Italy 2008
Polls, Web Visibility and Election Results.

Alessandro Bonazzi [*]   Paolo Brunori [†]   Riccardo Govoni [‡]
Giulio Isacco Lampronti [§]   Matteo Zandi [¶]

May, 2008

Abstract

The paper analyses the relationship between on-line media visibility, opinion polls, and election results in the April 14th 2008 Italian general elections. The average web visibility proved to be a good proxy of the preferences obtained for all coalitions. For the two main coalitions, the error can be considered negligible and performs, as a predictor, better than other traditional methods such as opinion polls. Our data show a second interesting result: party percentages recorded in polls for the two main coalitions, in the two months before the elections, tend to converge to our measure of web visibility. We are not proposing web visibility as means to forecast election results; however, the strong correlation observed seems to suggest that Italian web news correctly represents the political orientations of Italian society.

1   Introduction

Literature on relationships between mass media and voting behaviour suggests at least two points of view for interpreting a strong correlation between

---
[*]Universitá di Bologna
[†]Universitá di Bari
[‡]Associazione BayesFor
[§]Universitá di Firenze
[¶]Associazione BayesFor

media visibility and election results. The agenda setting theory claims that the media is able to a effect vote choices (McCombs, 2002). This effect has been recently measured by Della Vigna and Kaplan that showed how Fox News Channel expansion in some small towns in the US, between 1996 and 2000, can explain between 0.4 and 0.7% of the Republicans vote share gains in 2000 Presidential elections (Della Vigna and Kaplan, 2006). On the other hand, another part of the literature inverts the causality relationship and argues that the media tends to adapt itself to the opinions of media users; according to this perspective the media reinforces existing opinions rather than change them. In the model proposed by Gentzkow and Shapiro, for example, communication media compete to sell information services to consumers. Consumers have heterogeneous orientations and they want to receive information to confirm their opinions. Consequently the information the media produces tends to go along with this consumers desire. A third part of the literature tends to underline the complementarity of the two approaches, assuming that both of these effects are present in the relationship between visibility and election results, in a two-way process that has been effectively named reinforcing spirals (Slater, 2007). In the last years the growing relevance of the web as a source of information has attracted researchers attention. Nowadays the debate on media influence in shaping public opinion necessarily includes the web among information sources. However, very little empirical evidence has been produced. Some authors have focused on the blogsphere, which represents the main peculiarity of web communication (Gill, 2004); in a slightly different perspective we have considered on-line information sources as a good proxy for mass media as a whole, and we have exploited the relative computational simplicity to measure candidates visibility in this type of source. In a recent paper we have shown the strong correlation between electoral results and visibility in Italian web resources. We measured such cor- relation during the PD's Italian primary elections in October 2007 (Brunori et al., 2008) getting results consistent with other empirical findings obtained by Grippa and Del Vecchio (2008). In what follows, we present evidence from the March-April 2008 Italian electoral campaign. We show the existence of a strong relationship between: candidates and coalitions on-line visibility, opinion polls, and election results. In particular we show how the opinion polls in the two months before the election tend to converge to the average web visibility recorded in our database in the same period. Moreover average web visibility is found to represent vote shares with higher accuracy than opinion polls. The remainder of the article is organized

2

as follow: section 2 briefly presents the 2008 Italian electoral competition and election results. Section 3 explains the data we have used. Finally, section 4 presents the two main empirical findings about polls, votes, and internet visibility.

2   2008 Italian general elections

The Prodi government collapsed at the end of January 2008. Some days later, President Napolitano appointed Franco Marini, attempting to find agreement among parties to support an institutional government. On February 6th, President Napolitano dissolved the parliament and elections were called for April 13th and 14th. In Italian general elections citizens are asked to choose a party, parties are linked in coalitions, and each coalition supports a candidate for prime minister. The electoral law is based on proportional representation with a majority correction such that the coalition getting the relative majority gets 55% of the seats [1]. Given the electoral system there were only two possible winners: Berlusconi centre-right coalition and Veltroni centre-left coalition. There were 26 coalitions running for the Chamber of Deputies (Camera dei deputati) and 24 coalitions running for the Senate (Senato), however only 6 got more than 0.6% of the preferences and therefore the analysis considers only the 6 main coalitions. Table 1 reports the coalitions considered with the leaders name, the supporting parties, and the preferences obtained.

3   Data

The analysis considers three phenomena: internet visibility, polls percentage, and election results. Each variable is measured using data coming from a different source. Internet visibility comes from the Bayes-swarm database, polls predictions come from a number of research institutes, and election results are the average preferences obtained by coalitions in the two parliamentary houses .

---

[1]In the two weeks before the elections it is forbidden to publish political opinion polls

Table 1: Coalitions leader, parties, and votes

Coalitions leader	Parties	Camera	Senato	Average
Silvio Berlusconi	PdL, Lega Nord, MPA	46.81%	47.32%	47.07%
Walter Veltroni	PD, IdV	37.55%	38.01%	37.78%
Pier Ferdinando Casini	UdC	5.62%	5.56%	5.59%
Fausto Bertinotti	Sinistra Arcobaleno	3.04%	3.21%	3.13%
Daniela G. Santanché	La Destra-FT	2.43%	2.10%	2.27%
Enrico Boselli	PS	0.97%	0.87%	0.92%

Source:*Ministero dell'Interno*

3.1 Web visibility

Bayes-Swarm is a research project that aims to design and build an engine to extract information from internet sources (news portals, newspapers, news agencies and TV websites[2]). Thus far, it visits 98 sources once a day, which means around 200 web pages, mainly homepages and economical and political pages (the full list is reported in appendix A). Every page passes through a working process whose main steps are: (i) formatting tags and punctuation removal, (ii) conjunctions and articles removal, (iii) word roots extraction. Subsequently, the number of appearances of every word (word occurrences) is saved and stored in a database. Any word that happens to occur more than five times on a single day automatically enters the database. Thus, the database yields the number of appearances of a growing set of words on the sources we consider. From these, visibility time series trend graphs and correlations can be computed. Trends can then be linked to specific events. Bayes-Swarm saves and stores all the entire monitored web pages that are freely available on-line in its database at www.bayes-swarm.com. To analyze electoral campaign data we used 8 weeks of occurrences, from February 10th to April 12th. The time span is the official electoral campaign. Given the structure of the electoral competition, we have chosen to proxy coalition web visibility using both the leaders and parties names. The Silvio Berlus-

---

[2]A blog is also included among news sources given its well documented relevance in the Italian political scenario, that is www.beppegrillo.it

coni coalition web visibility is therefore obtained by adding the occurrences of *Berlusconi, Lombardo, Bossi, PdL, Lega* and *MPA*, the same method is applied for the other coalitions.

3.2   Opinion polls

We have found 73 polls published after February 10th and before March 31st[3]. We have selected only polls in which all 6 coalitions included in the analysis were reported, ending out with a time series of 62 polls published in 47 days. The full list of polls considered is reported in appendix B. Polls in the time series were often published in the same days. In order to have a time series a single measure of preferences obtained in each time period we have averaged polls percentage for each coalition in each day.

4   Empirical findings

4.1   Polls convergence to average web visibility

The first empirical finding concerns percentages recorded in opinion polls for the two main coalitions. The share of votes in the polls have a similar time pattern:

- they both show some degree of convergence and the variance tends to reduce with time. For instance, the polls standard deviation in the last two weeks is about half of the polls standard deviation in the first 6 weeks of the campaign (from 0.35 to 0.18 for Berlusconi and from 0.63 to 0.38 for Veltroni);

- the two time series show a statistically significant trend; running a linear regression of polls share on time we have found in both cases a significant coefficient for time to be added to a large fixed effect;

we can therefore define a convergence share for each coalition, that is what we expect the polls share to be in the days in which polls were not published, and in particular the day before the elections. We obtained two points of convergence for Berlusconi and Veltroni coalitions which are: 44.33% for

_____

[3]A blog is also included among news sources given its well documented relevance in the Italian political scenario, that is www.beppegrillo.it

Berlsuconi and 38.48% for Veltroni. These percentages are very close to average web visibility in the same period, the Berlusconi colalition had 45.23% of the web visibility (+0.9%) and Veltroni 39.54% (+1.06%). Figure 1 shows the two time series for Berlusconi and Veltroni coalitions polls share, the fitted values obtained with the linear regression models, and the average web visibility (slightly above the predicted share for both coalitions).
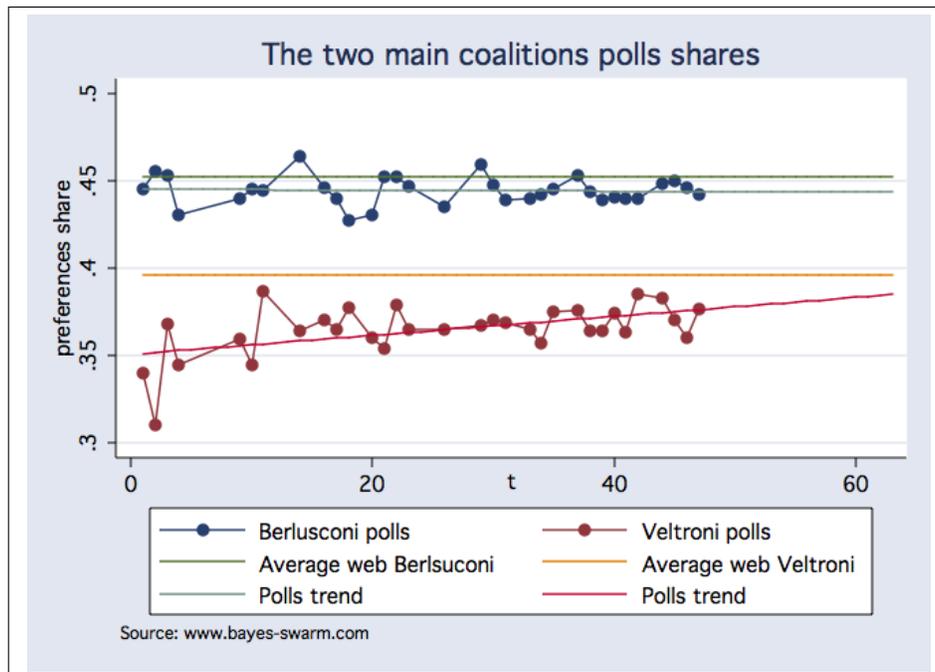

Figure 1: Berlusconi and Veltroni polls shares

4.2 Web visibility and elections result

The second empirical content of the analysis consists in the correlation between web visibility and election results. Table 2 reports average web visibilities, average poll results, and votes. As il possible to show average web visibility predicts vote share with a notable accuracy, only in one case the error is above 2%, in the Casini coalition, as noted, Casini had a lot of visibility at the beginning of the electoral campaign when he decided to run on its own instead of supporting Silvio Berlusconi. As shown in table 2, for the remaining 5 coalitions the error using web visibility to proxy electoral preferences

is rather law. Comparing polls and web visibility as votes predictors we notice that the average error is the same, however, the standard error measure, which weights larger errors more, is lower for web visibility. Moreover, web visibility is able to shed light on the most surprising election resultnamely, the sharp reduction in support for the extreme left-green coalition, that was believed to get more than twice the number of actual votes. This result is predicted with a negligible 0.7% error by the web visibility proxy.

Table 2: Coalitions leader, parties, and votes

Coalitions	Votes	Web visibility	Polls	Web - votes	Polls - votes
Berlusconi	47.07%	45.23%	44.43%	-1.8%	-2.6%
Veltroni	37.78%	39.54%	36.47%	1.8%	-1.3%
Casini	5.59%	9.92%	6.58%	4.3%	1.0%
Bertinotti	3.13%	2.45%	7.3%	-0.7%	4.2%
Boselli	0.92%	1.35%	1.36%	0.4%	0.4%
Santanché	2.27%	1.45%	2.3%	-0.8%	0.0%
Mean error				1.6%	1.6%
Standard error				2.2%	2.3%

Source:*Ministero dell'Interno and www.bayes-swarm.com*

5   Conclusion

In this brief paper we have presented two empirical findings about Italian Politicians web visibility, opinion polls, and election results. We have first shown how opinion polls tend to converge to values not far from average web visibility in the two months before elections. Secondly we have proven that average internet visibility and votes are so close that, *ex post*, internet visibility can be considered a better vote predictor then opinion polls. We believe that if on the one hand it is rather difficult to prove any causal relationship between the three phenomena, on the other hand the strong correlations we have measured suggest that a relationship exists and may deserve some attention.

References

- Brunori P., Zandi M., Bonazzi A., Govoni R. and Lampronti G. (2008), *Visibilit mediatica dei candidati alle primarie del Partito Democratico. Analisi con i dati di Bayes- Swarm*, Il Politico, N. 217 (1).

- Della Vigna S. e Kaplan E. (1991) *The Fox News Effect: Media Bias and Voting*, NBER Working Paper No. 12169.

- Gentzkow M. e Shapiro J. M. (2005) *Media Bias and Reputation* , NBER Working Paper No. 11664.

- Gill K. E. (2004) *How can we measure the influence of the blogosphere?*, Department of Communication, University of Washington at Seattle.

- Grippa F. and Del Vecchio P. (2008), *Take me to your Leader: Predicting Political Leadership using Social Network Metrics*, presented at Sunbelt - International Social Network Conference, Florida, USA.

- McCombs M. E. (2002) *The Agenda-Setting Role of the Mass Media in the Shaping of Public Opinion*, Mass Media Economics 2002 Conference, LSE, London, UK.

- Slater M. D. (2007) *Reinforcing Spirals: The Mutual Influence of Media Selectivity and Media Effects and Their Impact on Individual Behavior and Social Identity*, Communication Theory, N. 58.

Appendix A

Table 3: List of the websites monitored

1	espresso.repubblica.it	17	www.ilgiornale.it
2	ilrestodelcarlino.quotidiano.net	18	www.ilmanifesto.it
3	lescienze.espresso.repubblica.it	19	www.ilmessaggero.it
4	limes.espresso.repubblica.it	20	www.ilsecoloxix.it
5	micromega.repubblica.it	21	www.ilsole24ore.com
6	news.google.it	22	www.italiaoggi.it
7	qn.quotidiano.net	23	www.lavoce.info
8	www.adnkronos.com	24	www.liberazione.it
9	www.agenziaitalia.it	25	www.libero-news.it
10	www.agopress.info	26	www.megachip.info
11	www.ansa.it	27	www.nationalgeographic.it
12	www.avvenire.it	28	www.panorama.it
13	www.beppegrillo.it	29	www.repubblica.it
14	www.corriere.it	30	www.televideo.rai.it
15	www.euronews.net	31	www.unita.it
16	www.ilfoglio.it		

Appendix B

Table 4: List of the pools

Date	Company	Date	Company	Date	Company
27-mar	Crespi	14-mar	Ipr	29-feb	Ipr
27-mar	Swg	14-mar	Crespi	27-feb	Swg
27-mar	Digis	13-mar	Quaeris	26-feb	Demoskopea
27-mar	Gipieffe	14-mar	Gipieffe	25-feb	Crespi
26-mar	Euromedia	13-mar	Demopolis	25-feb	Ipsos
25-mar	Crespi	11-mar	Swg	23-feb	Euromedia
25-mar	Demoskopea	11-mar	Demoskopea	20-feb	Digis
24-mar	Ipsos	11-mar	Dinamiche	20-feb	Demos-Eurisko
22-mar	Ispo	11-mar	Ispo	19-feb	Ipr
21-mar	Quaeris	10-mar	Demos	18-feb	Swg
20-mar	Dinamiche	10-mar	Crespi	18-feb	Demoskopea
20-mar	Swg	10-mar	Ipsos	18-feb	Ipsos
20-mar	Gipieffe	9-mar	Digis	18-feb	Crespi
20-mar	Digis	6-mar	Ipr	13-feb	Swg
19-mar	Crespi	3-mar	Ipsos	12-feb	Crespi
18-mar	Demopolis	3-mar	Swg	12-feb	Fn and G
18-mar	Agron	3-mar	Demoskopea	12-feb	Ispo
17-mar	Ipsos	3-mar	Crespi	11-feb	Piepoli
17-mar	Demoskopea	2-mar	Digis	10-feb	Ipr
15-mar	Ispo	1-mar	Euromedia		