

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Guinaudeau, Benjamin; Jankowski, Michael

Working Paper Do Radical-Right Parties Use Descriptive Representation Strategically? A Replication of Weeks et al. (2023)

I4R Discussion Paper Series, No. 149

Provided in Cooperation with: The Institute for Replication (I4R)

Suggested Citation: Guinaudeau, Benjamin; Jankowski, Michael (2024) : Do Radical-Right Parties Use Descriptive Representation Strategically? A Replication of Weeks et al. (2023), I4R Discussion Paper Series, No. 149, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/302283

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

INSTITUTE for

No. 149 DISCUSSION PAPER SERIES

Do Radical-Right Parties Use Descriptive Representation Strategically? A Replication of Weeks et al. (2023)

Benjamin Guinaudeau Michael Jankowski

This paper received a response:

Weeks, A.C., B. M. Meguid, M.C. Kittilson, H. Coffé. 2024. Response to "Do Radical-Right Parties Use Descriptive Representation Strategically? A Replication of Weeks et al. (2023)." *IAR Discussion Paper Series* No. 150. Institute for Replication.

September 2024



I4R DISCUSSION PAPER SERIES

I4R DP No. 149

Do Radical-Right Parties Use Descriptive Representation Strategically? A Replication of Weeks et al. (2023)

Benjamin Guinaudeau¹, Michael Jankowski²

¹NYU's Center of Social Media and Politics, New York/USA ²Data Lab of the German Chancellery, Berlin/Germany

SEPTEMBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and metascientific work in the social sciences. Provided in cooperation with EconStor, a service of the <u>ZBW – Leibniz Information Centre for Economics</u>, and <u>RWI – Leibniz Institute for Economic Research</u>, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

		Editors	
Abel Brodeur	Anna Dreber	ool of Economics	Jörg Ankel-Peters
University of Ottawa	Stockholm Sch		RWI – Leibniz Institute for Economic Research
E-Mail: <u>joerg.peters@rwi-es</u>	sen.de	Hohenzollernstraße 1-3	www.i4replication.org
RWI – Leibniz Institute for E	conomic Research	45128 Essen/Germany	

Do Radical-Right Parties Use Descriptive Representation Strategically? A Replication of Weeks et al. (2023)

Benjamin Guinaudeau¹ and Michael Jankowski²

¹NYU's Center of Social Media and Politics, United States (bguinaudeau@gmail.com, benguinaudeau.com) ²Data Lab of the German Chancellery, Germany (michael.jankowski@bk.bund.de, michael-jankowski.de)

June 25, 2024

7 hen do populist radical-right parties (PRRP) foster the (descriptive) representation of women? In a recently published paper, Weeks et al. (2023) coin the concept of 'strategic descriptive representation'. When facing electoral struggles, PRRP would exploit the existing gender gap and strategically increase the descriptive representation of women to attract female votes and fare better in the election. Using data on 58 elections across 19 countries, the authors test their argument and find conclusive evidence supporting it. In this paper, we offer a replication of the study. First, we assess the numerical reproducibility of the published findings ('verification'). Second, we investigate the 'robustness' of the findings and evaluate the results under alternative model specifications. While our replication study identifies minor issues with the verification and some of the model specifications, it most importantly shows that the main results of the paper are driven by a single outlier. The paper's key finding is hence contingent on the inclusion of a single observation (French Front National in 2012), which is a questionable observation as it only elected two MPs, one of whom was a woman. Additionally, this woman's election was seemingly caused by a combination of idiosyncratic factors discussed in the study. Once the case is excluded from the analysis the key model parameter shrinks close to zero and loses its statistical significance. Accordingly, in light of our findings, there is no clear evidence supporting strategic descriptive representation and electoral pressures do not seem sufficient to encourage PRRP to increase their share of female representatives. Correcting this empirical finding has important implications for both understanding PRRP's electoral strategies and women's representation.

Keywords: descriptive representation \cdot women \cdot populist radical-right \cdot replication \cdot robustness small-N studies

Disclaimer: The views and contributions made in this paper are solely the views of the authors and should not be attributed to the German Federal Chancellery. Michael Jankowski's participation in this research is independent and not connected to his position at the Chancellery, nor does it reflect the policies or positions of the German government. This work was conducted in Michael Jankowski's personal capacity and during a period that does not conflict with his official duties.

Introduction

The persistent underrepresentation of women in parliaments worldwide represents a glaring inequality within liberal democracies (e.g., Celis and Erzeel, 2020). This disparity further sidelines women in politics, with numerous studies indicating that an increase in women's descriptive representation often leads to enhanced substantive representation (e.g., Wängnerud, 2009; Phillips, 2020). Therefore, deciphering the factors that perpetuate this gender imbalance is of paramount significance.

The ascent and sustained influence of populist radical-right parties (PRRPs) is frequently discussed as a potential contributor to this gender disparity (Coffé, 2018). PRRPs endorse traditional gender norms that essentially perpetuate structural gender inequalities (e.g., Mudde 2007, Chapter 5; De Lange and Mügge 2015; Akkerman 2015; Spierings 2020). This endorsement is reflected in their policy proposals, with PRRPs often opposing affirmative action measures designed to bolster women's representation (e.g., Caul, 2001; Kantola and Lombardo, 2021). Consequently, PRRPs are often labeled as male-dominated parties – *Männerparteien* –, as men predominantly represent and support them (on the gender gap in voting for PRRP, see, among various others, Immerzeel, Coffé, and Lippe, 2015; Erzeel and Rashkova, 2017; Spierings and Zaslove, 2017).

However, despite their anti-feminist positions and their perception as male-dominated parties, existing research highlights that there is considerable variation in the number of women represented by PRRP. Some of these parties achieve almost gender parity among their elected MPs, while other PRRP's groups remain largely dominated by males. How can this variation be explained? In a recent paper, Weeks, Meguid, Kittilson, and Coffé (in the following: WMKC; Weeks et al. 2023) coin the concept of 'strategic descriptive representation. In a nutshell, it expects PRRP to increase women's descriptive representation when two conditions apply simultaneously. First, the party has to face an electoral threat, operationalized as a decreasing vote share in the previous election. Second, the electoral gender gap has to be significant, i.e. the party is predominantly supported by male voters. When both of these conditions are met, WMKC expect PRRP to elect more women to parliament as a strategy to attract (new) female voters.

WMKC present evidence for the strategic descriptive representation hypothesis using a newly compiled dataset that encompasses past election outcomes, voter base gender composition, and the proportion of female legislators across multiple parties and elections in Europe. Consistent with their theoretical framework, WMKC's regression analysis shows that a high male-to-female voter ratio coupled with a drop in vote share from previous elections predicts a high proportion of female legislators for PRRP in subsequent elections. For other party families, these patterns are not observed.¹

This paper provides a replication of the above-mentioned study. We focus on two central aspects of replication, which Freese and Peterson (2017) call 'verification' and 'robustness'. Verification assesses the numerical reproducibility of the published findings using the same data and code as the authors. While this is a very narrow understanding of replication, it is still important as previous research has shown that the reproducibility of published findings cannot be taken for granted. Our verification analysis confirms that most results of WMKC can be reproduced using the provided data and code (Weeks et al., 2022). However, we document a problem with the computation of p-values which is anti-conservative and can be traced back to inaccuracies of the R package used for exporting the results. While the change in p-values is relatively small when adjusting for this problem, it affects whether the relevant interaction term is significant (i.e., p < 0.05) for some models.

More importantly, we also provide a comprehensive robustness analysis of WMKC's study. In general, robustness analyses encompass a wide range of steps that assess "whether key results are observed consistently across alternative analyses" (Freese and Peterson, 2017, p. 154). Here we proceed in two steps. Our first robustness analysis runs standard regression diagnostics on the published models for analyzing how well the models fit the data (Neumayer and Plümper, 2017). This analysis reveals the key finding of the regression analysis is fully contingent on a single data point (the case of Front National in 2012). Removing this outlier, the p-value of the key interaction term drastically increases from p < 0.05 to p > 0.7, and

¹With the exception of the christian democrats, for which a similar interaction effect is found. For a discussion of this finding see Weeks et al. (2023, p. 433).

the effect size diminishes to one-sixth of the published value under the most favorable model specification. In the second step of the robustness analysis, we scrutinize the sensitivity of the published results to alternative model specifications and variable operationalizations. In this second part, we identify that many characteristics of the model specification affect marginally the results. If the change in results is here small compared to the above-mentioned issue with the outlier, we nonetheless report them in this study as the main coefficient often becomes statistically non-significant under alternative modeling strategies.

In sum, our analyses suggest that the empirical data does not allow us to reject the null hypothesis and does not demonstrate the existence of strategic descriptive representation. Finally, we briefly discuss the theoretical consequences for the representation of women.

WMKC's Analyses and Verification of Results

WMKC compiles a comprehensive dataset incorporating multiple critical variables for a wide array of parties and elections in Europe. To empirically test PRRP's strategic descriptive representation, three variables are central. The primary outcome variable quantifies women's descriptive representation post-election (at time t) by the percentage of female legislators within a party. The first relevant independent variable is the change in national vote shares between the two previous elections (at time t - 1 and t - 2). Thus, negative values indicate that the party lost votes in the previous election. The second variable measures the gender composition of a party's voters in the previous election (at time t - 1). Specifically, the original study estimates the male-to-female voter ratio where values larger than 1 indicate a majority of male voters. In addition, the study includes several control variables such as the parties and countries.

The resulting datasets consist of 58 observations on which the main analysis is based. Given the nested structure of the data (countries, elections, and parties), multi-level regression analysis with log-likelihood is used. The original analyses are based on five models, that sequentially incorporate control variables. While the first model only includes the two key independent variables separately², the second model adds the theoretically important interaction term. For the third, fourth, and fifth models, the interaction is kept and additional control variables are added.

In Table 1 of WMKC, a significant interaction term (p < 0.05) is reported for Models 2 to 5. The interaction effect is also substantially relevant with an effect size ranging from $\beta = -0.66$ to $\beta = -0.87$, depending on the specific model specification. An AME plot further visualizes the strength and direction of the interaction effect (Figure 5 in WMKC). In general, the results of the regression analysis are consistent with WMKC's theoretical expectation of strategic descriptive representation.

We assessed whether the published results can be reproduced using the data and R code posted on APSR's Harvard Dataverse (Weeks et al., 2022). After installing all relevant packages, the code ran without errors (but with several warnings). All exported tables and figures are consistent with the published results. In the appendix to this paper, we provide code that automatically downloads all data and code from the Harvard Dataverse, installs all relevant packages, and then runs the code.

While the exported tables are consistent with the results displayed in the article, they are incorrect with regard to the estimated p-values. For example, Table 1 reports that the p-value of the interaction effect is always significant with p < 0.05. This is an artifact caused by the stargazer-package used for exporting the regression results. The stargazer-package (incorrectly) uses infinite degrees of freedom for estimating p-values.³ While this choice is certainly inconsequential for large enough samples, it can heavily impact the results of analyses run on smaller samples. Considering the small sample size (58 cases), this error is even more likely to lead to under-estimated p-values and the false rejection of the null hypothesis.

While it is challenging to estimate p-values for mixed-effect models, established approaches such as Satterthwaite method rely on a t-distribution resulting from approximated degrees of freedom (Elff et al., 2021, p. 417). This is also the default option in the lmerTest-package (Kuznetsova, Brockhoff, and Christensen, 2017) used in the paper.⁴ Using this approach, the

²Because the first model is of little relevance for the theoretical argument, we do not discuss it.

³stargazer's wrong behavior is only experienced with models fitted in 1me4. The alternative and popular texreg-package does the same.

⁴Notably, the replication code also contains the computations of the p-values using this method, but these

	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.55	-0.55	-0.63	-0.73	-0.60
	(0.49)	(0.48)	(0.43)	(0.43)	(0.43)
Lagged Change in Vote Share (B)	-0.17	0.84	0.90	0.99	1.15^{*}
	(0.18)	(0.54)	(0.49)	(0.51)	(0.51)
(A) X (B)		-0.66	-0.67^{*}	-0.76^{*}	-0.87^{**}
		(0.34)	(0.31)	(0.32)	(0.32)
Controls 3			Х	Х	Х
Controls 4				Х	Х
Controls 5					Х
AIC	432.39	430.72	423.28	424.75	425.77
BIC	444.75	445.15	439.76	445.35	456.67
Log Likelihood	-210.19	-208.36	-203.64	-202.37	-197.88
Num. obs.	58	58	58	58	58
Num. groups: party	22	22	22	22	22
Num. groups: country	19	19	19	19	19
Var: party (Intercept)	0.00	1.14	0.00	0.00	0.00
Var: country (Intercept)	65.97	57.15	56.16	49.58	23.91
Var: Residual	50.33	48.00	39.29	38.64	39.01

 Table 1: Interaction effect with adjusted p-values (see Table 1 in WMKC)

Notes: These models replicate the models from the original paper but report p-values adjusted for the actual number of degrees of freedom. The models reported in the paper assumed infinite degrees of freedom. ***p < 0.001; **p < 0.01; *p < 0.05

p-values in the original paper are slightly more conservative and not always p < 0.05 as we report in Table 1 (see Model 2, in which p > 0.05).⁵

Robustness Analysis

In contrast to verification, robustness tests assess whether results remain unchanged when the analysis is subjected to various perturbations or when alternative modeling strategies are employed. We first investigate whether the regression models fit the data well, mainly focusing

values are not the ones reported in the paper.

⁵A more general comment on the statistical significant of interaction effects: It (still) seems to be common practice to judge the presence of interaction effects based on the p-value of the interaction coefficient. However, it should be noted that an interaction can be present even when the interaction term is not significant (for details on this, see Brambor, Clark, and Golder 2006, p. 74). However, we follow here paper's approach and use statistical significance as the central criterion to judge the presence of interaction effects (see, e.g., Weeks et al. 2023, p. 433)

on the problem of outliers. Then we proceed with discussing several modeling strategies and the extent to which plausible alternative strategies yield non-significant results.

Outlier Analysis

When the sample size is small, the danger of biased estimates due to a few influential observations, i.e. outliers, strongly increases (see, e.g., Wooldridge, 2020, p. 317). The original study disclaims awareness of this problem and discusses outliers in the robustness tests with the aim of "check[ing] that extreme values [...] are not biasing" (Weeks et al., 2023, p. 429) the results. After excluding a case with a large value on the male-to-female voter ratio variable, the original study observes that the results remain stable. Instead of defining outliers as observations with large values on certain variables, an alternative approach understands outliers as observations with large values on certain variables. Instead, they are commonly understood as observations that have an undue impact on the regression coefficients. These cases do not necessarily involve extreme values and can be detected using standardized statistical tests – such as leverage or Cook's Distance (Cook, 1977) –. In essence, these tests estimate how strongly the removal of a certain observation affects the regression coefficients (for a thorough discussion of outliers with a special focus on multi-level models, see Nieuwenhuis, Grotenhuis, and Pelzer, 2012).

Thus, we estimate Cook's Distance for WMKC regression models to check for potential outliers. The results suggest that the regression models are strongly influenced by a few observations, particularly by the observation of the Front National in the French parliamentary election of 2012 (Cook's D > 1.5; see Appendix Table A1 for Cook's D values of all observations for all models). When we exclude this observation from the models, the key findings vanish as reported in Table 2. Specifically, the removal diminishes the interaction coefficient from .87 to .14 (Model 5). Likewise, the interaction coefficient is also no longer 'statistically significant'. In Model 2 the p-value was p = 0.055, but in the model without Front National in 2012 it is p > 0.99. For Model 5, the p-value increases from p = 0.008 to p = 0.71. Thus, the values are nowhere near conventional thresholds of significance.

Given these results, we can confidently claim that WMKC's results fully depend on a single

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.70	-0.70	-0.80	-0.82	-0.64
	(0.43)	(0.46)	(0.46)	(0.41)	(0.42)	(0.40)
Lagged Change in Vote Share (B)	1.15^{*}	-0.03	-0.02	-0.04	0.15	0.25
	(0.51)	(0.17)	(0.64)	(0.57)	(0.59)	(0.56)
(A) X (B)	-0.87^{**}		-0.00	0.05	-0.09	-0.14
	(0.32)		(0.43)	(0.39)	(0.40)	(0.38)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
AIC	425.77	417.45	419.45	409.84	412.46	410.88
BIC	456.67	429.71	433.75	426.18	432.89	441.52
Log Likelihood	-197.88	-202.73	-202.73	-196.92	-196.23	-190.44
Num. obs.	58	57	57	57	57	57
Num. groups: party	22	22	22	22	22	22
Num. groups: country	19	19	19	19	19	19
Var: party (Intercept)	0.00	0.00	0.00	0.00	0.00	0.00
Var: country (Intercept)	23.91	51.11	51.11	49.57	41.73	18.60
Var: Residual	39.01	45.68	45.68	35.21	36.08	34.76

 Table 2: Interaction effects without Front National 2012

Notes: These models replicate the models from the original paper after removing the observation for the FN 2012. The first column presents the original model 5 results (with the FN 2012) as a reference. ***p < 0.001; **p < 0.01; *p < 0.05

observation. Without Front National in 2012, there is no empirical evidence for the theorized interaction effect. In Figure 1 we display the absence of the interaction effect as AME plot. The left plot reproduces the AME plot published in WMKC, the right plot uses the same code but relies on the model with Front National in 2012 excluded (both plots are based on Model 5). Again, the plot further confirms the total absence of the interaction effect.





Note: Plot is based on model specification 5 which is the specification with the largest interaction term.

The massive impact of this observation calls for a closer investigation of what happened in 2012 with the Front National. If this observation was very well in line with the theoretical mechanism, it would justify its inclusion in the model. Also Wooldridge (2020, p. 317) argues that the "decision to keep or drop such [influential] observations in a regression analysis can be a difficult one". It is, thus, important to look at the case of the Front National in 2012 in more detail to get a better understanding of what information this case contributes to the analysis and whether it can serve as a representative case of how PRRP use the descriptive representation of women strategically.

When looking at the values of the key independent variables for Front National in 2012, the observation has relatively high (but not extreme) values on both variables. It has a lagged male-to-female ratio of 3.9 (4th highest value of the 58 cases) and a lagged change in vote share of -6.8 (3rd lowest value of the 58 cases). Thus, Front National in 2012 is in exactly the situation for which strategic descriptive representation expects PRRP to foster women's representation in future elections. In fact, with 50% women in parliament, Front National in 2012 had the highest value in the depend variable of all cases. At first sight, the case seems to fit the theoretical model very well.

However, a closer examination of the dependent variable indicates that the ostensibly high representation of women is attributable to the election of a single individual. In the 2012 election, the Front National secured only two seats: one by Gilbert Collard and the other by Marion Maréchal-Le Pen, the niece of Marine Le Pen. Consequently, the Front National's high proportion of 50% female legislators is not reflective of a broad support for women, but rather the outcome of having one of their two elected representatives being a woman. In essence, the results hinge not merely on the inclusion of a single data point in the regression model but also on the election of this particular woman. Had Marion Maréchal-Le Pen not been elected in 2012, the observed interaction effect would have been nonexistent.⁶

In sum, a more detailed analysis of the Front National in 2012 suggests that this case is not particularly representative of the assumed theoretical mechanism and should hence excluded from the analysis given its very high impact on the estimated parameters. Front National in 2012 did not elect *more women*; they elected *one woman*.

Modeling Strategies

While the previous section focused on one observation's disproportionate impact, the coming section concentrates on the various empirical specifications, including measurement strategy, chosen estimator, and model specification. We test the influence of these choices on the results of the full regression model.

⁶An in-depth analysis of Marion Maréchal-Le Pen's victory in her electoral district reveals that her success was rather fortuitous. She capitalized on the social-democratic candidate's refusal to back the right-wing UMP candidate in the second round of the election. Contrary to the social-democratic party leader's recommendation – which was to create a "cordon sanitaire" against the FN and support the UMP candidate to thwart Marion Maréchal-Le Pen's chances – the social-democratic candidate also ran in the second round. This split in the vote allowed Marion Maréchal-Le Pen to clinch the district with a relative majority of 42.09% (LePoint, 2012).

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60					
	(0.43)					
Lagged Change in Vote Share (B)	1.15^{*}	-0.14	1.81	1.89	2.34	2.51
	(0.51)	(0.18)	(1.56)	(1.45)	(1.50)	(1.43)
Lagged % Male Voters (A)	× ,	0.06	0.07	0.03	0.02	0.17
		(0.12)	(0.12)	(0.11)	(0.11)	(0.10)
(A) X (B)	-0.87^{**}		-0.03	-0.03	-0.04	-0.05
	(0.32)		(0.03)	(0.02)	(0.03)	(0.02)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
AIC	425.77	433.38	433.84	427.66	429.78	429.85
BIC	456.67	445.74	448.27	444.15	450.38	460.76
Log Likelihood	-197.88	-210.69	-209.92	-205.83	-204.89	-199.93
Num. obs.	58	58	58	58	58	58
Num. groups: party	22	22	22	22	22	22
Num. groups: country	19	19	19	19	19	19
Var: party (Intercept)	0.00	5.73	7.81	0.00	0.00	0.00
Var: country (Intercept)	23.91	53.69	50.65	57.24	53.05	5.22
Var: Residual	39.01	52.44	50.65	43.18	42.40	53.17

Table 3: Interaction effect when using lagged % of male voters

Notes: These models replicate the models from the original study but use the lagged percentage of male voters instead of the ratio of males over females. The column reference presents the original results from the model 5 as a reference. ***p < 0.001; **p < 0.01; *p < 0.05

Male-to-female voter ratio vs. share of male voters

The original study features mostly variables quantified as proportions, which aligns with standard practices. Notably, the dependent variable represents the *share* of women in parliament. Additionally, one of the key independent variables measures the lagged electoral success by shifts in vote *shares* between the two preceding elections. However, the male dominance in a party's voter composition is measured as the male-to-female voter *ratio*. Ratios can introduce non-linear distortions into the models so we test an alternative measurement strategy that relies on the share of males in a party's voter basis.

In light of this, we replicate the models using the share of male voters instead of the male-tofemale voter ratio. The revised outcomes, presented in Table 3, do not indicate any significant interaction effect when the share of male voters is used as the independent variable.

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.55	-0.55	-0.63	-0.72	-0.67
	(0.43)	(0.50)	(0.50)	(0.45)	(0.46)	(0.47)
Lagged Change in Vote Share (B)	1.15^{*}	-0.17	0.84	0.90	1.00	1.10
	(0.51)	(0.18)	(0.56)	(0.52)	(0.54)	(0.56)
(A) X (B)	-0.87^{**}		-0.66	-0.67^{*}	-0.76^{*}	-0.83^{*}
	(0.32)		(0.35)	(0.32)	(0.34)	(0.35)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
AIC	425.77	430.16	428.97	424.07	417.39	411.71
BIC	456.67	442.53	443.39	440.55	438.00	442.62
Log Likelihood	-197.88	-209.08	-207.48	-204.03	-198.70	-190.86
Num. obs.	58	58	58	58	58	58
Num. groups: party	22	22	22	22	22	22
Num. groups: country	19	19	19	19	19	19
Var: party (Intercept)	0.00	0.00	1.97	0.00	0.00	0.00
Var: country (Intercept)	23.91	71.06	61.48	60.83	55.22	41.91
Var: Residual	39.01	52.71	51.23	43.15	44.24	45.27

 Table 4: Interaction effect when using REML instead of log-likelihood

Notes: These models replicate the models from the original study but are estimated using restricted maximum likelihood instead of log-likelihood. The column reference presents the original results from model 5 as a reference. ***p < 0.001; **p < 0.01; *p < 0.05

Restricted Maximum Likelihood vs. Log-Likelihood

The original study estimates the multi-level model with the lmer function from the lme4 package. By default, this package employs restricted maximum-likelihood (REML) estimation, a method recommended by the literature on multi-level models (Elff et al., 2021). Instead, the original study estimates the model using a log-likelihood (LL) estimator. REML and LL can yield different results, especially when the sample size is small so we test an alternative specification that uses REML.

Table 4 presents the results using restricted maximum likelihood. Using REML marginally affects the results. While the coefficients slightly shrink and some p-values become larger (notably the interaction effect in Model 2 becomes non-significant at p < 0.05).

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.10	-0.26	-0.43	-0.37	-0.23
	(0.43)	(0.61)	(0.59)	(0.58)	(0.53)	(0.52)
Lagged Change in Vote Share (B)	1.15^{*}	-0.02	1.27	1.19	1.48^{*}	1.23*
	(0.51)	(0.25)	(0.67)	(0.65)	(0.59)	(0.59)
(A) X (B)	-0.87^{**}		-0.89^{*}	-0.80	-1.00^{*}	-0.91^{*}
	(0.32)		(0.43)	(0.42)	(0.38)	(0.37)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
FE Country	-	No	No	No	No	No
Num. obs.	58	58	58	58	58	58
\mathbb{R}^2		0.00	0.07	0.14	0.34	0.50
Adj. R ²		-0.04	0.02	0.08	0.26	0.38

 Table 5: Interaction effects based on OLS

***p < 0.001;**p < 0.01;*p < 0.05

Multi-Level Models vs. OLS

Using a multi-level modeling strategy allows the original study to account for the nested structure of the data by employing sophisticated cross-classified multi-level models with observations nested in parties and countries. While theoretically justified, given the data's nested structure, the limited number of observations prompts concerns regarding the suitability of such complex models. Indeed, the replication code triggers a warning in R: "boundary (singular) fit: see help('isSingular')", which may suggest issues of overfitting. In other words, the number of parameters is too high relative to the number of observations, and the estimation lacks degrees of freedom.

Although there is no definitive guideline on how to resolve this issue, juxtaposing the results with those obtained with a simple OLS can help increase trust in the estimates. OLS makes far fewer assumptions about the structure of the data, which is a great advantage when dealing with very small datasets. Notably, the original study performs this comparison in the appendix but does not include any country-fixed effects. Because the national context is very likely to influence both the dependent and independent variables, not accounting for the country-level introduces a confounding risk. Accordingly, we test two OLS specifications, one without fixed effects (following the paper's appendix) and one with country fixed effects.

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.75	-0.73	-0.73	-0.83	-0.87
	(0.43)	(0.54)	(0.53)	(0.49)	(0.50)	(0.46)
Lagged Change in Vote Share (B)	1.15^{*}	-0.21	0.57	0.73	0.45	0.60
	(0.51)	(0.19)	(0.63)	(0.58)	(0.63)	(0.59)
(A) X (B)	-0.87^{**}		-0.51	-0.58	-0.44	-0.45
	(0.32)		(0.39)	(0.36)	(0.39)	(0.37)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
FE Country	-	Yes	Yes	Yes	Yes	Yes
Num. obs.	58	58	58	58	58	58
\mathbb{R}^2		0.70	0.72	0.77	0.78	0.84
Adj. R ²		0.54	0.55	0.62	0.63	0.69

Table 6: Interaction effects based	on OLS with country	fixed-effects
------------------------------------	---------------------	---------------

***p < 0.001; **p < 0.01; *p < 0.05

The first series of OLS results are presented in Table 5. Generally, the results remain very stable and display robustness (only the interaction effect in model 3 becomes not significant (p > 0.05). However, once country fixed-effects are added as depicted in Table 6, not only do the coefficients become much smaller across all models, but they also all lose their statistical significance (p > .05).

Summary and Combination of Robustness Checks

This replication study identifies several issues that seriously question the empirical validity of the original paper's findings. Most critically, the main results are contingent upon a single observation. Other model specifications also impact the findings (but less drastically than the outlier issue). To encapsulate these findings, we undertake a comprehensive analysis by running a series of models that incorporate a range of different modeling choices. Specifically, we vary the following parameters:

- 1. Inclusion of Front National's 2012 data: included or excluded?
- 2. Measure of voter gender gap: male-to-female voter ratio or share of male voters?
- 3. Model specification: Multi-Level Model, Ordinary Least Squares (OLS) or OLS with coun-

try fixed-effects?

4. Estimation method: Log-Likelihood or Restricted Maximum-Likelihood? (applicable only if a multi-level model is estimated)





Note: The plot displays the p-values of the main interaction effect under different model specifications. Not a single specification obtains significant coefficients (p < 0.05) on all five models. Removing the observation of the Front National in 2012 ('FN 2012') removes any significant effect (lower panel of the plot). Significant effects are obtained only when using the male-to-female voter ratio and multi-level models or an OLS without fixed effects.

This approach allows us to evaluate the influence of each modeling choice, offering a more detailed insight into the robustness of the reported results. The combination of modeling choices results in 16 distinct models, and the p-values for the interaction effect in each are presented in Figure 2 and in Appendix A2. The data presented in the figure convey a conclusive trend. Excluding Front National's 2012 data results in the interaction effect being statistically insignificant across all model specifications. Similarly, despite the inclusion of the FN 2012,

using the male voter share instead of the male-to-female voter ratio led all models to yield non-significant results for the interaction. When using the FN 2012 and the male-to-female ratio, some specifications can yield significant effects, but not a single specification produces consistent significant effects across all five models.

Conclusion

Weeks et al. (2023) provide a new theoretical argument suggesting that PRRP elected more women when they lost votes in the previous election and were predominantly supported by male voters. The present replication casts serious doubt about the validity of the empirical findings. Most notably, the statistical evidence supporting the argument vanishes when a single observation is excluded. A closer inspection of this observation reveals that its idiosyncratic extreme values rely on the election of a single woman and do not fit the theoretical mechanism. Further modeling choices impact the results and yield more conservative and even sometimes null results.

In light of those results, the best interpretation of the available data is not to reject the null hypothesis. Strategic descriptive representation may be a reality but cannot be empirically captured with the proposed dataset. The number of observations is simply too low. As is well-known from previous research, testing interaction effects with sufficient power usually require a large number of observations (Gelman, Hill, and Vehtari, 2020, pp. 301–304). With low power, any observed 'significant effect' is likely to be biased (Gelman and Carlin, 2014). This means the true effect size is probably several times smaller than the observed effect size ('Type M' error) and there even is a high chance that the true effect is in the opposite direction ('Type S' error). Thus, properly testing strategic descriptive representation requires a much larger dataset to produce results with enough statistical power. Yet, the authors already did an impressive job at collecting the data from a large sample of European democracies and a very limited number of cases could be added even with unlimited resources. As a result, quantitatively testing the assumed mechanism proves to be very challenging.

In conclusion, we want to stress how relevant this research is for understanding the rela-

tionship between party competition and representation. The systematic under-representation of women (and other marginalized groups) in politics is an obvious inequality that needs to be better understood. Strategic descriptive representation suggests that political competition naturally encourages more descriptive representation of women, even for parties that defend traditional roles for women. Because the data does not show evidence in favor of strategic descriptive representation, it points out an inefficiency of party competition and calls even more for strong measures constraining parties to place more women in electable positions.

Disclaimer: The original replication stopped after the above conclusion. Following I4R's procedure, the authors had a chance to write a response and we added the next section, which addresses the most central arguments of their response.

Comments on WMKC's Response

We are grateful for the authors' detailed response and the effort invested in addressing our comments and clarifying some aspects of the paper. In this brief response, we consciously do not address all the points raised in the response. We do not want to start an indefinite back-and-forth with the authors and think that our point has been made clearly enough to allow readers to formulate their own judgment on the empirical evidence supporting strategic descriptive representation. Therefore, we briefly respond in the next section on the two most crucial points on which we disagree the most.

- Identification of Influential Cases.
- Relevance of the FN 2012 for the underlying theoretical argument.

Even after carefully reading the authors' response, we remain unconvinced that the empirical data contains any evidence of strategic descriptive representation.

Identification of Influential Cases

WMKC's critique centers on our method of removing only one influential case, the Front National in 2012, from the analysis. They qualify our method as "arbitrary" and suggest a more comprehensive approach that deletes at once all potentially influential cases to test the robustness of the regression model. Moreover, WMKC argue that influential cases should be identified at the cluster level, especially with nested data, identifying six country clusters as influential. We disagree with this view: as suggested by the literature, multilevel models should be robust to both the removal of one complete cluster or one observation.

Our methodology involved assessing all observations to determine if removing any single observation results in a significant change in the regression coefficients. This is consistent with robustness testing principles, where models should not yield substantially different estimates when single cases are omitted (Nieuwenhuis, Grotenhuis, and Pelzer, 2012, pp. 45–46).⁷ If the exclusion of a single observation is sufficient to drastically alter the conclusions drawn from a model, the model is not robust. This principle applies to both standard regression models as well as mixed-effect models. The only difference is that mixed effects models should be robust both to the exclusion of single observations *and* clusters of observations. Thus, excluding a single observation is standard practice for assessing the robustness of regression models.

Moreover, WMKC's argument for excluding all influential observations simultaneously contradicts standard practice, as influential observations are typically identified and tested one at a time to understand their individual impact on the model. When all cases are excluded simultaneously, it cannot be assessed how each observation drives the results of the model.

In addition to these general points on testing influential cases in regression models, the empirical approach in WMKC's response is flawed. Specifically, WMKC estimate Cook's D values *based on Model 5* to identify influential cases and then remove these cases from *Model 2*, which has a different model specification and no control variables.⁸ This is misleading because Cook's D is model-specific as it measures how the exclusion of observations affects all

⁷For a more detailed discussion, see Appendix D.

⁸The reason for this decision seems to be that Model 5 cannot be estimated when the six-country clusters are removed. While this is correct, it should be noted that Model 5 can be estimated when these cases are removed from the data if one of the control variables (measuring the electoral system in a country) is excluded. Alternatively, Models 3 and 4 can still be estimated despite removing those cases.

regression coefficients of a model. Consequently, the correct approach is to identify influential cases for each specific model and re-estimate each model with those cases excluded.

This has important consequences for the model results. As we demonstrate in Figure 3, the country clusters with a Cook's D value above the threshold of D > 4/58 are very different for each of the five model specifications. In other words, WMKC exclude several cases from Model 2 which are not influential in Model 2.





Note: Each dot represent the Cook's D value for every country cluster. Dotted vertical line is cut-off value for influential cases (defined as 4/N, where N is the number of country clusters, i.e. $4/19 \approx 0.21$). Clusters with values over the threshold are considered to be influential cases.

Hence, the correct approach would be to identify influential cases for each model specification (as displayed in Figure 3) and then to re-estimate each model with the respective cases excluded. Running these analyses confirms our initial findings that the relevant interac-

	Reference	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	0.87	0.22	-1.52^{*}	0.82
	(0.43)	(1.38)	(1.17)	(0.83)	(0.87)
Lagged Change in Vote Share (B)	1.15^{**}	0.60	0.65	0.07	0.01
	(0.51)	(0.72)	(0.60)	(0.40)	(0.41)
(A) X (B)	-0.87^{***}	-0.61	-0.54	-0.10	-0.19
	(0.32)	(0.51)	(0.43)	(0.28)	(0.30)
AIC	425.77	353.94	298.49	210.96	201.28
BIC	456.67	367.04	312.39	225.93	221.36
Log Likelihood	-197.88	-169.97	-141.24	-95.48	-86.64
Num. obs.	58	48	42	33	31
Num. groups: party	22	18	16	13	13
Num. groups: country	19	16	15	13	13
Var: party (Intercept)	0.00	3.98	0.00	0.00	0.47
Var: country (Intercept)	23.91	58.97	43.15	0.00	1.83
Var: Residual	39.01	40.31	28.09	19.08	13.65

Table 7:	Regression mo	odels of WN	MKC with	potentially	influential	country	clusters	excluded for	each r	nodel
	specification									

Notes: The models replicate the models from the original paper but remove for each model every country cluster that is identified as influential. The process of identifying influential cases is model-specific: for each model, we identify influential cases and remove them from the sample. As depicted in Figure 3, it means that we removed different clusters for each model. ***p < 0.01; **p < 0.05; *p < 0.1

tion term is substantially smaller and does not come close to conventional levels of statistical significance (see Table 7).

Even if we follow the response's (incorrect) approach and re-estimate all the models while removing the influential cases identified by WMKC, the interaction effect is only significant for Model 2.⁹. For Model 3, 4 and 5, the interaction effect vanishes (see Table 8). Model 2 is, thus, not representative for the other model specifications.

⁹As explained in the previous footnote, we estimate here an adapted version of Model 5 where one independent variable (measuring the type of electoral system) is dropped.

	Reference	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	0.99	0.28	0.20	0.82
	(0.43)	(0.98)	(0.87)	(0.91)	(0.87)
Lagged Change in Vote Share (B)	1.15^{**}	0.69	0.38	0.34	0.01
	(0.51)	(0.50)	(0.45)	(0.45)	(0.41)
(A) X (B)	-0.87^{***}	-0.69^{*}	-0.38	-0.36	-0.19
	(0.32)	(0.35)	(0.32)	(0.32)	(0.30)
AIC	425.77	217.26	210.58	214.16	201.28
BIC	456.67	227.30	222.05	228.50	221.36
Log Likelihood	-197.88	-101.63	-97.29	-97.08	-86.64
Num. obs.	58	31	31	31	31
Num. groups: party	22	13	13	13	13
Num. groups: country	19	13	13	13	13
Var: party (Intercept)	0.00	45.20	11.10	10.98	0.47
Var: country (Intercept)	23.91	17.96	38.52	37.89	1.83
Var: Residual	39.01	16.09	11.88	11.73	13.65

Table 8:	Regression	results when	1 potentially	[,] influential	country	clusters	from	Model 5	are e	excluded _.	from
	the other r	nodels									

Notes: These models replicate the models from the original study but remove the country clusters that are identified as influential cases based on Model 5. Hence, it follows the incorrect procedure proposed in the authors' response. ***p < 0.01; **p < 0.05; *p < 0.1

Front National in 2012

In their response, WMKC argue that the Front National in 2012 is crucial for understanding populism and the relationship between radical-right populism and gender representation. While we agree on its topical relevance, our concern with the case is statistical: the inclusion of Front National in 2012 skews results because it elected only two legislators, one of whom was female, resulting in an unusually high and uncertain value for the dependent variable. Small denominators introduce significant uncertainty, making it questionable whether the results should rely heavily on this case.

To illustrate this uncertainty, let's consider the following example. Suppose the actual latent probability of the Front National electing a woman is just 20%. Despite this low probability, there is still a 36% chance that, out of two elected legislators, at least one will be a woman.¹⁰ This demonstrates that, in more than one-third of the cases, we would observe a

¹⁰This probability is calculated by subtracting the probability of both legislators being men from 1. Given that the probability of electing a man is 0.8, the probability of electing two men is $0.8 \times 0.8 = 0.64$, and the probability of observing at least one elected woman is 1 - 0.64 = 0.36 or 36%.

considerably higher proportion of women elected, either 50% (P=.32) or 100% (P=.04) of women elected, than the actual underlying probability would suggest. In summary, proportions based on small sample sizes often fail to accurately reflect the true underlying value.

The outlier status of Front National in 2012 likely points to measurement problems of the latent construct due to the small number of elected MPs for the FN in 2012. Such influential observations can dramatically shift results; if the Front National had elected one additional legislator, the value of the dependent variable would drastically change¹¹, affecting the model's outcomes. This variability suggests that while the data point is not erroneous, it is unreliable, warranting caution in its inclusion in regression models.

In addition, WMKC also suggests that Front National's candidate selection in 2012 became more favorable to women, supporting the hypothesis of strategic descriptive representation. To empirically test this assumption in more detail, we collected all Front National candidates for each legislative election from 1997 to 2017¹². The results are presented below. Our systematic analysis of Front Nationals candidate data from 1997 to 2017 reveals that women were placed in less promising districts in 2012 than in 2007. Regression analysis shows a strong negative effect of candidate gender on electoral prospects in all French elections (Figure 4 and Table 9).

As a result, a more in-depth investigation of women's electoral prospect within the FN reveals no systematically better placement of female FN candidates in 2021, which does not support the concept of strategic descriptive representation.

¹¹With two legislators, the proportion of women was 50%. With three, it would have been either 33% or 66%. ¹²This data is publicly available for each election on the website of the French interior ministry https://www. data.gouv.fr/fr/datasets/

	Model 2002	Model 2007	Model 2012	Model 2017
(Intercept)	0.24	0.50^{*}	0.46^{*}	0.63**
	(0.27)	(0.22)	(0.23)	(0.21)
Vote Share in previous election	-36.79	-4.75^{**}	-12.00^{*}	-4.93^{***}
	(32.42)	(1.82)	(4.88)	(1.47)
AIC	142.38	767.58	722.31	779.59
BIC	147.61	776.22	730.83	788.28
Log Likelihood	-69.19	-381.79	-359.16	-387.80
Deviance	138.38	763.58	718.31	775.59
Num. obs.	101	556	523	568

Table 9: Effe	ct of FN's	previous	district s	uccess on	probability	of selecting	a women
---------------	------------	----------	------------	-----------	-------------	--------------	---------

Notes: Each column represent a different logistic regression. We model the gender of the candidate as a function of the vote share of the FN in a given district during the previous election. The negative coefficients suggest that women were less likely to be candidates in districts with higher vote share. Vote share during the previous election is here used as a proxy for electoral uncertainty. ***p < 0.001; **p < 0.01; *p < 0.05

Figure 4: Comparison of FN's candidate genders and previous electoral district success in the election of 2002 to 2017



Conclusion

Assessing the robustness of published work is essential for maintaining trust and confidence in the scientific publication process. We believe the issues raised in our initial replication report are significant for a broader audience and that our findings should be considered when interpreting the original results in WMKC. Notably, the disappearance of the central interaction effect after excluding a single observation raises concerns about the robustness of the original findings.

Highlighting these issues transparently is crucial so other scholars can evaluate their significance. This includes the possibility that others may find these issues less concerning. While we respectfully disagree with that interpretation, we remain convinced that such discussions are vital to scientific progress.

Bibliography

- Akkerman, Tjitske (2015). "Gender and the radical right in Western Europe: a comparative analysis of policy agendas". In: *Patterns of Prejudice* 49.1-2, pp. 37–60.
- Brambor, Thomas, William Roberts Clark, and Matt Golder (2006). "Understanding interaction models: Improving empirical analyses". In: *Political Analysis* 14.1, pp. 63–82.
- Caul, Miki (2001). "Political Parties and the Adoption of Candidate Gender Quotas: A Cross-National Analysis". In: *Journal of Politics* 63.4, pp. 1214–1229.
- Celis, Karen and Silvia Erzeel (2020). "Gender Equality". In: *The Oxford Handbook of Political Representation in Liberal Democracies*. Oxford University Press.
- Coffé, Hilde (2018). "Gender and the Radical Right". In: *The Oxford Handbook of the Radical Right*. Oxford University Press.
- Cook, R Dennis (1977). "Detection of influential observation in linear regression". In: *Technometrics* 19.1, pp. 15–18.
- De Lange, Sarah L and Liza M Mügge (2015). "Gender and right-wing populism in the Low Countries: ideological variations across parties and time". In: *Patterns of Prejudice* 49.1-2, pp. 61–80.

- Elff, Martin et al. (2021). "Multilevel Analysis with Few Clusters: Improving Likelihood-Based Methods to Provide Unbiased Estimates and Accurate Inference". In: *British Journal of Political Science* 51.1, pp. 412426.
- Erzeel, Silvia and Ekaterina R Rashkova (2017). "Still mens parties? Gender and the radical right in comparative perspective". In: *West European Politics* 40.4, pp. 812–820.
- Freese, Jeremy and David Peterson (2017). "Replication in Social Science". In: *Annual Review of Sociology* 43.1, pp. 147–165.
- Gelman, A., J. Hill, and A. Vehtari (2020). *Regression and Other Stories*. Cambridge University Press.
- Gelman, Andrew and John Carlin (2014). "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors". In: *Perspectives on Psychological Science* 9.6, pp. 641– 651.
- Immerzeel, Tim, Hilde Coffé, and Tanja Van der Lippe (2015). "Explaining the gender gap in radical right voting: A cross-national investigation in 12 Western European countries". In: *Comparative European Politics* 13, pp. 263–286.
- Kantola, Johanna and Emanuela Lombardo (2021). "Strategies of right populists in opposing gender equality in a polarized European Parliament". In: *International Political Science Review* 42.5, pp. 565–579.
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen (2017). "ImerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82.13, pp. 126.
- LePoint (June 13, 2012). "Vaucluse : le suppléant PS de Catherine Arkilovitch appelle à voter UMP". In: Le Point. URL: https://www.lepoint.fr/elections-legislatives/vauclusele-suppleant-ps-de-catherine-arkilovitch-appelle-a-voter-ump-13-06-2012-1472746_573.php (visited on 03/25/2022).

Mudde, Cas (2007). Populist Radical Right Parties in Europe. Cambridge University Press.

- Neumayer, Eric and Thomas Plümper (2017). *Robustness Tests for Quantitative Research*. Cambridge University Press.
- Nieuwenhuis, Rense, Manfred te Grotenhuis, and Ben Pelzer (2012). "influence.ME: Tools for Detecting Influential Data in Mixed Effects Models". In: *The R Journal* 4.2, pp. 38–47.

- Phillips, Anne (2020). "Descriptive Representation Revisited". In: *The Oxford Handbook of Political Representation in Liberal Democracies*. Oxford University Press.
- Spierings, Niels (2020). "Why gender and sexuality are both trivial and pivotal in populist radical right politics". In: *Right-wing populism and gender: European perspectives and beyond*.Ed. by G. Dietze and J. Roth, pp. 41–58.
- Spierings, Niels and Andrej Zaslove (2017). "Gender, populist attitudes, and voting: explaining the gender gap in voting for populist radical right and populist radical left parties". In: *West European Politics* 40.4, pp. 821–847.
- Wängnerud, Lena (2009). "Women in parliaments: Descriptive and substantive representation". In: *Annual Review of Political Science* 12, pp. 51–69.
- Weeks, Ana et al. (2022). Replication Data for: When do Männerparteien Elect Women? Radical Right Populist Parties and Strategic Descriptive Representation. Version V1. DOI: 10.7910/ DVN/SG55BJ.
- Weeks, Ana Catalano et al. (2023). "When Do Männerparteien Elect Women? Radical Right Populist Parties and Strategic Descriptive Representation". In: American Political Science Review 117.2, pp. 421–438.
- Wooldridge, Jeffrey M (2020). Introductory Econometrics: A Modern Approach. 7th. Cengage.

Appendices

A	Cook's distance for regression models	A1
B	P-values under different model specifications	B2
С	Share of Male Voters or Male/Female Voter Ratio (Added post-response)	C3
D	Outlier identification with multilevel models (Added post-response)	D4
E	R Script for Replication of Results	E5

A Cook's distance for regression models

Table A1: Cook's distance for WMKC's regression models

		Regre	ssion N	Iodels	
Dentes (Election ID	1	100,000	2010111	104010	
Party/Election ID	1	2	3	4	Э
France_Nat_201206	1.95	2.45	2.97	2.07	1.51
Greece_Gol_201501	4.92	1.39	1.02	0.58	1.21
Croatia_Cro_201609	0.56	0.35	0.47	1.13	0.74
Austria_Aus_200610	0.72	0.75	0.57	0.35	0.58
Greece_Ind_201501	0.58	0.66	0.62	0.36	0.21
Italy_Nor_201302	0.12	0.10	0.23	0.19	0.12
Slovenia_Slo_200010	0.05	0.07	0.21	0.14	0.12
Italy_Nat_200604	0.03	0.02	0.01	0.06	0.11
Denmark_Pro_199012	0.02	0.02	0.24	0.21	0.09
Greece_Ind_201509	0.03	0.24	0.30	0.14	0.09
Netherlands_Par_201209	0.05	0.06	0.06	0.04	0.08
Netherlands_Par_201703	0.20	0.19	0.12	0.07	0.07
Poland_Law_200710	0.00	0.25	0.18	0.04	0.07
Denmark_DPA_200711	0.15	0.14	0.17	0.13	0.06
Italy_Nat_199604	0.06	0.09	0.03	0.04	0.06
Switzerland_Sws_201110	0.01	0.02	0.04	0.13	0.06
Norway_Pro_201709	0.06	0.05	0.01	0.10	0.05
Denmark_DPA_201506	0.13	0.11	0.04	0.03	0.04
Luxembourg_Alt_200906	0.06	0.04	0.03	0.03	0.04
Switzerland_Sws_200310	0.10	0.01	0.01	0.07	0.04
Austria_Aus_199910	0.00	0.01	0.10	0.06	0.03
Denmark_DPA_200502	0.09	0.08	0.07	0.06	0.03
Finland_Tru_201504	0.36	0.22	0.09	0.08	0.03
Italy_Nor_200804	0.06	0.05	0.03	0.02	0.03
Luxembourg_Alt_201310	0.07	0.04	0.07	0.06	0.03
Slovakia_Slo_201603	0.16	0.08	0.03	0.00	0.03
Austria_Aus_201710	0.11	0.10	0.04	0.02	0.02
Finland_Tru_200703	0.19	0.09	0.04	0.03	0.02
France_Nat_201706	0.17	0.00	0.10	0.06	0.02
Italy_Nat_200105	0.05	0.04	0.02	0.02	0.02
Italy_Lea_201803	0.32	0.29	0.27	0.18	0.02
Luxembourg_Act_200406	0.06	0.08	0.02	0.02	0.02
Slovakia_Slo_201006	0.28	0.16	0.12	0.03	0.02
Sweden_Swe_201809	0.42	0.45	0.20	0.13	0.02
Austria_Aus_200809	0.02	0.01	0.01	0.01	0.01
Austria_Aus_201309	0.04	0.02	0.00	0.00	0.01
Bulgaria Nat 201405	0.62	0.59	0.48	0.29	0.01
Estopia Dro 201502	0.10	0.05	0.04	0.02	0.01
Estonia_Pro_201303	0.04	0.05	0.11	0.03	0.01
Italy Nor 200604	0.00	0.00	0.02	0.02	0.01
Italy_N01_200004	0.02	0.01	0.01	0.00	0.01
Norway Pro 200109	0.03	0.03	0.00	0.00	0.01
Norway Pro 200509	0.04	0.02	0.00	0.00	0.01
Norway Pro 201309	0.00	0.00	0.01	0.02	0.01
Romania Gre 200011	0.00	0.00	0.00	0.01	0.01
Bulgaria Nat 201305	0.02	0.00	0.01	0.01	0.00
Croatia Cro 201112	0.02	0.00	0.02	0.01	0.00
Denmark DPA 201109	0.00	0.00	0.01	0.00	0.00
Denmark Pro 199409	0.02	0.01	0.01	0.00	0.00
Finland Tru 201104	0.03	0.02	0.02	0.01	0.00
Italy Nor 200105	0.00	0.00	0.00	0.00	0.00
Norway Pro 200909	0.04	0.02	0.02	0.01	0.00
Poland Law 201110	0.00	0.02	0.02	0.00	0.00
Slovenia Slo 200809	0.00	0.01	0.00	0.00	0.00
Slovenia Slo 200410	0.09	0.08	0.07	0.05	0.00
Switzerland Sws 201510	0.02	0.01	0.00	0.00	0.00
Switzerland Sws 200710	0.00	0.00	0.00	0.00	0.00

B P-values under different model specifications

FN 2012?	Estimator	Gender gap	Model type	Model 2	Model 3	Model 4	Model 5
Yes	LL	M/F Ratio	multilevel	0.055	0.035	0.021	0.008
Yes	n.a.	M/F Ratio	ols	0.042	0.059	0.011	0.017
Yes	REML	M/F Ratio	multilevel	0.063	0.044	0.029	0.023
Yes	LL	% Male Voters	multilevel	0.216	0.170	0.101	0.070
Yes	n.a.	% Male Voters	ols	0.282	0.338	0.087	0.108
Yes	REML	% Male Voters	multilevel	0.234	0.190	0.124	0.119
Yes	n.a.	M/F Ratio	ols_fe	0.199	0.114	0.276	0.227
Yes	n.a.	% Male Voters	ols_fe	0.326	0.220	0.343	0.268
No	n.a.	M/F Ratio	ols	0.840	0.991	0.576	0.597
No	LL	M/F Ratio	multilevel	0.993	0.897	0.829	0.713
No	n.a.	% Male Voters	ols_fe	0.995	0.982	0.904	0.723
No	n.a.	% Male Voters	ols	0.891	0.770	0.734	0.746
No	REML	M/F Ratio	multilevel	0.994	0.902	0.839	0.776
No	n.a.	M/F Ratio	ols_fe	0.957	0.912	0.999	0.798
No	LL	% Male Voters	multilevel	0.915	0.861	0.916	0.864
No	REML	% Male Voters	multilevel	0.916	0.868	0.922	0.945

 Table A2: Adjusted p-values of interaction effect under different specifications

Note: Model 1 not reported because it does not contain the interaction effect. n.a. = not applicable.

C Share of Male Voters or Male/Female Voter Ratio (Added post-response)

In our initial study, we raised a concern about the decision to operationalize a male voter dominance based on the male-to-female voter ratio. We suggested that measuring the share of male voters might be more intuitive. WMKC responded that we misunderstood their oper-ationalization, clarifying that they estimated the share of male voters divided by the share of female voters, not the number of male voters divided by the number of female voters.

We acknowledge the difference in operationalization and recognize that our initial comment misunderstood this aspect. This confusion partly arose from WMKC's description of the variable, which often referred to it as an operationalization of a male-dominated voter base. WMKC's method, using the ratio of percentages, does not accurately measure a male or female voter base.

To make this point, let us consider the following example: 10 of 100 male voters voted for a PRRP (i.e., 10%) while 20 of 200 female voters voted for the PRRP (i.e., also 10%). 10% of both male and female voters support a PRRP and yield a M/F-Ratio of 1, despite observing a predominantly female electorate (two-thirds of voters being female). Instead of female dominance, it may be more intuitive to describe the M/F-Ratio as a measure of relative performance within two electorates.

Although the above example uses very different absolute numbers of male and female voters, most election surveys aim for gender balance, meaning our interpretation aligns closely with WMKC's approach. Formally, WMKC define the M/F Ratio as:

$$M/F \ Ratio_{PRRP} = \frac{\left(\frac{N_{Male \ Voters PRRP}}{N_{Male \ Voters}}\right)}{\left(\frac{N_{Female \ Voters PRRP}}{N_{Female \ Voters}}\right)}$$

If we assume a gender-balanced sample, which is the case in most electoral studies, it simplifies to:

$$M/F \ Ratio_{PRRP} = \frac{N_{Male \ Voters PRRP}}{N_{Female \ Voters PRRP}}$$

This aligns with our initial interpretation of the M/F-Ratio as representing the ratio of male to female voters, which can be transformed into percentages by:

 $\frac{M/F\ Ratio}{1+M/F\ Ratio}$

We acknowledge this is not WMKC's initial estimation, but it shows that in most cases, the differences between the ratio of percentages and absolute numbers are minimal unless the sample deviates significantly from gender parity.

D Outlier identification with multilevel models (Added postresponse)

Identifying influential cases in multi-level models requires careful consideration, especially when variables are measured at upper levels of a nested data structure. For example, when observations are nested within countries and variables are measured at the country level, excluding a single observation may not adequately check if results are influenced by the inclusion of a particular country. As (Nieuwenhuis, Grotenhuis, and Pelzer, 2012, p. 39) state: "To apply the logic of detecting influential data to generalized mixed effects models, one has to measure the influence of a particular higher level group on the estimates of a predictor measured at that level." This means that mixed-effect models should maintain stability not only when a single observation is excluded but also when a cluster of observations is removed. However, contrary to WMKC's response, this does not invalidate identifying influential observations at the individual level. Mixed effects models should be robust to the exclusion of both single observations and clusters. (Nieuwenhuis, Grotenhuis, and Pelzer, 2012, pp. 45-46) emphasize that identifying influential cases at the individual level in multi-level models is a valid approach. As none of the relevant variables of the interaction term is measured at the party- or country-level, there is no need to only assess robustness based on the exclusion of upper-level clusters.

E R Script for Replication of Results

Do Radical Right Parties Use Descriptive Representation Strategically? (R Script)

Michael Jankowski and Benjamin Guinaudeau

Document last compiled on: 2023-11-13

Replication of Weeks et al. "When do Männerparteien..." (APSR)

In this notebook, we document the *replication* and *additional analyses* of Weeks et al.'s paper "When Do Männerparteien Elect Women? Radical Right Populist Parties and Strategic Descriptive Representation" (doi: https://doi.org/10.1017/S0003055422000107; in the following: WMKC) published in the American Political Science Review (APSR).

Download Replication Material

The following code downloads the data and code provided by WMKC from the Harvard Dataverse. The download requires the dataverse-package (available from CRAN). After the download is completed, the code runs the R-file provided by WMKC which creates all results reported in WMKC. Please note: Several packages must be installed for the code of WMKC to run. Please also note that we do not report the output of this code here.

```
# Load package (install if necessary)
if(!require("pacman")) install.packages("pacman")
pacman::p_load(
  dataverse, tidyverse, directlabels,
  RColorBrewer, colorRamps, lme4,
  lmerTest, texreg, interplot,
  scales, readr, stargazer,
  prediction, pander, cowplot, xtable,
  patchwork
)
if(!require("dataverse")) install.packages("dataverse")
# Set dataverse server and doi for WMKC data
Sys.setenv("DATAVERSE_SERVER" = "dataverse.harvard.edu",
           "DATAVERSE_KEY" = "")
doi <- "doi:10.7910/DVN/SG55BJ"</pre>
# Download and save relevant files from dataverse
c("Replication_data.RData",
  "Replication code.R") %>%
  walk(~{
    .x %>%
      get_file(doi) %>%
      writeBin(.x)
```

})

Run replication code of WMKC
source("Replication_code.R")

ID Variable and Deletion of Excluded Cases

In the following, we create an id for each observation and delete all cases which contain missing values as they are excluded from the regression model. This creates the dataset on which WMKC run their regression models for PRRP.

```
df <- newd2 %>%
  dplyr::mutate(
    partyname = substr(partyname, 0, 3),
    id = paste(countryname, partyname, date, sep = "_")
) %>%
  dplyr::select(
    country, countryname, party, partyname, id,
    pfem_new2,
    lag1_mfcombined10, chgvotelagged,
    year, date, femaleleader2_lag, cabinet_party2_lag,
    lag1womenpar, tier1_avemag2, prop2, natquota,
    weurope
) %>%
  drop_na
```

Replication of WMKC's Regression Models

WMKC estimate five different models. The first model does not include the interaction, the second models adds the interaction term, the third model controls for time, the fourth and fifth models add further controls. In the following, we reproduce the regression results.

```
# Seed value used by WMKC
set.seed(02145)
```

```
# Formula for Model 1
form_model_1 <- "pfem_new2 ~ lag1_mfcombined10 + chgvotelagged + (1 | party ) + (1 | country)"</pre>
# Formula for Model 2
form_model_2 <- paste(form_model_1, "lag1_mfcombined10:chgvotelagged", sep = " + ")</pre>
# Formula for Model 3
form_model_3 <- paste(form_model_2, "year", sep = " + ")</pre>
# Formula for Model 4
form_model_4 <- paste(form_model_3, "femaleleader2_lag + cabinet_party2_lag", sep = " + ")</pre>
# Formula for Model 5
form_model_5 <- paste(form_model_4, "lag1womenpar + tier1_avemag2 +</pre>
                       as.factor(prop2) + natquota + weurope", sep = " + ")
# Use to ease readability and not print all control in the replication
regex_control_model_45 <- "Intercept|year|femaleleader2_lag|cabinet_party2_lag|
lag1womenpar|tier1_avemag2|prop2|natquota|weurope"
all_formulas <- c(
  form_model_1,
  form model 2,
  form_model_3,
 form_model_4,
  form_model_5
)
```

For this replication, we are going to use two main utility functions.

The first one get_estimate_and_p_values takes a fitted model and returns a data.frame containing (1) the model, (2) the coefficient for the interaction and (3) the corresponding p-values.

The second one get_interaction_plot takes a model and returns a graph representing the average marginal effect for the interaction.

```
get_estimate_and_p_values <- function(model){
  tmp <- summary(model)$coefficients
  estimate <- tmp[str_detect(rownames(tmp), "\\:"), "Estimate"]
  if(length(estimate) == 0) estimate <- NA_real_
    p_value <- tmp[str_detect(rownames(tmp), "\\:"), "Pr(>|t|)"]
  if(length(p_value) == 0) p_value <- NA_real_
  tibble::tibble(
    mod = list(model),
    estimate,
    p_value
  )
}</pre>
```

```
get_interaction_plot <- function(model, var1 = "lag1_mfcombined10"){</pre>
  interplot(m = model,
           var1 = var1,
           var2 = "chgvotelagged",
           hist=FALSE) +
   xlab("Vote Change (lagged)") +
   ylab("Estimated Coefficient for\n(lagged) M/F ratio") +
   theme_bw() +
   ggtitle("Estimated Coefficient of M/F ratio \non Women in Party by Vote Change") +
   theme(plot.title = element_text(face="bold")) +
   geom_hline(yintercept = 0, linetype = "dashed") +
   theme(text = element_text(size=16)) +
   lims(y = c(-30, 30))
}
shape_results <- function(models, title, omit.coef = NULL,</pre>
                         additional_rows = NULL, adjusted = T, label = ""){
  p_values <- models %>%
   map(~{
     if(adjusted){
       if(class(.x) == "lm"){
         degree_freedom <- summary(.x)$df[2]</pre>
       } else {
         degree_freedom <- parameters::dof_satterthwaite(.x)</pre>
       }
     } else {
       degree_freedom <- Inf</pre>
     }
     parameters::p_value_satterthwaite(.x, dof = degree_freedom)[["p"]]
   })
 additional_rows)
  custom_names <- tail(c("\tReference\t", paste("Model", 1:5)), length(models))</pre>
  custom.coef.map <- list(</pre>
    "lag1_mfcombined10" = "Lagged M/F-Ratio (A)",
    "chgvotelagged" = "Lagged Change in Vote Share (B)",
    "lag_mfshare" = "Lagged % Male Voters (A)",
    "lag1_mfcombined10:chgvotelagged" = "(A) X (B)",
    "lag_mfshare:chgvotelagged" = "(A) X (B)"
  )
  tex_output <- texreg(models,</pre>
                      omit.coef = omit.coef,
                      override.pvalues = p_values,
                      custom.model.names = custom_names,
                      custom.coef.map = custom.coef.map,
                      custom.gof.rows = custom_rows,
```

```
caption = title,
label = label,
caption.above = T,
booktabs = T,
float.pos = "hptb")
texreg_string <- capture.output(tex_output)
texreg_string <- texreg_string[!grep1("usepackage\\{booktabs\\}", texreg_string)]
cat(texreg_string, sep = "\n")
return(tex_output)
}
```

To replicate the results of the paper, we iterate over the list of formulas created previously. The paper reports p-value assuming infinite degrees of freedom. This underestimates the p-value, which in reality are larger than reported in the paper.

For the rest of this replication study, we always report p-values taking into account the degrees of freedom of the analysis using the Satterthwaite method.

```
## Warning in override(models = models, override.coef = override.coef, override.se
## = override.se, : p-values were provided using 'override.pvalues', but standard
## errors were not replaced!
```

Warning in override(models = models, override.coef = override.coef, override.se ## = override.se, : p-values were provided using 'override.pvalues', but standard ## errors were not replaced!

We also recreate the AME plot (Figure 5 in the paper).

```
set.seed(02145)
plot2 <- get_interaction_plot(all_mods$mod[[5]]) +
    labs(title = "WMKC")
plot2</pre>
```

	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.55	-0.55	-0.63	-0.73	-0.60
	(0.49)	(0.48)	(0.43)	(0.43)	(0.43)
Lagged Change in Vote Share (B)	-0.17	0.84	0.90	0.99	1.15^{*}
	(0.18)	(0.54)	(0.49)	(0.51)	(0.51)
(A) X (B)		-0.66^{*}	-0.67^{*}	-0.76^{*}	-0.87^{**}
		(0.34)	(0.31)	(0.32)	(0.32)
Controls 3			Х	Х	Х
Controls 4				Х	Х
Controls 5					X
AIC	432.39	430.72	423.28	424.75	425.77
BIC	444.75	445.15	439.76	445.35	456.67
Log Likelihood	-210.19	-208.36	-203.64	-202.37	-197.88
Num. obs.	58	58	58	58	58
Num. groups: party	22	22	22	22	22
Num. groups: country	19	19	19	19	19
Var: party (Intercept)	0.00	1.14	0.00	0.00	0.00
Var: country (Intercept)	65.97	57.15	56.16	49.58	23.91
Var: Residual	50.33	48.00	39.29	38.64	39.01

Table 1: Models Reported in WMKC (Table 1)

****p < 0.001; **p < 0.01; *p < 0.05

Table 2: Interaction effect with adjusted p-values (see Table 1 in WMKC)									
	Model 1	Model 2	Model 3	Model 4	Model 5				
Lagged M/F-Ratio (A)	-0.55	-0.55	-0.63	-0.73	-0.60				
	(0.49)	(0.48)	(0.43)	(0.43)	(0.43)				
Lagged Change in Vote Share (B)	-0.17	0.84	0.90	0.99	1.15^{*}				
	(0.18)	(0.54)	(0.49)	(0.51)	(0.51)				
$(A) \ge (B)$		-0.66	-0.67^{*}	-0.76^{*}	-0.87^{**}				
		(0.34)	(0.31)	(0.32)	(0.32)				
Controls 3			Х	Х	Х				
Controls 4				Х	Х				
Controls 5					Х				
AIC	432.39	430.72	423.28	424.75	425.77				
BIC	444.75	445.15	439.76	445.35	456.67				
Log Likelihood	-210.19	-208.36	-203.64	-202.37	-197.88				
Num. obs.	58	58	58	58	58				
Num. groups: party	22	22	22	22	22				
Num. groups: country	19	19	19	19	19				
Var: party (Intercept)	0.00	1.14	0.00	0.00	0.00				
Var: country (Intercept)	65.97	57.15	56.16	49.58	23.91				
Var: Residual	50.33	48.00	39.29	38.64	39.01				

Fable 2∙	Interaction	effect	with	adjusted	n-values	(see	Table	1 in	WMKC)	
Table 2.	interaction	enect	W1011	aujusteu	p-values	(see	Table	1 111	www.witco)	

***p < 0.001; ** p < 0.01; * p < 0.05



Robustness Tests

Cook's Distance

We run Cook's Distance analyses for identifying potential outliers.

% latex table generated in R 4.0.3 by xtable 1.8-4 package % Mon Nov 13 10:59:48 2023

```
cooks %>%
  pivot_longer(-id) %>%
  mutate(model = str_extract(name, "\\d+")) %>%
  mutate(id = fct_reorder(id, value, .fun = max)) %>%
  ggplot(aes(x = value, y = id, color = model)) +
  geom_point(position = position_dodge(width = .5)) +
  theme_minimal() +
  labs(color = "Model", y = "", x = "Cook's Distance")
```

I4R DP No. 149



Golden down exhibits high cook's distance for the first model that does not include the interaction. In contrast, the Front National 2012 has high Cook's distance no matter what model specification is considered. Therefore, we analyze FN 2012 in more detail.

id	$cooksd_1$	$cooksd_2$	$cooksd_3$	$cooksd_4$	$cooksd_5$
France_Nat_201206	1.95	2.45	2.97	2.07	1.51
$Greece_Gol_201501$	4.92	1.39	1.02	0.58	1.21
Croatia_Cro_201609	0.56	0.35	0.47	1.13	0.74
Austria_Aus_200610	0.72	0.75	0.57	0.35	0.58
$Greece_Ind_201501$	0.58	0.66	0.62	0.36	0.21
Italy_Nor_201302	0.12	0.10	0.23	0.19	0.12
Slovenia_Slo_200010	0.05	0.07	0.21	0.14	0.12
Italy_Nat_200604	0.03	0.02	0.01	0.06	0.11
Denmark_Pro_199012	0.02	0.02	0.24	0.21	0.09
$Greece_Ind_201509$	0.03	0.24	0.30	0.14	0.09
Netherlands_ Par_201209	0.05	0.06	0.06	0.04	0.08
Netherlands_ Par_201703	0.20	0.19	0.12	0.07	0.07
$Poland_Law_200710$	0.00	0.25	0.18	0.04	0.07
$Denmark_DPA_200711$	0.15	0.14	0.17	0.13	0.06
Italy_Nat_199604	0.06	0.09	0.03	0.04	0.06
Switzerland_Sws_201110	0.01	0.02	0.04	0.13	0.06
Norway_Pro_201709	0.06	0.05	0.01	0.10	0.05
$Denmark_DPA_201506$	0.13	0.11	0.04	0.03	0.04
$Luxembourg_Alt_200906$	0.06	0.04	0.03	0.03	0.04
$Switzerland Sws_200310$	0.10	0.01	0.01	0.07	0.04
Austria_Aus_199910	0.00	0.01	0.10	0.06	0.03
$Denmark_DPA_200502$	0.09	0.08	0.07	0.06	0.03
$Finland_Tru_201504$	0.36	0.22	0.09	0.08	0.03
Italy_Nor_200804	0.06	0.05	0.03	0.02	0.03
Luxembourg_Alt_201310	0.07	0.04	0.07	0.06	0.03
Slovakia_Slo_201603	0.16	0.08	0.03	0.00	0.03
Austria_Aus_201710	0.11	0.10	0.04	0.02	0.02
$Finland_Tru_200703$	0.19	0.09	0.04	0.03	0.02
France_Nat_201706	0.17	0.00	0.10	0.06	0.02
Italy_Nat_200105	0.05	0.04	0.02	0.02	0.02
Italy_Lea_201803	0.32	0.29	0.27	0.18	0.02
$Luxembourg_Act_200406$	0.06	0.08	0.02	0.02	0.02
Slovakia_Slo_201006	0.28	0.16	0.12	0.03	0.02
Sweden_Swe_201809	0.42	0.45	0.20	0.13	0.02
Austria_Aus_200809	0.02	0.01	0.01	0.01	0.01
Austria_Aus_201309	0.04	0.02	0.00	0.00	0.01
$Belgium_Fle_201405$	0.62	0.59	0.48	0.29	0.01
Bulgaria_Nat_201410	0.10	0.05	0.04	0.02	0.01
Estonia_Pro_201503	0.04	0.05	0.11	0.05	0.01
Estonia_Pro_201103	0.00	0.00	0.02	0.02	0.01
Italy_Nor_ 200604	0.02	0.01	0.01	0.00	0.01
Italy_Ita_199204	0.03	0.03	0.00	0.00	0.01
Norway_Pro_200109	0.04	0.02	0.00	0.00	0.01
Norway_Pro_200509	0.05	0.03	0.01	0.02	0.01
Norway_Pro_201309	0.00	0.00	0.00	0.01	0.01
$Romania_Gre_200011$	0.23	0.14	0.01	0.02	0.01
Bulgaria_Nat_201305	0.02	0.00	0.01	0.01	0.00
Croatia_Cro_201112	0.03	0.01	0.02	0.12	0.00
Denmark_DPA_201109	0.00	0.00	0.01	0.00	0.00
Denmark_Pro_199409	0.02	0.01	0.01	0.00	0.00
$Finland_Tru_201104$	0.03	0.02	0.02	0.01	0.00
Italy_Nor_200105	0.00	0.00	0.00	0.00	0.00
Norway_Pro_200909	0.04	0.02	0.02	0.01	0.00
Poland_Law_201110	0.00	0.02	0.02	0.00	0.00
Slovenia_Slo_200809	0.00	0.01	0.00	0.00	0.00
Slovenia_Slo_200410	0.09	0.08	0.07	0.05	0.00
Switzerland_Sws_201510	0.02	0.01	0.00	0.00	0.00
$Switzerland Sws_{200710}$	0.00	0.00	0.00	0.00	0.00

Removal of the Front National 2012

In this section, we simply re-estimate the models after excluding the FN 2012 observation.

```
set.seed(02145)
all_mods_wofn <- all_formulas %>%
 map_dfr(~{
   df %>%
     # Excluding the FN
     filter(id != "France_Nat_201206") %>%
     lmer(as.formula(.x), data = ., REML = FALSE) %>%
     get_estimate_and_p_values()
 })
## boundary (singular) fit: see help('isSingular')
table_2 <- c(all_mods$mod[5], all_mods_wofn$mod) %>%
  shape_results(title = "Interaction effects without Front National 2012",
                label = "table:wofn",
```

```
omit.coef = regex_control_model_45)
```

Warning in override(models = models, override.coef = override.coef, override.se ## = override.se, : p-values were provided using 'override.pvalues', but standard ## errors were not replaced!

eraction enec	us without	FIOID Nati	0mai 2012		
Reference	Model 1	Model 2 $$	Model 3	Model 4	Model 5
-0.60	-0.70	-0.70	-0.80	-0.82	-0.64
(0.43)	(0.46)	(0.46)	(0.41)	(0.42)	(0.40)
1.15^{*}	-0.03	-0.02	-0.04	0.15	0.25
(0.51)	(0.17)	(0.64)	(0.57)	(0.59)	(0.56)
-0.87^{**}		-0.00	0.05	-0.09	-0.14
(0.32)		(0.43)	(0.39)	(0.40)	(0.38)
Х			Х	Х	Х
Х				Х	Х
Х					Х
425.77	417.45	419.45	409.84	412.46	410.88
456.67	429.71	433.75	426.18	432.89	441.52
-197.88	-202.73	-202.73	-196.92	-196.23	-190.44
58	57	57	57	57	57
22	22	22	22	22	22
19	19	19	19	19	19
0.00	0.00	0.00	0.00	0.00	0.00
23.91	51.11	51.11	49.57	41.73	18.60
39.01	45.68	45.68	35.21	36.08	34.76
	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $

Table 5. Interaction energy without From Mational 20	Г	Table 3:	Interaction	effects	without	Front	National	201
--	---	----------	-------------	---------	---------	-------	----------	-----

*** p < 0.001; ** p < 0.01; * p < 0.05

Comparing effect sizes:

100*all_mods_wofn\$estimate/all_mods\$estimate

[1] NA 0.5627219 -7.5483508 11.5142480 16.0813812

Comparing P-values

all_mods<mark>\$</mark>p_value

[1] NA 0.054827152 0.034907578 0.020629973 0.008211174

all_mods_wofn\$p_value

[1] NA 0.9931741 0.8969154 0.8294891 0.7133342

We can also reproduce the conditional AME plot and combine it with the previous results. Both AME plots can be combined for a better comparison.



For the model with all control variables (Model 5), the reported estimates for the interaction term are more than 6 times larger when Front National in 2012 is included. The p-values of the models without the Front National in 2012 are nowhere near conventional levels of statistical significance.

REML vs LL

Now, we investigate wether using restricted maximum likelihood instead of log-likelihood has an influence on the results.

```
set.seed(02145)
all_mods_reml <- all_formulas %>%
map_dfr(~{
    lmer(as.formula(.x), data = df, REML = TRUE) %>%
    get_estimate_and_p_values()
})
```

```
## boundary (singular) fit: see help('isSingular')
all_mods$p_value
```

 $[1] \text{ NA } 0.054827152 \ 0.034907578 \ 0.020629973 \ 0.008211174 \\$

```
all_mods_reml$p_value
```

[1] NA 0.06328302 0.04378456 0.02936484 0.02255860

Warning in override(models = models, override.coef = override.coef, override.se ## = override.se, : p-values were provided using 'override.pvalues', but standard ## errors were not replaced!

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.55	-0.55	-0.63	-0.72	-0.67
	(0.43)	(0.50)	(0.50)	(0.45)	(0.46)	(0.47)
Lagged Change in Vote Share (B)	1.15^{*}	-0.17	0.84	0.90	1.00	1.10
	(0.51)	(0.18)	(0.56)	(0.52)	(0.54)	(0.56)
(A) X (B)	-0.87^{**}		-0.66	-0.67^{*}	-0.76^{*}	-0.83^{*}
	(0.32)		(0.35)	(0.32)	(0.34)	(0.35)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
AIC	425.77	430.16	428.97	424.07	417.39	411.71
BIC	456.67	442.53	443.39	440.55	438.00	442.62
Log Likelihood	-197.88	-209.08	-207.48	-204.03	-198.70	-190.86
Num. obs.	58	58	58	58	58	58
Num. groups: party	22	22	22	22	22	22
Num. groups: country	19	19	19	19	19	19
Var: party (Intercept)	0.00	0.00	1.97	0.00	0.00	0.00
Var: country (Intercept)	23.91	71.06	61.48	60.83	55.22	41.91
Var: Residual	39.01	52.71	51.23	43.15	44.24	45.27

Table 4: Interaction effect when using REML instead of log-likelihood

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

```
# clipr::write_clip(table_3)
```

```
plot_grid(
   plot2 + labs(title = "WMKC"),
   get_interaction_plot(all_mods_reml$mod[[5]]) +
    labs(y = "", title = "REML")
}
```



Male/Female Voter Ratio vs. Share of Male Voters

Now we estimate the models but use the percenatge of male voters as independent variable instead of the male-to-female voter ratio.

```
set.seed(02145)
# Ratio to percentages
df$lag_mfshare <- 100*df$lag1_mfcombined10/(1+df$lag1_mfcombined10)
all_mods_share <- all_formulas %>%
  str_replace_all("lag1_mfcombined10", "lag_mfshare") %>%
  map_dfr(~{
   lmer(as.formula(.x), data = df, REML = FALSE) %>%
     get_estimate_and_p_values()
 })
## boundary (singular) fit: see help('isSingular')
## boundary (singular) fit: see help('isSingular')
## boundary (singular) fit: see help('isSingular')
table_4 <- c(all_mods$mod[5], all_mods_share$mod) %>%
  shape_results(title = "Interaction effect when using lagged \\% of male voters",
                label = "table:shareratio",
                omit.coef = regex_control_model_45)
```

Warning in override(models = models, override.coef = override.coef, override.se ## = override.se, : p-values were provided using 'override.pvalues', but standard ## errors were not replaced!

		0	00			
	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60					
	(0.43)					
Lagged Change in Vote Share (B)	1.15^{*}	-0.14	1.81	1.89	2.34	2.51
	(0.51)	(0.18)	(1.56)	(1.45)	(1.50)	(1.43)
Lagged $\%$ Male Voters (A)		0.06	0.07	0.03	0.02	0.17
		(0.12)	(0.12)	(0.11)	(0.11)	(0.10)
(A) X (B)	-0.87^{**}		-0.03	-0.03	-0.04	-0.05
	(0.32)		(0.03)	(0.02)	(0.03)	(0.02)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	X					Х
AIC	425.77	433.38	433.84	427.66	429.78	429.85
BIC	456.67	445.74	448.27	444.15	450.38	460.76
Log Likelihood	-197.88	-210.69	-209.92	-205.83	-204.89	-199.93
Num. obs.	58	58	58	58	58	58
Num. groups: party	22	22	22	22	22	22
Num. groups: country	19	19	19	19	19	19
Var: party (Intercept)	0.00	5.73	7.81	0.00	0.00	0.00
Var: country (Intercept)	23.91	53.69	50.65	57.24	53.05	5.22
Var: Residual	39.01	52.44	50.65	43.18	42.40	53.17

Table 5: Interaction effect when using lagged % of male voters

*** p < 0.001; ** p < 0.01; * p < 0.05

In this case, there is no significant interaction effect. This can also be seen in the conditional AME plot.



Adding another scale for y, which will replace the existing scale.



Multi-Level vs. OLS

We now use simple OLS analysis for estimating the models. We run OLS without country FE and with country FE.

```
## Warning in override(models = models, override.coef = override.coef, override.se
## = override.se, : p-values were provided using 'override.pvalues', but standard
## errors were not replaced!
```

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.10	-0.26	-0.43	-0.37	-0.23
	(0.43)	(0.61)	(0.59)	(0.58)	(0.53)	(0.52)
Lagged Change in Vote Share (B)	1.15^{*}	-0.02	1.27	1.19	1.48^{*}	1.23^{*}
	(0.51)	(0.25)	(0.67)	(0.65)	(0.59)	(0.59)
(A) X (B)	-0.87^{**}		-0.89^{*}	-0.80	-1.00^{*}	-0.91^{*}
	(0.32)		(0.43)	(0.42)	(0.38)	(0.37)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
FE Country	-	No	No	No	No	No
AIC	425.77					
BIC	456.67					
Log Likelihood	-197.88					
Num. obs.	58	58	58	58	58	58
Num. groups: party	22					
Num. groups: country	19					
Var: party (Intercept)	0.00					
Var: country (Intercept)	23.91					
Var: Residual	39.01					
\mathbb{R}^2		0.00	0.07	0.14	0.34	0.50
Adj. \mathbb{R}^2		-0.04	0.02	0.08	0.26	0.38

Table 6: Interaction effects based on OLS

***p < 0.001; **p < 0.01; *p < 0.05

set.seed(02145)

all_mods_olsfe <- all_formulas %>%

```
str_replace_all("\\(1 \\| party \\) \\+ \\(1 \\| country\\)", "countryname") %>%
map_dfr(~{
```

Warning in override(models = models, override.coef = override.coef, override.se ## = override.se, : p-values were provided using 'override.pvalues', but standard ## errors were not replaced!

Table 7:	Interaction	effects	based	on OLS	with	country	^r fixed-e	effects
	ו	Roferon		[ode] 1	Mod	പാ	Andel 3	Model

	Reference	Model 1	Model 2	Model 3	Model 4	Model 5
Lagged M/F-Ratio (A)	-0.60	-0.75	-0.73	-0.73	-0.83	-0.87
	(0.43)	(0.54)	(0.53)	(0.49)	(0.50)	(0.46)
Lagged Change in Vote Share (B)	1.15^{*}	-0.21	0.57	0.73	0.45	0.60
	(0.51)	(0.19)	(0.63)	(0.58)	(0.63)	(0.59)
(A) X (B)	-0.87^{**}		-0.51	-0.58	-0.44	-0.45
	(0.32)		(0.39)	(0.36)	(0.39)	(0.37)
Controls 3	Х			Х	Х	Х
Controls 4	Х				Х	Х
Controls 5	Х					Х
FE Country	-	Yes	Yes	Yes	Yes	Yes
AIC	425.77					
BIC	456.67					
Log Likelihood	-197.88					
Num. obs.	58	58	58	58	58	58
Num. groups: party	22					
Num. groups: country	19					
Var: party (Intercept)	0.00					
Var: country (Intercept)	23.91					
Var: Residual	39.01					
\mathbb{R}^2		0.70	0.72	0.77	0.78	0.84
$Adj. R^2$		0.54	0.55	0.62	0.63	0.69

Summary of all Model Specifications

We now combine the robustness tests. We combine the following modeling choices:

- Inclusion of Front National 2012: Yes vs. No
- Regression model: Multilevel, OLS+FE, OLS
- Gender composition of voters: Male-Female Ratio or Share of Male Voters in t-1
- Estimation method: Log-Likelihood or REML (only applies to Mixed-Effect Models)

```
# Set up the specification grid
specifications <- crossing(
  fn_included = c(F, T),
  model_type = c("multilevel", "ols", "ols_fe"),
  mf_ratio = c(F, T),
  reml = c(F, T)
) %>%
  # Excluding REML spec with OLS
  filter(model_type == "multilevel" | !reml)
```

The function below takes a specification and returns a data-frame containing the model, estimates and p_values.

```
get_models_specification <- function(df_tmp,</pre>
                                      fn_included = T,
                                      mf_ratio = T,
                                      model_type = "multilevel",
                                      reml = F){
  if(!fn_included){
    df_tmp <- df_tmp %>%
      filter(id != "France_Nat_201206")
  }
  if(!mf ratio){
    all_formulas <- all_formulas %>%
      str_replace_all("lag1_mfcombined10", "lag_mfshare")
  }
  if(model_type == "ols"){
    all_formulas <- all_formulas %>%
      str_replace_all("\\+ \\(1 \\| party \\) \\+ \\(1 \\| country\\)",
                      "")
  }
  if(model_type == "ols_fe"){
    all_formulas <- all_formulas %>%
      str_replace_all("\\(1 \\| party \\) \\+ \\(1 \\| country\\)",
                       "countryname")
  }
  all_formulas %>%
    imap_dfr(~{
      if(model_type == "multilevel"){
        mod <- lmer(as.formula(.x), data = df_tmp, REML = reml)</pre>
```

```
} else {
        mod <- lm(as.formula(.x), data = df_tmp)</pre>
      }
      mod %>%
        get_estimate_and_p_values() %>%
        mutate(name = paste("Model", .y))
    }) <mark>%>%</mark>
    mutate(
      fn_included = fn_included, reml = reml, model_type = model_type, mf_ratio
    )
}
final_models <- specifications %>%
 pmap_dfr(get_models_specification, df_tmp = df)
## boundary (singular) fit: see help('isSingular')
final_pvalues <- final_models %>%
  dplyr::select(-mod, -estimate) %>%
  pivot_wider(
    names_from = name,
    values_from = c(p_value)
 ) %>%
```

```
arrange(`Model 5`) %>%
dplyr::select(-`Model 1`)
```

final_pvalues

```
## # A tibble: 16 x 8
      fn_included reml model_type mf_ratio `Model 2` `Model 3` `Model 4` `Model 5`
##
                                  <lgl>
##
      <lgl>
                  <lgl> <chr>
                                                <dbl>
                                                          <dbl>
                                                                    <dbl>
                                                                              <dbl>
##
   1 TRUE
                 FALSE multilevel TRUE
                                               0.0548
                                                         0.0349
                                                                   0.0206
                                                                            0.00821
## 2 TRUE
                  FALSE ols
                                   TRUE
                                               0.0424
                                                         0.0588
                                                                   0.0107
                                                                            0.0175
## 3 TRUE
                 TRUE multilevel TRUE
                                               0.0633
                                                         0.0438
                                                                   0.0294
                                                                            0.0226
## 4 TRUE
                  FALSE multilevel FALSE
                                               0.216
                                                         0.170
                                                                   0.101
                                                                            0.0697
## 5 TRUE
                  FALSE ols
                                  FALSE
                                               0.282
                                                         0.338
                                                                   0.0870
                                                                            0.108
## 6 TRUE
                  TRUE multilevel FALSE
                                               0.234
                                                         0.190
                                                                   0.124
                                                                            0.119
## 7 TRUE
                  FALSE ols_fe
                                  TRUE
                                               0.199
                                                         0.114
                                                                   0.276
                                                                            0.227
## 8 TRUE
                                               0.326
                                                                            0.268
                  FALSE ols_fe
                                  FALSE
                                                         0.220
                                                                   0.343
## 9 FALSE
                  FALSE ols
                                               0.840
                                                         0.991
                                                                   0.576
                                                                            0.597
                                   TRUE
                  FALSE multilevel TRUE
## 10 FALSE
                                               0.993
                                                         0.897
                                                                   0.829
                                                                            0.713
## 11 FALSE
                  FALSE ols_fe
                                  FALSE
                                               0.995
                                                         0.982
                                                                   0.904
                                                                            0.723
## 12 FALSE
                  FALSE ols
                                   FALSE
                                               0.891
                                                         0.770
                                                                   0.734
                                                                            0.746
## 13 FALSE
                  TRUE multilevel TRUE
                                               0.994
                                                         0.902
                                                                   0.839
                                                                            0.776
## 14 FALSE
                  FALSE ols fe
                                  TRUE
                                               0.957
                                                         0.912
                                                                   0.999
                                                                            0.798
## 15 FALSE
                  FALSE multilevel FALSE
                                               0.915
                                                         0.861
                                                                   0.916
                                                                            0.864
                  TRUE multilevel FALSE
## 16 FALSE
                                               0.916
                                                         0.868
                                                                   0.922
                                                                            0.945
xtable_7 <- final_pvalues %>%
  transmute(
      `FN 2012?` = ifelse(fn_included, "Yes", "No"),
      `Estimator` = ifelse(reml, "REML", "LL"),
      `Gender gap` = ifelse(mf_ratio,"M/F Ratio", "% Male Voters"),
      `Model type` = model_type,
      across(matches("Model \\d"))
    ) %>%
  mutate(Estimator = ifelse(`Model type` == "multilevel", Estimator, "n.a.")) %>%
  xtable::xtable(caption = "Adjusted p-values of interaction
                 effect under different specifications",
                 digits = 3, label = "multispec")
xtable_7 %>%
  xtable::print.xtable("latex", booktabs = TRUE,
                       caption.placement = "top",
                       include.rownames = F)
```

% latex table generated in R 4.0.3 by xtable 1.8-4 package % Mon Nov 13 10:59:57 2023

```
final_pvalues %>%
pivot_longer(matches("Model \\d")) %>%
mutate(
   published = (model_type == "multilevel" & !reml & mf_ratio & fn_included) |
    (model_type == "ols" & !reml & mf_ratio & fn_included),
   fn = ifelse(fn_included, "With FN 2012", "Without FN 2012"),
   model_type = case_when(
      model_type == "ols" ~ "OLS",
      model_type == "ols_fe" ~ "OLS with FE",
      T ~ "Multilevel"
```

FN 2012?	Estimator	Gender gap	Model type	Model 2	Model 3	Model 4	Model 5
Yes	LL	M/F Ratio	multilevel	0.055	0.035	0.021	0.008
Yes	n.a.	M/F Ratio	ols	0.042	0.059	0.011	0.017
Yes	REML	M/F Ratio	multilevel	0.063	0.044	0.029	0.023
Yes	LL	% Male Voters	multilevel	0.216	0.170	0.101	0.070
Yes	n.a.	% Male Voters	ols	0.282	0.338	0.087	0.108
Yes	REML	% Male Voters	multilevel	0.234	0.190	0.124	0.119
Yes	n.a.	M/F Ratio	ols_fe	0.199	0.114	0.276	0.227
Yes	n.a.	% Male Voters	ols_fe	0.326	0.220	0.343	0.268
No	n.a.	M/F Ratio	ols	0.840	0.991	0.576	0.597
No	LL	M/F Ratio	multilevel	0.993	0.897	0.829	0.713
No	n.a.	% Male Voters	ols_fe	0.995	0.982	0.904	0.723
No	n.a.	% Male Voters	ols	0.891	0.770	0.734	0.746
No	REML	M/F Ratio	multilevel	0.994	0.902	0.839	0.776
No	n.a.	M/F Ratio	ols_fe	0.957	0.912	0.999	0.798
No	LL	% Male Voters	multilevel	0.915	0.861	0.916	0.864
No	REML	% Male Voters	multilevel	0.916	0.868	0.922	0.945

Table 8: Adjusted p-values of interaction effect under different specifications

```
),
 reml = ifelse(model_type %in% c("OLS", "OLS with FE"),
                "",
                paste0("(", ifelse(reml, "REML", "LL"), ")")),
 mf_ratio = ifelse(mf_ratio, "M/F Ratio", "% Male Voters"),
 id = glue::glue("{model_type} {reml} with {mf_ratio}"),
 id = ifelse(published, paste0(id,"*"), id),
 id = tidytext::reorder_within(id, value, within = fn, .fun = min),
 sig = ifelse(value < 0.05, "Sig", "Not sig")</pre>
) %>%
ggplot(aes(x = value, y = id, color = name, shape = sig)) +
geom_point(position = position_dodge(width = .6), size = 2) +
facet_wrap(~fn, ncol = 1, scales = "free_y") +
geom_vline(xintercept = 0.05, linetype = 4) +
theme_minimal(base_size = 14) +
labs(x = "p-value (based on Satterthwaite method)",
    y = "Specification",
    color = "",
    shape = "") +
viridis::scale_color_viridis(discrete = T) +
tidytext::scale_y_reordered() +
scale_shape_manual(values = c("Sig" = 8, "Not sig" = 16)) +
scale_x_continuous(expand = c(0, 0), limits = c(0, 1.01))
```



Session Info

pander(sessionInfo())

R version 4.0.3 (2020-10-10)

Platform: x86_64-apple-darwin17.0 (64-bit)

locale: $en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||C||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US.UTF-8||en_US$

attached base packages: stats, graphics, grDevices, utils, datasets, methods and base

other attached packages: patchwork(v.1.1.3), xtable(v.1.8-4), cowplot(v.1.1.1), pander(v.0.6.5), prediction(v.0.3.14), stargazer(v.5.2.3), scales(v.1.2.1), interplot(v.0.2.3), arm(v.1.13-1), MASS(v.7.3-53), abind(v.1.4-5), texreg(v.1.38.6), lmerTest(v.3.1-3), lme4(v.1.1-28), Matrix(v.1.2-18), colorRamps(v.2.3.1), RColorBrewer(v.1.1-3), directlabels(v.2023.8.25), lubridate(v.1.9.3), forcats(v.1.0.0), stringr(v.1.5.0), dplyr(v.1.1.3), purr(v.1.0.2), readr(v.2.1.4), tidyr(v.1.3.0), tibble(v.3.2.1), ggplot2(v.3.4.4), tidyverse(v.2.0.0), dataverse(v.0.3.13), pacman(v.0.5.1) and here(v.1.0.1)

loaded via a namespace (and not attached): nlme(v.3.1-149), insight(v.0.19.6), httr(v.1.4.7), rprojroot(v.2.0.4), SnowballC(v.0.7.1), numDeriv(v.2016.8-1.1), tools(v.4.0.3), backports(v.1.4.1), utf8(v.1.2.4), R6(v.2.5.1), colorspace(v.2.1-0), withr(v.2.5.2), gridExtra(v.2.3), tidyselect(v.1.2.0), curl(v.5.1.0), compiler(v.4.0.3), textshaping(v.0.3.6), cli(v.3.6.1), labeling(v.0.4.3), bayestestR(v.0.13.1), checkmate(v.2.3.0), quadprog(v.1.5-8), systemfonts(v.1.0.4), digest(v.0.6.33), minqa(v.1.2.4), rmarkdown(v.2.25), pkgconfig(v.2.0.3), htmltools(v.0.5.7), fastmap(v.1.1.1), highr(v.0.10), rlang(v.1.1.2), rstudioapi(v.0.15.0), generics(v.0.1.3), farver(v.2.1.1), jsonlite(v.1.8.7), interactionTest(v.1.2), tokenizers(v.0.3.0), magrittr(v.2.0.3), stringi(v.1.7.12), yaml(v.2.3.7), grid(v.4.0.3), lattice(v.0.20-41), splines(v.4.0.3), hms(v.1.1.3), knitr(v.1.45), pillar(v.1.9.0), boot(v.1.3-25), glue(v.1.6.2), evaluate(v.0.23), tidytext(v.0.4.1), data.table(v.1.14.8), vctrs(v.0.6.4), nloptr(v.2.0.0), tzdb(v.0.4.0), gtable(v.0.3.4), datawizard(v.0.9.0), xfun(v.0.41), coda(v.0.19-4), janeaustenr(v.1.0.0), viridisLite(v.0.4.2), ragg(v.1.2.2) and timechange(v.0.2.0)

Appendix Bibliography

Nieuwenhuis, Rense, Manfred te Grotenhuis, and Ben Pelzer (2012). "influence.ME: Tools for Detecting Influential Data in Mixed Effects Models". In: *The R Journal* 4.2, pp. 38–47.