Citizen Forecasts of the 2021 German Election – Online Appendix

Andreas E. Murr, University of Warwick
Michael S. Lewis-Beck, University of Iowa

18 June 2021

Characteristics of the Politbarometer Samples

Before 1988, the survey was collected using face-to-face interviews. The population consisted of West Germans eligible to vote. These surveys, with the time unit of calendar month, were stratified with stages according to the ADM-Mastersample. After 1988 several changes occurred. First, they changed to telephone interviews, meaning the population to be sampled changed to West Germans eligible to vote who lived in a private household with a telephone connection. The respondents were selected using a two-step procedure based on a randomized last digit; they randomly sampled households and then they selected the person with the most recent birthday. Second, from 1990 on the population included East Germans. Typically, the survey was stratified by East (about 500 respondents) and West (about 1,500 respondents), but sometimes the sampling actually took place for the country as a whole. By far the most common stratification was East and West. Finally, they also changed the time unit of the survey. In 1994, 1998, and from 2002 onwards the time unit was the calendar week. Overall, then, they have been essentially probability samples drawn from a list of households. (For our analysis, we weighted the survey responses by the representative weight.)

Table A1 below shows the lead time in months for each *Politbarometer* survey that contains the expectation question, by election year. Further, the cells show the sample size.

Table A1. Number of respondents in *Politbarometer* surveys that collected citizen forecasts across elections and by lead time in months.

Lead time in months	1980	1983	1987	1990	1994	1998	2002	2005	2009	2013	2017
1	0	0	999	0	2081	3350	4769	3088	4579	4948	3443
2	1518	1042	960	938	914	1162	4469	5979	4356	4669	3229
3	0	952	944	971	1947	1169	2407	2999	2723	0	3087
4	1042	0	1023	939	0	1132	2411	2996	1539	0	1508
5	1473	1439	0	913	1945	1164	2455	1538	1594	0	1483
6	973	0	1012	0	1865	1210	2393	0	0	0	0
7	1051	0	1005	904	1933	1207	2410	0	1467	0	0
8	1009	0	977	929	1973	1136	2250	0	0	0	0
9	1051	0	948	909	1941	1223	1599	0	0	0	0
10	1102	0	933	906	1957	1147	1584	0	0	0	0
11	0	0	951	870	842	1138	1526	0	0	0	0
12	947	0	960	906	1906	1203	1515	0	0	0	0
13	0	0	982	864	0	1144	0	0	0	0	0
14	0	0	941	870	1885	1123	0	0	0	0	0
15	0	0	979	0	0	0	0	0	0	0	0
16	0	0	978	0	0	0	0	0	0	0	0
17	0	0	925	0	1873	0	0	0	0	0	0
18	0	0	0	0	1899	0	0	0	0	0	0
19	0	0	0	0	877	0	0	0	0	0	0

Lead time

Table A2 shows the jackknifed prediction errors by lead times of one, two, three and five months. The benchmark is a lead time of 2 months for which all elections are available. The errors are shown in the second column. For the remaining lead times one election (3 months) or two elections (1 and 5 months) are unavailable. To compare these lead times with the benchmark of 2 months, we hence estimated the models only on the available data. As the set of elections varies across comparisons, we wouldn't recommend comparing accuracy measures across sets of elections. However, we can compare accuracy measures between lead times within a given set of elections.

For instance, the last two columns of Table A2 compare the lead times of two and five months. For all parties except the SPD the lead time of 2 months is more accurate than the lead time of five months. Overall, the mean RMSE is 2.8 for two months and 3.7 for 5 months. Hence, a lead time of 2 months is more accurate than one of 5 months. For the remaining lead times of one or three months the accuracy measures are similar to the one of two months.

Table A2. Leave-one-out prediction error by lead time (in %). Note that no survey with a lead time of 1 month was available in 1983 and 1990, of 3 months in 2013, and of 5 months in 1987 and 2013. Surveys with a lead time of 2 months were available for all elections. The predicted values are normalized, if necessary, to add up to 100%.

| | All elections | Without | | | | | |
| | | 1983 & 1990 | | 2013 | | 1987 & 2013 | |
Lead time in months	2	2	1	2	3	2	5
CDU/CSU	3.4	3.4	3.7	3.8	5.5	4.3	6.5
SPD	4.5	5.6	3.9	2.3	1.5	2.4	1.9
FDP	1.8	2.4	2.0	1.0	1.7	1.1	1.9
Grüne	1.9	1.1	1.4	1.9	1.8	1.8	2.0
Linke	1.7	1.7	4.0	1.7	1.4	1.7	2.7
Others	6.2	7.3	5.5	5.2	5.2	5.5	6.9
Mean	3.2	3.6	3.4	2.6	2.9	2.8	3.7

Predicted versus actual values

Figure A1 shows the jackknifed predicted versus actual values for each party. The dashed lines indicate perfect predictions, that is when actual and predicted values are the same. Ideally, all predictions fall on that line, but this ideal is, of course, unrealistic. To assess the predictions, we look for any systematic over- or under-predictions ("bias") as well as how tightly the predictions cluster ("variance"). Overall, the plots show little evidence of bias or high variance (with the exception of Others who show the most spread). Both facts are reflected in the RMSE values reported in the main text.
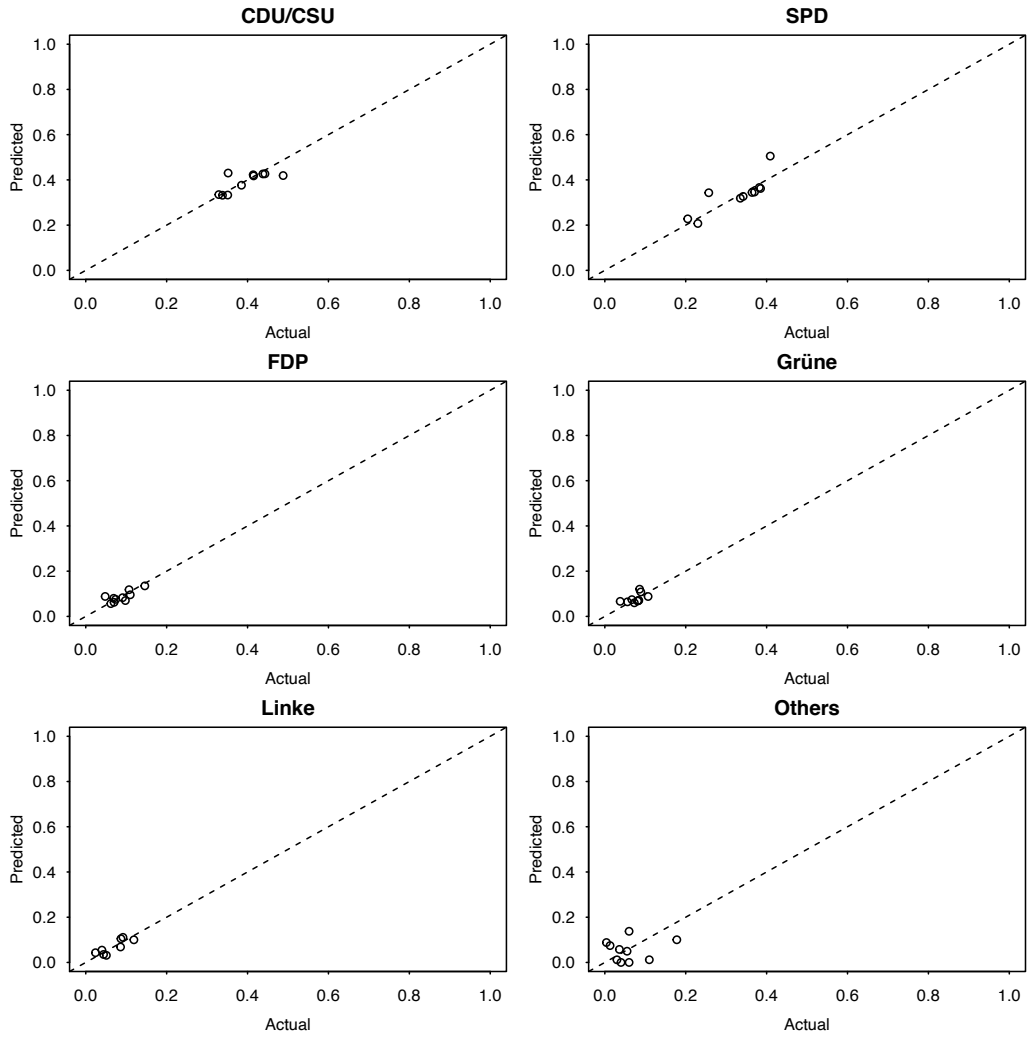
Figure A1. Scatterplot of predicted versus actual values. The predicted values come from the leave-one-out cross-validation and are normalized, if necessary, to add up to 100%.