Gender Biases: Evidence from a Natural Experiment in French Local Elections

Jean-Benoît Eyméoud* Paul Vertier†

This version: February 22, 2017

Abstract

Women are under-represented in politics. In this paper, we test one of the frequently given explanations for this puzzle: gender-bias from voters. We use a natural experiment during the French *Départementales* elections of 2015: for the first time in this country, candidates had to run by pairs, which had to be gender-balanced. The order of the candidates on the ballot was determined by alphabetical order, thus making the order of appearance of male and female candidates as-good-as-random. This setting allows us isolating gender biases from selection effects. Our main result is that there exists a negative gender bias affecting right-wing candidates, who receive less votes when the female candidate appears first on the ballot. The effect is sizable and corresponds to about 10% of a standard deviation in relative shares of votes. Gender biases against right-wing candidates are amplified when the female candidate is retired, and reverted when the female candidate is particularly young. However, we do not find evidence of gender-biases against candidates of other parties.

*Sciences Po (Department of Economics, LIEPP), 28 rue des Saints-Pères, 75007 Paris, France; jean-benoit.eymeoud@sciencespo.fr

†Sciences Po (Department of Economics, LIEPP), 28 rue des Saints-Pères, 75007 Paris, France; paul.vertier@sciencespo.fr

[0]We thank Quoc-Ahn Do, Sergei Guriev, Emeric Henry, Elise Huillery, Etienne Wasmer as well as participants to the LIEPP Graduate Student Seminar for useful comments.

1 Introduction

The lack of representation of women in politics is a commonly acknowledged fact. As emphasized by Deber (1982), four main channels are likely to explain it. First, women might have lower political ambitions than men, and thus select less into politics (Lawless and Fox (2005)). Secondly, women might be placed by parties in tougher political races, where their chances of victory are lower (Thomas and Bodet (2013)). Thirdly, women might not receive equal campaign contributions and political support. Finally, female candidates might face sexism from voters. This paper tackles this last hypothesis.

Numbers of studies on the topic yielded contradictory results. Indeed, while male and female candidates have similar success rates in elections in aggregate vote data (Darcy and Schramm (1977), Seltzer et al. (1997), McElroy and Marsh (2009)), gender stereotypes have been proven, through both surveys and lab experiments, to shape the choices of different categories of population with respect to gender (Koch (2002),Mo (2015)).

A key question in this context is to disentangle selection effects from preferences over female candidates. Indeed, in real-world and survey data, unobserved characteristics which might drive both the probabilities of running as a candidate and of winning the election selection process are likely to play a role, and to affect estimates of gender-biases. Conversely, while laboratory experiments are helpful to disentangle such effects (and the potential channels behind observed biases), they are hardly represent real-world situation.

In this paper, we contribute to disentangling selection effects from preferences over female candidates in a real-world setting. To do so, we use a unique feature of the French *Départemental* [1] elections of 2015: for the first time in the history of French elections, candidates ran by pairs, which necessarily had to be gender-balanced. Therefore, each pair of candidates included a man and a woman (each with a substitute of the same gender as them). Crucially, within each pair, the order of appearance of the candidates on the ballot was determined by alphabetical order: this order determines only the place of the candidate's name on the ballot. As we argue, such a setting yields a close from random allocation of the order of gender on the ballot, and allows us exploring whether pairs where the woman appears first on the ballot perform differently from

[1] The *Département* is a French territorial unit gathering numerous competences in terms of schooling, public infrastructures, culture, sports...

pairs where the man appears first. Importantly, upon casting their ballot, voters can only opt for one of the different pairs of candidates, so that for every pair, each male and female candidate receives the same number of votes. If a pair is elected, both candidates are appointed to the same seat in the *Conseil Général* (the *Département* assembly where the elected candidates are seating).

Such a setting is particularly interesting for several reasons. First, it enables having exactly as many female and female candidates - so that in the end, strict parity is respected. Secondly, while we show that the characteristics of male and female candidates are on average different, candidates characteristics hardly predict whether the male or the female candidate is first on the ballot. The effect we measure is therefore unlikely to be affected by selection biases, since it consists in comparing how on average identical candidates fare when the male or the female candidate is first on the ballot.

Our main result is that on average, pairs where the female candidate appears first are not adversely affected. Yet, we find strong heterogeneity depending on the party and characteristics of the candidates. Indeed, our results suggest that right-wing pairs where the female candidate is first on the ballot lose about 10% of a standard deviation in relative share of votes in the first round, while on average this is not the case of pairs from other parties. However, party is not the only determinant of gender biases. Indeed, we find that among right-wing pairs, having the female candidate first on the ballot is especially detrimental when she is retired. To the contrary, it is especially beneficial when the female candidate is particularly young (younger than 30).

These results therefore suggest that both partisan preferences and candidates characteristics are likely to shape gender-biases. How to interpret them ? First, gender stereotypes and cues are unlikely to be the main drivers of our results. Indeed, if they were to play a role, we would expect opposite heterogeneous effects based on age for the right-wing parties. Indeed, assuming that younger candidates signal less conservative points of views, we would expect less support for young females in the right-wing electorate.

Secondly, these results can only be rationalized by behaviors implying incomplete information about how the election is ran. Indeed, because the fates of both candidates on a ballot are tied, if all voters knew perfectly the rules of the elections, we would not find any treatment effect. Therefore, a possible explanation for the differences between right-wing and other candidates would be that right-wing voters are on average less informed about the format of the election, and thus assume that the first candidate on the ballot is the *main* candidate. This argument is reinforced by the fact that the *Départementales* elections of 2015 were the first binominal elections to be held in France. However, while it could explain the difference of treatment effects between right-wing candidates and other candidates (especially since no noticeable heterogeneous treatment effect is found for other candidates), it can hardly explain the heterogeneities with respect to age among right-wing voters.

Therefore, a conservative interpretation of our results is that gender biases affect *at least* right-wing candidates, and that these gender-biases overlap with age-biases, whereby older female candidates are sanctioned, and younger female candidates are beneficiary. However our results cannot allow us to assert that gender-biases do not exist for candidates of other parties, since the absence of results might be driven by differences in the understanding of the rules of the election.

The remainder of the paper is structured as follows. Section 2 presents the related literature on the topic. Section 3 describes the institutional setting and the data we use. We provide descriptive statistics and various balance-checks showing that selection into the treatment is unlikely. Section 4 describes our estimation strategy. Section 5 gathers our main empirical results. Section 6 performs some robustness checks and Section 7 concludes.

2 Literature Review

Three main methodologies were used to evaluate potential voter-biases against women: analyses of aggregate votes, survey analysis and laboratory experiments. Most of the aggregate votes analyses find that male and female candidates have equal success rates in elections, thus arguing that voters do not have gender biases (Darcy and Schramm (1977), Seltzer et al. (1997), McElroy and Marsh (2009)). Some studies even argue that women might have an electoral advantage compared to men (Black and Erickson (2003), Borisyuk et al. (2007)), and that after their first election, they are at least as likely to be reelected as men (Shair-Rosenfield and Hinojosa (2014)). Milyo and Schosberg (2000) even argue that the barriers to entry faced by women makes female incumbent of higher quality than male incumbents, resulting in an advantage for female incumbents. The main issue with such studies is that they are unlikely to fully control for the selection process leading to the observed political competition. This is especially true if male and female candidates are likely to differ in unobserved characteristics which might drive both their probabilities of running as a candidate and of winning the election.

Alternatively, survey studies allow analyzing more deeply the reasons why certain voters might prefer candidates of a certain gender. Using surveys, several studies argue that voter biases are marginal compared to partisan preferences (Dolan (2004), Dolan (2014), Hayes and Lawless (2016)). Though, some types of individuals might have preferences over female politicians: McDermott (1998) and Burrell (1995) find that women are more likely to prefer female candidates, while Dolan (1998) finds that minorities and elderly are more likely to vote for women. McDermott (1997) argues that liberal voters are more likely to prefer female candidates. Such preferences are likely to be driven by gender stereotypes (Koch (2002)). In particular, in a context of low information, the gender of the candidate can be interpreted by the voters as signals about the ideology of the candidates: McDermott (1998) shows that female candidates are typically perceived as more liberal and more dedicated to honest government. However, surveys have obvious drawbacks, since respondents' answers might be affected by characteristics of the interviewer, such as her gender (Huddy et al. (1997), Flores-Macias and Lawson (2008), Pino et al. (2011), Benstead (2013)), religion Blaydes and Gillum (2013) or language Lee and Pérez (2014)). Furthermore, as in the case of real aggregate data studies, survey analysis is likely to suffer from endogeneity biases, and to be contingent on the characteristics of the candidates and of the elections.

Finally, laboratory experiments are able to cleanly disentangle potentially concurring hypotheses, while dealing with selection biases. Using such a setting, Leeper (1991) showed that even when women candidates emit "masculine" message, voters attribute them "feminine" characteristics. Huddy and Terkildsen (1993) showed that gender-based expectations over policies were more related to gender-traits stereotypes than to gender-beliefs stererotypes. King and Matland (2003) show that biases against women are likely to depend on partisan preferences, while Mo (2015) shows that both explicit and implicit attitude against women shape the probability of voting for female candidates. Yet, while these studies allow disentangling more accurately the mechanisms leading to potential gender-biases, they are hardly likely to represent real-world election settings.

This paper aims at addressing the question of gender-biases using another kind of tool, namely a natural experiment. Using natural or quasi-natural experiments to analyze gender-related issues is a growing trend. Many applications were used to assess the impact of women in office. In a seminal study, using a randomized experiment, Chattopadhyay and Duflo (2004) showed that public goods investments were affected by the gender of city councils' heads in India, and skewed towards the needs of their gender. Ferreira and Gyourko (2014), using a Regression Discontinuity Design on close elections between male and female candidates in US municipalities, do not find different policies between male and female mayors. They do find, however, that female mayors have higher unobserved political skills, since they have a greater incumbency advantage. More recently, Brollo and Troiano (2016) showed thanks to a similar regression discontinuity-design that female mayors in Brazil are less likely to be corrupted and hire less temporary public employees. Contrarily, to Ferreira and Gyourko (2014), they find that women are less likely to be reelected than men, and that they attract less contributions during the political campaigns.

But recent contributions also make use of natural experiments to analyze potential biases against women. Beaman et al. (2009) show, using gender quotas for leadership positions in Indian villages, that exposure to female leadership changes the attitude of voters, and become less gender-biased. Pino et al. (2011) shows evidence of a slightly negative gender bias in Chilean congressional elections, using separated voting booths between men and women: overall, right-wing female voters vote less for female candidates than right-wing male voters, while left-wing female voters vote slightly more for female candidates than left-wing male voters. Such results emphasize a specific type of gender-bias (namely the differential propensity of male and female voters to vote for female candidates) and are a line with the idea that men and women have different policy preferences, and that gender represents a signal for policy preferences. Bagues et al. (2016) show that the presence of women in academic committees does not affect the quantity and quality of hired female candidates. We contribute to this literature by using a unique natural experiment which allows isolating gender-biases from selection effects, and which allows characterizing which types of individuals and parties are most affected by them.

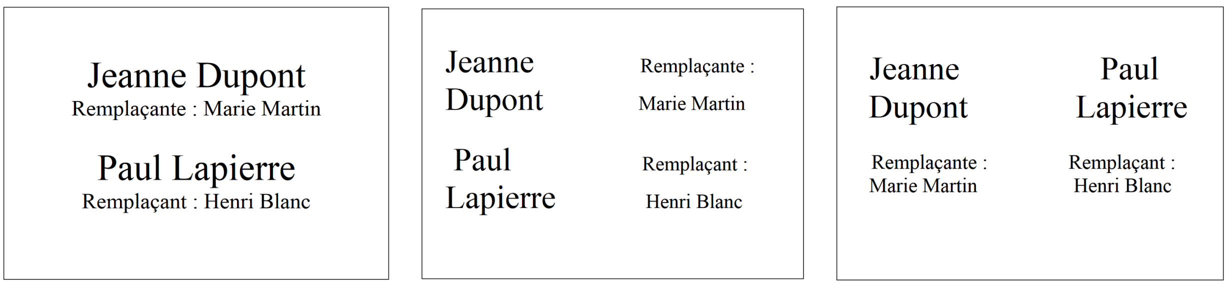
3 Institutional Framework and Data

3.1 Institutional Framework

This study relies on data from the 2015 French departmental elections, which took place on March 22nd and March 29th. Departmental councellors were elected in 2,054 *cantons* (subdivisions of the *départements*). In each canton, lists ran by pairs which necessarily had to be balanced. Each candidate of a pair had to have a substitute of the same sexe as her. Overall, 9,097 pairs of candidates ran for office.

Within each list, the order of the candidates on the ballot was determined by alphabetical order. Such a requirement is imposed by the article L.191 of the French electoral legislation. The rules for printing electoral ballots are also stringent: it must be printed in only color on a blank sheet of format 105x148 mm, weight between 60 and 80 grams per square meter, be in landscape format. For each candidate, the name of its substitute must be written right after its name, using a smaller font. According to the articles L.66, L.191, R.66-2, R.110 and R.111 of the electoral code, any ballot not respecting these requirement is considered as null. Figure 1 shows examples of compliant ballots, as communicated by the Ministry of Interior.

Figure 1: Examples of valid ballots



Importantly, the ballots on the day of the election are the only ones to be subject to these requirements, which do not affect campaign advertisement leaflets or electoral posters. Campaign leaflets should be printed on a 210x297 mm sheet, and should not combine the colors blue, white and red (which are the colors of the French flag). Similarly, electoral posters should be printed on sheets of a maximum size corresponding to 594x841 mm, and should not be white. Each candidate has the right to one advertising location per municipality. The order of advertising location between lists is randomly drawn in the *préfecture*, and this order also corresponds to the order in which the ballots are displayed at the polling station.

3.2 Data and Descriptive Statistics

3.2.1 General population of candidates

For this analysis, we retrieved information about all the pairs of candidates from the Ministry of Interior. Our database includes information on age, gender, previous political experience, political affiliation and socioprofessional categories of each of these candidates. We match these informations with sociodemographic information at the canton-level.

As indicated in Table 1, overall, 5% of female candidates are incumbent, while it is the case of 19% of male candidates. Female candidates are about 51 years old, while male candidates are about 52 years old. The majority of candidates are either private sector employees (27% of female candidates and 23% of male candidates) or retired (21% of female candidates and 26% of male candidates). On average the number of candidates per *canton* is equal to 4.7. 28% of candidates were left-wing, a number which is comparable to the share of right-wing candidates. 14% of candidates were classified as extreme-left, while 22% were classified as extreme-right. [2]

Overall, out of 28 variables tested, only have unbalances significant at the 10% level: women in intermediary professions and in liberal occupations are over-represented in tandems where the female candidate appears first, while women in other occupations are under-represented. Similarly, right-wing candidates are over-represented among the tandems where the female candidate is at the top of the ballot.

In order to confirm the idea that the treatment assignment is as-good-as-random, we compare the distributions of surname initials across gender. Figure 2 plots the frequency of each surname initial for male and female candidates of all parties: the distributions are strikingly similar. In

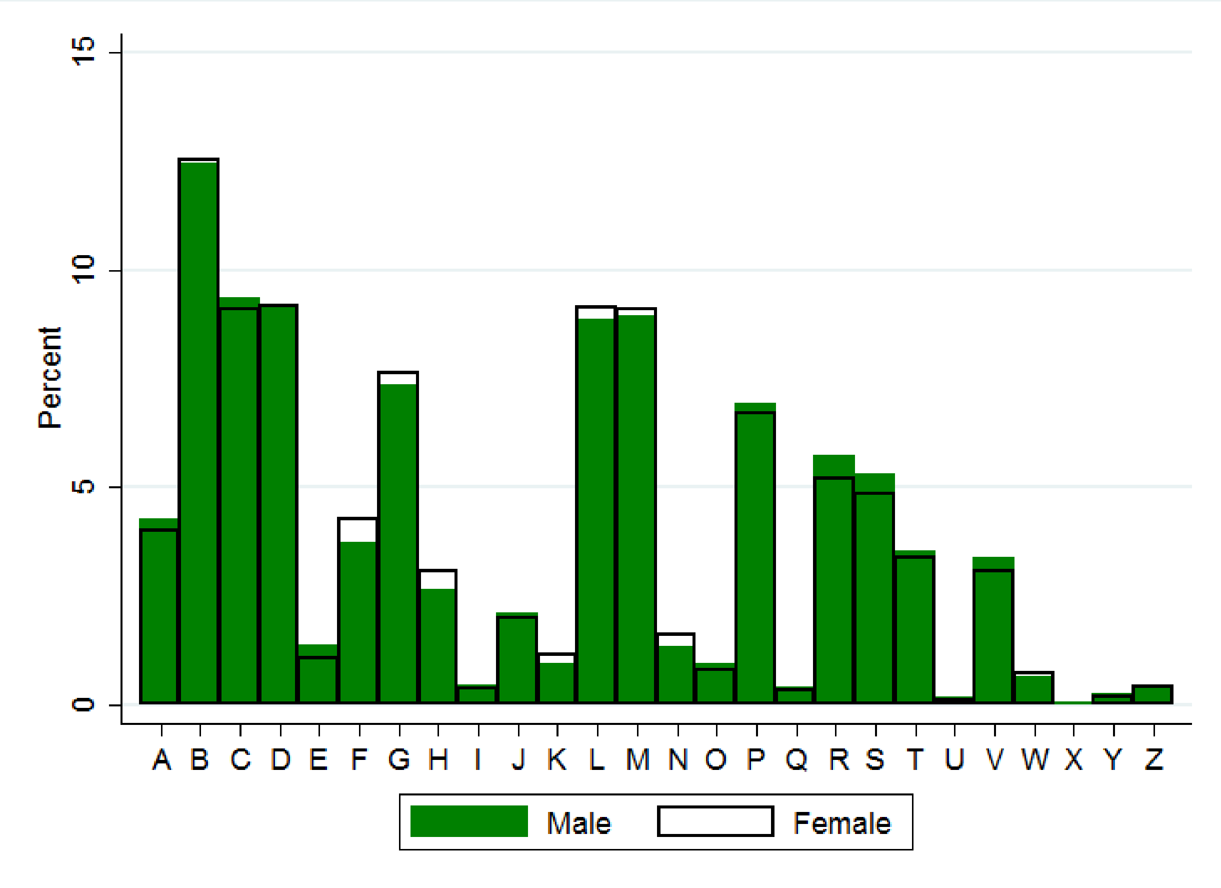
[2] We classified as extreme-left the lists labeled as *Communists, Extreme-Left, Front de Gauche and Parti de Gauche*. We classified as left-wing the lists labeled as *Parti Socialiste, Union de la Gauche, Radicraux de Gauche and Divers Gauche*. We classified as right-wing the lists labeled as *MoDem, Union du Centre, Union des Démocrates et des Indépendants, Debout La France, Divers-Droite, Union des Droites, UMP*. Finally we classified as extreme-right the lists labeled as *Front National and Extreme Droite*.

Table 1: Balance check: all candidates

	Man First	N	Woman First	N	Diff	T-Test
Incumbent (W)	.0520926	4492	.0522193	4596	-.0001267	-.0271596
Incumbent (M)	.1903384	4492	.1869017	4596	.0034367	.4186859
Age (W)	51.51536	4492	51.30614	4596	.2092249	.8267642
Age (M)	52.58037	4492	52.4876	4596	.0927672	.3420194
Farmer (W)	.0175868	4492	.020235	4596	-.0026482	-.9262169
Intermediary Profession (W)	.0532057	4492	.0615753	4596	-.0083696	-1.714534
Private Sector Employee (W)	.2749332	4492	.2832898	4596	-.0083566	-.887832
Liberal Occupation (W)	.0598842	4492	.0765883	4596	-.0167041	-3.156798
Education Occupation (W)	.1162066	4492	.1133594	4596	.0028471	.4257015
Civil Servant(W)	.1184328	4492	.1153177	4596	.0031151	.4621274
Public Firm Worker (W)	.0398486	4492	.0387293	4596	.0011193	.2745836
Other Occupation(W)	.1077471	4492	.0913838	4596	.0163633	2.606527
Retired (W)	.2121549	4492	.1995213	4596	.0126336	1.489525
Farmer (M)	.0338756	4487	.0339573	4594	-.0000817	-.0215
Intermediary Profession (M)	.0944952	4487	.0968655	4594	-.0023703	-.3838479
Private Sector Employee (M)	.2302206	4487	.2392251	4594	-.0090044	-1.01209
Liberal Occupation (M)	.0759973	4487	.0814105	4594	-.0054132	-.957541
Education Occupation (M)	.1067528	4487	.101872	4594	.0048808	.7608051
Civil Servant(M)	.1049699	4487	.0966478	4594	.0083221	1.317202
Public Firm Worker (M)	.0403388	4487	.0380932	4594	.0022456	.5512173
Other Occupation(M)	.0537107	4487	.0537658	4594	-.0000551	-.0116321
Retired (M)	.259639	4487	.2581628	4594	.0014761	.1605409
Number of Candidates	4.74266	4496	4.760487	4601	-.0178267	-.7048989
Randomly Drawn Position	2.841415	4496	2.880678	4601	-.0392635	-1.248039
Unemployment Rate (2012)	.1353706	4461	.1371932	4573	-.0018226	-1.670633
Abstention Rate	.4893835	4496	.4898833	4601	-.0004998	-.3825752
Share of rural cities	.4646494	4496	.4664214	4601	-.001772	-.2140984
Extreme Left	.1383452	4496	.1364921	4601	.0018531	.2566498
Left	.2811388	4496	.2701587	4601	.0109801	1.17182
Right	.287589	4496	.3088459	4601	-.0212569	-2.215854
Extreme Right	.2137456	4496	.210389	4601	.0033565	.3915314

Table 4, we perform different tests of equal disributions, and we cannot reject the null hypothesis that the two distributions are similar.

Figure 2: Distribution of surname initials across gender



3.2.2 Samples of analysis

In this section, we present descriptive statistics on the samples we will use in our main analysis. As we argue in the next section, in order to identify causal effects of the treatment, our estimation needs to satisfy the *Stable Unit Treatment Value Assumption* (SUTVA), which states that the potential outcomes of a unit are not affected by the treatment status of another unit. This hypothesis is likely to be violated if we consider altogether several candidates from a same *canton*. Indeed, let us assume that the treatment affects negatively a given pair of candidates. One can therefore imagine that the votes they lost positively affected another pair of candidates from the same canton (especially if the voters reacting to the treatment are non-partisan).

In order to avoid such a scenario, we run an analysis on different samples of candidates having the same partisan affiliation, and being the sole candidate of this party in their *canton*. [3]. Such subsamples meet the *SUTVA* assumption: while it is possible that these candidates are affected by the treatment status of candidates of other parties, they cannot be affected by the treatment status of other units in the sample.

Using these subsamples, if we focus only on intra-party imbalances, the picture remains broadly similar. To test more systematically for imbalances, we regress the dummy variable indicating whether the female candidate is listed first on the whole set of individual characteristics, as well as on some *canton* characteristics. Results of these regressions are reported in Table 2. Overall, for all parties, the characteristics of the candidates explain very few (if any) of the variance of the treatment variable, and they are not jointly significant. No imbalances are found for extreme-left candidates. Among left-wing candidates, women are slightly less likely to be on the top of the ballot when they are in tandem with a male incumbent, but more so when they are retired or associated with an intermediary profession (coefficients significant at the 10% level only). Among right-wing pairs, women are more likely to be first on the ballot if they work in intermediary professions or liberal occupation. Finally, among extreme-right candidates, younger female candidate are more likely to be on the top of the ballot. So do female candidates who are retired, civil servants, or working in liberal and intermediary professions.

Additionally, we provide evidence that the political competition faced by treated and untreated candidates of each political affiliation are broadly similar. Table 3 gathers the results of regressions, where we explain the treatment by the average characteristics of the political opponents of the considered candidates. Here again, we find that the characteristics of the opponent explain a tiny part of the variance in treatment, and the set of covariates is never jointly significant. Some small imbalances seem to exist nonetheless. In particular, among extreme-left candidates, those facing more female candidates working in education are more likely to be treated. Similarly, among left-wing candidates, those facing more male incumbents and male private sector employees are less likely to be treated. Finally, among the extreme-right candidates, those facing older female candidates, a greater number of male civil servants or of women working in

[3] By an abuse of language, we hereafter call "parties" the broad categorizations of *extreme-left, left-wing, right-wing and extreme-right* candidates, described above

Table 2: Determinants of the treatment (Restricted Samples)

Treatment	XLeft	Left	Right	XRight
Incumbent (W)	0.150	0.000	-0.023	
	(1.20)	(0.01)	(0.48)	
Incumbent (M)	0.054	-0.056	-0.051	
	(0.88)	(1.86)*	(1.57)	
Age (W)	-0.001	-0.001	0.002	-0.002
	(0.60)	(0.71)	(0.93)	(2.12)**
Age (M)	0.002	0.001	-0.002	-0.000
	(1.00)	(0.73)	(1.00)	(0.31)
Intermediary Profession (W)	-0.029	0.142	0.093	0.089
	(0.24)	(1.52)	(1.65)*	(1.85)*
Private Sector Employee (W)	0.023	0.065	0.036	0.052
	(0.44)	(1.34)	(0.86)	(1.56)
Liberal Occupation (W)	0.021	0.090	0.093	0.174
	(0.25)	(1.39)	(1.74)*	(2.54)**
Education Occupation (W)	0.004	0.074	-0.015	0.034
	(0.07)	(1.39)	(0.26)	(0.57)
Civil Servant(W)	-0.009	0.067	-0.010	0.105
	(0.17)	(1.28)	(0.19)	(1.77)*
Retired (W)	-0.079	0.095	-0.019	0.140
	(1.43)	(1.69)*	(0.38)	(3.30)***
Intermediary Profession (M)	0.124	0.140	0.026	-0.076
	(0.97)	(1.94)*	(0.48)	(1.58)
Private Sector Employee (M)	0.078	0.005	0.020	-0.052
	(1.49)	(0.10)	(0.44)	(1.21)
Liberal Occupation (M)	0.065	0.052	0.002	0.002
	(0.68)	(0.80)	(0.04)	(0.03)
Education Occupation (M)	0.037	0.082	-0.025	-0.081
	(0.63)	(1.52)	(0.39)	(1.36)
Civil Servant(M)	-0.013	0.047	-0.039	-0.054
	(0.21)	(0.90)	(0.67)	(0.88)
Retired (M)	0.006	0.033	0.055	-0.043
	(0.11)	(0.63)	(1.12)	(0.89)
Number of Candidates	0.001	-0.017	-0.012	0.002
	(0.07)	(1.04)	(0.69)	(0.12)
Randomly Drawn Position	0.012	0.015	0.003	0.003
	(1.15)	(1.37)	(0.23)	(0.33)
Unemployment Rate (2012)	-0.232	0.225	0.509	-0.429
	(0.62)	(0.66)	(1.38)	(1.44)
Abstention Rate	0.011	0.381	0.186	-0.072
	(0.03)	(1.24)	(0.63)	(0.29)
Share of rural cities	0.001	0.042	0.049	-0.000
	(0.02)	(0.88)	(1.03)	(0.01)
Adj. R2	-0.00	-0.00	-0.00	0.00
F	1.01	0.93	0.98	1.12
N	1,187	1,334	1,389	1,882

OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. The outcome is a variable equal to one if the female candidate is first on the ballot and zero otherwise. Robust Standard-Errors. T-Stats between brackets.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

educations are less likely to be treated. All these unbalances except the last one, are significant only at the 10% level.

Similarly, as Figure 3 indicates, for all partisan affiliations, the distributions of surname initials seem to be similar across genders. Here again, the tests of equal of equal distributions generally cannot reject the null hypothesis that the distributions are identical across gender (Table 4). The only exception is for the left-wing candidates, were the distribution seem slightly different : the Kolmogorov Test and the Mann-Whitney-Wilcoxon test reject the hypothesis of equal distributions at the 10% level. But as Figure 3b shows, this difference seems mainly driven by an over-representation of women with names beginning by the letter B, and is unlikely to represent a more general manipulation of the treatment.

These results suggest that if any selection into the treatment based on the characteristics of the candidates or on the characteristics of the political opponents exists, it is of low magnitude.

4 Estimation strategy

Our main estimation strategy aims at analyzing whether candidates lose or gain from having the female candidate first on the ballot.

In an initial specification, we test whether, on average, the electoral performances of tandems where the female candidate is first on the ballot are different from those where the male candidate is first. Identification takes place within the potential outcomes framework from the Rubin Causal Model, where we assume two potential outcomes for each unit i - $Y_i(0)$ and $Y_i(1)$ - and the causal effect of the program on the unit i is defined as $\tau_i = Y_i(1) - Y_i(0)$. The actual observed outcome is defined as such:

$$Y_i^{obs} = \begin{cases} Y_i(0) & if \quad T_i = 0 \\ Y_i(1) & if \quad T_i = 1 \end{cases}$$

In this framework, the Average Treatment Effect is defined as $ATE = \mathbb{E}[Y_i(1) - Y_i(0)]$. A naive estimate of this quantity is given by $\overline{Y_1^{obs}} - \overline{Y_0^{obs}}$. In general, such a quantity is unbiased under the *Stable Unit Treatment Value Assumption* (SUTVA) and the *complete randomization assumption* .

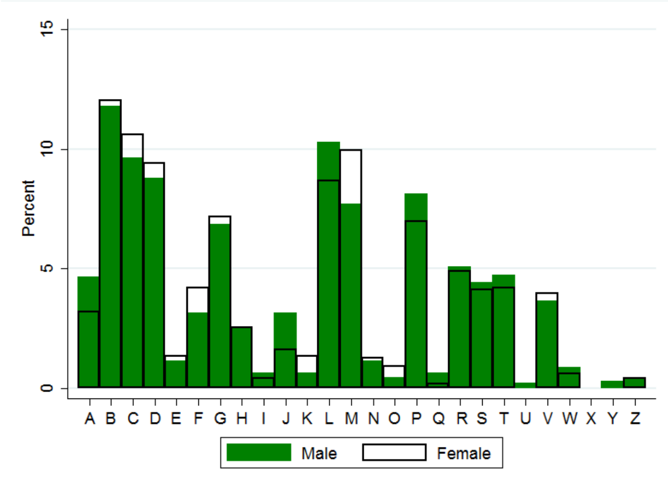
Table 3: Characteristics of political opponents (Restricted Samples)

Treatment	XLeft	Left	Right	XRight
Share Treated opponents	0.048	0.024	-0.002	-0.016
	(0.86)	(0.50)	(0.05)	(0.39)
Share XLeft opponents		0.203	-0.007	-0.013
		(1.61)	(0.06)	(0.12)
Share Left opponents	-0.161		-0.012	0.051
	(1.13)		(0.11)	(0.53)
Share Right opponents	-0.201	0.189		-0.034
	(1.42)	(1.55)		(0.33)
Share XRight opponents	-0.234	0.155	0.071	
	(1.32)	(1.12)	(0.58)	
Mean Incumbent (W)	0.003	0.110	0.048	-0.039
	(0.03)	(0.95)	(0.43)	(0.48)
Mean Incumbent (M)	0.065	-0.129	0.010	-0.027
	(0.82)	(1.68)*	(0.13)	(0.48)
Mean Age (W)	0.001	0.003	-0.001	-0.004
	(0.20)	(1.32)	(0.29)	(1.75)*
Mean Age (M)	0.000	0.000	0.000	0.003
	(0.03)	(0.06)	(0.13)	(1.55)
Mean Intermediary Profession (W)	0.106	-0.044	0.061	-0.058
	(0.84)	(0.41)	(0.54)	(0.57)
Mean Intermediary Profession (M)	-0.069	-0.061	-0.034	-0.017
	(0.61)	(0.60)	(0.32)	(0.18)
Mean Private Sector Employee (W)	0.102	0.039	0.024	-0.063
	(1.18)	(0.54)	(0.33)	(0.94)
Mean Private Sector Employee (M)	0.019	-0.144	0.024	0.058
	(0.19)	(1.75)*	(0.30)	(0.81)
Mean Liberal Occupation (W)	-0.078	-0.030	0.139	0.035
	(0.64)	(0.27)	(1.23)	(0.38)
Mean Liberal Occupation (M)	-0.051	-0.077	0.093	0.020
	(0.43)	(0.73)	(0.78)	(0.23)
Mean Education Occupation (W)	0.205	0.059	0.058	-0.180
	(1.84)*	(0.61)	(0.65)	(2.31)**
Mean Education Occupation (M)	-0.086	0.010	0.047	0.109
	(0.68)	(0.10)	(0.49)	(1.30)
Mean Civil Servant(W)	0.022	0.045	0.076	-0.066
	(0.20)	(0.47)	(0.83)	(0.86)
Mean Civil Servant(M)	0.018	-0.039	-0.054	-0.146
	(0.15)	(0.37)	(0.55)	(1.79)*
Mean Retired (W)	0.143	-0.061	0.076	0.012
	(1.38)	(0.69)	(0.88)	(0.15)
Mean Retired (M)	-0.073	-0.044	0.015	-0.056
	(0.71)	(0.50)	(0.18)	(0.74)
Adj.R2	-0.01	-0.00	-0.01	0.00
F	0.70	0.76	0.25	1.16
N	1,188	1,333	1,389	1,892

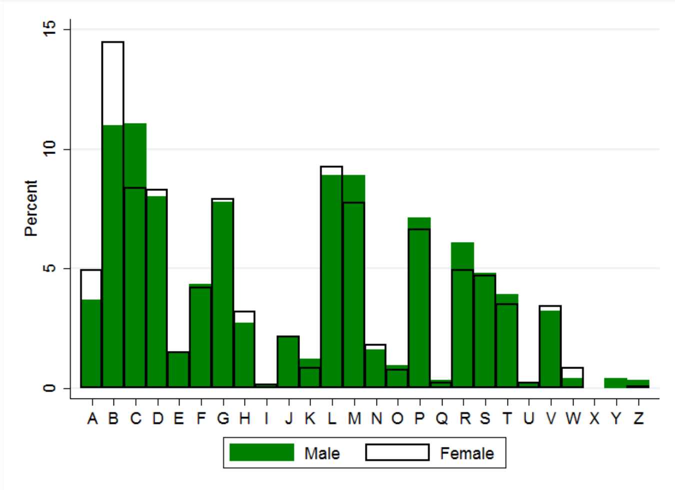
OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. The outcome is a variable equal to one if the female candidate is first on the ballot and zero otherwise. Robust Standard-Errors. T-Stats between brackets.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

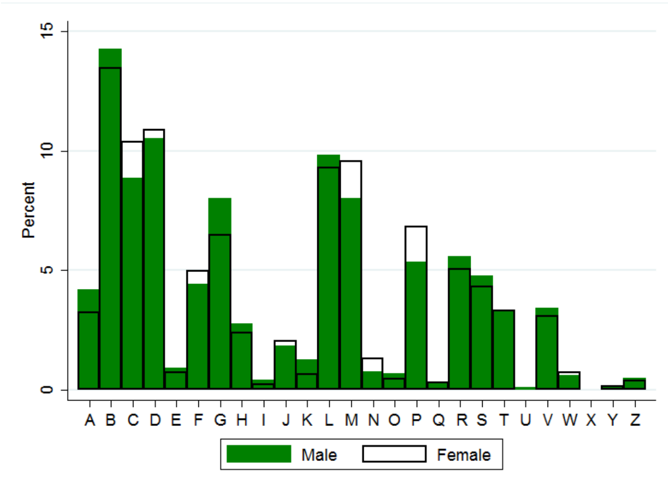
Figure 3: Distribution of surname initials across parties (Restricted Sample)



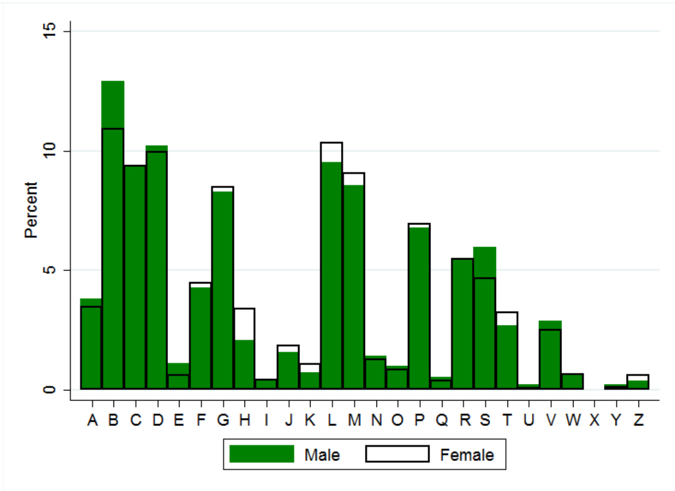
(a) Distribution: Extreme Left



(b) Distribution: Left



(c) Distribution: Right



(d) Distribution: Extreme Right

Table 4: Tests of equal distributions of surnames initials

P-Value	All	Extreme-Left	Left	Right	Extreme-Right
KS	0.211	0.782	0.094*	0.855	0.377
Median	0.320	0.774	0.132	0.622	0.474
MWW	0.385	0.652	0.0546*	0.583	0.372

The table presents the P-values of three tests of equal distributions: Kolmogorov-Smirnov (KS), non-parametric test of equality of medians (Median), and Mann-Whitney-Wilcoxon rank-sum test (MWW)
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

As explained above, the *SUTVA* is likely to be violated if we do not restrict our analysis to a sample of observations which cannot interact with each other (meaning that the treatment status of one observation will not affect the outcome of any other unit). To do so, we therefore restrict our analysis to candidates who are the only ones to represent their party in the *canton*.

The second assumption states that both the potential outcomes and the covariates are independent from the treatment. Formally, the condition writes as such:

$$T_i \perp (Y_i(0), Y_i(1), X_i)$$

In our setting, the treatment-assignment is based on a procedure which is supposedly as-good-as-random, since the order of the candidates (and hence the place of the female candidate) on the ballot is determined by alphabetical order. However, as have shown in the last section, while the treatment assignment is hardly affected by candidates' characteristics, the covariates are not systematically perfectly balanced across treatment status. In our setting, it therefore seems more plausible to assume the milder assumption of *unconfoundedness*, which states that the potential outcomes and the treatment are independent after controlling for covariates potentially affecting them. Formally, this assumption writes:

$$T_i \perp (Y_i(0), Y_i(1))|X_i$$

Our baseline OLS specification is therefore the following:

$$Y_i = \alpha + \beta T_i + \delta X_i + \epsilon_i \tag{1}$$

where Y_i is an outcome variable indicating the electoral performance of tandem i, T_i is the treatment variable, which is equal to 1 if the female candidate in tandem i is first on the ballot and 0 otherwise, X_i is a set of candidates characteristics, and ϵ_i is an error term. In such a setting, heterogeneous effects can be estimated by interacting a subset of the control variables X_i and $X_i^{'}$ with the treatment T_i.

This specification does not model how the electoral performance of a pair of candidates depends on the characteristics of the other candidates. This is problematic for the estimation of the treatment effect only to the extent that, on average, a treated pair faces candidates with different characteristics. However, as shown in Table 3, on average, treated and non-treated pairs of candidates face similar electoral competition. In Section 6, we also control for the average characteristics of the opponents of the considered, and in the robustness-check section we estimate the effect of the treatment by comparing the results of the different candidates pairwise.

5 Results

5.1 OLS estimation

In this section, we present our main results, by estimating equation (1). In order to do so, we compare the scores received by candidates in the first round of the election in the control and in the treatment group. Note that in this setting, the number of candidates is not identical in each *canton*, and the scores of tandems facing different are therefore not directly comparable. In order to make the electoral performances comparable across different number of candidates, we demean and standardize the share of votes received by each tandem within each stratum of number of candidates.

Does the order of the candidates affect electoral the electoral performances of the tandem ? Table 5 summarizes the estimates of such an average treatment effect across several specifications. The first panel reports results without any controls. The second panel reports results controlling for individual characteristics as well as unemployment rate, share of rural municipalities and abstention rate in the *canton*. The third panel involves the same control variables, but interacts the characteristics of male and female candidates. Finally, the last panel is similar to the second one, but also includes the average characteristics of the political opponents. Overall, the results suggest that the performances of extreme-left, left-wing and right-wing pairs are not affected by the order of the candidates. However, right-wing tandems lose a sizeable share of votes if

Table 5: Results: OLS Specification

No Controls	XLeft	Left	Right	XRight
Woman first	0.039	-0.020	-0.158	0.005
	(1.09)	(0.44)	(3.50)***	(0.15)
R^2	0.00	0.00	0.01	0.00
N	1,188	1,341	1,389	1,893
Indiv. Controls	XLeft	Left	Right	XRight
Woman first	0.005	0.044	-0.098	0.016
	(0.18)	(1.16)	(2.50)**	(0.50)
R^2	0.47	0.35	0.29	0.11
N	1,187	1,334	1,387	1,882
Interacted Indiv. Controls	XLeft	Left	Right	XRight
Woman first	0.007	0.051	-0.096	0.025
	(0.25)	(1.33)	(2.38)**	(0.79)
R^2	0.49	0.41	0.33	0.15
N	1,187	1,334	1,387	1,882
Indiv Controls and Opponents Char.	XLeft	Left	Right	XRight
Woman First	0.009	0.021	-0.101	0.015
	(0.35)	(0.61)	(2.81)***	(0.49)
R^2	0.54	0.49	0.44	0.20
N	1,191	1,336	1,398	1,914

OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. The outcome is the share of votes received by each candidate in the first round of the election, demeaned and standardized within each stratum of number of candidates. The first panel includes no covariates. The second panel includes age, socioprofessional categories and incumbency status of male and female candidates, as well as the unemployment rate, the abstention rate and the share of rural municipalities in the *canton*. The third specification includes the same variable, but interacts the age of men and women, the socioprofessional categories of men and women, and the incumbency status of men and women. The fourth specification is similar to the second specification, but adds the average individual characteristics of male and female opponents (age, socioprofessional categories, incumbency status) of the considered candidate). Robust Standard-Errors. T-Stats between brackets.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

the female candidate is first. Estimates of the loss range between 10% and 15% of a standard deviation in relative vote share. Such results suggest that right-wing voters have a negative bias against female candidates. The magnitude of the coefficient is similar across the specifications (and especially stable in all the specifications including covariates), suggesting that the inclusion of covariates hardly affects the general pattern.

5.2 Heterogeneity analysis

In order to better understand the determinants of this negative bias against female candidates, we interact the treatment with several characteristics of the candidates. Results of these estimations are gathered in Tables 6, 7 and 8. All specifications control for non-interacted candidates'

Table 6: Heterogeneity: Retired Candidate

Interaction with woman retired	XLeft	Left	Right	XRight
Woman First	0.018	0.038	-0.063	0.039
	(0.56)	(0.88)	(1.47)	(1.09)
Woman Retired	-0.005	-0.148	0.030	-0.053
	(0.12)	(1.98)**	(0.37)	(0.90)
Woman Retired*Woman First	-0.050	0.016	-0.246	-0.078
	(0.91)	(0.18)	(2.46)**	(1.10)
R^2	0.47	0.34	0.29	0.11
N	1,187	1,334	1,387	1,882
Interaction with man retired	XLeft	Left	Right	XRight
Woman First	-0.001	0.030	-0.097	-0.008
	(0.04)	(0.64)	(2.13)**	(0.22)
Man Retired	-0.064	-0.176	-0.144	-0.099
	(1.52)	(2.61)***	(1.95)*	(1.79)*
Woman First*Man Retired	0.011	0.029	0.014	0.093
	(0.19)	(0.37)	(0.15)	(1.35)
R^2	0.46	0.34	0.28	0.11
N	1,187	1,334	1,387	1,882

OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. The outcome is the share of votes received by each candidate in the first round of the election, demeaned and standardized within each stratum of number of candidates. Each panel controls for age, socioprofessional categories and incumbency status of male and female candidates, as well as the unemployment rate, the abstention rate and the share of rural municipalities in the *canton*. Robust Standard-Errors. T-Stats between brackets.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

characteristics, and the same sociodemographic variables as in the baseline specification. In the upper panel of Table 6, we interact the treatment with a dummy variable indicating whether the female candidate is retired, while in the second panel, we interact the treatment with a dummy variable indicating whether the male candidate is retired. We observe that, on average extreme-left and extreme-right candidates do not have lower relative performances if their female candidate is retired, irrespectively of its position on the ballot (*Woman Retired* and *Woman Retired*Woman First* are insignificant). Conversely,a left-wing pairs involving a retired female is lower by about 15% of a standard deviation in relative score, but here again, this effect is not significantly different if the retired female candidate is first on the ballot.

However, among right-wing candidates, pairs involving retired female candidates are underperforming only if the female candidate is listed first on the ballot : among the pairs where the female candidate is retired, the treatment effect is lower by -24.6% of a standard deviation compared to the effect among pairs where the woman is not retired (which is equal to -6.3% of a standard deviation). This suggests that overall, among pairs involving retired female, the treat-

Table 7: Heterogeneity: Young Candidate

Interaction with woman below 30	XLeft	Left	Right	XRight
Woman First	0.008	0.047	-0.113	0.013
	(0.29)	(1.24)	(2.85)***	(0.38)
Woman Under 30	-0.058	-0.072	-0.289	-0.040
	(0.65)	(0.38)	(1.78)*	(0.58)
Woman First*Woman Under 30	-0.051	-0.119	0.429	0.038
	(0.51)	(0.51)	(1.76)*	(0.42)
R^2	0.47	0.35	0.30	0.11
N	1,187	1,334	1,387	1,882
Interaction with man below 30	XLeft	Left	Right	XRight
Woman First	0.003	0.045	-0.102	0.016
	(0.11)	(1.17)	(2.57)**	(0.47)
Man Under 30	-0.054	-0.155	-0.408	0.011
	(0.92)	(1.21)	(2.68)***	(0.18)
Woman First*Man Under 30	0.015	-0.061	0.056	-0.003
	(0.16)	(0.37)	(0.25)	(0.04)
R^2	0.47	0.35	0.30	0.11
N	1,187	1,334	1,387	1,882

OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. The outcome is the share of votes received by each candidate in the first round of the election, demeaned and standardized within each stratum of number of candidates. Each panel controls for age, socioprofessional categories and incumbency status of male and female candidates, as well as the unemployment rate, the abstention rate and the share of rural municipalities in the *canton*. Robust Standard-Errors. T-Stats between brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ment effect is of about -30% of a standard deviation in relative score. Strikingly, no such interactions are found when we interact the treatment with a dummy variable indicating whether the male candidate is retired: overall, for all parties, pairs involving retired men in second position on the ballot have lower performances (from -6.4% of a standard deviation among extreme-left candidates (not significant) to -17.6% of a standard deviation among left-wing candidates), but this effect is not different if the male candidate is first on the ballot.

Table 7 runs the same kind of analysis, but interacting the treatment respectively with dummies indicating whether the female and male candidates are below 30 [4]. No effect is found on any of these variables for the extreme-left, left-wing and extreme-right candidates. However, among right-wing candidates, we observe that pairs involving young female candidates under-perform when the latter are listed second on the ballot (among the lists where female candidates are listed second, the relative score of those involving a woman under 30 is inferior by 29% of a standard deviation, significant at the 10% level), while they over-perform when the female candidates

[4]Note that the results broadly hold when we consider other age thresholds for "young" candidates

Table 8: Heterogeneity: Incumbency

Interaction with woman incumbent	XLeft	Left	Right
Woman First	0.010	0.034	-0.104
	(0.38)	(0.83)	(2.52)**
Woman Incumbent	1.570	0.535	0.543
	(7.02)***	(6.83)***	(5.99)***
Woman First*Woman Incumbent	-0.405	0.080	0.065
	(1.11)	(0.73)	(0.49)
R^2	0.47	0.35	0.29
N	1,187	1,334	1,387
Interaction with man incumbent	XLeft	Left	Right
Woman First	0.020	0.070	-0.103
	(0.75)	(1.32)	(2.10)**
Man Incumbent	1.458	0.882	0.718
	(10.00)***	(15.13)***	(11.20)***
Woman First*Man Incumbent	-0.092	-0.061	0.006
	(0.49)	(0.79)	(0.07)
R^2	0.41	0.30	0.26
N	1,187	1,334	1,387

OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. The outcome is the share of votes received by each candidate in the first round of the election, demeaned and standardized within each stratum of number of candidates. Each panel controls for age, socioprofessional categories and incumbency status of male and female candidates, as well as the unemployment rate, the abstention rate and the share of rural municipalities in the *canton*. Robust Standard-Errors. T-Stats between brackets.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

are listed first on the ballot : among pairs where the female candidate is younger than 30 the treatment effect is greater by 43% of a standard deviation compared to the effect found among pairs where the female candidate is older than 30. This suggests that among pairs involving female candidates under 30, the treatment effect is of about 30% of a standard deviation in relative scores. Conversely, among right-wing pairs where the woman is listed second on the ballot, those where the male candidate is younger than 30 under-score by about 41% of a standard deviation. However, this effect is not statistically different if the female candidate is listed first on the ballot.

Finally, in Table 8, we run the same analysis, interacting the treatment with dummy variables indicating whether the female and male candidates are incumbent (we exclude the sample of extreme-right candidates, which contains no female incumbent, and only one male incumbent). We find strong positive incumbency effects for both men and women, but they seem to be independent from the place of the candidate on the ballot. Reciprocally, the negative treatment effect found among right-wing pairs does not seem to be affected by the incumbency status of

Table 9: Results: OLS on Full Sample

No Interaction	(1)	(2)	(3)
Woman first	0.003	-0.003	-0.000
	(0.17)	(0.21)	(0.01)
R^2	0.00	0.44	0.45
N	9,095	9,016	9,016
Controls	N	Y	Interacted
Interaction with Right-Wing	(1)	(2)	(3)
Woman First	0.023	0.025	0.027
	(1.01)	(1.24)	(1.31)
Right	0.640	0.453	0.445
	(18.00)***	(14.35)***	(14.17)***
Woman First*Right	-0.109	-0.097	-0.092
	(2.30)**	(2.32)**	(2.23)**
R^2	0.07	0.29	0.30
N	9,095	9,016	9,016
Controls	N	Y	Interacted

OLS Regressions. All regressions are ran on the full population of candidates. The outcome is the share of votes received by each candidate in the first round of the election, demeaned and standardized within each stratum of number of candidates. The first column includes no covariates. The second column includes age, socioprofessional categories and incumbency status of male and female candidates, as well as the unemployment rate, the abstention rate and the share of rural municipalities in the *canton*. The third column includes the same variable, but interacts the age of men and women, the socioprofessional categories of men and women, and the incumbency status of men and women. Standard-Errors are clustered at the *canton* level. T-Stats between brackets.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

the male and female candidates, thus suggesting that the gender-bias occurs regardless of the political experience of the candidate.

Overall, this heterogeneity analysis suggests that the gender-bias existing among right-wing voters depends on the characteristics of the candidates: in particular, pairs involving retired women receive less votes if the latter are listed first on the ballot, and pairs involving young women receive more votes if the latter are listed second on the ballot, while we do not observe such patterns for pairs involving retired and young men.

6 Alternative Specifications and Robustness Checks

In this last section, we provide evidence that our main results are not an artifact of our sample of selection or of our choice of modelling.

6.1 Full Sample

In Table 9, we run the same baseline model on the full population of candidates. While in such setting we cannot exclude that the *SUTVA* is violated (even though, in Table 3, we have provided evidence that treated and untreated pairs face on average a similar number of treated opponents), it provides consistent evidence that our main estimates are not an artifact of our sample selection. The upper panel of the table reports average treatment effects on the population of candidates in three specifications (without controls, with individual and sociodemographic controls, and with an interacted version of these controls). We find no evidence of treatment effects whatsoever.

However, when we interact the treatment with a dummy indicating that the pair of candidates is from the right-wing, we find a strongly negative interaction term, suggesting that the treatment effect is lower by 10% of a standard deviation among right-wing candidates, compared to the treatment effects among other candidates. Since the average treatment effect among other candidates is equal to 2.5% of a standard deviation in relative scores, this suggests that the treatment effects among right-wing candidates is roughly equal to 7.5% of a standard deviation in relative scores - a figure which is in line with our previous estimates.

6.2 Dyadic Estimation

Finally, in order to take into account more thoroughly the structure of the political competition (even though we have provided evidence that treated and untreated candidates broadly face the same number and types of political opponents), we compute for each candidate the difference between his relative score and the relative score of each of his opponents. We then regress the relative score between the considered candidate and the considered opponent on their respective characteristics and treatment statuses.

Formally, we therefore run the following estimation:

$$Y_{ij} = \alpha + \beta T_i + \gamma T_j^{'} + \delta X_i + \nu X_j^{'} + \epsilon_{ij} \tag{2}$$

where Y_{ij} is the difference between the relative score of candidate i and the relative score of candidate j, T_i is the treatment status of candidate i, $T_j^{'}$ is the treatment status of candidate j, $X_j^{'}$ is a set of characteristics of candidate j, and ϵ_{ij} is an error term.

Table 10: Results: Dyadic Specification

No Controls	XLeft	Left	Right	XRight
Woman First	0.063	-0.032	-0.163	-0.003
	(1.34)	(0.54)	(2.69)***	(0.08)
Woman First (Opponent)	0.017	0.015	-0.045	-0.001
	(0.48)	(0.35)	(1.05)	(0.03)
R^2	0.00	0.00	0.00	0.00
N	4,458	4,423	4,361	6,733
Indiv Controls	XLeft	Left	Right	XRight
Woman First	0.007	0.023	-0.101	-0.012
	(0.21)	(0.55)	(2.19)**	(0.28)
Woman First (Opponent)	0.013	0.053	-0.028	0.001
	(0.50)	(1.73)*	(0.90)	(0.03)
R^2	0.43	0.52	0.50	0.37
N	4,446	4,416	4,344	6,698
Interacted Indiv. Controls	XLeft	Left	Right	XRight
Woman First	0.017	0.039	-0.110	-0.002
	(0.49)	(0.92)	(2.35)**	(0.05)
Woman First (Opponent)	0.012	0.051	-0.029	-0.009
	(0.46)	(1.71)*	(0.92)	(0.34)
R^2	0.46	0.56	0.53	0.40
N	4,446	4,416	4,344	6,698

OLS Regressions. Each subsample considers only the candidates who are the only ones of the considered party in the *canton* where they run. For each of these candidates, we compute the difference between their relative share of votes and the relative share of votes of any of their opponents, which is the outcome variable. The first column includes no covariates. The second column includes age, socioprofessional categories and incumbency status of male and female candidates and opponents, as well as the unemployment rate, the abstention rate and the share of rural municipalities in the *canton*. The third column includes the same variable, but interacts the age of men and women, the socioprofessional categories of men and women, and the incumbency status of men and women (both for the considered candidate and the considered opponent). Standard-Errors are clustered at the *canton level*. T-Stats between brackets.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

The results of this estimation are gathered in Table 10. Here again, we consider three types of regressions, including respectively no controls, controls for individual and sociodemographic characteristics, and an interacted version of these characteristics. The results look very similar to the main estimation: we do not find any treatment effect for extreme-left, left-wing and right-wing candidates, but we do a find a negative treatment effect for right-wing candidates, corresponding to between 10% and 16% of a standard deviation. However, we hardly find significant effects for the treated status of the opponent.

7 Conclusion

Among the numerous reasons which might explain why women are under-represented in politics, gender-bias is frequently considered as a potential candidate. While several pieces of research argue that gender-biases are unlikely to play a role, isolating such effects using actual electoral data can prove complicated, due to the presence of selection effects.

In this paper, we isolate gender-biases from selection effects using a natural experiment in France. Using the fact that the candidates of the *Départementales* elections of 2015 had to run for the first time by gender-balanced pairs, and considering that the, we argue that the gender of the first candidate on the ballot was determined in as-good-as-random manner.

While such a framework does not allow us testing the effect of such biases on the stock of women elected in office, since it is by construction equal to 50%, it allows disentangling cleanly selection effects and potential biases, since we compare pairs of candidates which are on average very similar, but which differ only in the order of male and female candidates on the ballot.

In this framework, we provide evidence of a sizable gender-bias affecting right-wing candidates: among them, the pairs where the female candidate was first on the ballot saw their relative score in the first round decrease by about 10% of a standard deviation. This negative effect was amplified when female candidates were retired, but reversed when the female candidates were young. Such results are likely to be triggered by voters who were not fully aware of the rules of the elections, and who mistakenly assumed that the first candidate was the main candidate. Accordingly, while we find evidence of gender-biases against right-wing candidates, the absence of evidence concerning the candidates of other parties does not necessarily imply that they are not also affected by gender biases (since our results might also reflect a better understanding of the electoral rules among the voters of other parties).

Therefore, while the results of this paper suggest that gender-biases against female candidates do exist, it would be particularly interesting to test whether such effects persist in the next occurrences of these elections.

References

Bagues, M., Sylos-Labini, M., and Zinovyeva, N. (2016). Do gender quotas pass the test? evidence from academic evaluations in italy.

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

Benstead, L. J. (2013). Effects of interviewer–respondent gender interaction on attitudes toward women and politics: Findings from morocco. *International Journal of Public Opinion Research*, page edt024.

Black, J. H. and Erickson, L. (2003). Women candidates and voter bias: do women politicians need to be better? *Electoral Studies*, 22(1):81–100.

Blaydes, L. and Gillum, R. M. (2013). Religiosity-of-interviewer effects: Assessing the impact of veiled enumerators on survey response in egypt. *Politics and Religion*, 6(03):459–482.

Borisyuk, G., Rallings, C., and Thrasher, M. (2007). Women in english local government, 1973–2003: getting selected, getting elected. *Contemporary Politics*, 13(2):181–199.

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

Burrell, B. C. (1995). *A Woman's Place is in the House: Campaigning for Congress in the Feminist Era*. University of Michigan Press.

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

Darcy, R. and Schramm, S. S. (1977). When women run against men. *Public Opinion Quarterly*, 41(1):1–12.

Deber, R. B. (1982). " the fault, dear brutus": Women as congressional candidates in pennsylvania. *The Journal of Politics*, 44(2):463–479.

Dolan, K. (1998). Voting for women in the" year of the woman". *American Journal of Political Science*, pages 272–293.

Dolan, K. (2014). *When Does Gender Matter?: Women Candidates and Gender Stereotypes in American Elections*. Oxford University Press, USA.

Dolan, K. A. (2004). *Voting for women: How the public evaluates women candidates*. Westview Pr.

Ferreira, F. and Gyourko, J. (2014). Does gender matter for political leadership? the case of us mayors. *Journal of Public Economics*, 112:24–39.

Flores-Macias, F. and Lawson, C. (2008). Effects of interviewer gender on survey responses: Findings from a household survey in mexico. *International Journal of Public Opinion Research*, 20(1):100–110.

Hayes, D. and Lawless, J. L. (2016). *Women on the run: gender, media, and political campaigns in a polarized Era*. Cambridge University Press.

Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P. J., and Pugliani, P. (1997). The effect of interviewer gender on the survey response. *Political Behavior*, 19(3):197–220.

Huddy, L. and Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, pages 119–147.

King, D. C. and Matland, R. E. (2003). Sex and the grand old party: An experimental investigation of the effect of candidate sex on support for a republican candidate. *American Politics Research*, 31(6):595–612.

Koch, J. W. (2002). Gender stereotypes and citizens' impressions of house candidates' ideological orientations. *American Journal of Political Science*, pages 453–462.

Lawless, J. L. and Fox, R. L. (2005). *It takes a candidate: Why women don't run for office*. Cambridge University Press.

Lee, T. and Pérez, E. O. (2014). The persistent connection between language-of-interview and latino political opinion. *Political Behavior*, 36(2):401–425.

Leeper, M. S. (1991). The impact of prejudice on female candidates: An experimental look at voter inference. *American Politics Quarterly*, 19(2):248–261.

McDermott, M. L. (1997). Voting cues in low-information elections: Candidate gender as a social information variable in contemporary united states elections. *American Journal of Political Science*, pages 270–283.

McDermott, M. L. (1998). Race and gender cues in low-information elections. *Political Research Quarterly*, 51(4):895–918.

McElroy, G. and Marsh, M. (2009). Candidate gender and voter choice: Analysis from a multimember preferential voting system. *Political Research Quarterly*.

Milyo, J. and Schosberg, S. (2000). Gender bias and selection bias in house elections. *Public Choice*, 105(1-2):41–59.

Mo, C. H. (2015). The consequences of explicit and implicit gender attitudes and candidate quality in the calculations of voters. *Political Behavior*, 37(2):357–395.

Pino, F. et al. (2011). Is there gender bias among voters? evidence from the chilean congressional elections. In *Working Paper*.

Seltzer, R., Newman, J., and Leighton, M. V. (1997). *Sex as a political variable: Women as candidates and voters in US elections*. Lynne Rienner Publishers.

Shair-Rosenfield, S. and Hinojosa, M. (2014). Does female incumbency reduce gender bias in elections? evidence from chile. *Political Research Quarterly*, 67(4):837–850.

Thomas, M. and Bodet, M. A. (2013). Sacrificial lambs, women candidates, and district competitiveness in canada. *Electoral Studies*, 32(1):153–166.