Verts	Left
Extrême Droite	Far-right
Extrême Gauche	Far-left
France Insoumise	Far-left
Front National	Far-right
Les Républicains	Right
Modem	Liberal
Radicaux de Gauche	Left
Régionalistes	Other
République en Marche	Liberal
Parti Socialiste	Left
Union des Démocrats et Indépendants	Liberal

2022 Parliamentary Elections	
Political label	**Political orientation**
Divers	Other
Divers Centre	Other
Divers Droite	Right
Divers Gauche	Left
Divers Extrême-droite	Far-right
Divers Extrême-gauche	Far-left
Droite Souveraniste	Far-right
Écologiste	Left
Ensemble	Liberal
Extrême-droite	Far-right
Les Républicains	Right
Nupes	Left
Radicaux de Gauche	Left
Réconquête	Far-right
Régionalistes	Other
Rassemblement National	Far-right
Union des Démocrats et Indépendants	Liberal

B Data

B.1 Political platforms

I transform the pdf versions of the manifestos into text using optical character recognition: Tesseract.

Table B1: Sampling frame

	First round			Second round		
Year	Races	Total candidates	Platforms collected	Races	Total candidates	Platforms collected
1981	491	2644	2452	333	658	649
1988	577	2820	2585	452	893	830
1993	577	5180	4071	490	977	956
1997	577	6205	2851	553	1170	1068
2017	577	6714	4666	482	964	741
2022	577	4990	3932	454	909	828

Notes: The table indicates the number of races, total candidates, and the manifestos collected for each legislative election included in the data set.

Figure B.1: Florence Blatrix-Contat political platform during 1st round 2017 legislative elections - 1st page

Source: RegardsCitoyens (https://github.com/regardscitoyens)

Translation: Florence BLATRIX-CONTAT Your deputy Michel FONTAINE deputy On the left to make France succeed

Madam, Sir, On May 7, the French people chose a new President of the Republic. The legislative elections of June 11 and 18 will decide the future of France; they will be an opportunity to choose the Republic we want. I am standing for election with the determination to make France a success, to make this five-year term a success. The majority resulting from this vote will have to act for social justice, solidarity, equal opportunities and ecological transition. For this, the left must be strong in the National Assembly. Tomorrow, in the Assembly, I will present a constructive and demanding left. I will ensure the defense of social achievements, respect for social dialogue and will fight for a growth model compatible with the preservation of the planet. With my substitute, Michel FONTAINE, we live, work and invest in this constituency. We will put our experience at your service to develop in this territory: employment, public services, health, solidarity, but also the cultural and sporting activities essential to social ties. As a rural elected official, I know how much public services must be preserved in each municipality. On the ground as in the Assembly, I will put all my energy to accompany and support the projects of our territory and to help those who encounter difficulties. Sunday, June 11, I am counting on your support. You can count on my determination and dedication.

51 years old, married, 3 children I live in my native village in Drom, in the heart of Revermont. My farming grandparents and my parents passed on to me their attachment to these lands; above all, they taught me that school was a means of emancipation and social advancement. After studying accounting, I became a teacher. Associate of economics-management, I teach economics and law. Elected since 1995 in my municipality, then in 2015 in the Auvergne-Rhône-Alpes Region, I am convinced that public action and the will make it possible to to advance. In our communities, I am committed to working with everyone, beyond partisan divisions; it is thanks to the gathering of elected officials that projects can succeed.

Figure B.2: Florence Blatrix-Contat political platform during 1st round 2017 legislative elections - 2nd page

Michel FONTAINE

Premier-adjoint de la Ville de Bourg-en-Bresse Président de l'Agglomération de 2008 à fin 2016 et depuis Vice-président au développement économique.

Je vis sur Bourg-en-Bresse depuis plus de 40 ans, j'ai partagé ma vie professionnelle entre le lycée Carriat et ma société FontainePicard.

J'ai deux enfants et vis maritalement. J'ai assumé des responsabilités associatives avant de devenir un élu local.

En m'engageant aux côtés de Florence Blatrix-Contat, je fais le choix de la réussite dans une grande fidélité à mes convictions.

NATIONALEMENT, je serai une députée constructive mais exigeante vis-à-vis des réformes proposées par le président de la République, fidèle aux valeurs qui fondent mon engagement politique à Gauche.

Je soutiendrai

→ **La moralisation de la vie publique et l'interdiction du cumul** de plus de trois mandats successifs pour permettre le renouvellement

→ Les projets visant à concilier **efficacité économique et justice sociale** ainsi qu'une action forte en matière de **formation pour lutter contre le chômage**

→ La priorité donnée à l'école primaire pour que chaque enfant puisse réussir et le soutien à l'autonomie des jeunes par une **revalorisation des bourses**

→ Les mesures en faveur du pouvoir d'achat pour les plus modestes avec **l'augmentation du minimum vieillesse et de la prime d'activité**

→ **La transition écologique et énergétique** : des investissements dans la croissance verte et les économies d'énergie, l'interdiction des perturbateurs endocriniens

Je serai vigilante

→ Sur la justice fiscale : **je m'opposerai à la baisse de l'Impôt Sur la Fortune et à l'augmentation de la CSG** qui entraînera une diminution des pensions pour plus de 8 millions de retraités

→ Sur le dialogue social et les droits des salariés : **le dialogue social doit être renforcé** et le Code du travail doit rester un socle protecteur élevé pour les salariés

→ Sur les services publics : « Ils sont le patrimoine de ceux qui n'en ont pas » ; je m'opposerai à une nouvelle baisse du nombre de fonctionnaires et des **dotations des collectivités**

LOCALEMENT, je défendrai les projets de nos territoires : fibre optique, investissement dans l'université, dans les activités sportives et culturelles, lutte contre les déserts médicaux et maintien des services en milieu rural.

« Ce quinquennat sera réussi si la France n'oublie personne en route. Pour cela, nous avons besoin d'une gauche forte. Je sais que Florence et Michel portent ces valeurs. Je vous appelle à les soutenir les 11 et 18 juin prochains. » **Jean-François DEBAT**

Vu le candidat - Agence TOUT&POSSIBLE - Imprimerie du Centre - Bourg-en-Bresse

Candidate de la gauche, des démocrates et des écologistes, je compte sur vous dès le 1[er] tour.

PS Social-écologie · europe écologie les Verts · PRG Parti Radical de Gauche · ude Union Démocrates Écologistes

www.florenceblatrix2017.com - florence.blatrix.contat2017@gmail.com

florenceblatrix @FlorenceBlatrix

(continuation)

Translation: I will be vigilant: On tax justice: I will oppose the reduction in Wealth Tax and the increase in the general social contribution which will lead to a reduction in pensions for more than 8 million retirees. On social dialogue and the rights of employees: social dialogue must be strengthened and the Labor Code must remain a strong protective base for employees. On public services: "They are the heritage of those who have none"; I will oppose a further drop in the number of civil servants and local authority grants.
Locally, I will defend the projects of our regions: fibre optics, investment in universities, sports and cultural activities, combating medical deserts and maintaining services in rural areas.
Michel FONTAINE First Deputy of the City of Bourg-en-Bresse President of the Agglomeration from 2008 to the end of 2016 and after Vice-President for Economic Development. I have lived in Bourg-en-Bresse for more than 40 years, I shared my professional life between the Carriat high school and my company Fontaine Picard. I have two children and live together. I assumed associative responsibilities before becoming a local elected official. By committing myself alongside Florence Blatrix-Contat, I am choosing success with great loyalty to my convictions.
≪This five-year term will be successful if France does not forget anyone on the way. For that, we need a strong left. I know that Florence and Michel carry these values. I call on you to support them on June 11 and 18.≫ Jean-François DEBAT Candidate of the left, the democrats and the ecologists, I am counting on you from the 1st round.
www.florenceblatrix2017.com - florence.blatrix.contat2017@gmail.com florenceblatrix @FlorenceBlatrix Vu le candidat - Agence TOUT&POSSIBLE - Imprimerie du Centre - Bourg-en-Bresse

B.2 Text pre-processing

For all types of text data, I perform the following pre-processing procedures.

I remove a list of words containing party names, party acronyms, parliamentary titles, and terms describing blocs of parties. I also eliminate first and last names.

I pre-process the content of the corpus following standard practices in natural language processing: remove punctuation and numbers, convert all letters to lower-case, lemmatize each word, and restrict the vocabulary to words used by at least 1% and 50% of the documents using spacy French version 3.5.0. In the case of legislative debates, I restrict the vocabulary to 0.5% and 50% of the documents. Last, I convert words from Latin-1 to UTF-8 for three reasons. Given that the OCR sometimes does not detect accents, it permits to reduce error; second, written questions for the ninth legislature are in UTF-8; third, it permits to save memory.

Some politicians opt for using the party platform instead of a personalized platform; this is common in small parties (e.g., Rassemblement National and Green parties) but an infrequent practice among well-established parties (e.g., socialist and republican parties). I do not include manifestos similar to the party platform in the training datasets. Keeping duplicate measures introduces the problem of multicollinearity, and it will make these manifestos count more. I include them in the regression model.

C Methodology

C.1 Tone

Table C1: Tone of the political platforms by ideology

	Tone 1st round		Tone 2nd round	
	Mean	Sd	Mean	Sd
far-left	-0.623	1.091	0.214	0.870
left	0.169	0.945	0.485	0.854
liberal	0.335	0.522	0.549	0.643
right	0.387	0.846	0.370	0.825
far-right	-0.453	0.830	-0.826	0.708
other	-0.051	0.827	0.474	0.642
Total	-0.056	1.003	0.336	0.877

Figure C.1: Noël Joseph manifesto during 1st round 1993 legislative elections - most positive manifesto

Source: Electoral archives of CEVIPOF SciencesPo, `https://archive.org/details/EL193_L_1993_03_062_11_1_PF_01`

Translation: French Republic - 11th constituency of Pas-de-Calais Legislative Elections March 1993

”Our will, to serve you, can be summed up in three words: fight, protect, advance, by mastering progress, because life, for all, must be a constant march towards greater equality and happiness.”

March 21th, I agree with Marcel CABIDDU to: Be Everyone’s Deputy. Acting together to create and maintain jobs. Maintain social protection. Defend national education and public services. Extending democracy to everyday life. Ensuring solidarity between the city and the countryside. Living in brotherhood.

Noel Joseph Mayor of Beuvry Regional Councilor Deputy Knight of the Legion of Honour Officer of the National Order of Merit

Marcel Cabiddu Mayor of Wingles General Counsel

Socialist Party candidate

Figure C.2: Frédéric Berger manifesto during 1st round 1993 legislative elections - most negative manifesto

Source: Electoral archives of CEVIPOF SciencesPo, `https://archive.org/details/EL195_L_1993_03_073_02_1_PF_06/mode/2up`

Translation: French Republic - 2nd Circumscription of Savoy Legislative Elections of March 21, 1993
Therese MINO-MATOT SHEPHERD wife born SOUVY Substitute
Frederic Berger Candidate
Vote Shepherd
I say what I do, I do what I say.
1) Suppression of the current political class to end corruption. 2) Really fight against unemployment, by removing the tax racket, which will allow 100,000 companies to hire 10 unemployed people, that is a total of ONE MILLION unemployed people! 3) Addressing injustice: With a real housing program applying the ideas of Abbé Pierre. Fight against the injustice of the courts and the tax authorities. Fight against all forms of racism and intolerance. Fight against the injustice of the victims of contaminated blood.
TO TRY BERGER is to ADOPT IT.
See the Candidates

C.2 Left-right dimension

I start with a matrix C_t whose rows correspond to candidates and whose columns correspond to distinct words. This means that an element c_{wt} gives the frequency of word w in a platform/document d. As in Taddy (2013), Gentzkow et al. (2019) and Cagé et al. (n.d.), I use a multinomial inverse regression model (MNIR):

$$c_{wt} \sim \text{MN}(q_{wd}, m_d) \text{ with } q_{wd} = \frac{\text{exp}[\alpha_w + \varphi_w E_d]}{\sum_{l=1}^{W} \text{exp}[\alpha_l + \varphi_l E_d]} \tag{5}$$

$m_{wd} = \sum_d c_{wd}$ denotes the total amount of words by a candidate in document d. The text-generating process is fully characterised by the verbosity of m_d and the probability $q_{wd}(\cdot)$ of writing each word. α_d is a scalar parameter capturing the baseline popularity of word w in document d. φ_w is a scalar parameter capturing the effect of being a left or right candidate on the propensity to use word w. E_d is a dummy variable equal to one if d is a left candidate as opposed to a right candidate.

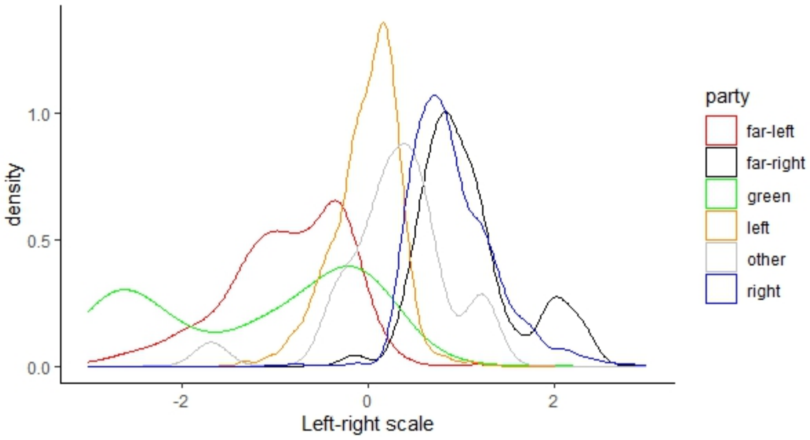
I am confronted with two methodological challenges to compute an accurate estimate of left-right. First, I have a problem of finite-sample bias that arises because the number of words a candidate can choose is large in relation to the platform content, so many words are mostly written by one party and others by chance. Second, in this multinomial logit model, the large number of choices and parameters makes standard approaches to estimation computationally infeasible. To address these issues, as proposed by Gentzkow et al. (2019), I control bias through penalization, using a gamma-lasso procedure and a Poisson approximation to the multinomial logit likelihood to permit distributed computing. The Poisson regression has a negative log-likelihood proportional to

$$l(\alpha_w, \varphi_w) = \sum_{d=1}^{N} [m_d \text{exp}(\alpha_w + \varphi_w E_d) - c_{wd}(\alpha_w + \varphi_w E_d)] \tag{6}$$

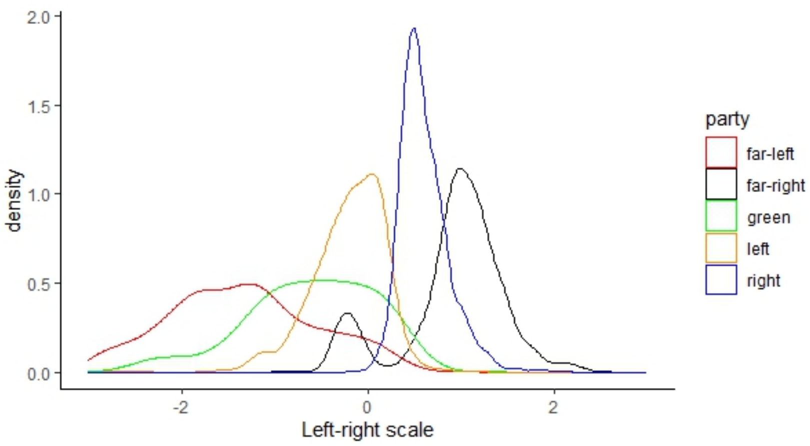
In high-dimensional regression, it is helpful to regularize estimation through a penalty on coefficient size to avoid over-fit and stabilize estimation. I apply the gamma-lasso procedure described in Taddy (2015):

$$\hat{\alpha}_w, \hat{\varphi_w} = \text{argmin} l(\alpha_d, \varphi_d) + N\lambda^{-1} \text{log}(1 + \lambda|\varphi_w|) \tag{7}$$

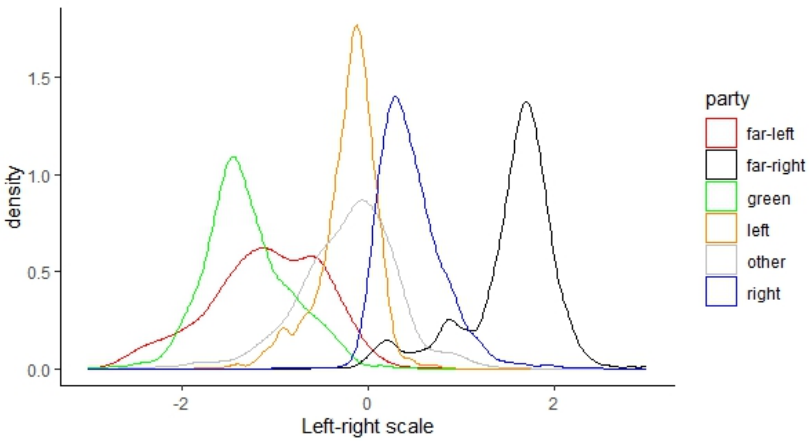
Kernel density of left-right score by political ideology



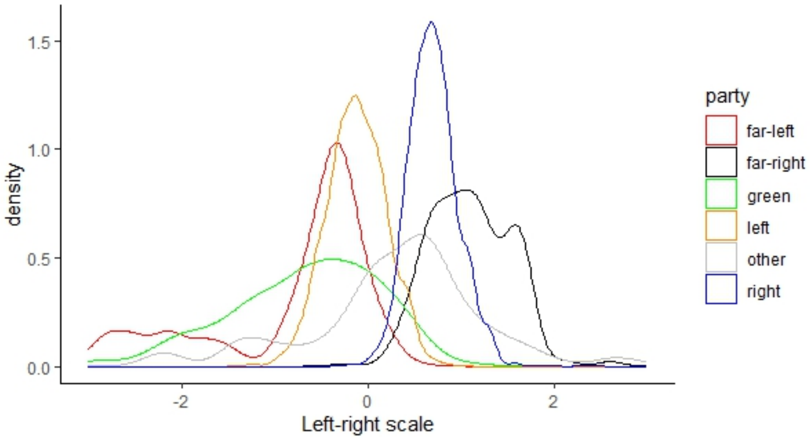
1981



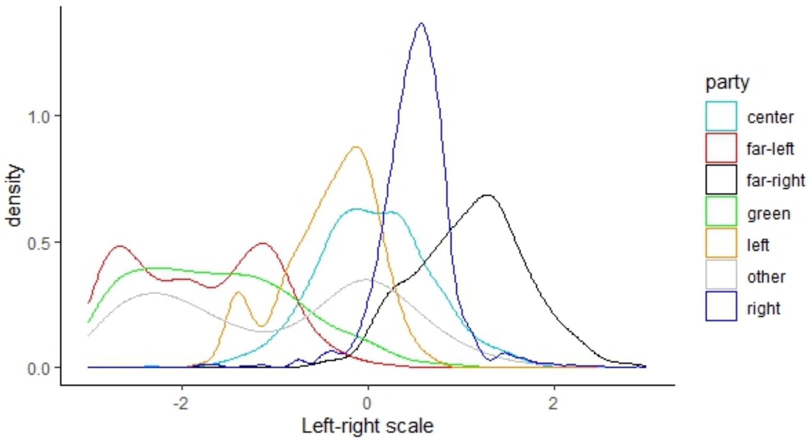
1988



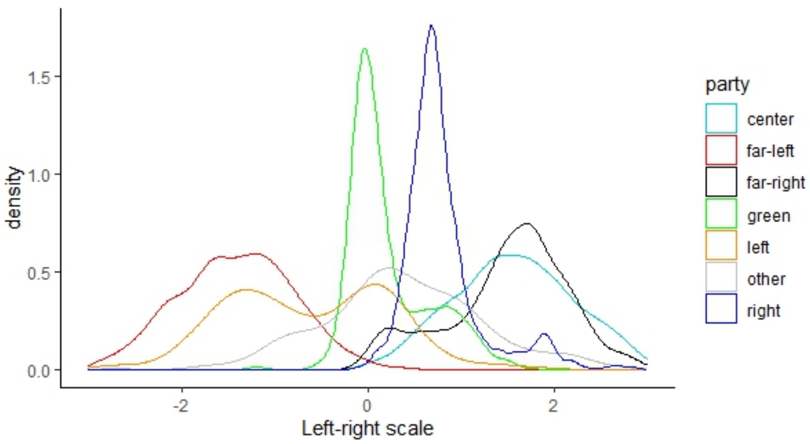
1993



1997



2017



2022

Table C2: Left-right words per election-year

1981		1988	
Left	Right	Left	Right
productivism	heaviness	over-armament	same
crushed	rampart	pacifist	anti-socialist
fatally	*socialo*	anti-racist	pride
vitalize	devaluation	apartheid	clergy
controllable	marxist	franckly	tax specialist
employers	socialo-comunist	add	bureaucratic
feminist	counterweight	bomb	deception
biomass	reassuring	cuts	reoccupy
interim	cell	remove	tighten
firmly	control	kanak	traffic

1993		1997	
Left	Right	Left	Right
disarmament	birth rate	austerity	*socialo*
gift	hateful	rediscussion	corrupt
progressif	granted	progressive	hassle
recovered	reevaluate	maneuver	incompetence
capitalist	*socialo*	secularism	civil solidarity pact
over-armament	car driver	ultra-liberal	bureaucracy
break up	reassure	radical	expel
sustainable	weight	speculative	clandestin
productivity	pinay	trade unionist	assassin
productivist	deterioration	boost	warning

2017		2022	
Left	Right	Left	Right
bourgeoisie	patriot	bifurcation	schengen
capitalist	islamist	capitalist	savoie
austerity	quotient	petroleum	patriot
xenophobic	clandestin	related	ardenne
capitalism	deserved	worker	var
employers	wine maker	renault	hunt
exchange	embryo	economist	islamism
trade unionist	islamism	employers	stratum
revocatory	wind turbine	institut radiation protection (irsn) nuclear safety	wine maker
progessif	alsatian	dividend	perfectly

Notes: The tables, translated into English, show the words with the highest left and right-wing partisan scores per election year. The scores are obtained by fitting a multinomial regression of word frequency in the documents on an indicator variable equal to one if the candidate is from a right-wing party instead of a left-wing party. For the years 1997, 2017, and 2022, the debates of the corresponding legislative term are also used to improve efficiency. Each year was estimated separately.

C.3 Topic classification - Seeded LDA

Latent Dirichlet Allocation (LDA) (Blei et al. (2003)) has been the most widely used topic model (e.g. Hansen, McMahon, and Prat (2018), Weigel (2020), Djourelova (2023)). LDA is an unsupervised method that assumes that documents are composed of words that help determine the topics and maps documents to a list of topics by assigning each word in the document to different topics. The assignment is in terms of conditional probability estimates. Under LDA, a document, d, is generated under the following hierarchical process:

- For each topic k draw a multinomial over words $\phi \sim Dirichlet(\beta)$.
- For each document d:
  - Draw a multinomial over topics $\theta \sim Dirichlet(\alpha)$.
  - For each word w_{N_d}:
    - Draw a topic $Z_{N_d} \sim Mult(\theta_D)$,

* Draw a word: $w \sim Mult(\phi_{Z_{d,w}})$.

When the number of documents is not large, the method is not efficient and topics tend to be difficult to interpret. A potential solution is to transform the method into a semi-supervised. Seeded LDA (Lu et al. (2011)) permits to define topics *a priori* through seeded words, before fitting the model.[20] Lu et al. (2011) specify a combined conjugate prior for each seed word, w, in $\phi \sim Dirichlet(\beta + C_w)$, where C_w is a pseudo-count added to the topic to which w belongs. In case there is no prior knowledge for a word w, $C_w = 0$. With a sample obtained via Gibbs sampling, the topic-word distribution phi_k is approximated, for each topic k and the document-topic distribution, θ_d, for each document d.

An alternative to the seeded LDA could be the Correlation Explanation (CorEx) model of Gallagher, Reing, Kale, and Ver Steeg (2017), and implemented by Djourelova, Durante, and Martin (2021). However, this method forces to choose an anchor strength. The anchor strength controls how much weight CorEx puts toward maximizing the mutual information between the seeded words and their respective topics. The authors encourage users to experiment with the anchor strength and determine the values that best suit their needs. Seeded LDA does not need an anchor strength.

Seeded words were gathered from a simple LDA fitted on the same corpus. For the case of manifestos, I start to print the list of selected words in the document. Then, I classify the most obvious words into their respective topics and use them as "seeded words".

Seeded Lda was run in RStudio using the "seededlda" package version 0.9.1 (Watanabe, Xuan-Hieu, and Watanabe (2023)).

C.3.1 List of seeded words per topic - Political platforms

Economy, Employment & Social Security: disability pension (*aah*), money (*argent*), austerity (*austérité*), budget (*budget*), budgetary (*budgétaire*), unemployment (*chômage*), unemployed (*chômeur*), business (*commerce*), trade (*commercer*), competition (*competition*), competitiveness (*competivité*), competition (*concurrence*), general social contribution (*csg*), deficit (*déficit*), tax exemption (*défiscalisation*), reduce taxation (*défiscaliser*), economy (*économie*), economic (*économique*), hire (*employer*), employer (*employeur*), entrepreneur (*entrepreneur*), entrepreneurship (*entrepreneuriat*), entreprise (*firm*), exportation (*exportation*), finance (*finance*), financing (*financement*), fund (*financer*), financial (*financier*), fiscal (*fiscal*), tax (*fiscalité*), gatt (*gatt*), tax (*impôt*),

[20] Some previous applications: Curini and Vignoli (2021), Fraccaroli and Pizzigolotto (2021) and Ash, Krümmel, and Slapin (2023). For a detailed explanation of the differences between the LDA and the seeded LDA, see Watanabe and Baturo (2023).

industry (*industrie*), industrial (*industriel*), inflation (*inflation*), wealth tax (*isf*), khomri (*khomri*), monetary (*monetaire*), worker (*ouvrier*), boss (*patron*), employer (*patronal*), bosses (*patronat*), poor (*pauvre*), poverty (*pauvreté*), pension (*pension*), sme (*pme*), small medium industry (*pmi*), precarious (*précaire*), precarity (*précarité*), privatisation (*privatisation*), privatise (*privatiser*), price (*prix*), companies register (*rcs*), recession (*récession*), reform (*réforme*), retirement (*rétraite*), income (*revenu*), income of active solidarity (*rsa*), social security scheme (*rsi*), wage (*salaire*), salary (*salariale*), employee (*salarié*), minimum wage (*smic*), rate (*taux*), tax (*tax*), taxation (*taxation*), tax (*taxer*), worker (*travailleur*), *uberisation*.

Environment: agrarian (*agricole*), farmer (*agriculteur*), agriculture (*agriculture*), agro, agribusiness (*agroalimentaire*), animal (*animal*), bio (*bio*), biodiversity (*biodiversité*), biological (*biologique*), carbon (*carbone*), fuel (*carburant*), carbide (*carbure*), climat (*climate*), climatic (*climatique*), water (*eau*), ecology (*écologie*), ecological (*écologique*), environment (*écologiste*), energetic (*énergétique*), energy (*énergie*), environment (*environnement*), green (*environnemental*), wind (*éolien*), species (*éspece*), forest (*fôret*), nuclear (*nucléaire*), fishing (*pêche*), fisher (*pêcheur*), programme for the endorsement of forest certification (*pefc*), pesticide (*pesticide*), petrol (*petrole*), planet (*planète*), polluting (*polluant*), pollute (*polluer*), pollution (*pollution*), recycling (*recyclage*), recycle (*recycler*), vegetarian (*végétarien*), winegrower (*viticulteur*), viticulture (*viticulture*).

Health & Education: academic (*académique*), class (*classe*), college (*collège*), doctor (*docteur*), school (*école*), educator (*éducateur*), educational (*éducatif*), education (*éducation*), educate (*éduquer*), establishment of accommodation for dependent old persons (*ehpad*), student (*élève*), endocrine (*endocrinien*), childhood (*enfance*), child (*enfant*), confinement (*enfermement*), teacher (*enseignant*), teaching (*enseignement*), teach (*enseigner*), study (*étude*), student (*étudiant*), study (*étudier*), training (*formation*), hospital (*hôpital*), hospitable (*hospitalier*), college (*lycée*), sick (*malade*), disease (*maladie*), maternity (*maternité*), doctor (*médecin*), medecine (*médecine*), medical (*médicale*), medication (*médicament*), patient (*patient*), teacher (*professeur*), blood (*sang*), sanitary (*sanitaire*), health (*santé*), science (*science*), scientific (*scientifique*), academic (*scolaire*), hiv (*sida*), care (*soin*), universitary (*universitaire*), university (*université*).

Security, Justice & Foreign Policy: africa (*afrique*, germany (*allemagne*), american (*americain*), weapon (*arme*), armed (*armée*), weapons (*armement*), asylum (*asile*), brussels (*bruxelle*), clandestin (*clandestine*), crime (*crime*), criminal (*criminel*), cybercrime (*cybercriminalité*), delinquency (*déliquant*), delinquent (*délinquant*), offence (*délit*), drug (*drogue*), foreigner (*étranger*), europe (*europe*), european (*européen*), border (*frontière*), policeman (*gendarme*), war (*guerre*), immigration (*immigration*), immigrant (*immigré*), insecurity (*insecurité*), maastricht (*maastricht*), magistrat (*mag-*

istrate), world (*monde*), worldwide (*mondial*), nationality (*nationalité*), otan, sentence (*peine*), penal (*pénal*), police (*police*), police (*policier*), prison (*prison*), security (*sécurité*), terrorism (*terrorisme*), terrorist (*terroriste*), treaty (*traité*), court (*tribunal*), ukraine (*ukraine*), victim (*victime*), violence (*violence*).

Local: canton (*canton*), cantonal (*cantonal*), municipal (*communal*), communitarian (*communautaire*), community (*communauté*), municipality (*commune*), decentralisation (*décentralisation*), decentralise (*décentraliser*), departmental (*départementale*), desert (*désert*), desertification (*désertification*), inhabitant (*habitant*), intercommunal (*intercommunal*), municipal (*municipal*), municipality (*municipalité*), region (*région*), regional (*régional*), rural (*rural*), rurality (*ruralité*), land (*terrain*), territory (*territoire*), territorial (*territorial*), city (*ville*), area (*zone*).

Politics: antisocialist (*antisocialiste*), assembly (*assemblée*), campaign (*campagne*), candidature (*candidature*), centrist (*centriste*), coalition (*coalition*), cohabitation (*cohabitation*), constitution (*constitution*), democrat (*démocrate*), democracy (*démocratie*), democratisation (*démocratisation*), sunday (*dimanche*), dissolution (*dissolution*), dissolve (*dissoudre*), right (*droite*), voter (*électeur*), elective (*électif*), electoral (*électoral*), elected (*élu*), inhibit (*empêcher*), left (*gauche*), gaulliste, holland, majority (*majoritaire*), majority (*majority*), presidency (*présidence*), president (*président*), presidential (*présidentiel*), reelection (*reélection*), reelect (*reélire*), republican (*républicain*), republic (*république*), senate (*sénat*), senator (*sénateur*), socialism (*socialisme*).

C.3.2 Top 10 words per topic

Economy, employment & social security: firm (*entreprise*), retirement (*retraite*), economic (*économique*), unemployment (*chômage*), economy (*économie*), tax (*impôt*), wage (*salaire*), reform (*réforme*), worker (*travailleur*), employee (*salarié*).

Environment: environment (*environnement*), energy (*énergie*), ecology (*écologie*), environment (*écologiste*), agriculture (*agriculture*), ecological (*écologique*), farmer (*agriculteur*), agrarian (*agricole*), nuclear (*nucleaire*), water (*eau*).

Health & education: child (*enfant*), health (*santé*), school (*école*), education (*éducation*), training (*formation*), class (*classe*), hospital (*hôpital*), academic (*scolaire*), teaching (*enseignement*), medical (*médicale*).

Security, justice & foreign policy: europe (*europe*), security (*sécurité*), world (*monde*), european (*européen*), immigration (*immigration*), fight against (*lutter contre*), fight (*lutter*), foreigner (*étranger*), police (*police*), insecurity (*insecurité*).

Local: territory (*territoire*), city (*ville*), region (*région*), rural (*rural*), municipal (*municipal*), regional (*régional*), land (*terrain*), inhabitant (*habitant*), defend (*défendre*), municipality (*commune*).

Politics: majority (*majorité*), left (*gauche*), president (*président*), right (*droite*), assembly (*assemblée*), republic (*république*), elected (*élu*), voter (*électeur*), presidential (*présidentiel*), sunday (*dimanche*).

Other 1: councillor (*counseiller*), general (*général*), general councillor (*counseiller général*, trust (*confiance*), child (*enfant*), future (*avenir*), council (*conseil*), married (*marié*), department (*département*), freedom (*liberté*).

Other 2: citizen (*citoyen*), make (*falloir*), no (*non*), society (*societé*), other (*autre*), man (*homme*), right (*droit*), enter (*entrer*), today (*aujour*), live (*vivre*).

C.4 Descriptive statistics

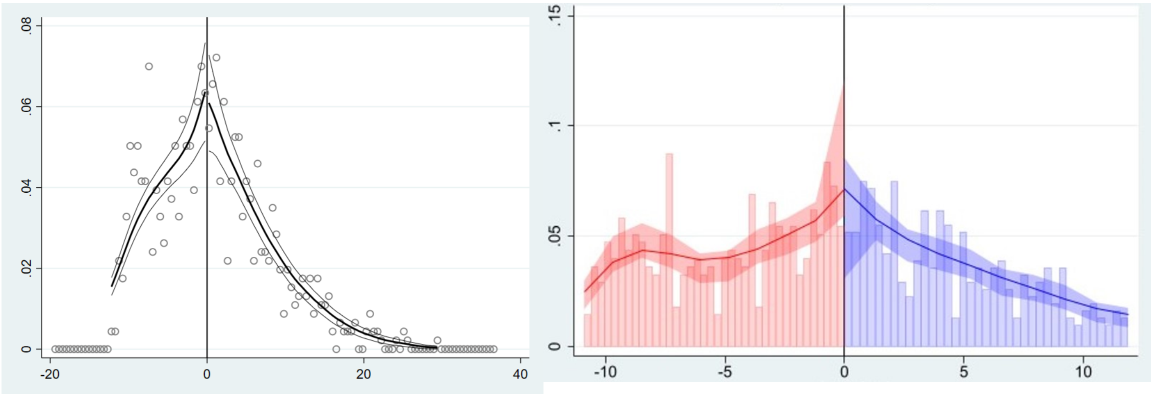
Table C3: Differences between elected and non-elected politicians of the same gender - Summary statistics

	Women			Men		
	Elected	Not elected	Difference	Elected	Not elected	Different
% votes	26.85	20.10	6.75***	33.35	25.30	8.06***
Number words	0.08	0.01	0.07	-0.39	-0.47	0.08***
Tone	0.44	0.14	0.31***	0.47	0.18	0.29***
Left right	0.49	0.55	-0.06	0.25	0.42	-0.17***
Similar to party	0.23	0.01	0.22***	0.05	-0.13	0.18***
Similar to women	0.30	0.11	0.19***	0.05	-0.16	0.21***
Economy & employment	-0.56	-0.10	-0.46***	-0.60	-0.35	-0.25***
Environment	-0.19	-0.01	-0.17***	-0.29	-0.25	-0.04***
Health & education	1.31	0.49	0.82***	0.19	-0.06	0.26***
Security & foreign policy	0.63	0.55	0.08	-0.04	-0.05	0.01
Local politics	0.13	0.13	0.00	-0.09	-0.21	0.12***
National politics	0.01	-0.00	0.01	0.70	0.56	0.14***
Observations	583	692	1275	2390	2331	4721

C.5 Sharp regression discontinuity design

Figure C.3: Manipulation testing: Most voted female - 2nd most voted male

(a) McCrary (2008) (b) Cattaneo et al. (2018)



Notes. Figures (a) and (b) represent the density test for races where only one candidate (or none) obtained the 12.5%; the margin is the difference between the most-voted woman and the second most-voted man. Figure (a) represents the McCrary density test; discontinuity estimate b: -0.096 (s.e. 0.330). Figure (b) represents the Cattaneo et al. (2018) manipulation test; p-value 0.139 (not reject the null hypothesis of no manipulation).

I conduct placebo tests to examine whether there is discontinuity at the threshold for any of the variables used to predict treatment. I first provide information about the construction of each variable. If the information is missing, it is because the name of the dependent variable is self-explanatory.

Platform available: dummy equal to 1 if the manifesto for the 1st and 2nd round is available, 0 if not.

Votes: number of votes obtained divided by the number of enrolled voters.

Number candidates: number of candidates running in the electoral district.

Number female: number of female candidates running in the electoral district.

Victory margin: margin between the most voted and the second most voted candidate.

Number far-left, left, right, far-right: Number of candidates of the respective ideology.

Sum left/right: sum of the vote share in all left/right candidates.

Number words: total number of words in the manifesto.

Tone, Left right, Similar to party, Similar to women, Similar to opponent, Economy & employment, Environment, Health & education, Security & foreign policy, Local politics, National politics: explained in subsections 5.1 to 5.4.

Far-left, left, liberal, right, far-right: a dummy equal to 1 if the politician is classified as belonging to that ideology, 0 if not.

Table C4: Balancing tests (1st round)

	(1) Platform available	(2) Votes	(3) Turnout	(4) Number candidates	(5) Number female	(6) Enrolled voters	(7) Victory margin
Woman	-0.030	-0.578	-2.771	-0.277	0.028	-4,731	-0.308
	(0.099)	(0.551)	(2.389)	(0.712)	(0.521)	(4,078)	(1.413)
Observations	1008	1008	1008	1008	1008	1008	1008
Eff. number of obs	348	368	319	272	316	350	318
Robust p-value	0.910	0.331	0.229	0.982	0.788	0.206	0.744
Polyn. order	1	1	1	1	1	1	1
Bandwidth	3.722	3.924	3.283	2.741	3.221	3.731	3.269
Outcome mean	0.764	11.468	54.382	11.644	3.991	77904	6.492

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C5: Balancing tests - electoral district characteristics (1st round)

	(1) Number far-left	(2) Number left	(3) Number right	(4) Number far-right	(5) Sum left	(6) Sum right
Woman	0.486	-0.431	0.120	-0.304	-1.419	-2.997
	(0.305)	(0.310)	(0.245)	(0.191)	(1.429)	(2.519)
Observations	1008	1008	1008	1008	1008	1008
Eff. number of obs	206	423	351	319	354	261
Robust p-value	0.063	0.155	0.546	0.115	0.343	0.161
Polyn. order	1	1	1	1	1	1
Bandwidth	1.984	4.557	3.736	3.292	3.797	2.612
Outcome mean	2.164	3.051	1.924	1.906	20.205	22.478

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C6: Balancing tests - differences in ideology of female and male candidates

	(1) Far-left	(2) Left	(3) Liberal	(4) Right	(5) Far-right
Woman	0.061	0.040	-0.020	-0.075	0.015
	(0.057)	(0.091)	(0.089)	(0.086)	(0.095)
Observations	1017	1017	1017	1017	1017
Eff. number of obs	448	433	382	349	362
Robust p-value	0.435	0.826	0.769	0.587	0.908
Polyn. order	1	1	1	1	1
Bandwidth	4.911	4.741	4.016	3.616	3.853
Outcome mean	0.057	0.306	0.095	0.280	0.231

This table reports the results of t-tests investigating whether there are significant differences in the ideological alignment of male and female candidates in the second-round elections. Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C7: Balancing tests - differences in ideology of female and male opponents

	(1) Far-left	(2) Left	(3) Liberal	(4) Right	(5) Far-right
Woman	0.031	-0.137	0.125	0.099	-0.114
	(0.048)	(0.103)	(0.130)	(0.095)	(0.072)
Observations	1017	1017	1017	1017	1017
Eff. number of obs	352	312	279	325	303
Robust p-value	0.460	0.182	0.238	0.443	0.113
Polyn. order	1	1	1	1	1
Bandwidth	3.687	3.135	2.827	3.315	3.063
Outcome mean	0.188	0.268	0.268	0.269	0.268

This table reports the results of t-tests investigating whether there are significant differences in the ideological alignment of male and female opponents in the second-round elections. Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C8: Balancing tests - platforms characteristics (1st round)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Woman	0.047	0.440*	0.332	0.332*	0.066
	(0.218)	(0.246)	(0.229)	(0.189)	(0.060)
Observations	835	835	835	835	788
Eff. number of obs	334	150	322	328	272
Robust p-value	0.807	0.038	0.285	0.073	0.293
Polyn. order	1	1	1	1	1
Bandwidth	4.861	1.964	4.439	4.528	4.107
Outcome mean	0.077	0.269	-0.115	0.035	0.681

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C9: Balancing tests - platforms topics (1st round)

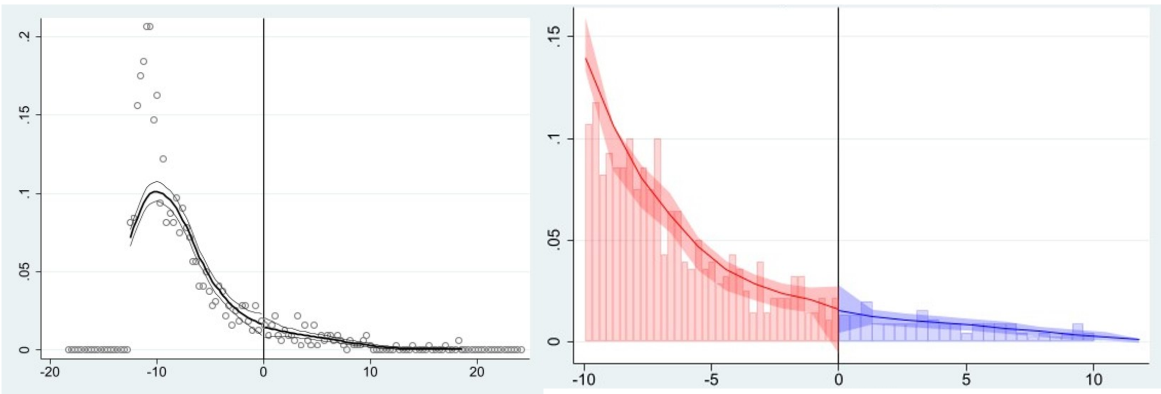
	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Woman	-0.436	0.143	0.767*	0.173	-0.645*	0.469*
	(0.377)	(0.144)	(0.441)	(0.412)	(0.372)	(0.277)
Observations	835	835	835	835	835	835
Eff. number of obs	160	272	264	349	207	192
Robust p-value	0.168	0.588	0.072	0.908	0.267	0.061
Polyn. order	1	1	1	1	1	1
Bandwidth	2.095	3.791	3.634	5.083	2.836	2.504
Outcome mean	-0.066	-0.180	-0.007	0.225	0.871	-0.181

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

C.6 Fuzzy regression discontinuity design

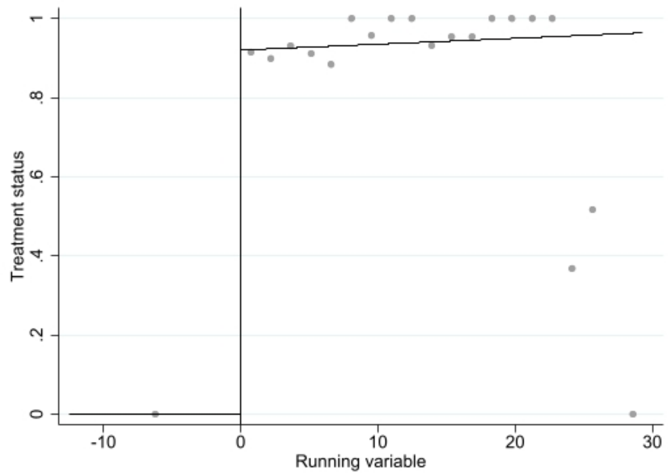
Figure C.4: Manipulation testing: Most voted women - 12.5

(a) McCrary (2008) (b) Cattaneo et al. (2018)



Notes. This figure tests for a jump in the density of the running variable. The solid line represents the density of the running variable. Thin lines represent the confidence intervals. Figures (a) and (b) represent the density test for races where a third candidate (woman) obtained the minimum 12.5% of votes (and two male candidates also passed the threshold). Figure (a) represents the McCrary density test; discontinuity estimate b: -0.046 (s.e. 0.148). Figure (b) represents the Cattaneo et al. (2018) manipulation test; p-value 0.600 (not reject the null hypothesis of no manipulation).

Figure C.5: First stage



Notes. Dots represent the local averages of the treatment status (y-axis). Averages are calculated within bins of the running variable (x-axis). The running variable (the qualifying margin of a woman in a race where at least two candidates obtained 12.5% of the votes) is measured as percentage points. Continuous lines are a linear fit.

I conduct placebo tests to examine whether there is a discontinuity at the threshold for any variables used to predict treatment. Information about the construction of each variable is in subsection C.5. The dependent variables refer only to the first round.

Table C10: Balancing tests (1st round)

	(1) Platform available	(2) Votes	(3) Turnout	(4) Number candidates	(5) Number female	(6) Enrolled voters	(7) Victory margin
Treatment	-0.130	-0.804	-2.873	-0.322	0.043	-3,959	-0.322
	(0.084)	(1.335)	(2.363)	(0.701)	(0.519)	(4,075)	(1.368)
Observations	876	876	875	876	876	875	876
Eff. number of obs	282	295	326	274	310	351	327
Robust p-value	0.120	0.478	0.228	0.951	0.767	0.292	0.777
Polyn. order	1	1	1	1	1	1	1
Bandwidth	2.886	3.034	3.404	2.793	3.196	3.742	3.424
Outcome mean	0.137	4.483	53.066	12.609	4.396	78996	6.504

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C11: Balancing tests - electoral district characteristics (1st round)

	(1) Number far-left	(2) Number left	(3) Number right	(4) Number far-right	(5) Sum left	(6) Sum right
Treatment	0.388	-0.498	0.140	-0.289	-1.899	-2.520
	(0.297)	(0.334)	(0.248)	(0.192)	(1.548)	(2.390)
Observations	876	876	876	876	876	876
Eff. number of obs	216	372	341	319	350	280
Robust p-value	0.115	0.131	0.474	0.136	0.243	0.223
Polyn. order	1	1	1	1	1	1
Bandwidth	2.074	3.968	3.550	3.283	3.736	2.876
Outcome mean	2.300	3.284	2.055	1.929	19.137	21.986

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C12: Balancing tests - platforms characteristics (1st round)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Treatment	0.029	0.166	0.452***	0.052	0.001
	(0.219)	(0.190)	(0.173)	(0.277)	(0.084)
Observations	762	762	762	762	728
Eff. number of obs	378	293	375	274	263
Robust p-value	0.760	0.411	0.022	0.754	0.922
Polyn. order	1	1	1	1	1
Bandwidth	4.904	3.529	4.869	3.219	3.348
Outcome mean	0.138	-0.187	-0.118	0.042	0.700

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C13: Balancing tests - platforms topics (1st round)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Treatment	-0.131	-0.158	0.340	0.455	-0.215	0.047
	(0.222)	(0.120)	(0.680)	(0.412)	(0.239)	(0.143)
Observations	762	762	762	762	762	762
Eff. number of obs	316	261	265	277	306	296
Robust p-value	0.439	0.206	0.509	0.392	0.401	0.656
Polyn. order	1	1	1	1	1	1
Bandwidth	3.904	3.044	3.076	3.311	3.808	3.593
Outcome mean	0.240	0.124	0.026	0.025	0.262	-0.306

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table C14: Balancing tests - differences in ideology between male candidates running against a woman and males running only against other male candidates

	(1) Far-left	(2) Left	(3) Liberal	(4) Right	(5) Far-right
Treatment	0.041	-0.227**	0.064	0.125	-0.037
	(0.054)	(0.099)	(0.117)	(0.096)	(0.078)
Observations	876	876	876	876	876
Eff. number of obs	295	286	325	319	310
Robust p-value	0.360	0.028	0.537	0.266	0.732
Polyn. order	1	1	1	1	1
Bandwidth	3.021	2.916	3.377	3.297	3.157
Outcome mean	0.188	0.269	0.048	0.167	0.165

This table reports the results of t-tests investigating whether there are significant differences in the ideological alignment of treated and non-treated male candidates in the second-round elections. Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The independent variable is a dummy equal to 1 if the candidate is a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

D Results

D.1 Baseline results: Marginally eligible women versus marginally eligible men

D.1.1 Political platforms: Positioning and topic content

Table D15: Differences between female and male 2nd round political candidates (independently of adversary)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Woman	0.135	-0.267	0.353*	0.636**	0.796***
	(0.176)	(0.221)	(0.202)	(0.272)	(0.305)
Observations	788	788	788	784	743
Eff. number of obs	258	304	314	322	294
Robust p-value	0.369	0.220	0.099	0.023	0.015
Polyn. order	2	2	2	2	2
Bandwidth	3.870	4.647	4.942	5.155	5.066
Outcome mean	-0.320	0.049	0.194	-0.181	-0.201

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 2, and the optimal bandwidths are derived under the MSERD procedure. All the dependent variables are standardized. In column 1, the outcome is the number of words used in the platform. For column 2, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 3 is explained in 5.2. The methodology to obtain columns 4 and 5 is explained in 5.4.

Table D16: Differences between female and male 2nd round political candidates (independently of adversary)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Woman	0.249	-0.049	-0.293	0.745*	-0.274	-0.180
	(0.224)	(0.064)	(0.220)	(0.396)	(0.276)	(0.294)
Observations	788	788	788	788	788	788
Eff. number of obs	278	304	438	258	270	242
Robust p-value	0.235	0.493	0.183	0.052	0.437	0.398
Polyn. order	2	2	2	2	2	2
Bandwidth	4.109	4.692	7.547	3.859	3.991	3.524
Outcome mean	-0.153	-0.196	-0.088	0.110	0.392	0.298

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 2, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

D.1.2 Can campaign financing explain gender differences in platforms?

Table D17: Gender differences in campaign financing

	(1) Total expenditures	(2) Total revenues	(3) Party contribution	(4) Private donations	(5) Personal contribution
Panel A: Left candidates					
Woman	0.034	-0.565*	-0.193	0.103	-0.037
	(0.366)	(0.335)	(0.235)	(0.354)	(0.136)
Observations	346	231	344	344	344
Eff. number of obs	125	88	107	148	125
Robust p-value	0.996	0.070	0.291	0.709	0.674
Polyn. order	1	1	1	1	1
Bandwidth	3.450	3.556	2.776	4.414	3.482
Outcome mean	1.188	-0.253	0.192	0.364	0.587
Panel B: Right candidates					
Woman	-0.258	-0.245	-0.037	-0.046	-0.121
	(0.203)	(0.214)	(0.079)	(0.141)	(0.123)
Observations	776	776	775	775	775
Eff. number of obs	274	266	247	258	279
Robust p-value	0.277	0.363	0.616	0.874	0.352
Polyn. order	1	1	1	1	1
Bandwidth	4.010	3.920	3.626	3.820	4.136
Outcome mean	1.183	1.238	0.194	0.339	0.605

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. Each outcome uses the number of registered voters as the denominator.

Table D18: Gender differences in campaign financing

	(1) Total expenditures	(2) Total revenues	(3) Party contribution	(4) Private donations	(5) Personal contribution
Woman	-0.074	-0.040	-0.015	0.120	-0.099
	(0.301)	(0.302)	(0.127)	(0.239)	(0.106)
Observations	1135	1135	1132	1132	1132
Eff. number of obs	500	508	529	565	466
Robust p-value	0.718	0.827	0.799	0.538	0.249
Polyn. order	2	2	2	2	2
Bandwidth	5.040	5.140	5.589	6.121	4.585
Outcome mean	1.204	1.289	0.181	0.369	0.593

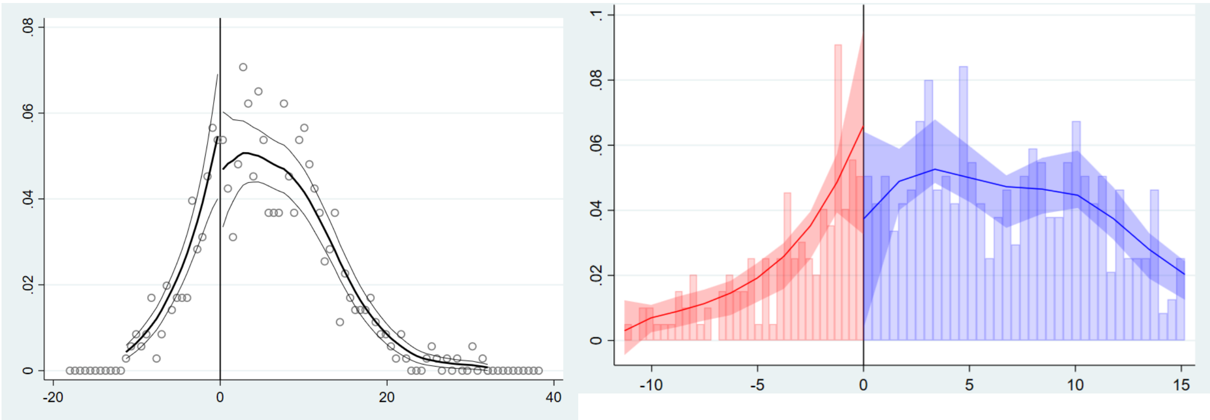
Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman running in the second round instead of a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 2, and the optimal bandwidths are derived under the MSERD procedure. Each outcome uses the number of registered voters as the denominator.

D.2 Do women adapt to the gender of their opponent?

D.2.1 Platform content: positioning and topic content

Figure D.6: Manipulation testing

(a) McCrary (2008) (b) Cattaneo et al. (2018)



Notes. This figure tests for a jump in the density of the running variable. The solid line represents the density of the running variable. Thin lines represent the confidence intervals. Figures (a) and (b) represent the density test for races where a woman competes in the second round against a man instead of competing against a woman. Figure (a) represents the McCrary density test; discontinuity estimate b: -0.228 (s.e. 0.220). Figure (b) represents the Cattaneo et al. (2018) manipulation test; p-value 0.187 (not reject the null hypothesis of no manipulation).

I conduct placebo tests to examine whether there is a discontinuity at the threshold for any variables used to predict treatment. Information about the construction of each variable is in subsection C.5. The dependent variables refer only to the first round.

Table D19: Balancing tests (1st round)

	(1) Platform available	(2) Votes	(3) Turnout	(4) Number candidates	(5) Number female	(6) Enrolled voters	(7) Victory margin
Treatment	-0.025	0.194	-0.351	-0.187	-0.073	5,930	0.183
	(0.111)	(1.406)	(3.120)	(1.026)	(0.551)	(4,709)	(1.408)
Observations	579	579	579	579	579	579	579
Eff. number of obs	205	204	165	146	163	205	230
Robust p-value	0.840	0.677	0.890	0.886	0.884	0.258	0.752
Polyn. order	1	1	1	1	1	1	1
Bandwidth	3.846	3.782	3.088	2.755	3.029	3.790	4.523
Outcome mean	0.758	3.732	48.507	13.659	6.764	82108	7.252

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest is a woman competing in the second round with a man instead of a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table D20: Balancing tests - electoral district characteristics (1st round)

	(1) Number far-left	(2) Number left	(3) Number right	(4) Number far-right	(5) Sum left	(6) Sum right
Treatment	0.537	-0.490	0.353	-0.739**	-1.581	-0.265
	(0.350)	(0.391)	(0.379)	(0.315)	(1.794)	(2.294)
Observations	579	579	579	579	579	579
Eff. number of obs	163	129	144	116	199	205
Robust p-value	0.135	0.186	0.303	0.014	0.485	0.999
Polyn. order	1	1	1	1	1	1
Bandwidth	3.035	2.464	2.707	2.086	3.693	3.811
Outcome mean	2.371	3.569	2.045	2.302	15.136	17.475

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest is a woman competing in the second round with a man instead of a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table D21: Balancing tests - platforms characteristics (1st round)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Treatment	0.173	-0.478*	0.183	0.552***	0.183***
	(0.410)	(0.246)	(0.321)	(0.204)	(0.067)
Observations	493	493	493	493	481
Eff. number of obs	165	187	146	164	143
Robust p-value	0.644	0.071	0.611	0.013	0.010
Polyn. order	1	1	1	1	1
Bandwidth	3.554	3.981	3.112	3.487	3.083
Outcome mean	0.111	-0.280	-0.075	0.150	0.765

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest is a woman competing in the second round with a man instead of a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table D22: Balancing tests - platforms topics (1st round)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Treatment	0.060	-0.139	1.671	-0.120	-0.654*	-0.233*
	(0.400)	(0.163)	(1.226)	(0.557)	(0.397)	(0.126)
Observations	493	493	493	493	493	493
Eff. number of obs	146	131	126	193	140	119
Robust p-value	0.996	0.361	0.181	0.723	0.142	0.072
Polyn. order	1	1	1	1	1	1
Bandwidth	3.160	2.864	2.712	4.220	2.982	2.540
Outcome mean	0.286	0.236	0.142	0.170	0.494	-0.365

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest is a woman competing in the second round with a man instead of a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table D23: Balancing tests - differences in ideology between women running against a man and women running against a woman

	(1) Far-left	(2) Left	(3) Liberal	(4) Right	(5) Far-right
Treatment	-0.009	-0.049	0.118	-0.093	0.042
	(0.008)	(0.102)	(0.182)	(0.085)	(0.129)
Observations	576	576	576	576	576
Eff. number of obs	127	169	159	167	200
Robust p-value	0.319	0.616	0.544	0.360	0.743
Polyn. order	1	1	1	1	1
Bandwidth	2.411	3.209	2.972	3.191	3.759
Outcome mean	0.168	0.258	0.072	0.145	0.186

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest is a woman competing in the second round with a man instead of a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table D24: Balancing tests - differences in the ideology of women's opponents

	(1) Far-left	(2) Left	(3) Liberal	(4) Right	(5) Far-right
Treatment	0.081	-0.108	-0.052	0.122	-0.033
	(0.107)	(0.124)	(0.134)	(0.130)	(0.096)
Observations	576	576	576	576	576
Eff. number of obs	151	241	212	188	226
Robust p-value	0.361	0.440	0.834	0.511	0.746
Polyn. order	1	1	1	1	1
Bandwidth	2.863	4.700	4.020	3.492	4.461
Outcome mean	0.170	0.257	0.258	0.259	0.257

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. The unit of observation is the candidate. The variable of interest is a woman competing in the second round with a man instead of a woman. Each column reports the results from a separate local polynomial regression. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table D25: Impact of a marginal presence of a man on female candidates in the 2nd round (polynomial of order 2)

	(1) No. words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Treatment	-0.126	-0.193	0.356*	0.334	0.197
	(0.289)	(0.197)	(0.194)	(0.274)	(0.283)
Observations	465	465	465	461	453
Eff. number of obs	131	143	188	229	221
Robust p-value	0.452	0.404	0.071	0.318	0.670
Polyn. order	2	2	2	2	2
Bandwidth	2.937	3.187	4.359	5.499	5.437
Outcome mean	0.180	0.282	0.439	0.215	0.282

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman competing in the second round with a man instead of a woman. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All the dependent variables are standardized. In column 1, the outcome is the number of words used in the platform. For column 2, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 3 is explained in 5.2. The methodology to obtain columns 4 and 5 is explained in 5.4.

Table D26: Impact of a marginal presence of a man on female candidates in the 2nd round (polynomial of order 2)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
Treatment	0.239	-0.027	0.339	0.584	-0.350	-0.404
	(0.173)	(0.048)	(0.377)	(0.395)	(0.249)	(0.291)
Observations	465	465	465	465	465	465
Eff. number of obs	184	205	225	187	221	184
Robust p-value	0.133	0.658	0.420	0.227	0.212	0.160
Polyn. order	2	2	2	2	2	2
Bandwidth	4.144	4.774	5.195	4.269	5.112	4.059
Outcome mean	-0.295	-0.080	1.171	0.711	0.486	-0.294

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman competing in the second round with a man instead of a woman. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

D.3 Do men adapt to the presence of a woman in the race?

Table D27: Impact of a marginal presence of a woman on male candidates in the 2nd round (without controlling for left-right in the first round)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
Panel A: All candidates					
Treatment	0.006	-0.182	0.124	0.128	0.131
	(0.128)	(0.149)	(0.130)	(0.134)	(0.136)
Observations	2495	2495	2495	2429	2252
Eff. number of obs	619	487	484	633	596
Robust p-value	0.980	0.217	0.474	0.417	0.375
Polyn. order	1	1	1	1	1
Bandwidth	4.548	3.673	3.603	4.806	4.938
Outcome mean	-0.189	0.216	0.304	-0.063	0.696
Panel B: Left candidates					
Treatment	1.085***	-0.368	-0.024	-0.359	-0.551
	(0.394)	(0.332)	(0.233)	(0.438)	(0.389)
Observations	923	923	923	918	748
Eff. number of obs	110	94	124	76	65
Robust p-value	0.004	0.291	0.785	0.255	0.111
Polyn. order	1	1	1	1	1
Bandwidth	2.707	2.433	3.221	1.979	2.170
Outcome mean	-0.099	0.251	0.326	-0.105	-0.146
Panel C: Right candidates					
Treatment	-0.200	-0.122	-0.043	0.212	0.280
	(0.165)	(0.149)	(0.143)	(0.143)	(0.176)
Observations	1561	1561	1561	1501	1504
Eff. number of obs	363	466	320	418	324
Robust p-value	0.196	0.497	0.608	0.155	0.115
Polyn. order	1	1	1	1	1
Bandwidth	3.830	4.941	3.242	4.413	3.429
Outcome mean	-0.172	0.246	0.328	-0.073	-0.101

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. The dependent variables of columns 2-8 are standardized. In column 1, the outcome is the number of votes obtained divided by the number of voters enrolled in the district. In column 2, the outcome is the number of words used in the manifesto. For columns 3 and 4, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 5 is explained in 5.2. The methodology to obtain columns 6-9 is explained in 5.4.

Table D28: Impact of a marginal presence of a woman on male candidates in the 2nd round

	No. words	Tone	Left right	Similar to party	Similar to women
		Panel A: All candidates			
Treatment	-0.006	-0.300	-0.083	0.025	0.033
	(0.161)	(0.203)	(0.114)	(0.221)	(0.219)
Observations	2495	2495	2495	2429	2252
Eff. number of obs	912	589	825	585	571
Robust p-value	0.847	0.105	0.477	0.874	0.925
Polyn. order	2	2	2	2	2
Bandwidth	6.670	4.308	6.181	4.337	4.681
Outcome mean	-0.214	0.246	0.341	-0.076	0.696
		Panel B: Left candidates			
Treatment	1.853**	-0.384	-0.309	-0.620	-1.373**
	(0.734)	(0.398)	(0.288)	(0.496)	(0.552)
Observations	923	923	923	918	748
Eff. number of obs	120	134	132	127	97
Robust p-value	0.008	0.454	0.346	0.188	0.017
Polyn. order	2	2	2	2	2
Bandwidth	3.177	3.701	3.567	3.340	3.330
Outcome mean	-0.117	0.216	0.307	-0.049	0.703
		Panel C: Right candidates			
Treatment	-0.329	-0.198	-0.131	0.248	0.328
	(0.262)	(0.213)	(0.117)	(0.191)	(0.231)
Observations	1561	1561	1561	1501	1504
Eff. number of obs	423	570	538	582	448
Robust p-value	0.216	0.345	0.332	0.261	0.191
Polyn. order	2	2	2	2	2
Bandwidth	4.388	6.083	5.734	6.371	4.938
Outcome mean	-0.181	0.238	0.344	-0.083	0.696

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. The dependent variables of columns 2-8 are standardized. In column 1, the outcome is the number of votes obtained divided by the number of voters enrolled in the district. In column 2, the outcome is the number of words used in the manifesto. For columns 3 and 4, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 5 is explained in 5.2. The methodology to obtain columns 6-9 is explained in 5.4.

Table D29: Impact of a marginal presence of a woman on male candidates in the 2nd round

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
			Panel A: All candidates			
Treatment	0.069	-0.082**	-0.181	-0.078	0.028	0.084
	(0.118)	(0.039)	(0.181)	(0.238)	(0.173)	(0.216)
Observations	2495	2495	2495	2495	2495	2495
Eff. number of obs	988	540	654	739	775	779
Robust p-value	0.566	0.028	0.282	0.618	0.739	0.683
Polyn. order	2	2	2	2	2	2
Bandwidth	7.089	3.994	4.887	5.491	5.776	5.808
Outcome mean	-0.383	-0.221	0.574	0.047	0.162	0.271
			Panel B: Left candidates			
Treatment	0.833**	-0.173	-0.184	-0.378***	0.043	0.149
	(0.414)	(0.129)	(0.140)	(0.141)	(0.285)	(0.404)
Observations	923	923	923	923	923	923
Eff. number of obs	139	129	157	124	148	178
Robust p-value	0.040	0.155	0.188	0.017	0.774	0.754
Polyn. order	2	2	2	2	2	2
Bandwidth	3.774	3.447	4.117	3.277	3.933	4.587
Outcome mean	-0.338	-0.207	0.572	0.051	0.259	0.231
			Panel C: Right candidates			
Treatment	-0.135	-0.051	-0.173	-0.013	0.056	0.063
	(0.172)	(0.032)	(0.248)	(0.304)	(0.204)	(0.237)
Observations	1561	1561	1561	1561	1561	1561
Eff. number of obs	423	436	453	543	573	603
Robust p-value	0.342	0.138	0.469	0.870	0.742	0.885
Polyn. order	2	2	2	2	2	2
Bandwidth	4.385	4.559	4.775	5.805	6.198	6.434
Outcome mean	-0.351	-0.228	0.566	0.060	0.138	0.292

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman being present in the second round in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 2, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

Table D30: Impact of a marginal presence of a woman on male candidates in the 2nd round (restricting to races with only two candidates)

	(1) Number words	(2) Tone	(3) Left right	(4) Similar to party	(5) Similar to women
	Panel A: All candidates				
Treatment	-0.072	-0.075	-0.171	0.266*	0.243
	(0.166)	(0.152)	(0.106)	(0.155)	(0.153)
Observations	2271	2271	2271	2207	2049
Eff. number of obs	435	565	395	561	569
Robust p-value	0.650	0.752	0.092	0.168	0.145
Polyn. order	1	1	1	1	1
Bandwidth	3.718	4.565	3.220	4.653	5.130
Outcome mean	-0.152	0.245	0.326	-0.066	0.695
	Panel B: Left candidates				
Treatment	1.702***	-0.564	-0.277	-0.259	-0.397
	(0.458)	(0.420)	(0.284)	(0.436)	(0.234)
Observations	828	828	828	824	666
Eff. number of obs	75	92	69	75	46
Robust p-value	0.000	0.210	0.256	0.363	0.044
Polyn. order	1	1	1	1	1
Bandwidth	2.340	2.920	2.192	2.321	1.895
Outcome mean	-0.089	0.251	0.320	-0.079	0.700
	Panel C: Right candidates				
Treatment	-0.271	-0.073	-0.183*	0.235	0.292
	(0.172)	(0.176)	(0.093)	(0.171)	(0.185)
Observations	1432	1432	1432	1373	1383
Eff. number of obs	334	386	363	334	327
Robust p-value	0.106	0.818	0.061	0.226	0.133
Polyn. order	1	1	1	1	1
Bandwidth	3.849	4.333	4.087	3.895	3.852
Outcome mean	-0.170	0.241	0.307	-0.068	0.701

Robust standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. The dependent variables of columns 2-8 are standardized. In column 1, the outcome is the number of votes obtained divided by the number of voters enrolled in the district. In column 2, the outcome is the number of words used in the manifesto. For columns 3 and 4, the procedure to obtain the outcome is explained in 5.1. The methodology to obtain column 5 is explained in 5.2. The methodology to obtain columns 6-9 is explained in 5.4.

Table D31: Impact of a marginal presence of a woman on male candidates in the 2nd round (restricting to races with only two candidates)

	(1) Economy & employment	(2) Environment	(3) Health & education	(4) Security & foreign policy	(5) Local politics	(6) National politics
			Panel A: All candidates			
Treatment	0.036	-0.016	-0.076	0.034	-0.013	0.102
	(0.120)	(0.027)	(0.138)	(0.152)	(0.144)	(0.192)
Observations	2271	2271	2271	2271	2271	2271
Eff. number of obs	514	525	561	795	565	459
Robust p-value	0.710	0.428	0.627	0.885	0.889	0.775
Polyn. order	1	1	1	1	1	1
Bandwidth	4.158	4.247	4.495	6.413	4.554	3.840
Outcome mean	-0.345	-0.226	0.582	0.069	0.204	0.229
			Panel B: Left candidates			
Treatment	0.663	-0.092	-0.239*	-0.154*	0.223	-0.078
	(0.405)	(0.105)	(0.133)	(0.079)	(0.327)	(0.407)
Observations	828	828	828	828	828	828
Eff. number of obs	90	59	63	59	59	68
Robust p-value	0.087	0.298	0.060	0.105	0.490	0.753
Polyn. order	1	1	1	1	1	1
Bandwidth	2.708	1.733	1.986	1.843	1.742	2.123
Outcome mean	-0.346	-0.179	0.680	-0.007	0.242	0.246
			Panel C: Right candidates			
Treatment	-0.064	-0.030	-0.123	0.040	0.057	0.078
	(0.116)	(0.026)	(0.183)	(0.213)	(0.168)	(0.240)
Observations	1432	1432	1432	1432	1432	1432
Eff. number of obs	377	296	363	432	383	290
Robust p-value	0.668	0.244	0.547	0.958	0.600	0.838
Polyn. order	1.000	1.000	1.000	1.000	1.000	1.000
Bandwidth	4.257	3.336	4.092	4.938	4.299	3.235
Outcome mean	-0.346	-0.207	0.583	0.056	0.217	0.245

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (a woman being present in the second round) is instrumented by the assignment variable. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

E After elections: Do elected females debate differently from elected males?

E.1 Legislative debates

The Conference of Presidents sets the time allocated to groups and independent MPs. The speaking time of the committees and the Government is not limited. As time is globalized, most speeches are not subject to any time limit (such as speeches on procedural motions, an article, or an amendment). All speeches by MPs are deducted from the group's time. Parliamentary speeches are essential to signal policy stands to electors, but they are also strategically used to mobilize media attention regarding specific topics.

E.1.1 List of seeded words per topic - Debates

Justice: justice (*justice*), sentence (*peine*), judiciary (*judiciaire*), minister justice (*garde sceaux*).

Health: health (*santé*), medical (*médical*), doctor (*médecin*), hospital (*hôpital*), treatment (*soin*), medicament (*médicament*), patient (*patient*), health insurance (*assurance maladie*), disease (*maladie*), sanitary (*sanitaire*), cancer (*cancer*), hospitable (*hospitalier*), medecine (*médecine*), sick (*malade*).

Employment: employment (*emploi*), work (*travail*), employee (*salarié*), employer (*employeur*), unemployment (*chômage*), wage (*salaire*), worker (*travailleur*).

Taxes: fiscal (*fiscal*), tax (*impôt*), tax (*taxe*), taxation (*fiscalité*), finance (*finance*).

Social security: retirement (*retraite*), reform (*réforme*), contribution (*cotisation*), social security (*sécurité social*), pension (*pension*).

Agriculture: agricultural (*agricole*), agriculture (*agriculture*), farmer (*agriculteur*), alimentary (*alimentaire*).

Environment: energy (*énergie*), environment (*environnement*), water (*eau*), ecological (*écologique*), energetic (*énergétique*), nuclear (*nucléaire*), electricity (*électricité*), green (*environnemental*), vehicle (*véhicule*), ecology (*écologie*), gas (*gaz*), pollution (*pollution*).

Foreign policy: european (*européen*), europe (*europe*), european union (*union européen*), international (*international*), worldwide (*mondial*), germany (*allemagne*), treaty (*traité*), united states (*états unis*), german (*allemand*), world (*monde*).

Finance: budget (*budget*), spending (*dépense*), budgetary (*budgétaire*), deficit (*déficit*).

Security: safety (*sécurité*), insecurity (*insécurité*), police (*police*), police (*policier*), violence (*violence*).

Education: child (*enfant*), school (*école*), young (*jeune*), student (*élève*), academic (*scolaire*), parent (*parent*), formation (*formation*), teacher (*enseignant*), research (*recherche*), teaching (*enseignement*), national education (*éducation national*), student (*étudiant*), university (*université*), class (*class*), professor (*professeur*).

Local: territory (*territoire*), department (*departement*), region (*région*), collectivity (*collectivité*), local (*local*), territorial (*territorial*), territorial collectivity (*collectivité territorial*), inhabitant (*habitant*), regional (*régional*), departmental (*départemental*), territorial planning (*aménagement territorial*), city (*ville*), municipal (*municipal*), rural (*rural*), public service (*service public*).

E.1.2 Top 10 words per topic

Economy: firm (*entreprise*), economy (*économie*), public (*public*), competition (*concurrence*), service (*service*), industrial (*industriel*), industry (*industrie*), business (*commerce*), market (*marché*), society (*société*).

Justice: law (*droit*), justice (*justice*), propose (*proposer*), provision (*disposition*), case (*cas*), commission (*commission*), act (*agir*), give (*donner*), sentence (*peine*), provide (*prevoir*).

Health: health (*santé*), treatment (*soin*), sanitary (*sanitaire*), medical (*médical*), doctor (*médecin*), disease (*maladie*), hospital (*hôpital*), patient (*patient*), hospitable (*hospitalier*), health insurance (*assurance maladie*).

Employment: work (*travail*), employment (*emploi*), employee (*salarié*), firm (*entreprise*), social (*social*), unemployment (*chômage*), wage (*salaire*), employer (*employeur*), contract (*contrat*), worker (*travailleur*).

Taxes: fiscal (*fiscal*), tax (*impôt*), finance (*finance*), tax (*taxe*), euro (*euro*), firm (*entreprise*), rate (*taux*), taxation (*fiscalité*), revenue (*revenu*), measure (*mesurer*).

Social security: reform (*réform*), retirement (*retraite*), social (*social*), social security (*sécurité social*), contribution (*cotisation*), person (*personne*), pension (*pension*), regime (*régime*), system (*système*), age (*âge*).

Agriculture: agricultural (*agricole*), agriculture (*agriculture*), farmer (*agriculteur*), public (*public*), service (*service*), product (*produit*), market (*marché*), alimentary (*alimentaire*), price (*prix*), sector (*secteur*).

Environment: energy (*énergie*), environment (*environnement*), water (*eau*), energetic (*énergétique*), transport (*transport*), gas (*gaz*), green (*environnemental*), nuclear (*nucléaire*), ecology (*écologie*), electricity (*électricité*).

Foreign policy: european (*européen*), world (*monde*), france (*france*), europe (*europe*), country (*pays*), french (*français*), international (*international*), european union (*union européen*), politics (*politique*), union (*union*).

Finance: budget (*budget*), spending (*dépense*), euro (*euro*), budgetary (*budgétaire*), million (*million*), public (*public*), deficit (*déficit*), credit (*crédit*), billion (*milliard*), year (*année*).

Internal security: security (*sécurité*), police (*police*), law (*droit*), violence (*violence*), person (*personne*), against (*contre*), fight (*lutter*), judge (*juger*), freedom (*liberté*), penal (*pénal*).

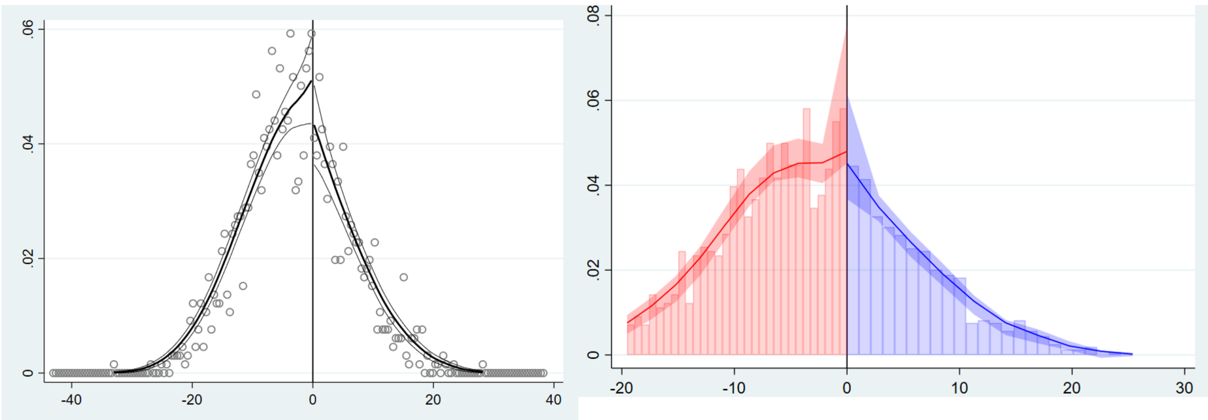
Education: child (*enfant*), young (*jeune*), formation (*formation*), school (*école*), student (*élève*), woman (*femme*), parent (*parent*), academic (*scolaire*), teaching (*enseignement*), student (*étudiant*).

Local: territory (*territoire*), collectivity (*collectivité*), local (*local*), region (*région*), department (*département*), territorial (*territorial*), city (*ville*), regional (*régional*), municipality (*commune*), public service (*service public*).

Other: want (*vouloir*), see (*voir*), no (*non*), nothing (*rien*), hour (*heure*), propose (*proposer*), thing (*chose*), politics (*politique*), understand (*comprendre*), today (*aujourd*).

Figure E.7: Manipulation testing

(a) McCrary (2008) (b) Cattaneo et al. (2018)



Notes. This figure tests for a jump in the density of the running variable. The solid line represents the density of the running variable. Thin lines represent the confidence intervals. Figures (a) and (b) represent the density test for mixed-gender races where a woman wins against the second most-voted man. Figure (a) represents the McCrary density test; discontinuity estimate b: -0.155 (s.e. 0.116). Figure (b) represents the Cattaneo et al. (2018) manipulation test; p-value 0.258 (not reject the null hypothesis of no manipulation).

E.1.3 Balance tests

I conduct placebo tests to examine whether there is a discontinuity at the threshold for any variables used to predict treatment. Data about the legislative debates is between 1998 and 2022. Given that platforms for the 2002, 2007, and 2012 elections are missing, I do not run balance tests for the platforms' characteristics.

Table E32: Balancing tests (1st round)

	(1) Votes	(2) Turnout	(3) Number candidates	(4) Number female	(5) Enrolled voters	(6) Victory margin
Woman	0.636	1.615	1.013**	0.534	-2,998	-0.210
	(0.888)	(1.473)	(0.473)	(0.336)	(2,265)	(0.521)
Observations	1319	1319	1319	1319	1319	1319
Eff. number of obs	569	540	796	587	594	525
Robust p-value	0.431	0.286	0.048	0.176	0.253	0.715
Polyn. order	1	1	1	1	1	1
Bandwidth	5.193	4.986	7.594	5.336	5.458	4.810
Outcome mean	20.210	57.353	12.378	5.185	77160	6.425

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman being elected in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

Table E33: Balancing tests - electoral district characteristics (1st round)

	(1) Number far-left	(2) Number left	(3) Number right	(4) Number far-right	(5) Sum left	(6) Sum right
Woman	-0.041	0.491**	0.217	0.147	1.497	1.388
	(0.167)	(0.201)	(0.163)	(0.130)	(1.306)	(1.342)
Observations	1319	1319	1319	1319	1319	1319
Eff. number of obs	685	640	696	767	586	637
Robust p-value	0.746	0.023	0.195	0.273	0.210	0.342
Polyn. order	1	1	1	1	1	1
Bandwidth	6.457	5.878	6.579	7.323	5.330	5.849
Outcome mean	2.208	3.242	1.847	2.088	22.107	25.112

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman being elected in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 1, and the optimal bandwidths are derived under the MSERD procedure.

E.1.4 Results

To ensure comparability with the results in Section 6, I sum topics employment, taxes, social security and finance, and refer to them as economy & employment. I sum topics agriculture and environment, and refer to them as environment. Similarly, I sum topics health & education. Finally, I also sum security, justice and foreign policy and refer to them as security & foreign policy. I classify a document as a specific topic if the highest value refers to that topic.

Table E34: Differences between female and male MPs during parliamentary work - legislative debates (polynomial of order 2)

	(1) Debates	(2) Economy & employment	(3) Environment	(4) Health & education	(5) Security & foreign policy	(6) Local politics
			Panel A: All politicians			
Woman	-0.044	0.006	-0.067	0.842***	-0.221	-0.094
	(0.144)	(0.198)	(0.218)	(0.233)	(0.204)	(0.199)
Observations	1319	1319	1319	1319	1319	1319
Eff. number of obs	740	859	814	669	818	828
Robust p-value	0.913	0.978	0.675	0.000	0.349	0.618
Polyn. order	2	2	2	2	2	2
Bandwidth	7.049	8.478	7.876	6.331	7.971	8.062
Outcome mean	0.028	0.004	-0.021	0.061	-0.016	-0.050
			Panel B: Left politicians			
Woman	0.148	-0.171	0.155	1.202***	-0.406	-0.509*
	(0.280)	(0.449)	(0.277)	(0.404)	(0.352)	(0.282)
Observations	501	501	501	501	501	501
Eff. number of obs	231	272	327	217	261	312
Robust p-value	0.524	0.716	0.572	0.004	0.228	0.085
Polyn. order	2	2	2	2	2	2
Bandwidth	6.742	8.142	10.075	6.436	7.727	9.527
Outcome mean	0.028	0.004	-0.021	0.061	-0.016	-0.050
			Panel C: Right politicians			
Woman	0.051	0.018	-0.343	0.761**	-0.038	0.256
	(0.157)	(0.271)	(0.348)	(0.339)	(0.348)	(0.347)
Observations	794	794	794	794	794	794
Eff. number of obs	326	439	349	444	459	342
Robust p-value	0.566	0.988	0.282	0.039	0.964	0.401
Polyn. order	2	2	2	2	2	2
Bandwidth	4.509	6.209	4.935	6.378	6.544	4.775
Outcome mean	0.028	0.004	-0.021	0.061	-0.016	-0.050

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman being elected in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 2, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.

E.2 Written questions

A deputy writes written questions to a minister; only those relating to the government's general policy are made to the prime minister. They must be summarily drafted and be limited to the elements strictly essential to understanding the question. They must not contain any imputation of a personal nature concerning third parties. In addition, the principle of the separation of powers and the irresponsibility of the Head of State prohibits the author of a written question from questioning the acts of the President of the Republic. There is no limit to the number of questions or their timing.

Written questions are available on the National Assembly website.[21] I web-scraped their content, the author, the ministry interrogated, and the day of the question.

Example of written question: Question no 104122 of the 14th legislature (2012-2017). Mr. Yves Jégo draws the attention of the Minister of State, Minister of the Interior, regarding the access to the Parisian telecom galleries which, over more than 45 kilometres, are accessible by unsecured hatches. In times of terrorist attacks, the absence of drastic security measures for these underground networks seems to him to be a particularly proven risk, particularly around the Ministry of the Interior. He would like to know what measures the Government intends to take within the framework of the Vigipirate plan to guarantee that this network is inaccessible to unauthorized persons.

E.2.1 List of seeded words per topic - Questions

Economy: firm (*entreprise*), economy (*économie*), industry (*industrie*), retail (*commerce*), craft (*artisanat*).

Security & justice: justice (*justice*), sentence (*peine*), judiciary (*judiciaire*), minister justice (*garde sceaux*), court (*tribunal*), prison (*pénitentiaire*), lawyer (*avocat*), security (*sécurité*), insecurity (*insécurité*), police (*police*), police (*policier*), violence (*violence*), terrorism (*terrorisme*), terrorist (*terroriste*), victim (*victime*), military (*militaire*).

Health: health (*santé*), medical (*médical*), doctor (*médecin*), hospital (*hôpital*), treatment (*soin*), medicament (*médicament*), patient (*patient*), health insurance (*assurance maladie*), disease (*maladie*), sanitary (*sanitaire*), cancer (*cancer*), hospitable (*hospitalier*), medecine (*médecine*), sick (*malade*).

Employment: employment (*emploi*), work (*travail*), employee (*salarié*), employer (*employeur*), unemployment (*chômage*), wage (*salaire*), worker (*travailleur*).

Social security: retirement (*retraite*), reform (*réforme*), contribution (*cotisation*), social security (*sécurité social*), pension (*pension*).

Agriculture: agricultural (*agricole*), agriculture (*agriculture*), farmer (*agriculteur*),

[21] https://www2.assemblee-nationale.fr/recherche/questions

alimentary (*alimentaire*).

Environment: energy (*énergie*), environment (*environnement*), water (*eau*), ecological (*écologique*), energetic (*énergétique*), nuclear (*nucléaire*), electricity (*électricité*), green (*environnemental*), vehicle (*véhicule*), ecology (*écologie*), gas (*gaz*), pollution (*pollution*).

Foreign policy: european (*européen*), europe (*europe*), european union (*union européen*), international (*international*), worldwide (*mondial*), germany (*allemagne*), treaty (*traité*), united states (*états unis*), german (*allemand*), world (*monde*).

Finance: budget (*budget*), spending (*dépense*), budgetary (*budgétaire*), deficit (*déficit*), fiscal (*fiscal*), tax (*impôt*), tax (*taxe*), taxation (*fiscalité*), finance (*finance*).

Education: child (*enfant*), school (*école*), young (*jeune*), student (*élève*), academic (*scolaire*), parent (*parent*), formation (*formation*), teacher (*enseignant*), research (*recherche*), teaching (*enseignement*), national education (*éducation national*), student (*étudiant*), university (*université*), class (*class*), professor (*professeur*).

Local: territory (*territoire*), department (*departement*), region (*région*), collectivity (*collectivité*), local (*local*), territorial (*territorial*), territorial collectivity (*collectivité territorial*), inhabitant (*habitant*), regional (*régional*), departmental (*départemental*), territorial planning (*aménagement territorial*), city (*ville*), public service (*service public*).

E.2.2 Top 10 words per topic

Economy: firm (*entreprise*), economy (*économie*), industry (*industrie*), retail (*commerce*), craft (*artisanat*), minister economy (*ministre economie*), small (*petit*), sector (*secteur*), activity (*activité*), economic (*économique*).

Security & justice: security (*sécurité*), justice (*justice*), report (*rapport*), victim (*victime*), military (*militaire*), minister justice (*garde sceau*), veteran (*ancien*), proposition (*proposition*), state (*état*), fighter (*combattant*).

Health: health (*santé*), disease (*maladie*), treatment (*soin*), medical (*médical*), doctor (*médecin*), patient (*patient*), sanitary (*sanitaire*), social (*social*), hospitable (*hospitalier*), medicament (*médicament*).

Employment: work (*travail*), employment (*emploi*), employee (*salarié*), person (*personne*), social (*social*), help (*aider*), disability (*handicap*), professional (*professionnel*), employer (*employeur*), situation (*situation*).

Social security: retirement (*retraite*), reform (*réforme*), pension (*pension*), social (*social*), social security (*sécurité social*), contribution (*cotisation*), capacity (*pouvoir*), situation (*situation*), system (*régime*), take (*prendre*).

Agriculture: agriculture (*agriculture*), agricultural (*agricole*), farmer (*agriculteur*), minister agriculture (*ministre agriculture*), fishing (*pêche*), product (*produit*), alimentary (*alimentaire*), rural (*rural*), sector (*filière*), government (*gouvernement*).

Environment: energy (*énergie*), transport (*transport*), vehicle (*véhicule*), environment (*environnement*), ecology (*écologie*), water (*éau*), development (*developpement*), durable (*durable*), permit (*permettre*), equipment (*équipement*).

Foreign policy: european (*européen*), france (*france*), french (*français*), country (*pays*), international (*international*), foreigner (*étranger*), law (*droit*), state (*état*), european union (*union européen*), air france (*af*).

Finance: finance (*finance*), budget (*budget*), fiscal (*fiscal*), tax (*taxe*), tax (*impôt*), euro (*euro*), housing (*logement*), spending (*dépense*), rate (*taux*), budgetary (*budgétaire*).

Education: child (*enfant*), formation (*formation*), teaching (*enseignement*), national education (*éducation national*), young (*jeune*), research (*recherche*), student (*élève*), national (*national*), academic (*scolaire*), education (*éducation*).

Local: territory (*territoire*), collectivity (*collectivité*), public (*public*), local (*local*), territorial (*territorial*), department (*département*), service (*service*), region (*région*), municipality (*commune*), territorial collectivity (*collectivité territorial*).

Other: law (*loi*), article (*article*), decree (*décret*), application (*application*), provision (*disposition*), legislation (*code*), may (*pouvoir*), relative (*relatif*), provide (*prevoir*), law (*droit*).

E.2.3 Results

To ensure comparability with the results in Section 6, I sum topics economy, employment, social security and finance, and refer to them as economy & employment. I sum topics agriculture and environment, and refer to them as environment. Similarly, I sum topics health & education. Finally, I also sum topics security & justice and foreign policy and refer to them as security & foreign policy. I classify a document as about a specific topic if the highest value refers to that topic.

Table E35: Differences between female and male MPs during parliamentary work - written questions

	(1) Questions	(2) Economy & employment	(3) Environment	(4) Health & education	(5) Security & foreign policy
		Panel A: All politicians			
Woman	-0.188	-0.364**	-0.208	0.537***	0.059
	(0.140)	(0.122)	(0.168)	(0.167)	(0.170)
Observations	1436	1436	1436	1436	1436
Eff. number of obs	680	766	675	808	735
Robust p-value	0.240	0.009	0.230	0.005	0.738
Polyn. order	1	1	1	1	1
Bandwidth	5.787	6.635	5.743	7.017	6.398
Outcome mean	-0.002	-0.149	0.111	0.020	0.008
		Panel B: Left politicians			
Woman	-0.097	-0.476**	-0.151	0.769***	-0.098
	(0.263)	(0.227)	(0.260)	(0.295)	(0.306)
Observations	544	544	544	544	544
Eff. number of obs	277	267	273	235	232
Robust p-value	0.674	0.059	0.613	0.011	0.621
Polyn. order	1	1	1	1	1
Bandwidth	7.163	6.908	7.022	5.921	5.883
Outcome mean	-0.002	-0.149	0.111	0.020	0.008
		Panel C: Right politicians			
Woman	-0.191	-0.341*	-0.208	0.448**	0.143
	(0.162)	(0.183)	(0.260)	(0.225)	(0.297)
Observations	868	868	868	868	868
Eff. number of obs	439	366	307	470	362
Robust p-value	0.328	0.148	0.487	0.110	0.586
Polyn. order	1	1	1	1	1
Bandwidth	5.745	4.819	3.910	6.271	4.733
Outcome mean	-0.002	-0.149	0.111	0.020	0.008

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significance at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest is a woman being elected in alternative to a man. Separate polynomials are fitted on each side of the threshold. The polynomial order is 2, and the optimal bandwidths are derived under the MSERD procedure. All dependent variables are standardized.