

When Men Run Against Women: Political Speeches During Campaigns

Marli Fernandes*

Nova School of Business and Economics

Preliminary and incomplete
August 5, 2023

Abstract

How the entry of a woman changes a political campaign? I combine computational text analysis with a regression discontinuity design in two-round French legislative elections, to study how male politicians adapt their political program when they directly compete with a woman. I find that male politicians tend to change the topics covered during the campaign, although left and right-wing politicians tend to do it in different ways.

1 Introduction

In 2022, women were 27% of the candidates in US primaries for Senate, House or governor.¹ In the same year, women were 44% of the candidates in French legislative elections. Despite the substantial increase in female political participation over the past 40 years, women are still underrepresented in political campaigns. A large literature has been advancing several hypotheses for females under-representation in politics. However, before women can influence the outcome of an election, they are likely to affect the way in which candidates conduct their campaigns, the ideas that they highlight and how they describe their opponent.

This article is the first to study the effects of female presence on political competitor's discourse. I focus on two-round French legislative elections. The research exploits the fact that in both electoral districts, where women barely make to the second round and those who just miss the qualification threshold are arguably

*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 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.

¹Source: Center for American Women and Politics, State Election Officials

comparable in observed and unobserved characteristics. I combine a regression discontinuity design with computational text analysis to investigate whether the female presence affected the content of political opponents' communication during the election.

I use a comprehensive candidates manifestos dataset 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 manifestos, which the State prints and sends to all registered voters a few days before the election. Politicians use these manifestos to provide information to their constituents about themselves, their program and appeal to vote.

I find that left-wing male politicians tend to talk more about the economy and less on health and education when they compete with a woman. On the other side, right-wing male politicians tend to talk more about politics, but less on environment, health and education and local issues; furthermore, they tend to use a more negative political platform.

2 Literature Review

These results contribute to several strands of the literature. First, the paper contributes to the large 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), Kanthak and Woon (2015)). 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 is 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 in tournaments individuals tend to compete less with same-sex opponents. 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, differences in competition by gender during political campaigns.

The second strand relates to the entry of women into politics. A large literature analysis the effects on gender quotas on the quality of politicians and poli-

cymakers and find a significant increase in male's competence (Baltrunaite, Bello, Casarico, and Profeta (2014), Besley, Folke, Persson, and Rickne (2017), Júlio and Tavares (2017)) and diversity in office (Barnes and Holman (2020)). Other studies demonstrate that gender quotas have limited effects when voters are biased against woman (Le Barbanchon and Sauvagnat (2022)), parties nominate women in less winnable districts (Lippmann (2021)) or because women tend to have a higher chance of resign (Lassébie (2020)). I contribute to this literature by showing that men tend to adapt to the presence of a women in a campaign.

Third, the paper is also related to the impact of politicians' identity. The Median Voter Theorem (Downs (1957)) assumes that the parties' unique objective is to win elections, and as a consequence, 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 et al. (2020) and Di Tella, Kotty, Le Pennec, and Pons (2023) demonstrate empirically the convergence mechanism underlying the median voter theorem. However, candidates only converge to a certain extent. 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.

3 Institutional Context

3.1 French Parliamentary Elections

The sample includes parliamentary elections. The parliamentary elections elects 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 candidate votes and 25 percent of the registered citizens. In most part of the elections, no candidate is able to be elected in the first round, and a second round is held one week later. The second round is decided by simple plurality: the candidate that obtains 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 around. However, if only one of the candidates (or none) meets the threshold, the two candidates that obtain the largest vote share are allowed to proceed to the second round.

3.2 Candidate Manifestos

Electoral manifestos (*trans. professions de foi*) are a document produced by a candidate for an election to present his program and ideas, that he plans to commit once elected. They are an important part of French electoral campaigns, and together with the posters and ballots they represent one of the three main parts of official electoral advertising.

Candidates are responsible for printing these manifestos, and their costs can be reimbursed by the state 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 for the second round (if it happens).²

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.

4.2 Candidate Manifestos

Candidate manifestos for the 1988 and 1993 elections were digitized by the Arch-elec project (Gaultier-Voituriez (2016)).³ Until 1993 the CEVIPOF collected manifestos each election with the support of the government. Each departmental administration mailed the manifestos distributed in their district. Unfortunately, this practise finished in 1993, and for that reason manifestos between 2002 and 2012 are not available.

Manifestos 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 in their website⁴, I manually digitized missing manifestos at CEVIPOF (Sciences Po), the National

²These manifestos are only mailed if they obey to these rules: they must have a maximum size of 210x297 millimeters and weight between 60 and 80 grams per square meter (Electoral law, article R29); they cannot combine the three colors of the French flag, except if they are part of the party's emblem.

³They are available at <https://archive.org/details/archiveselectoralesducevipof/>

⁴They are available at: <https://programme-candidats.interieur.gouv.fr/>

Assembly website also shares the manifestos of all elected MPs⁵, some missing manifestos were also found on several local news or candidates websites.

For 2022, manifestos come from the Ministry of Interior website, and several local news and candidates websites.

4.3 Legislative work

I collect information on the legislative work of MPs from two sources. First, I web-scrape legislative debates.⁶

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.⁷

For further details on these two data sources, see the Appendix A.2 and A.3.

4.4 Data Preparation

I pre-process manifestos' 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. Furthermore, I restrict the vocabulary to words used by at least 0.5% and 50% of the manifestos, for each election year separately.

5 Methodology

5.1 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 like 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));

⁵They are available at: <https://www.assemblee-nationale.fr/dyn/decouvrir-l-assemblee/histoire/barodet2/recueil-des-professions-de-foi-de-la-15eme-legislature>

⁶They are available at: <https://www.assemblee-nationale.fr/dyn/15/comptes-rendus/seance>

⁷They are available at: <https://www2.assemblee-nationale.fr/recherche/questions>

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, Butler, Preece, 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 clearly endogenous, and they also might affect the type of women present in the race and how opponents respond to their presence. If one wants to study how male politicians respond to the entry of a woman into a political campaign, a quasi-experiment is necessary to infer causality.

To answer this question, 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 for a small margin. At the cutoff, the female presence is orthogonal to voters’ and politicians’ characteristics.

Running variable. The running variable is the margin of qualification. 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 the 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.

Evaluation framework. Qualified third candidates can drop off the race between the two rounds, making the regression discontinuity design fuzzy. I define the running variable R as the qualifying margin of the candidate in the first round, the assignment variable D as a dummy equal to 1 if a woman qualifies for the second round ($R \geq 0$) and 0 otherwise ($R < 0$), and the treatment variable T as a dummy equal to 1 if a woman is present in the second round and 0 otherwise. I call compliers the districts in which a woman qualifies ($D = 1$) and runs in the second round ($T = 1$). I evaluate the impact of a woman being present in complier

districts with the following specification:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 R_i + \beta_3 R_i T_i + \beta_4 X_i + \mu_i \quad (1)$$

where Y_i is the outcome of interest in district i . X_i is a vector of first-round independent variables.⁸. These controls are not necessary for identification but improve efficiency. T_i is instrumented with D_i as shown in the following first-stage equation:

$$T_i = \alpha_0 + \alpha_1 D_i + \alpha_2 R_i + \alpha_3 R_i D_i + \alpha_4 X_i + \epsilon_i \quad (2)$$

Following Calonico, Cattaneo, Farrell, and Titiunik (2019), I use a non-parametric approach, which amounts to fitting two linear regressions on districts close to the left and right of the threshold. I use the mean squared error optimal bandwidths selection procedure proposed by Calonico, Cattaneo, and Titiunik (2014) in its covariate-adjusted version (Calonico et al. (2019)). This procedure is data-driven, implying that bandwidth size varies with the outcome under consideration.⁹

Identifying assumption and robustness. The identifying assumption is that the win margin of qualification cannot be precisely manipulated. To assess its plausibility, I first test for continuity of the density of the running variable. Intuitively, if election outcomes were manipulated, the distribution of the running variable would be discontinuous at the threshold, with an excess mass to its right. I test the validity of the assumption, using McCrary (2008) test and checking 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. As Figure B.1 in the Appendix demonstrates there is no jump in any of the cutoffs.

Second, the first stage is very strong. Figure B.2 in the Appendix plots the treatment against the running variable.

Third, the main implication of the identifying assumption is that districts' characteristics are continuous at the threshold. I run balancing tests for: number of enrolled voters, candidates, turnout, sum of votes for left and right-wing candidates in the first round and, importantly, for the dependent variables. Tables C4, B2 and B3 show the regression discontinuity estimates of the effect of having a woman in the second round on first round electoral district characteristics. Of the 22 regressions, the coefficient on the treatment variable is significant at 10% level in only one case, which is in the expected range. Pre-treatment characteristics of the districts are balanced.

⁸Control variables are presented in Tables C4, B2 and B3

⁹I use the Stata package `rdrobust` (Calonico, Cattaneo, Farrell, and Titiunik (2017)).

5.2 Topic classification

In my analysis, I try to understand if the entry of a woman into campaign affects the topics covered by male politicians during the campaign. To compute the distribution of topics over each document, I rely on topic modelling techniques, 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 the words observed in the document of a corpus are generated by latent topics. 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 is generally produced by LDA.

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

In the Appendix section B I provide further details on the method, the seed words and the top words for each topic.

5.3 Measuring tone

I compute a quantitative measure of the sentiment using a dictionary-based approach.¹⁰ The tone of each manifesto is computed as follows:

$$Tone_i = \frac{W_i^P - W_i^N}{W_i^T} \quad (3)$$

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 number of total words within manifesto i .

To perform the tone assessment, I use the LSDFr dictionary (Duval and Pétry (2016)), which was created by political communication researchers.¹¹

¹⁰In the case of tone, I do not restrict my analysis to words used by at least 0.5% and 50% of the documents.

¹¹The authors also show that their dictionary over-performs the LIWC (Piolat, Booth, Chung,

In the Appendix, I demonstrate that radical parties tend to have more negative manifestos as expected (Table D1) and show the most positive and negative manifesto (Figures D.1 and D.2).

6 Results

Table 1: Impact of the presence of a woman in the 2nd round - all parties

| | (1) No. words | (2) Economy & Employment | (3) Environment | (4) Health & Education | (5) Local | (6) Security & Foreign policy | (7) Politics | (8) Tone |
|--------------------|--------------------|--------------------------------|--------------------|------------------------------|-------------------|-------------------------------------|------------------|--------------------|
| Treatment | 16.387 (35.696) | 0.022 (0.014) | -0.006* (0.003) | -0.004 (0.025) | -0.026 (0.020) | -0.003 (0.017) | 0.030 (0.026) | -0.222* (0.134) |
| Observations | 2568 | 2568 | 2568 | 2568 | 2568 | 2568 | 2568 | 2568 |
| Eff. number of obs | 670 | 652 | 449 | 593 | 852 | 607 | 723 | 692 |
| Robust p-value | 0.605 | 0.213 | 0.082 | 0.755 | 0.279 | 1.000 | 0.323 | 0.106 |
| Polyn. order | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Bandwidth | 4.116 | 4.017 | 2.609 | 3.649 | 5.477 | 3.769 | 4.472 | 4.281 |
| Outcome mean | 624.875 | 0.107 | 0.023 | 0.136 | 0.189 | 0.105 | 0.185 | 0.255 |

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significant at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (the woman presence 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.

6.1 Heterogeneity

Table 2: Impact of the presence of a woman in the 2nd round - all left-wing parties

| | (1) No. words | (2) Economy & Employment | (3) Environment | (4) Health & Education | (5) Local | (6) Security & Foreign policy | (7) Politics | (8) Tone |
|--------------------|------------------------|--------------------------------|--------------------|------------------------------|-------------------|-------------------------------------|------------------|-------------------|
| Treatment | 596.265** (234.309) | 0.132*** (0.038) | -0.009 (0.010) | -0.016** (0.007) | -0.020 (0.031) | 0.005 (0.010) | 0.012 (0.054) | -0.113 (0.222) |
| Observations | 931 | 931 | 931 | 931 | 931 | 931 | 931 | 931 |
| Eff. number of obs | 102 | 140 | 117 | 150 | 209 | 146 | 168 | 151 |
| Robust p-value | 0.006 | 0.002 | 0.365 | 0.026 | 0.525 | 0.650 | 0.848 | 0.574 |
| Polyn. order | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Bandwidth | 1.981 | 2.724 | 2.271 | 3.115 | 4.420 | 2.992 | 3.731 | 3.127 |
| Outcome mean | 626.633 | 0.109 | 0.023 | 0.141 | 0.199 | 0.096 | 0.190 | 0.258 |

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significant at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (the woman presence 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.

Davids, and Pennebaker (2011)).

Table 3: Impact of the presence of a woman in the 2nd round - all right-wing parties

| | (1) No. words | (2) Economy & Employment | (3) Environment | (4) Health & Education | (5) Local | (6) Security & Foreign policy | (7) Politics | (8) Tone |
|--------------------|-----------------------|--------------------------------|----------------------|------------------------------|----------------------|-------------------------------------|--------------------|---------------------|
| Treatment | -118.436* (60.757) | -0.002 (0.017) | -0.012*** (0.004) | -0.010*** (0.003) | -0.094*** (0.031) | -0.021 (0.047) | 0.111** (0.055) | -0.562** (0.262) |
| Observations | 1258 | 1258 | 1258 | 1258 | 1258 | 1258 | 1258 | 1258 |
| Eff. number of obs | 234 | 315 | 195 | 178 | 314 | 254 | 261 | 231 |
| Robust p-value | 0.033 | 0.995 | 0.010 | 0.001 | 0.004 | 0.786 | 0.079 | 0.046 |
| Polyn. order | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Bandwidth | 3.240 | 4.372 | 2.755 | 2.483 | 4.359 | 3.501 | 3.717 | 3.216 |
| Outcome mean | 640.088 | 0.109 | 0.022 | 0.144 | 0.201 | 0.103 | 0.190 | 0.260 |

Standard errors are in parenthesis. Statistical significance is computed based on the robust p-value and ***, ** and * indicate significant at 1, 5 and 10, respectively. Each column reports the results from a separate local polynomial regression. The variable of interest (the woman presence 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.

7 Conclusion

References

- 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.
- Baltrunaite, A., Bello, P., Casarico, A., & Profeta, P. (2014). Gender quotas and the quality of politicians. *Journal of Public Economics*, 118, 62–74.
- Barber, M., Butler, D. M., Preece, J., et al. (2016). Gender inequalities in campaign finance. *Quarterly Journal of Political Science*, 11(2), 219–48.
- Barnes, T. D., & Holman, M. R. (2020). Gender quotas, women's representation, and legislative diversity. *The Journal of Politics*, 82(4), 1271–1286.
- 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.
- 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.
- 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.
- Besley, T., Folke, O., Persson, T., & Rickne, J. (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American economic review*, 107(8), 2204–2242.
- 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.
- Cagé, J., Le Pennec, C., & Mougin, E. (2021). Corporate donations and political

rhetoric: Evidence from a national ban.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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 ve république, du papier au numérique. *Histoire@ Politique*, 3, 213–220.
- 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.
- 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.
- 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.
- Júlio, P., & Tavares, J. (2017). The good, the bad and the different: can gender quotas raise the quality of politicians? *Economica*, 84(335), 454–479.
- Kanthak, K., & Woon, J. (2015). Women don't run? election aversion and candidate entry. *American journal of political science*, 59(3), 595–612.
- Lassébie, J. (2020). Gender quotas and the selection of local politicians: Evidence from french municipal elections. *European Journal of Political Economy*, 62, 101842.
- 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., et al. (2020). *Strategic campaign communication: Evidence from 30,000 candidate manifestos* (Tech. Rep.). Monash University, SoDa Laboratories.
- Lippmann, Q. (2021). Are gender quotas on candidates bound to be ineffective? *Journal of Economic Behavior & Organization*, 191, 661–678.
- 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 estimator. *Journal of Econometrics*, 142(2), 601–615.

- tinuity design: A density test. *Journal of econometrics*, 142(2), 698–714.
- 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.
- Pons, V., Tricaud, C., & Tricaud, C. (2019). The large effects of a small win: How past rankings shape the behavior of voters and candidates. *NBER working paper*(w26599).
- 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.
- 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 Data

A.1 Campaign manifestos

According to the Ipsos - CEVIPOF 2022 Presidential electoral survey, the main sources of candidates information are: 38% of the electors follow the television, 15% the internet (many candidates share their manifestos online), 13% newspapers, 12% the manifestos received in the mail box, 6% use other sources around them, and 16% do not inform. Relatively to the manifestos for the years 1988, 1993 and 1997, internet was not yet widely available and used by politicians. Furthermore, per election there are 577 constituencies and an average of xxxx candidates, making impossible for television to give coverage to all candidates. As a consequence, it is likely that in legislative elections manifestos have a bigger importance than in presidential elections.

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

Some politicians opt for using the party platform instead of a personalised platform, this is common in small parties (e.g. Rassemblement National and green parties), but a very rare practice among well-established parties (e.g socialist and republican party). I do not include manifestos that are very 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.

A.2 Legislative debates

I web-scraped the speeches from the Assemblée Nationale website covering the 1998-2022 period. I do not include the years of 2001 and end of the 11th legislature (2002), because the website during this period is design in a different way and is more difficult to webscrape. I restrict my analysis to politicians who were elected, 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)).¹² I also remove speeches with less than 5 words, these speeches tend to be small reactions to the speech of an opponent and are not informative in terms of group or topic. I aggregate speeches so that a document captures all speeches by

¹²I obtain the list of procedural phrases from the following websites: <https://www2.assemblee-nationale.fr/dcouvrir-l-assemblee/folder/lexique> and <https://www.assemblee-nationale.fr/connaissance/lexique.asp>.

Figure A.1: Florence Blatrix-Contat manifesto 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 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 A.2: Florence Blatrix-Contat manifesto 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



Véritable 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.

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

a given speaker in one agenda of a plenary session; I remove aggregate speeches with less than 15 words, very short speeches are not ideal to detect group differences or topics.

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

A.3 Written questions

Written questions are written by a deputy to a minister; only those relating to the general policy of the Government 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 with regard to 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 in the number of questions or in their timing.

Written questions are available at the National Assembly website.¹³ 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.

A.4 Text pre-processing

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

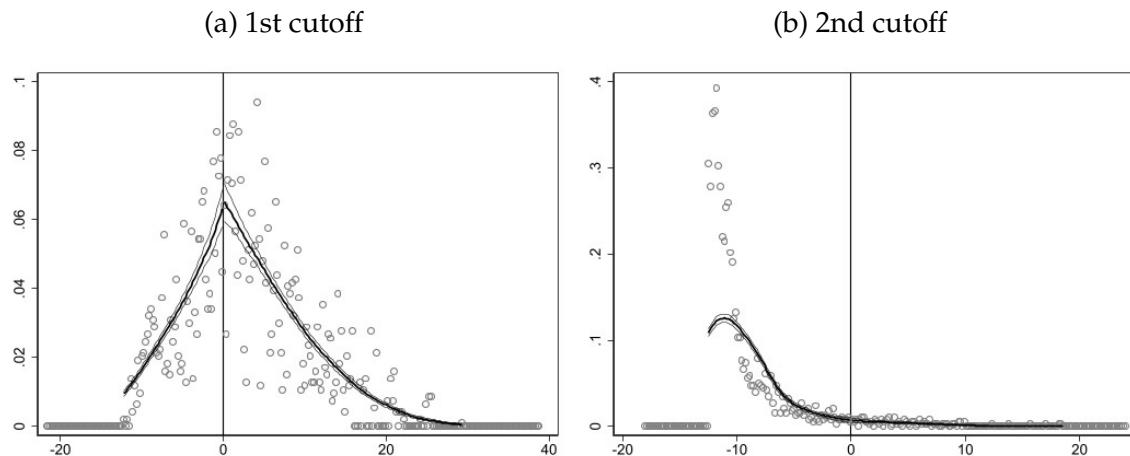
¹³<https://www2.assemblee-nationale.fr/recherche/questions>

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

I pre-process the content of the corpus following standard practices in natural language processing: remove punctuation, numbers, convert all letters to lower-case, lemmatize each word and restrict the vocabulary to words used by at least 0.5% and 50% of the documents using spacy french version 3.5.0. Last, I convert words from Latin-1 to UTF-8 for two reasons. Given that the OCR sometimes does not detect accents, it permits to reduce error; second, written questions for the 9th legislature are in UTF-8; third, it permits to save memory.

B Methodology

Figure B.1: McCrary test of the density of the running variable

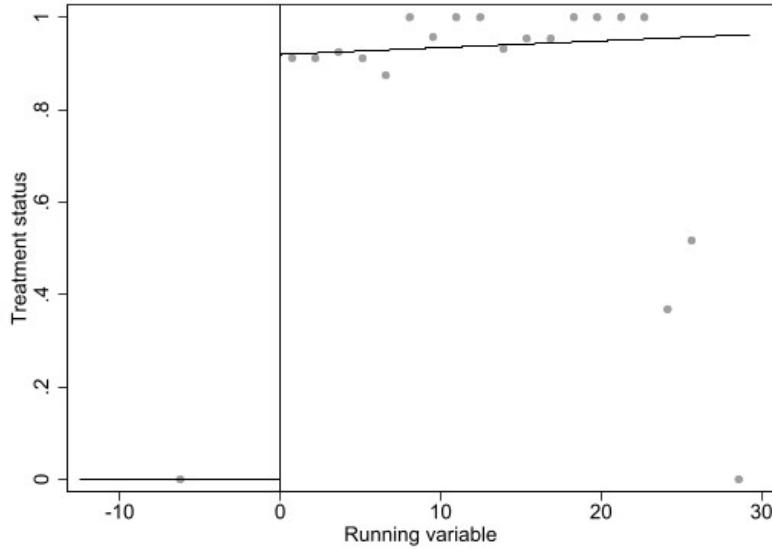


Notes. This figure tests for a jump in the density of the running variable. The solid line represents the density of the running variable. Thin lines represent the confidence intervals. Figure (a) represents the density test for races where a third candidate (woman) obtained the minimum 12.5% of votes (and two male candidates also passed the threshold). Discontinuity estimate b: 0.020 (s.e. 0.062). Figure (b) represents 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. Discontinuity estimate b: -0.072 (s.e. 0.337).

C 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

Figure B.2: 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 is measured as percentage points. Continuous lines are a linear fit.

Table B1: Balancing tests

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------------|-------------------|-------------------|------------------|------------------|-----------------------|-------------------|
| | Turnout | N. candidates | N. women | N. men | Enrolled voters | Victory margin |
| Treatment | -2.023 (2.326) | -0.118 (0.837) | 0.045 (0.478) | 0.091 (0.592) | -5,568 (3,967.110) | -0.130 (1.529) |
| Observations | 1998 | 1998 | 1998 | 1998 | 1998 | 1998 |
| Eff. number of obs | 576 | 375 | 495 | 353 | 499 | 604 |
| Robust p-value | 0.336 | 0.896 | 0.847 | 0.675 | 0.139 | 0.941 |
| Polyn. order | 1 | 1 | 1 | 1 | 1 | 1 |
| Bandwidth | 4.215 | 2.663 | 3.658 | 2.456 | 3.729 | 4.391 |
| Outcome mean | 54.112 | 12.438 | 4.240 | 8.141 | 78425 | 8.070 |

The table shows regression discontinuity estimates of the effect of having a woman in the second round of the election on first-round electoral district characteristics. Robust standard errors in parenthesis.

(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 $\theta \sim Dirichlet(\alpha)$.

Table B2: Balancing tests (continuation)

| | (1) N. far-left | (2) N. left | (3) N. center | (4) N. right | (5) N. far-right | (6) Nother | (7) All left | (8) All right |
|--------------------|--------------------|-------------------|------------------|------------------|---------------------|------------------|-------------------|-------------------|
| Treatment | 0.252 (0.281) | -0.368 (0.311) | 0.042 (0.155) | 0.165 (0.253) | -0.126 (0.202) | 0.126 (0.385) | -3.073 (1.869) | -0.007 (2.221) |
| Observations | 1998 | 1998 | 1998 | 1998 | 1998 | 1998 | 1998 | 1998 |
| Eff. number of obs | 351 | 529 | 488 | 432 | 468 | 482 | 434 | 636 |
| Robust p-value | 0.247 | 0.227 | 0.674 | 0.434 | 0.539 | 0.567 | 0.114 | 0.845 |
| Polyn. order | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Bandwidth | 2.447 | 3.885 | 3.563 | 3.117 | 3.404 | 3.482 | 3.150 | 4.740 |
| Outcome mean | 2.324 | 3.275 | 0.733 | 2.072 | 1.925 | 2.061 | 19.396 | 23.368 |

The table shows regression discontinuity estimates of the effect of having a woman in the second round of the election on first-round electoral district characteristics. Robust standard errors in parenthesis. Columns (1)-(6) represent the number of candidates per ideology. Columns (7) and (8) the vote share on all left and right candidates, respectively.

Table B3: Balancing tests - manifestos 1st round

| | (1) N. words | (2) Economy | (3) Environment | (4) Health | (5) Local | (6) Defense | (7) Politics | (8) Tone |
|--------------------|--------------------|-------------------|--------------------|------------------|-------------------|------------------|------------------|------------------|
| Treatment | -1.067 (62.164) | -0.022 (0.031) | -0.020* (0.011) | 0.070 (0.063) | -0.053 (0.058) | 0.016 (0.058) | 0.005 (0.014) | 0.001 (0.171) |
| Observations | 2649 | 2649 | 2649 | 2649 | 2649 | 2649 | 2649 | 2649 |
| Eff. number of obs | 641 | 813 | 603 | 696 | 561 | 559 | 525 | 630 |
| Robust p-value | 0.860 | 0.539 | 0.101 | 0.229 | 0.573 | 0.995 | 0.657 | 0.998 |
| Polyn. order | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Bandwidth | 3.890 | 4.977 | 3.559 | 4.191 | 3.227 | 3.221 | 3.028 | 3.830 |
| Outcome mean | 715.386 | 0.282 | 0.087 | 0.049 | 0.312 | 0.215 | 0.052 | -0.148 |

The table shows regression discontinuity estimates of the effect of having a woman in the second round of the election on first-round electoral district characteristics. Robust standard errors in parenthesis.

– 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.¹⁴ 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 ϕ_{ik} is approximated, for each topic k and the document-topic distribution, θ_d , for each document d .

¹⁴Some 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).

An alternative to the seeded LDA could be the Correlation Explanation (CorEx) model of Gallagher, Reing, Kale, and Ver Steeg (2017), and implemented by Djourellova, Durante, and Martin (2021). However, this method forces to choose an anchor strength. The anchor strength controls how much weight CorEx puts towards maximizing the mutual information between the seeded words and their respective topics. The authors encourage users to experiment with the anchor strength and determine what values are best for 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 also complemented with words that tend to be frequent and select them into their respective topics.

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

C.1 List of seeded words per topic - Manifestos

Economy & Employment: employment (*emploi*), wage (*salaire*), tax (*impôt*), worker (*travailleur*), unemployment (*chômage*), income tax (*impôt revenu*), minimum wage (*smic*), pension (*pension*), retirement (*retraite*), work (*travail*), economy (*economie*), economic (*economique*), firm (*entreprise*), industry (*industrie*), business (*commerce*), craft (*artisanat*), buying power (*pouvoir achat*), finance (*finance*), fiscal (*fiscal*), budgetary (*budgétaire*), deficit (*déficit*), finance (*financier*), tax specialist (*fiscaliste*), taxation (*fiscalité*), investment (*investissement*), finance (*financier*), financing (*financement*), small micro firms (*pme tpe*), small firms (*pme*).

Environment: environment (*environnement*), green (*environnemental*), fuel (*carburant*), petroleum (*pétrole*), energetic (*énergétique*), ecological transition (*transition écologique*), energy (*énergie*), fossil (*fossile*), carbon (*carbone*), ecological (*écologique*), ecology (*écologie*), animal (*animal*), planet (*planète*), nuclear (*nucléaire*), biodiversity (*biodiversité*), nature (*nature*), renewable energy (*énergie renouvelable*), pollution (*pollution*), agriculture (*agriculture*).

Local: rural (*rural*), rural territory (*territoire rural*), rural town (*commune rural*), rural area (*milieu rural*), rurality (*ruralité*), rural area (*zone rural*), public service (*service public*), territorial collectivity (*collectivité territoriale*), local collectivity (*colléctivité local*), medical desert (*désert médical*), desertification (*désertification*), medical desertification (*désertification médicale*), municipality (*commune*), region (*région*), proximity (*proximité*), city (*ville*), municipal (*municipal*), territory (*territoire*).

Health & Education: health (*santé*), patient (*patient*), doctor (*médecin*), disease (*maladie*), sanitary (*sanitaire*), hospitable (*hospitalier*), medicament (*médicament*), hos-

pital (*hôpital*), covid (*covid*), disability (*handicap*), disabled (*handicapé*), disability pension (*aah*), medicine (*médecine*), cancer (*cancer*), nurse (*infirmier*), student (*élève*), academic (*scolaire*), school (*école*), child (*enfant*), training (*formation*), student (*étudiant*), teaching (*enseignement*), education (*éducation*), teacher (*enseignant*), professor (*professeur*), university (*université*), college (*collège*), university (*universitaire*), class (*classe*).

Security, Justice & Foreign Policy: police (*police*), safety (*sécurité*), victim (*victime*), military (*militaire*), drug (*drogue*), insecurity (*insécurité*), terrorism (*terrorisme*), terrorist (*terroriste*), criminal (*délinquant*), sentence (*peine*), immigration (*immigration*), immigrant (*immigré*), immigrate (*immigrer*), migrant (*migrant*), migratory (*migratoire*), clandestin (*clandestine*), asylum (*asile*), refugee (*refugié*), nationality (*nationalité*), foreigner (*étranger*), border (*frontière*), islamism (*islamisme*), islamist (*islamiste*), europe (*europe*), european (*européen*), international (*international*), world (*monde*), worldwide (*mondial*), war (*guerre*), russia (*russie*), european union (*union européenne*), maastricht (*maastricht*), european treaty (*traité européen*).

Politics: president (*président*), mp (*député*), national assembly (*assemblée nationale*), president (*président république*), elected (*élu*), presidential (*présidentiel*), voter (*électeur*), sunday (*dimanche*), turn (*tour*), assembly (*assemblée*), right (*droite*), left (*gauche*), vote (*voter*), democracy (*démocratie*).

C.1.1 Top 10 words per topic

Economy & Employment: firm (*entreprise*), retirement (*retraite*), economic (*économique*), unemployment (*chômage*), economy (*economie*), tax (*impôt*), wage (*salaire*), buying power (*pouvoir achat*), finance (*financier*), right (*droit*).

Environment: environment (*environnement*), ecology (*écologie*), energy (*énergie*), agriculture (*agriculture*), ecological (*écologique*), environmental (*écologiste*), nuclear (*nucléaire*), nature (*nature*), pollution (*pollution*), energetic (*énergétique*).

Local: territory (*territoire*), city (*ville*), rural (*rural*), public service (*service public*), municipal (*municipal*), proximity (*proximité*), region (*région*), defend (*défendre*), municipality (*commune*), support (*soutenir*).

Health & Education: child (*enfant*), health (*santé*), school (*école*), education (*éducation*), training (*formation*), majority (*majorité*), class (*class*), hospital (*hôpital*), academic (*scolaire*), student (*élève*).

Security, Justice & Foreign Policy: europe (*europe*), security (*sécurité*), european (*européen*), world (*monde*), immigration (*immigration*), international (*international*), foreigner (*étranger*), insecurity (*insécurité*), state (*état*), police (*police*).

Politics: right (*droite*), left (*gauche*), president (*président*), assembly (*assemblée*), elected (*élu*), national assembly (*assemblée national*), presidential (*présidentiel*), voter (*électeur*), sunday (*dimanche*), turn (*tour*).

Other 1: citizen (*citoyen*), other (*autre*), society (*société*), men (*homme*), should (*falloir*), parti (*parti*), no (*non*), interest (*intérêt*), today (*aujour*), together (*ensemble*).

Other 2: councillor (*conseiller*), majority (*majorité*), general (*général*), trust (*confiance*), solidarity (*solidarité*), men (*homme*), action (*action*), general councillor (*conseiller général*), should (*falloir*), council (*council*).

C.2 Summary statistics

Table C4: Summary statistics

| | 1st Round | | | 2nd Round | | |
|---------------------------|-----------|--------|------------|-----------|--------|------------|
| | Female | Male | Difference | Female | Male | Different |
| No. words | 708.84 | 685.83 | -23.01*** | 665.08 | 564.52 | -100.56*** |
| Economy & Employment | 0.23 | 0.18 | -0.05*** | 0.14 | 0.10 | -0.04*** |
| Environment | 0.09 | 0.07 | -0.02*** | 0.03 | 0.02 | -0.01*** |
| Health & Education | 0.05 | 0.03 | -0.02*** | 0.22 | 0.07 | -0.15*** |
| Local | 0.19 | 0.14 | -0.05*** | 0.18 | 0.13 | -0.06*** |
| Security & Foreign policy | 0.13 | 0.14 | 0.01*** | 0.08 | 0.10 | 0.03*** |
| Politics | 0.14 | 0.14 | 0.00 | 0.16 | 0.20 | 0.05*** |
| Observations | 7859 | 20693 | 28552 | 858 | 4713 | 5571 |

C.3 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 (*allemande*), treaty (*traité*), united states (*États-Unis*), german (*allemand*), world (*monde*).

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

Internal 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 nationale*), student (*étudiant*), university (*université*), class (*classe*).

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

C.3.1 Top 10 words per topic

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

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 (*mesure*).

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

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

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*), right (*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*), housing (*logement*).

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

D Tone

Table D1: Tone of the manifestos by ideology

| | Tone 1st round | | Tone 2nd round | |
|-----------|----------------|-------|----------------|-------|
| | Mean | Sd | Mean | Sd |
| far-left | -0.698 | 1.035 | 0.141 | 0.946 |
| left | 0.157 | 0.964 | 0.471 | 0.876 |
| center | 0.422 | 0.502 | 0.620 | 0.584 |
| right | 0.427 | 0.860 | 0.420 | 0.831 |
| far-right | -0.428 | 0.826 | -0.831 | 0.722 |
| other | -0.073 | 0.832 | 0.562 | 0.680 |
| Total | -0.058 | 1.002 | 0.347 | 0.893 |

E Political orientations

I allocate candidates into seven political orientations (far-left, left, center, right, far-right and other). I use the party classifications from the Chapel Hill Expert Survey (CHES) and when missing from the ParlGov and Pons, Tricaud, and Tricaud (2019). I consider the party positioning on a scale between 0 (left) and 1 (right) and family classification.

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

*Notre volonté, pour vous servir,
se résume en trois mots :
combattre, protéger, avancer,
en maîtrisant le progrès, car la vie,
pour tous, doit être une marche constante
vers plus d'égalité et de bonheur*

**Le 21 Mars,
je m'engage avec Marcel CABIDDU à :**

- Etre le Député de Tous
- Agir ensemble pour la création et le maintien des emplois
- Maintenir la protection sociale
- Défendre l'éducation nationale et les services publics
- Elargir la démocratie au quotidien
- Assurer la solidarité entre la Ville et la Campagne
- Vivre en Fraternité

Noël JOSEPH
Maire de Beuvry
Conseiller Régional
Député
Chevalier de la Légion d'Honneur
Officier de l'Ordre National du Mérite

Marcel CABIDDU
Maire de Wingles
Conseiller Général

Vus, les candidats

Noël JOSEPH
Candidat du Parti Socialiste

*Ouvrons, ensemble, à nos Enfants
les Portes de la Vie ...*

[Signature]

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



Source: Electoral archives of CE-

VIPOF 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 business. 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 tainted blood.

TO TRY BERGER is to ADOPT IT.

See the Candidates

| 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 | Center |
| 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 | Center |
| Union pour un Mouvement Populaire | Right |
| Les Verts | Left |

| 2012 Parliamentary Elections | |
|-----------------------------------|-----------------------|
| Political label | Political orientation |
| Alliance Centriste | Center |
| Autres | Other |
| Centre pour la France | Center |
| 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 | Center |
| 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 Démocrates et Indépendants | Liberal |

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

F Computation of left-right axis

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 manifesto/document d . As in Taddy (2013), Gentzkow et al. (2019) and Cagé, Le Pennec, and Mougin (2021), I use a multinomial inverse regression model (MNIR):

$$c_{wt} \sim \text{MN}(q_{wd}, m_d) \text{ with } q_{wd} = \frac{\exp[\alpha_w + \varphi_w E_d]}{\sum_{l=1}^W \exp[\alpha_l + \varphi_l E_d]} \quad (4)$$

$m_{wd} = \sum_d c_{wd}$ 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.

To compute an accurate estimate of left-right, I am confronted with two methodological challenges. First, I have a problem of finite-sample bias that arises because the number of words that a candidate can choose is large in relation to the manifesto 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 penalisation, using a gamma-lasso procedure, and a Poisson approximation to the multinomial logit likelihood to permit distributed computing. The Poisson regression has negative log likelihood proportional to

$$l(\alpha_w, \varphi_w) = \sum_{d=1}^N [m_d \exp(\alpha_w + \varphi_w E_d) - c_{wd}(\alpha_w + \varphi_w E_d)] \quad (5)$$

In high-dimensional regression it is useful 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):

$$\hat{\alpha}_w, \hat{\varphi}_w = \text{argmin}_{\alpha, \varphi} l(\alpha, \varphi) + N\lambda^{-1}\log(1 + \lambda|\varphi_w|) \quad (6)$$

F.1 Kernel density of left-right score by political ideology