How to Poll Runoff Elections

Online Supplementary Materials

Forthcoming in *Public Opinion Quarterly*

Peter Selb (corresponding author)
University of Konstanz
Department of Politics and Public Administration
peter.selb@uni.kn

Sascha Göbel
University of Konstanz
Graduate School of Decision Sciences
sascha.goebel@uni-konstanz.de

Romain Lachat
Sciences Po Paris
CEVIPOF
romain.lachat@sciencespo.fr

Contents

Appendix A The 2017 French presidential election

In this section, we describe the legal framework and campaign context. French presidential elections take place every five years. They are based on a two-round system where voters cast a single vote for a candidate on election day in one of approximately 65,000 polling stations (*bureaux de vote*). Eligible voters are affiliated to the polling station closest to their address, as indicated on their voter ID. Early voting is not permitted, which enhances the coverage of exit polls (Mokrzycki, Keeter and Kennedy, 2009). If no candidate receives an absolute majority of the votes in round 1, the two candidates who received the most votes face each other in a runoff.

The first round of the voting took place on April 23, 2017. The election was unusual in several respects. It was widely expected that Marine Le Pen, leader of the anti-immigrant and anti-European Front National, would qualify for the runoff. Virtually every poll conducted since the previous election had placed her as one of the top two contenders. The election was thus largely perceived as a litmus test for the rise of populism in Western democracies. Another novel feature of this election was the rejection of the established parties. French politics had long been dominated by two mainstream parties: the Socialists on the left and the moderate right-wing Republicans. Since the popularity of then-president François Hollande had reached low levels, many expected the election would be won by the candidate of the Republicans. However, the election campaign took several unexpected turns

and led to a very competitive race between five main contenders. In addition to Le Pen, these were the Socialist Benoît Hamon, the Republican François Fillon, and two outsiders, Jean-Luc Mélenchon of the far-left movement La France Insoumise (Unsubmissive France) and the centrist Emmanuel Macron of En Marche! (On the Move!). Both the Socialist and Republican parties chose their candidate in an open primary election, organized jointly with other smaller parties. The Republican primary election took place in November 2016 and resulted in the surprise victory of François Fillon, a former Prime Minister with strong social conservative and economic liberal views. Soon after, Fillon's candidacy was marred by accusations of misuse of public funds. This led to Fillon being placed under formal investigation later in the campaign. In January 2017 Benoît Hamon came off as the winner of the Socialist primary. Himself a proponent of the left wing of his party, Hamon faced strong competition from Jean-Luc Mélenchon, who, benefiting from the unpopularity of François Hollande and the more general mood of defiance against established parties, campaigned on a left-wing populist platform and managed to garner substantial support. In the centre of the political spectrum, Emmanuel Macron rose from being an outsider to be a key candidate, attracting both centre-left citizens who saw the platforms of both Hamon and Mélenchon as too far-left, and right-wing voters disillusioned by the scandals surrounding the Fillon candidacy. Macron's standing in the polls increased steadily during the campaign. In the end, he won the plurality of the votes in the first round of the election with 24.0% of the

vote, followed by Marine Le Pen (21.3%), François Fillon (20.0%), Jean-Luc Mélenchon (19.6%), and Benoît Hamon trailing behind (6.4%). In the runoff on May 7, 2017, Macron defeated Le Pen by a landslide, with 66.1% against 33.9% of the valid votes. Turnout was at 77.8% in round 1 and at 74.6% in round 2, yet the proportion of blank and invalid ballots increased from 2.6% in round 1 to 11.5% in round 2.

Appendix B Analytic potential

The data produced offer analytic potential beyond its original purpose of election forecasting. On the one hand, the candidate ratings we collected for predicting the runoff can also be used to inquire whether the winner of the runoff is the 'Condorcet winner', that is, would have won the head-to-head contests against all the other candidates, too. The question is relevant from a decision-theoretic point of view (Gehrlein, 1983), and it has an obvious bearing on the fairness and legitimacy of an election result. For instance, Abramson (2007) found that president-elect Nicolas Sarkozy was *not* the Condorcet winner of the 2007 presidential race. On the other hand, unfolding models can be used to uncover the latent ideological space underlying the respondents' candidate ratings. Once again, such analyses are important for the understanding of political competition and voting behavior.

Hypothetical runoffs to determine the Condorcet winner

Hypothetical runoffs are easily constructed from our data using the above redistribution logic. Table B1 presents estimates for the actual runoff and the five most likely hypothetical runoffs involving the four strongest candidates from round 1: Emmanuel Macron, Marine Le Pen, François Fillon and Jean-Luc Mélenchon. Macron leads all the runoff scenarios and is therefore the likely Condorcet winner of the 2017 election. According to our data a head-to-head contest between Macron and Mélenchon would have been sub-

stantively closer than the actual race against Le Pen, though. Both Fillon and Mélenchon would have beaten Le Pen, and Mélenchon would have won out over Fillon.

Table B1. Mock runoff predictions for the four strongest candidates from round 1.

	Mélenchon	Fillon	Le Pen
Macron	55.0 (1.3)	67.7 (1.4)	66.0 (1.6)
Mélenchon	—	59.4 (1.7)	61.7 (1.8)
Fillon	—	—	55.8 (1.5)

Note: Estimated vote shares pertain to row candidates. The prediction of the actual runoff is printed in grey. Standard errors in parentheses incorporate uncertainty at both sampling stages as well as imputation uncertainty.

Spatial analysis of candidate ratings

The redistribution of first-round votes among runoff candidates proceeded from the Downsian notion of utility differentials, yet we have been silent about where candidate utilities come from. Spatial models of political competition consider candidate utilities as loss functions of the distance between voters and candidates in a latent policy space (Hinich and Munger, 1997). A very general form of such models may be written as

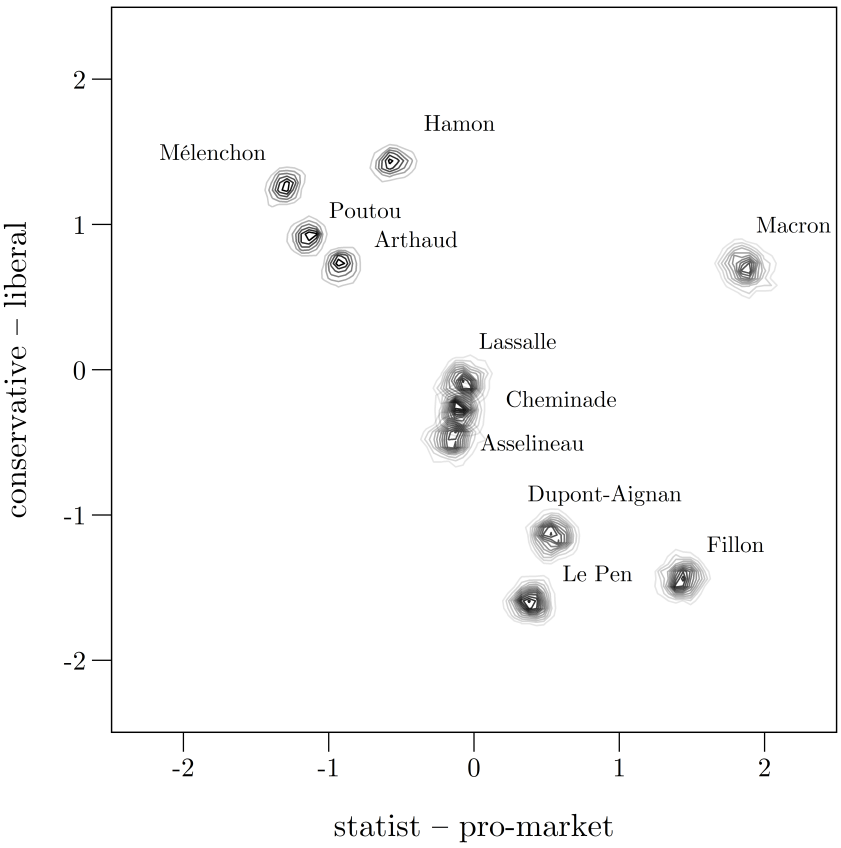
$$\mu_{ij} = \alpha_j - v_i f(\beta_{jd} - \theta_{id}),$$

where μ_{ij} is voter i's expected utility from candidate j, θ_i is i's position on

ideological dimension d, β_j is j's position, f is some loss function, v_i is the weight i attributes to the ideological component, and α_j is a non-spatial utility component specific to j. Shikano and Käppner (2017) propose a Bayesian method to estimate the parameters of this generic model based on candidate ratings x_{ij}, which are given a normal distribution, $x_{ij} \sim N(\mu_{ij}, \sigma^2)$. To demonstrate the analytic potential of our data for examining the structure of political competition from the voters' perspective, we exemplarily fit a two-dimensional model with a quadratic loss function to the ratings of the eleven candidates competing in round 1. Figure B1 gives posterior estimates of the candidates' positions, β_j. The resulting pattern is consistent with expert accounts of the 2017 presidential election. The horizontal axis can be interpreted in terms of the classical economic divide, with Hamon and Mélenchon to the left and the pro-market candidates to the right. The vertical dimension runs from culturally liberal positions on top to conservative positions at the bottom. The most striking feature of the spatial representation is the location of Emmanuel Macron. While all other candidates are aligned on a single dimension running from an economic left and culturally liberal pole to an economic right and culturally conservative pole, Macron stands out as the only candidate combining culturally and economically liberal positions. French voters, it seems, predominantly agreed on the need for economic reform while they rejected the culturally conservative positions of candidates like Fillon and Le Pen. Note that the expected utility model may also be used for the imputation of missing candidate ratings instead of the chained

equations approach we employed.

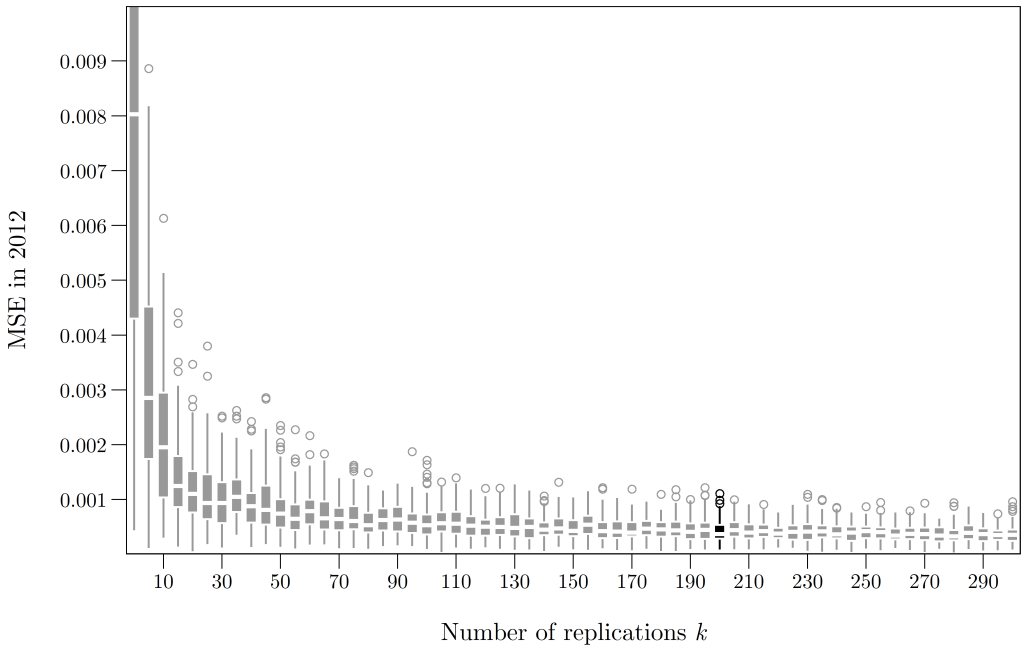
Figure B1. Contour plot of the estimated candidate positions.



Note: Posterior estimates of the candidates' positions. 2,000 iterations after 1,000 burn-ins in three MCMC chains. Convergence tests and information criteria suggest that a two-dimensional model provides a better fit to the data than a uni-dimensional model.

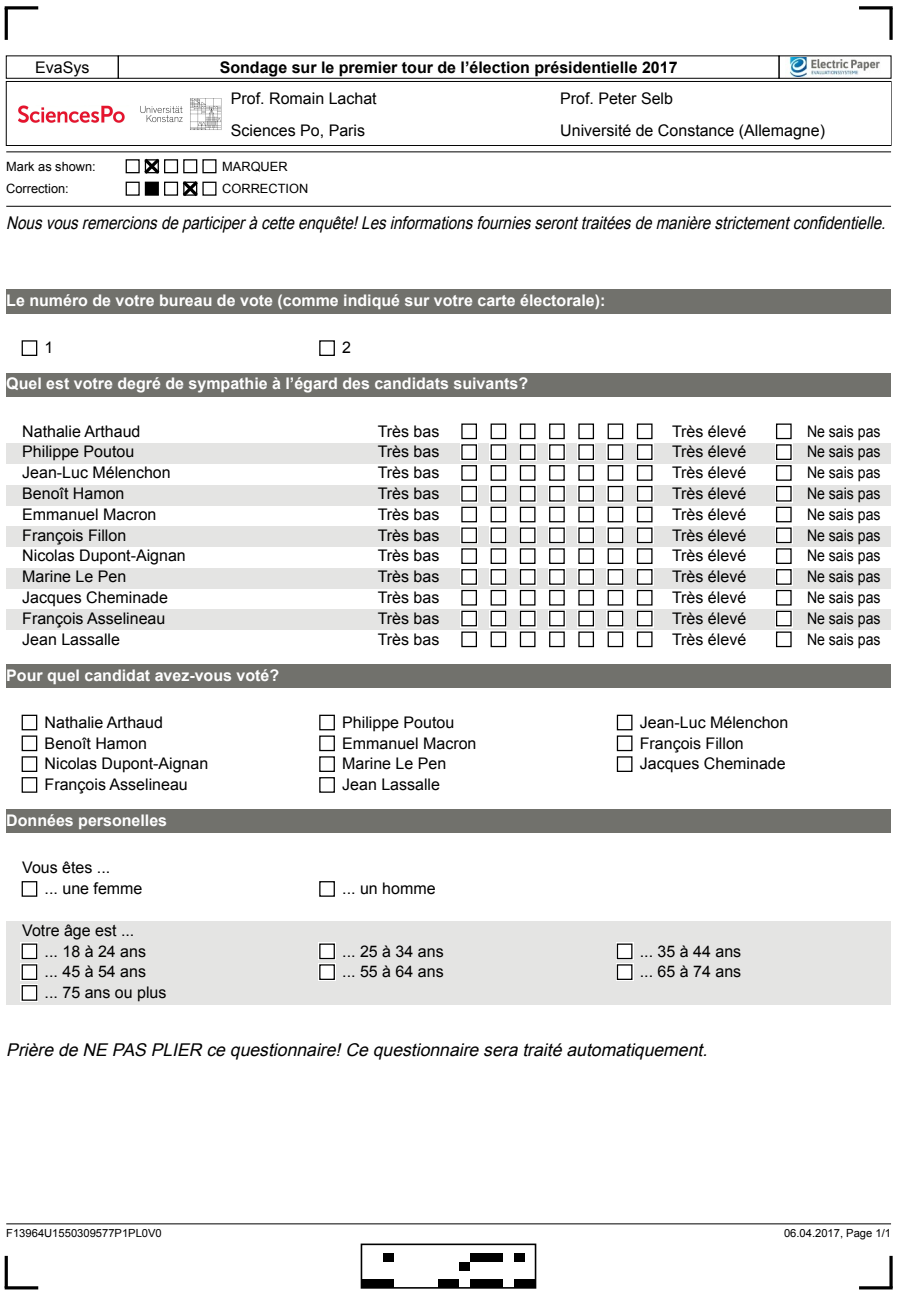
Appendix C Additional figures and tables

Figure C1. Choice of number of replications k.



Note: Each whisker represents the quartiles of the distribution of 2012 averaged squared error values calculated for 100 samples of size $m = 20$, each drawn from k PPS samples of polling stations. The number of replications k is varied from 1 to 300. The black whisker indicates the value of $k = 200$ that we used.

Figure C2. Example questionnaire, bureau de vote 1, Eygalière, Département Bouches-du-Rhône.



EvaSys | **Sondage sur le premier tour de l'élection présidentielle 2017** | Electric Paper

SciencesPo — Universität Konstanz

Prof. Romain Lachat, Sciences Po, Paris

Prof. Peter Selb, Université de Constance (Allemagne)

Mark as shown: ☐ ☒ ☐ ☐ ☐ MARQUER

Correction: ☐ ■ ☐ ☒ ☐ CORRECTION

Nous vous remercions de participer à cette enquête! Les informations fournies seront traitées de manière strictement confidentielle.

Le numéro de votre bureau de vote (comme indiqué sur votre carte électorale):

☐ 1 ☐ 2

Quel est votre degré de sympathie à l'égard des candidats suivants?

Nathalie Arthaud	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Philippe Poutou	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Jean-Luc Mélenchon	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Benoît Hamon	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Emmanuel Macron	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
François Fillon	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Nicolas Dupont-Aignan	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Marine Le Pen	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Jacques Cheminade	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
François Asselineau	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas
Jean Lassalle	Très bas ☐ ☐ ☐ ☐ ☐ ☐ ☐	Très élevé	☐ Ne sais pas

Pour quel candidat avez-vous voté?

☐ Nathalie Arthaud	☐ Philippe Poutou	☐ Jean-Luc Mélenchon
☐ Benoît Hamon	☐ Emmanuel Macron	☐ François Fillon
☐ Nicolas Dupont-Aignan	☐ Marine Le Pen	☐ Jacques Cheminade
☐ François Asselineau	☐ Jean Lassalle	

Données personelles

Vous êtes ...

☐ ... une femme ☐ ... un homme

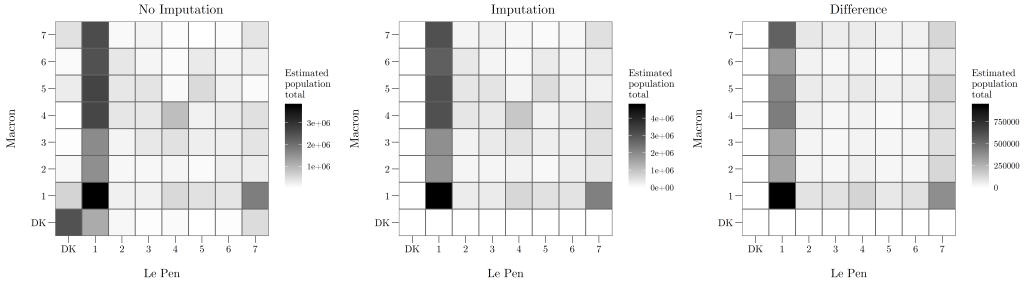
Votre âge est ...

☐ ... 18 à 24 ans	☐ ... 25 à 34 ans	☐ ... 35 à 44 ans
☐ ... 45 à 54 ans	☐ ... 55 à 64 ans	☐ ... 65 à 74 ans
☐ ... 75 ans ou plus		

Prière de NE PAS PLIER ce questionnaire! Ce questionnaire sera traité automatiquement.

F13964U1550309577P1PL0V0 06.04.2017, Page 1/1

Figure C3. Bivariate distribution of Macron and Le Pen ratings before and after imputation.



Note: Rating scales range from 1 (very unfavorable) to 7 (very favorable). Category DK includes both 'Don't knows' and 'No answers'. Cells contain estimated voter totals. No imputation results (left panel), imputation results (center panel), difference between no imputation and imputation results (right panel).

Table C1. Standard PPS versus balanced samples: detailed simulation results.

		All PPS samples				Balanced samples				
Candidate	Population parameter	Expected value	MSE	Bias	Variance	Expected value	MSE	Bias	Variance	Variance ratio
Macron	24.01	23.75	2.12	−0.26	2.05	24.20	0.87	0.19	0.84	0.41
Le Pen	21.30	21.69	5.13	0.39	4.97	21.07	0.99	−0.23	0.94	0.19
Fillon	20.01	19.81	3.92	−0.20	3.88	19.98	1.01	−0.03	1.01	0.26
Mélenchon	19.58	19.59	2.58	0.01	2.58	19.58	0.83	0.00	0.83	0.32
Hamon	6.36	6.28	0.38	−0.07	0.38	6.43	0.19	0.07	0.19	0.50
Dupont-Aignan	4.70	4.84	0.24	0.14	0.23	4.76	0.14	0.06	0.13	0.58
Lassalle	1.21	1.23	0.10	0.02	0.10	1.19	0.08	−0.02	0.08	0.79
Poutou	1.09	1.09	0.02	0.00	0.02	1.08	0.02	−0.01	0.02	0.78
Asselineau	0.92	0.91	0.02	−0.01	0.02	0.91	0.02	−0.01	0.02	1.02
Arthaud	0.64	0.64	0.01	0.00	0.01	0.63	0.01	−0.01	0.01	0.79
Cheminade	0.18	0.18	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.96

Table C2. Sampling information

Department	Municipality	Polling station ID	Voters (2012)	Inclusion probabilities			Sampling interval	Sample size	Refusals (%)	Misses (%)
				initial	empirical	stage 2				
Bouches-du-Rhône	Eygalières	1	682	0.00039	0.00040	0.08798	5	73	29.6	2.8
Bouches-du-Rhône	Salon-de-Provence	13	1,023	0.00058	0.00048	0.05865	8	87	33.8	3.6
Côtes-d'Armor	Trégastel	2	743	0.00042	0.00028	0.08075	5	63	38.7	8.4
Haute-Garonne	Avignonet-Lauragais	2	254	0.00014	0.00016	0.23622	1	100	26.6	1.4
Haute-Garonne	Castelginest	1	909	0.00052	0.00068	0.06601	3	76	24.1	23.4
Haute-Garonne	Pinsaguel	2	880	0.00050	0.00044	0.06818	6	86	19.8	28.7
Gironde	Pessac	39	857	0.00049	0.00040	0.07001	15	74	23.1	13.7
Indre	Valençay	1	972	0.00055	0.00080	0.06173	6	47	47.2	9.3
Maine-et-Loire	Charcé-Saint-Ellier-sur-Aubance	1	494	0.00028	0.00016	0.12146	2	100	24.5	3.6
Meurthe-et-Moselle	Ludres	3	806	0.00046	0.00048	0.07444	3	73	43.4	2.2
Moselle	Terville	6	524	0.00030	0.00032	0.11450	8	98	22.6	5.8
Oise	Bailleval	1	937	0.00053	0.00044	0.06403	3	100	27,3	0.7
Orne	Tinchebray	2	982	0.00056	0.00060	0.06110	6	53	32.5	1.3
Métropole de Lyon	Lyon	512	869	0.00050	0.00048	0.06904	6	82	27.2	5.0
Haute-Savoie	Cluses	2	549	0.00031	0.00040	0.10929	10	69	48.1	0.7
Seine-Maritime	Aumale	1	1,355	0.00077	0.00068	0.04428	5	43	50.4	15.7
Var	Toulon	36	403	0.00023	0.00020	0.14888	6	81	39.0	8.4
Hauts-de-Seine	Rueil-Malmaison	30	747	0.00043	0.00056	0.08032	15	57	52.6	5.2
Seine-Saint-Denis	Montreuil	3	595	0.00034	0.00032	0.10084	4	81	28.5	5.7
Seine-Saint-Denis	Noisy-le-Grand	29	539	0.00031	0.00028	0.11132	2	97	36.4	7.5

Table C3. Predicted and official first- and second-round results using empirical inclusion probabilities.

	Round 1				Round 2	
Step	Macron	Le Pen	Fillon	Mélenchon	Macron	Le Pen
Sample	24.9	19.8	18.5	21.7	69.0	31.0
	(1.4)	(2.2)	(2.0)	(1.9)	(3.0)	(3.0)
Survey	27.3	13.7	15.6	25.9	73.4	26.6
	(1.5)	(1.7)	(1.8)	(2.8)	(3.3)	(3.3)
Poststratification	–	–	–	–	64.2	35.8
	–	–	–	–	(1.4)	(1.4)
Imputation	–	–	–	–	66.2	33.8
	–	–	–	–	(1.8)	(1.8)
Official	24.0	21.3	20.0	19.6	66.1	33.9

Note: Empirical inclusion probabilities are based on rejective sampling with 25,000 balanced samples. Sample: estimated vote shares based on official results from the sample of polling stations. Survey: reported vote shares (Round 1), shares of redistributed votes (Round 2) from the exit poll. Poststratification: shares of redistributed votes, re-weighted by national round 1 results. Imputation: shares of redistributed votes based on imputed candidate ratings, re-weighted by official round 1 results. Official: official election results. Jackknife standard errors in parentheses.

Table C4. Predicted redistribution of first-round votes using empirical inclusion probabilities.

Candidate	Macron	Le Pen	Abstention
Macron	8,483,166	173,180	0
	(107,398)	(107,398)	(0)
Le Pen	774,834	6,903,658	0
	(255,445)	(255,445)	(0)
Fillon	4,020,347	1,799,052	1,393,596
	(460,366)	(268,396)	(301,762)
Mélenchon	4,293,783	490,655	2,275,513
	(247,587)	(130,745)	(192,053)
Hamon	1,638,143	132,347	520,799
	(112,791)	(56,058)	(114,621)
Dupont-Aignan	582,028	561,732	551,238
	(91,160)	(144,065)	(135,055)
Lassalle	239,963	46,558	148,780
	(64,557)	(34,844)	(53,874)
Poutou	199,563	90,900	104,041
	(77,106)	(62,928)	(64,957)
Asselineau	52,737	198,248	81,562
	(38,094)	(55,316)	(41,199)
Arthaud	132,721	13,370	86,292
	(44,073)	(21,233)	(50,509)
Cheminade	12,400	14,726	38,461
	(14,443)	(16,978)	(20,859)
Total	20,429,685	10,424,426	5,200,282
	(731,521)	(492,787)	(447,802)
Official	20,743,128	10,638,475	4,672,790

Note: Empirical inclusion probabilities are based on rejective sampling with 25,000 balanced samples. Entries are estimated vote totals transferred from the first-round candidates in the rows to the runoff candidates in the columns. Our method does not discriminate between abstention, blank and invalid votes, they are all subsumed in the column Abstention. Jackknife standard errors in parentheses.

Table C5. Predicted and official first-round results for all candidates.

Step	Macron	Le Pen	Fillon	Mélenchon	Hamon	Dupont-Aignan	Lassalle	Poutou	Asselineau	Arthaud	Cheminade
Sample	25.0	19.8	18.3	21.5	6.9	4.4	1.5	1.1	0.9	0.4	0.2
	(1.3)	(2.0)	(1.9)	(1.8)	(0.7)	(0.4)	(0.5)	(0.1)	(0.2)	(0.1)	(0.1)
Survey	27.1	13.5	15.3	26.3	8.3	4.6	1.7	0.8	1.5	0.5	0.4
	(1.4)	(1.6)	(1.8)	(2.7)	(1.3)	(0.8)	(0.5)	(0.2)	(0.4)	(0.2)	(0.2)
Official	24.0	21.3	20.0	19.6	6.4	4.7	1.2	1.1	0.9	0.6	0.2

Note: Sample: Estimated candidate vote shares using official results from our sample of polling stations. Survey: Reported vote shares (Round 1) from our exit poll. Standard errors (jackknife method) in parentheses. Official: official election results.

Appendix D Software Statement

Table D1. R packages.

crayon (Csárdi and Gaslam, 2017)	openxlsx (Walker and Braglia, 2017)
data.table (Dowle et al., 2017)	pacman (Rinker et al., 2017)
doParallel (Calaway et al., 2018)	plyr (Wickham, 2011)
dplyr (Wickham et al., 2017)	pps (Gambino, 2012)
extrafont (Chang, 2014)	sampling (Tillé and Matei, 2016)
foreach (Calaway, Microsoft Corporation and Weston, 2017)	skalunfold (Shikano and Käppner, 2017)
ggplot2 (Wickham, 2009)	stringr (Wickham, 2017)
magrittr (Bache and Wickham, 2014)	survey (Lumley, 2004)
mice (van Buuren and Groothuis-Oudshoorn, 2011)	tidyr (Wickham, Henry and RStudio, 2019)

Note: All code was run under Windows 10 x86-64 using R version 3.4.2. (R Development Core Team, 2017).

References

Abramson, Paul R. 2007. "The French presidential election of 2007: was Sarkozy the Condorcet winner?" *French Politics* 5(3):287–291.

Bache, Stefan M. and Hadley Wickham. 2014. *magrittr: A Forward-Pipe Operator for R.* R package version 1.5.
URL: *https://CRAN.R-project.org/package=magrittr*

Calaway, Rich, Microsoft Corporation and Steve Weston. 2017. *foreach: Provides Foreach Looping Construct for R.* R package version 1.4.4.
URL: *https://CRAN.R-project.org/package=foreach*

Calaway, Rich, Microsoft Corporation, Steve Weston and Dan Tenenbaum. 2018. *doParallel: Foreach Parallel Adaptor for the 'parallel' Package.* R package version 1.0.14.
URL: *https://CRAN.R-project.org/package=doParallel*

Chang, Winston. 2014. *extrafont: Tools for using fonts.* R package version 0.17.
URL: *https://CRAN.R-project.org/package=extrafont*

Csárdi, Gábor and Brodie Gaslam. 2017. *crayon: Colored Terminal Output.* R package version 1.3.4.
URL: *https://CRAN.R-project.org/package=crayon*

Dowle, Matt, Arun Srinivasan, Jan Gorecki, Tom Short, Steve Lianoglou and

Eduard Antonyan. 2017. *data.table: Extension of 'data.frame'*. R package version 1.10.4.

URL: *https://CRAN.R-project.org/package=data.table*

Gambino, Jack G. 2012. *pps: Functions for PPS sampling*. R package version 0.94.

URL: *https://CRAN.R-project.org/package=pps*

Gehrlein, William V. 1983. "Condorcet's paradox." *Theory and Decision* 15(2):161–197.

Hinich, Melvin J and Michael C Munger. 1997. *Analytical politics*. Cambridge University Press.

Lumley, Thomas. 2004. "Analysis of complex survey samples." *Journal of Statistical Software* 9(8):1–19.

URL: *https://www.jstatsoft.org/article/view/v009i08*

Mokrzycki, Michael, Scott Keeter and Courtney Kennedy. 2009. "Cell-phone-only voters in the 2008 exit poll and implications for future noncoverage bias." *Public Opinion Quarterly* 73(5):845–865.

R Development Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

URL: *https://www.r-project.org/*

Rinker, Tyler, Dason Kurkiewicz, Keith Hughitt, Albert Wang and Jim Hester. 2017. *pacman: Package Management Tool.* R package version 0.4.6.
URL: *https://CRAN.R-project.org/package=pacman*

Shikano, Susumu and Konstantin Käppner. 2017. "Identification of ideological space with valence based on feeling thermometer scores.".

Tillé, Yves and Alina Matei. 2016. *sampling: Survey sampling.* R package version 2.8.
URL: *https://CRAN.R-project.org/package=sampling*

van Buuren, Stef and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate imputation by chained equations in R." *Journal of Statistical Software* 45(3).
URL: *https://www.jstatsoft.org/article/view/v045i03*

Walker, Alexander and Luca Braglia. 2017. *openxlsx: Read, write and edit XLSX files.* R package version 4.0.17.
URL: *https://CRAN.R-project.org/package=openxlsx*

Wickham, Hadley. 2009. *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York.
URL: *http://ggplot2.org/*

Wickham, Hadley. 2011. "The split-apply-combine strategy for data analysis." *Journal of Statistical Software* 40(1):1–29.
URL: *https://www.jstatsoft.org/article/view/v040i01*

Wickham, Hadley. 2017. *stringr: Simple, consistent wrappers for common string operations*. R package version 1.2.0.

URL: *https://CRAN.R-project.org/package=stringr*

Wickham, Hadley, Lionel Henry and RStudio. 2019. *tidyr. Tidy messy data*. R package version 1.0.0.

URL: *https://CRAN.R-project.org/package=tidyr*

Wickham, Hadley, Romain Francois, Lionel Henry and Kirill Müller. 2017. *dplyr: A grammar of data manipulation*. R package version 0.7.1.

URL: *https://CRAN.R-project.org/package=dplyr*