Does fake news affect voting behaviour?

A quasi-experiment on the 2018 Italian elections in Trentino Alto-Adige

Michele Cantarella [1] Nicolò Fraccaroli [2] Roberto Volpe [3]

[1]University of Modena and Reggio Emilia, European Central Bank

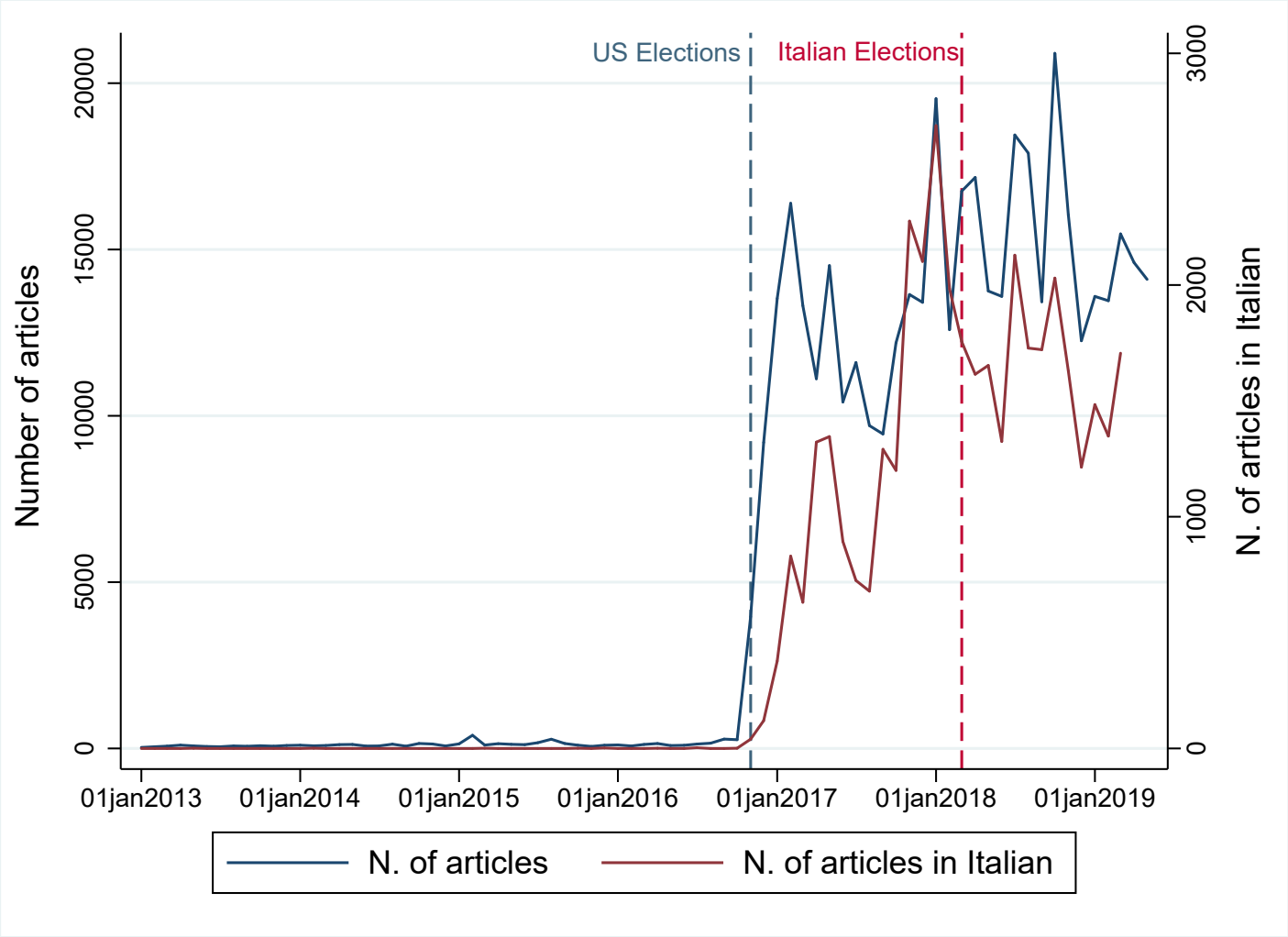
[2]University of Rome Tor Vergata, Bank of England

[3]Italian Ministry of Economic Development

4 November 2019

Modelling with Big Data and Machine Learning
Bank of England - King's College, London (United Kingdom)

The salience of fake news grew around elections



Note: number of newspaper articles in all languages (blue line) and in Italian (red line) containing the term 'fake news'. Source: Factiva.

Motivation

- *Fake News: intentionally fabricated misinformation characterised by politically-charged content.*

- *Correlation* between exposure to fake news and populist voting
  - Allcott e Gentzkow (2017): Trump voters are more prone to believe fake news
  - Gunther et al. (2019): voters who shifted from Obama to Trump tend to believe more fake news

  ...but no *causation*

- Lack of research on the Italian case. The 2018 election suggests a link between fake news and populism (Giglietto et al. 2018):
  - Diffusion of non-traditional and "biased" sources of information on social networks
  - "Biased" sources close to M5s e Lega had a higher reach wrt incumbent (PD) and other anti-establishment (FdI, neofascists, PaP,...) parties

Research Question

Does Fake News Affect Voting Behaviour?

- Two plausible hypotheses:
  - H_0: fake news influences voters in favour of anti-establishment parties
  - H_1: fake news reaches voters who already support anti-establishment parties $\Rightarrow$ fake news has no effect on voters
- Ideal setting: lab-experiment on similar voters, some of whom are treated, the others are not $\Rightarrow$ difficult and costly

Our aim is to identify the direction of causation

Why Trentino-Alto Adige? (1)

- **Intuition:** the German-speaking community of Alto Adige/Südtirol is an ideal control group in an experiment
- The two autonomous provinces of Trento and Bolzano/Bozen are *similar* in terms of...
  - Population (ca. 500 000 ab) and population density
  - Income
  - Degree of legislative and fiscal autonomy
- ... but have a fundamental *difference*:
  - About 2/3 of the Alto Adige/Südtirol population has German as its mothertongue, whereas Trentino is uniformly Italian-speaking

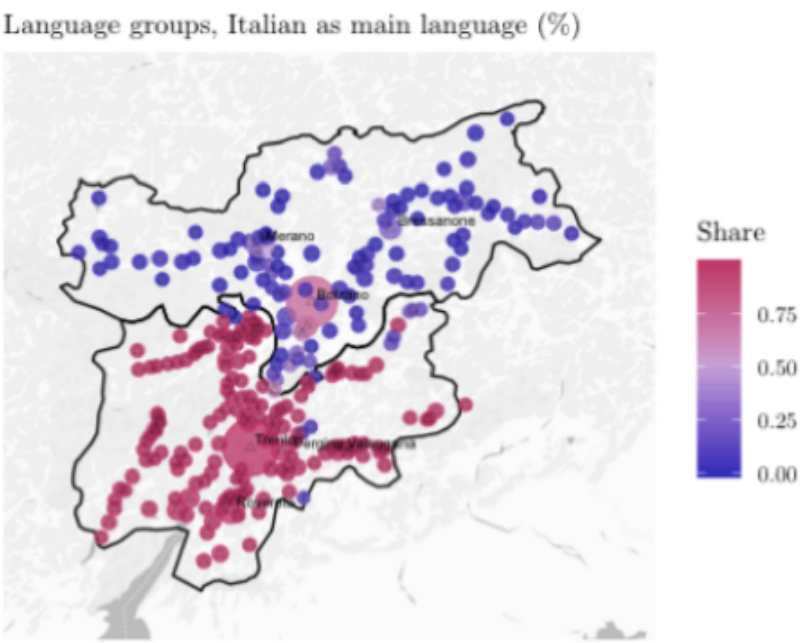
Why Trentino-Alto Adige? (2)

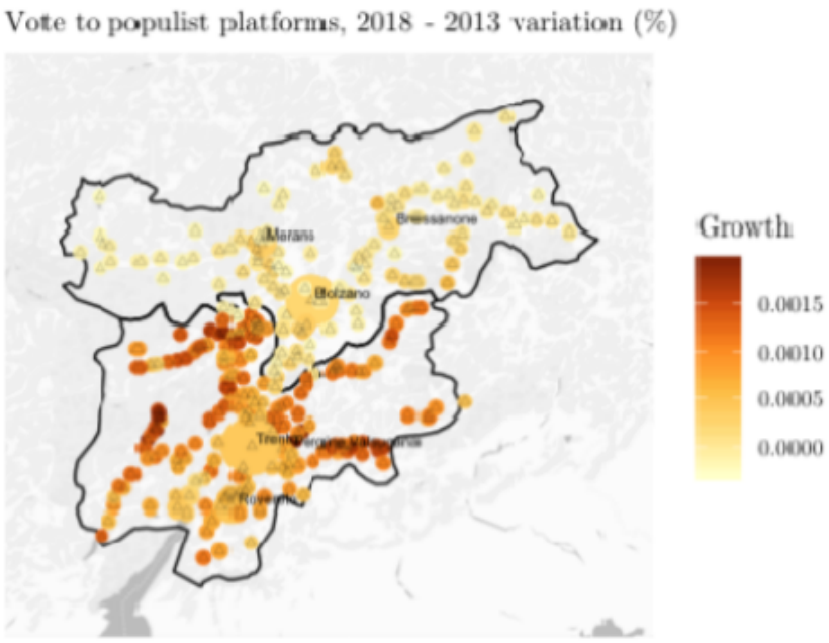
Our assumptions:

- German-speakers are in a **separate mediatic filter bubble**
  - Functional bilinguism is not spread (<10% DE students has a C1 level of IT, and viceversa) (Ebner, 2016)
  - In many Alto Adige municipalities >95% has German as mothertongue
- **Disseminators are rational agents:** producing fake news in German for Italy has a low marginal effect in terms of electoral returns

It follows that the German-speaking voters are less exposed to fake news on Italian politics (as they are in Italian)

Linguistic groups and populism in Trentino-Alto Adige





Source: Authors' elaborations on electoral data and on data on social media posts by political parties.

Votes in Bolzano/Bozen

Model

OLS: $Y_{ti} = \alpha + (F_{ti}P_t)\lambda + F_{ti}\delta + P_t\zeta + X'_{ti}\gamma + e_{ti}$

DID: $Y_{ti} = \alpha + (Z_{ti}P_t)\lambda + Z_{ti}\delta + P_t\zeta + X'_{ti}\gamma + e_{ti}$

"First stages":

(A.1) $F_{ti} = \alpha_1 + (Z_{ti} \times P_t)\lambda_1 + Z_{ti}\delta_1 + P_t\zeta_1 + X'_{ti}\gamma_1 + e_{1ti}$

(A.2) $(F_{ti} \times P_t) = \alpha_2 + (Z_{ti} \times P_t)\lambda_2 + Z_{ti}\delta_2 + P_t\zeta_2 + X'_{ti}\gamma_2 + e_{2ti}$

Final Regression:

(B.1) $Y_{ti} = \alpha_3 + (F_{ti} \hat{\times} P_t)\lambda_3 + \hat{F}_{ti}\delta_3 + P_t\zeta_3 + X'_{ti}\gamma_3 + e_{3ti}$

- Y_{ti} is the product of populism and electoral result in municipality i in election $t \in [2013; 2018]$
- F_{ti} is the number of voters exposed to fake news for each i and t
- Z_{ti} is the share (%) of Italian-mothertongue voters
- P_t is a time dummy
- X'_{ti} is a vector of controls (*Electorate size, Income per capita, Number of broadband connections, Number of abstentions and invalid votes*)

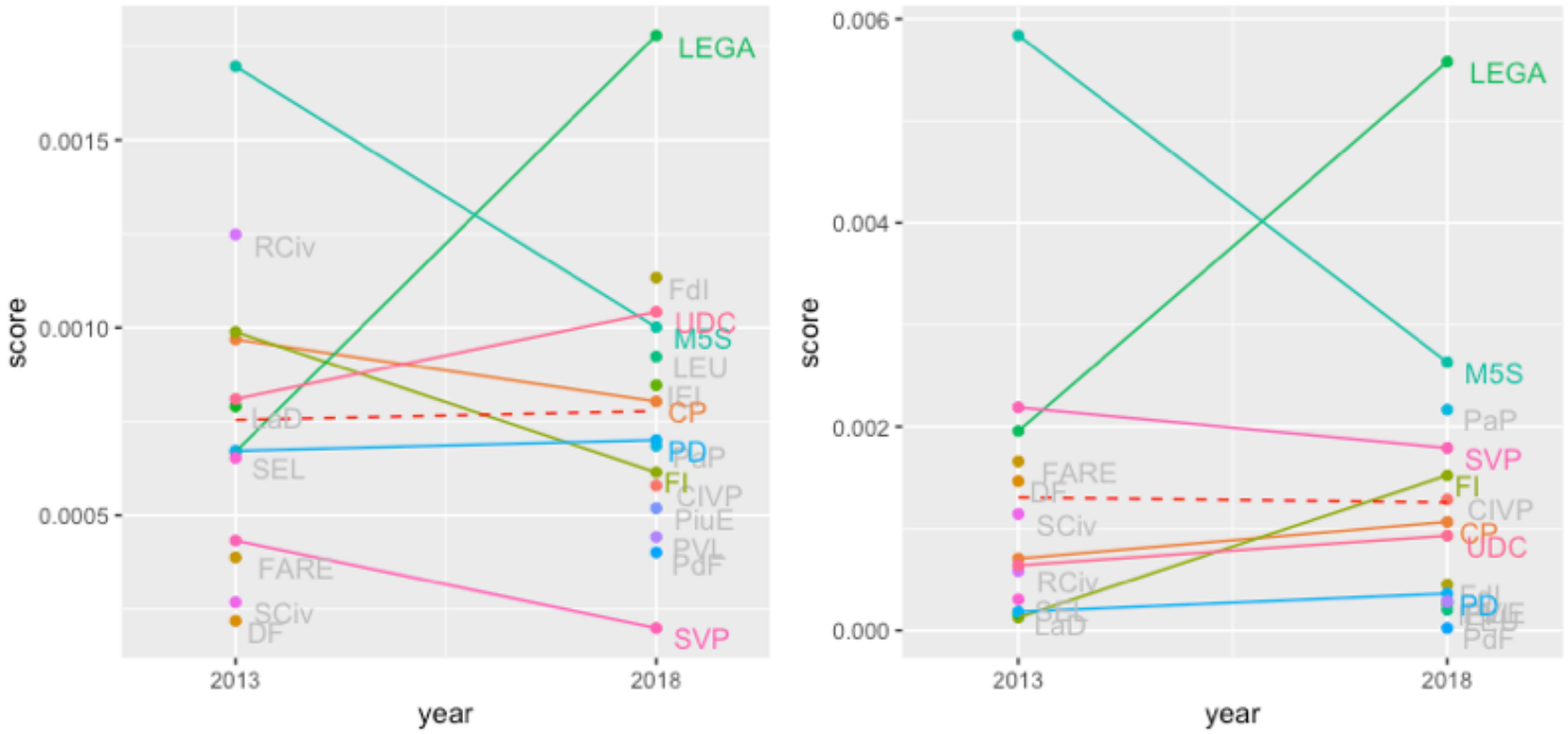
Empirical Strategy

- **Main approach**
  - Linguistic groups are exploited as sources of random allocation of exposure to fake news
    - ⇒ Instrumental variable regression
- **Two methodological issues:**
  - Different but parallel voting traditions between linguistic groups
    - ⇒ Difference in differences (variation 2013-2018)
  - Parties running in 2013 e 2018 are not always the same
    - ⇒ Populism as continuous variable
    - ⇒ Control for the variation in abstentions

Measuring Populism

- Defining a party as 'populist' in an objective manner is not easy (multiple dimensions: communication, ideology, political economy...)
- Given the focus on fake news, we look at populism as *political rhetoric*
- We construct two **indicators of populism**:
  - Anti-establishment rhetoric: words like *caste* or *elite*;
  - Assertive tone: number of exclamation marks.
- We measure the frequency of these *text bags* in the Facebook posts of all parties and their leaders before the 2013 and 2018 elections
- In this way, we obtain a score of populism for each party + leader for each election

Degree of populism by party (and leader) in the 2013 and 2018 Italian elections



Fonte: Elaboration by the authors based on social media posts of political parties and their leaders. The left figure refers to the scores obtained with the populist text-bag, while the right figurecomputes the same score as the frequency of exclamation marks in the text. Parties in grey have only took partin one of the two elections. The red dashed line refers to the election specific average.

Data

- **Italian Ministry of the Interior:** electoral results by municipality
- **Astat and census:** population and linguistic groups
- **Italian Ministry of Economy and Finance:** income per capita (based on IRPEF tax)
- **ACT (2013) and AgCom (2018):** high-speed internet penetration
- **Facebook:** social media posts (more than 20k) during the electoral campaign
- **Facebook Audience Insights Tool:** fake news penetration by municipality $\Rightarrow$ number of likes to disseminator pages flagged by debunkers, by municipality

Results: OLS estimates

Table 1: OLS estimates of the effect of misinformation on populist vote

Table I

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Exposed to fake news	-0.002	-0.003***	-0.003***	-0.003***
	(0.002)	(0.001)	(0.000)	(0.000)
Exposed to fake news × Year of election	0.002*	0.003***	0.003***	0.003***
	(0.001)	(0.000)	(0.000)	(0.000)
Year of election	0.778***	0.211**	0.114	0.728***
	(0.073)	(0.100)	(0.107)	(0.077)
Italian speaking voters	0.002***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Broadband connections		0.002***	0.002***	-0.000
		(0.000)	(0.000)	(0.000)
Income per capita (natural log)			1.251***	0.167
			(0.352)	(0.161)
Electorate size				0.002***
				(0.000)
Abstentions and invalid votes				-0.000
				(0.000)
Observations	584	584	584	584
Adjusted R-squared	0.934	0.992	0.992	0.997

Notes: Robust standard errors in parentheses. OLS estimates for the effect of misinformation on populist vote (natural log). Populist scores computed using the 'Assertive' text bag. *p<.05; **p<.01; ***p<.001

Results: instrumental variable estimates

Table 2: 2SLS estimates of the effect of misinformation on populist vote

	DiD	First Stages		2SLS
VARIABLES		Exposure	Interaction	
	(1)	(2)	(3)	(4)
Italian speaking voters	-0.0001	0.131**	-0.009	
	(0.0001)	(0.054)	(0.023)	
Italian speaking voters × Year of election	0.0005***	0.297***	0.455***	
	(0.0000)	(0.039)	(0.057)	
Year of election	0.4336***	-207.135***	-332.732***	0.778***
	(0.0689)	(33.100)	(46.819)	(0.098)
Broadband connections	-0.0000	0.021	0.013	-0.000
	(0.0001)	(0.079)	(0.043)	(0.000)
Electorate size	0.0019***	0.022	-0.019	0.002***
	(0.0001)	(0.053)	(0.051)	(0.000)
Income per capita (natural log)	0.4947***	-143.422	-37.960	0.491
	(0.1908)	(171.636)	(124.660)	(0.299)
Abstentions and invalid votes	-0.0003	-0.047	0.063	-0.000
	(0.0003)	(0.143)	(0.158)	(0.000)
Exposed to fake news				-0.000
				(0.001)
Exposed to fake news × Year of election				0.001***
				(0.000)
Observations	584	584	584	584
Adjusted R-squared	0.9972	0.931	0.930	0.994
Partial R-squared		0.342	0.397	
F-Test		0.714	0.830	

*N*otes: Standard errors robust to clustering by municipality in parentheses.

Conclusions and Discussion (1)

- Once exposure to fake news is randomly assigned, our results show a **negligible** of fake news on voting
  - For each additional voter who likes a FB page disseminator in a municipality, the populist content of parties gain increases of 0.001 points
- OLS estimates display a positive correlation between populism and fake news ⇒ voters **self-select** into "misinformation bubbles" (*ex-ante* preferences)
- If fake news proliferate in specific echo chambers, it is due to the **confimation bias** ⇒ rise of populism due to other factors

Conclusions and Discussion (2)

Why does fake news have such a low effect? Three hypotheses:

1- Disseminators underestimate the ability of voters to recognise hoaxes
2- Voters are indifferent to the number of facts in support of / against their preferences
3- Echo chambers influence voters' political preferences and perception of reality without the need of fake news

$\Rightarrow$ As long as one voter is able to recognise a fake news, the marginal effect of a true news will always be higher of the one of a fake news perceived as true

Validity of our results

- **Internal validity:**
  - Our results do *not* imply that fake news are unable to generate externalities, or that they may have problematic social implications
  - Individual behaviour remains unexplored: survey- or esperiment-based studies may highlight the unobservable dynamic at municipal level
  - We cannot exclude that fake news influences voters' shifts across populist platforms
- **External validity:**
  - Results for Trentino may not be valid for other contexts: studies on other regions/countries may shed light on different local dynamics
    - $\Rightarrow$ If so, need to identify the local drivers of fake news' effectiveness (ability to identify fake news, political culture, etc.)

Future patterns of research

1. Machine-learning + cosine similarity approaches to measure the 'rhetoric affinity' to fake news through topic analysis
2. Lab-experiment to identify the social and economic determinants of individual access to "misinformation bubbles" $\Rightarrow$ Survey
3. Expand our analysis to the national context exploiting broadband as instrument
4. Study the other social costs of misinformation

Thank you for your attention!1!!

Populist score and dependent variable

$$Popscore_{tp} = (Freq_{tpd} / T_{tpd}) * 10$$

$$Y_{ti} = \sum_{p=1}^{P} Popscore_{tp} * Votes_received_{tpi}$$

- *Freq* is the number of words matched for each text bag in each Facebook post d by each party/leader $p = \{1, ..., P\}$ three months ahead of each election $t \in [2013, 2018]$.
- T is the total number of words contained in each FB post d.
- *Votes_received* is the number of votes received by each party p in each election t in each municipality i.
- *Populism* measures the electoral weight of each party p in each election t in each municipality i.

Back

IV limited to the growth of Lega and Five Star Movement

Table 3: 2SLS estimates of the effect of misinformation on vote to LEGA and M5S (binary outcome)

	(1)	(2)	(3)	(4)
	Language group		IT connectivity	
VARIABLES	DiD	2SLS	DiD	2SLS
Italian speaking voters	0.17663***		0.20725***	0.03842
	(0.00838)		(0.01061)	(0.27479)
Italian speaking voters × Year of election	0.11989***			
	(0.00928)			
Broadband connections	0.01619	-0.02315	0.06160***	
	(0.01265)	(0.06440)	(0.01656)	
Broadband connections × Year of election			0.13135***	
			(0.00804)	
Year of election	73.24935***	247.33200***	45.93550***	233.22003**
	(13.78798)	(46.84445)	(12.43112)	(111.98957)
Electorate size (natural log)	61.15233***	348.66490***	85.44525***	302.21239
	(12.16195)	(86.66988)	(11.66156)	(293.99039)
Income per capita (natural log)	10.53376	-408.42314***	-29.45080	-341.87453
	(37.38124)	(118.24482)	(36.52034)	(445.24376)
Abstentions and invalid votes	0.05411	-0.03984	-0.28461***	-0.07156
	(0.04848)	(0.18060)	(0.05609)	(0.21393)
Exposed to fake news		1.31970***		0.99183
		(0.35714)		(1.74009)
Exposed to fake news × Year of election		-0.56535**		-0.34184
		(0.25948)		(1.07813)
Observations	584	584	584	584
Adjusted R-squared	0.994	0.910	0.995	0.932

Summary statistics

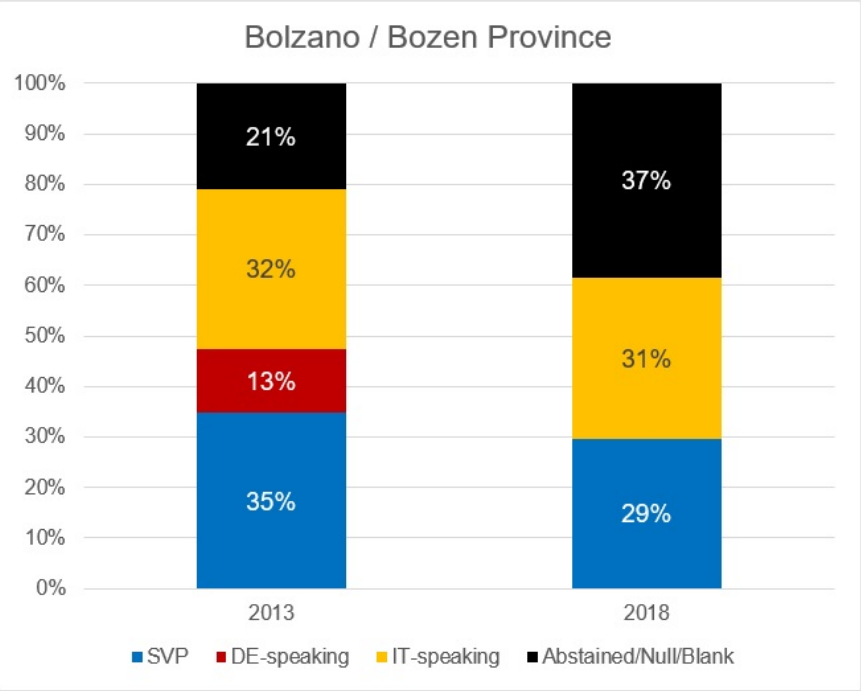
Table 4: Summary statistics by province and year

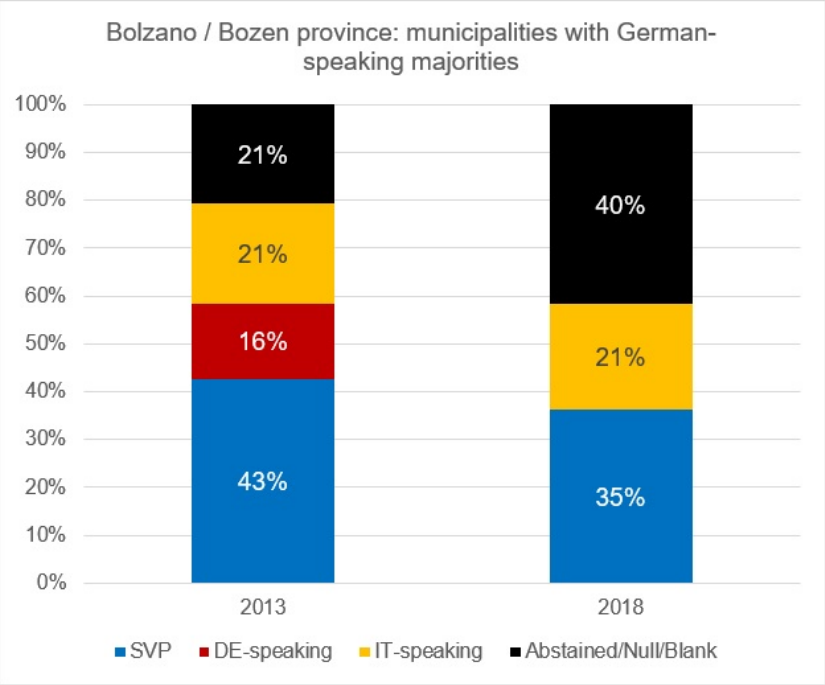
	(1) Bolzano/Bozen			(2) Trento		
	2013	2018	Average	2013	2018	Average
Populist score (total)	5.972	6.526	6.249	4.123	5.899	5.011
	(13.376)	(15.823)	(14.621)	(12.074)	(15.465)	(13.883)
Exposed to fake news	148.424	376.876	262.650	199.858	521.874	360.866
	(869.421)	(2366.432)	(1782.496)	(1206.976)	(3459.335)	(2592.059)
Broadband connections	2469.032	3007.955	2738.494	1987.041	2217.626	2102.334
	(7542.113)	(7661.528)	(7590.388)	(6763.404)	(7030.532)	(6889.395)
Italian speaking voters	879.493	907.630	893.562	2173.312	2226.701	2200.007
	(5321.100)	(5435.457)	(5366.947)	(6837.192)	(6963.666)	(6890.933)
Income per capita	18721.313	20692.652	19706.982	17193.012	18434.036	17813.524
	(2534.504)	(2728.883)	(2807.309)	(2188.215)	(1891.333)	(2134.696)
Electorate size	3245.284	3332.853	3289.069	2276.602	2330.716	2303.659
	(7548.803)	(7683.594)	(7600.120)	(6986.644)	(7112.244)	(7039.726)
Abstentions and invalid votes	677.569	1345.802	1011.685	524.455	649.989	587.222
	(1668.889)	(2502.587)	(2148.627)	(1499.638)	(1966.976)	(1747.624)
Observations	232			352		

*N*otes: Mean coefficients, standard deviations in parentheses.

Back

Electoral results in the province of Bolzano and in municipalities with German-speaking majority

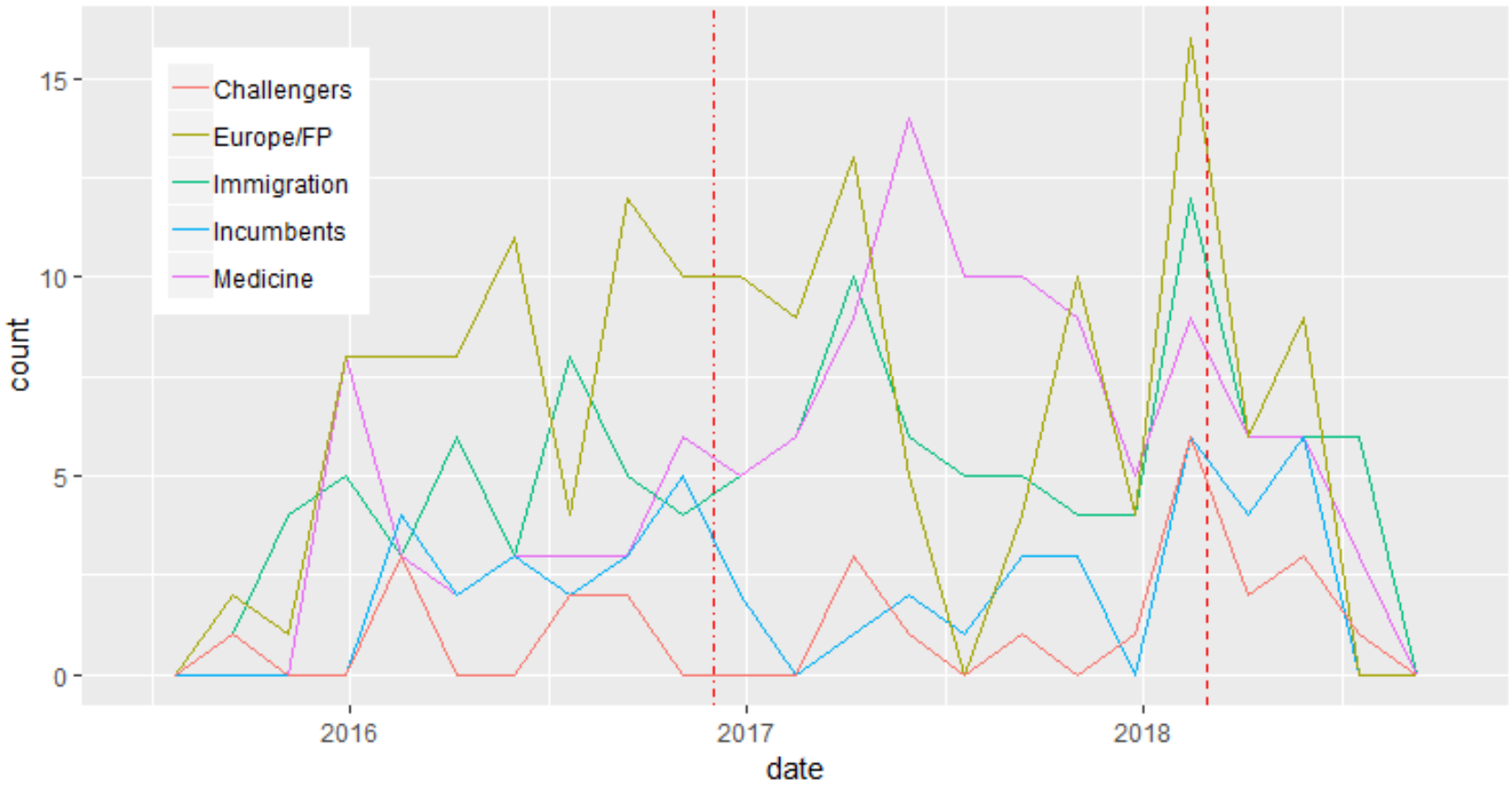




Source: own elaboration on data from the Italian Ministry of the Interior. Note: Shares are computed as the ratio of number of votes over total votes (including abstentions, blank and null votes) in the province and in German-majority villages respectively. The only German-speaking party different from the SVP (red section) is Die Freiheitlichen (DF), which did not run in the 2018 election.

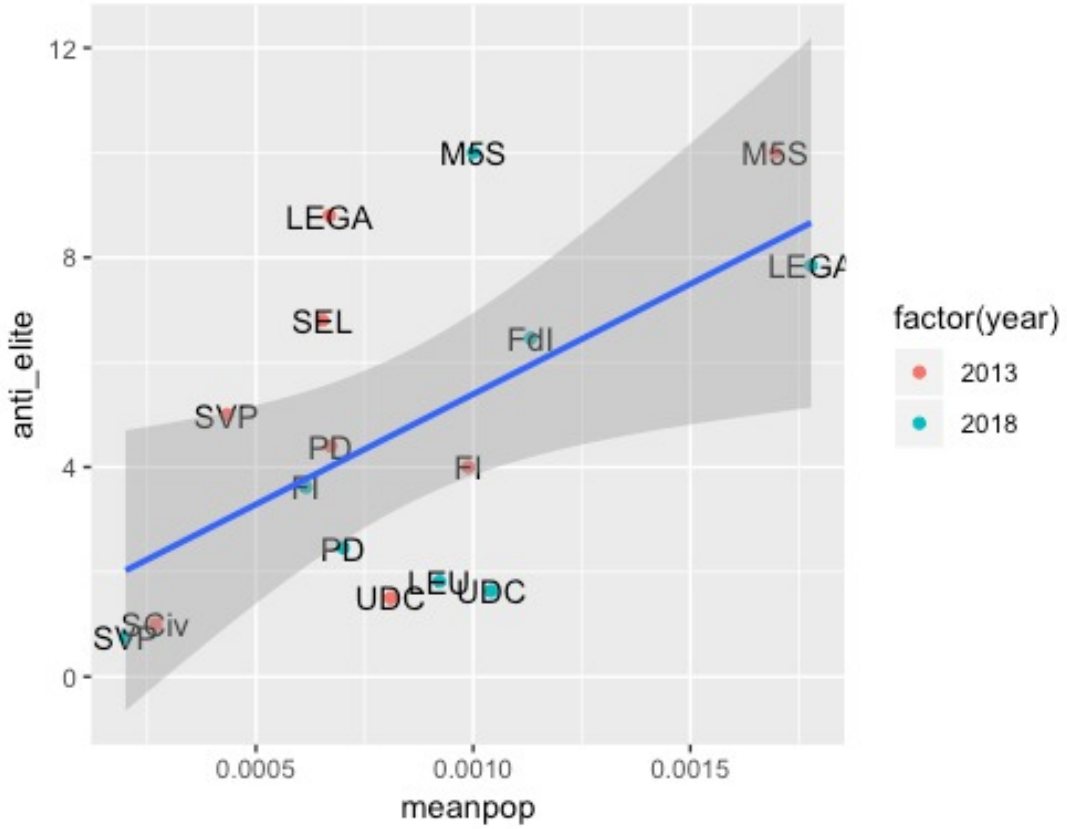
Back

Fake News Topics



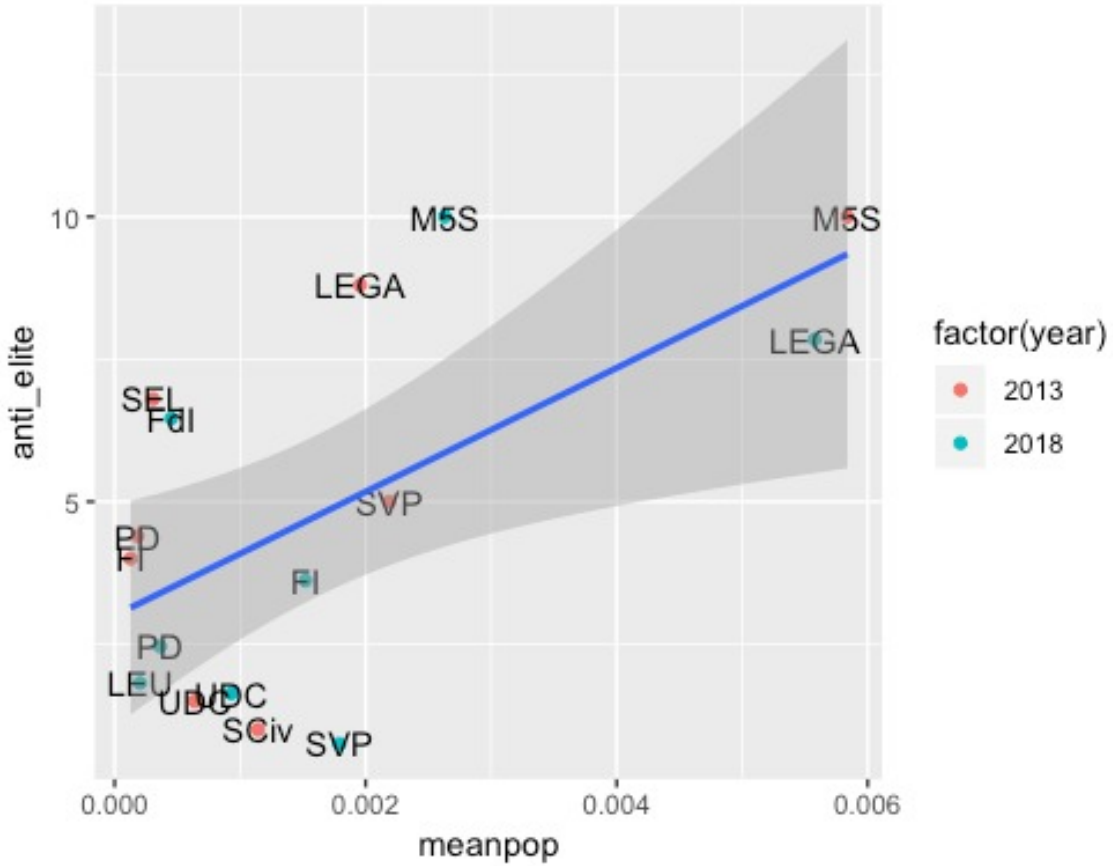
Bimonthly frequency of pieces of fake news by topic. The dot-dashed and dashed red lines indicate the date of the 2016 Constitutional Referendum and the 2018 General Elections respectively.

Correlation between populist Anti-establishment bag and CHES data



Text analysis scores of social media posts from parties and their leaders during the 2013 and 2018 elections campaigns (x-axis) and their relationship with the scores on the variable 'People vs the Elites' from the Chapel Hill Expert Survey for year 2014 (y-axis). Higher values on the y-axis correspond to higher salience of anti-establishment and anti-elite rhetoric on a scale from 0 to 10. Scores are computed using the 'Anti-establishment/ aggressive' text bag.

Correlation between populist 'Assertive' bag and CHES data



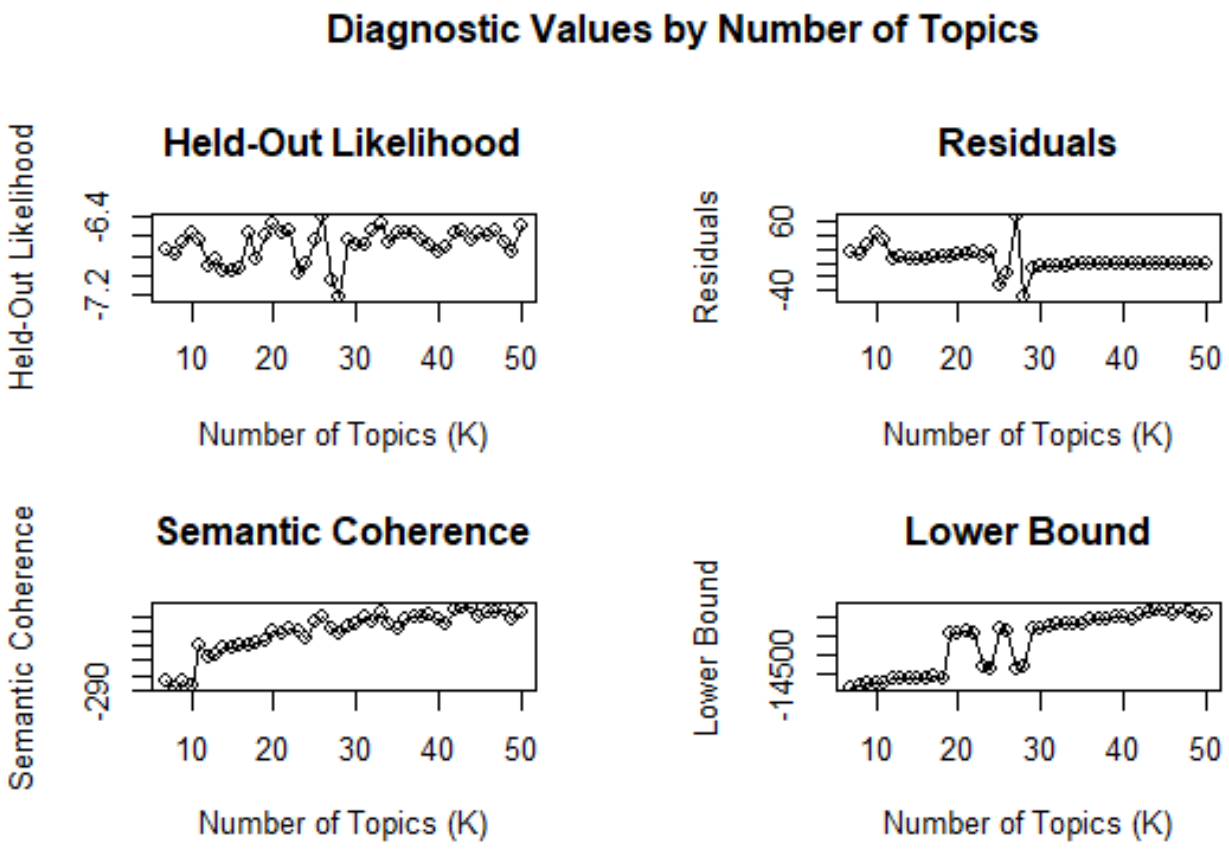
Text analysis scores of social media posts from parties and their leaders during the 2013 and 2018 elections campaigns (x-axis) and their relationship with the scores on the variable 'People vs the Elites' from the Chapel Hill Expert Survey for year 2014 (y-axis). Higher values on the y-axis correspond to higher salience of anti-establishment and anti-elite rhetoric on a scale from 0 to 10. Scores are computed using the 'Assertive' text bag.

Word in Fake News by Frequency

Structural Topic Model: Anti-Establishment

Structural Topic Model: Migration

Structural Topic Model Diagnostic: K=26



Facebook Audience Insights Tool

Back