

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/154353>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Citizen Forecasts of the 2021 German Election

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

Accepted for publication at *PS: Political Science and Politics*.

Andreas E. Murr is Associate Professor in Quantitative Political Science in the Department of Politics and International Studies at the University of Warwick. His research interests are election forecasting, party leadership elections, and political methodology. He may be reached at a.murr@warwick.ac.uk.

Michael S. Lewis-Beck is F. Wendell Miller Distinguished Professor of Political Science at the University of Iowa. He has authored or co-authored over 300 articles and books, including *Economics and Elections*, *The American Voter Revisited*, *French Presidential Elections*, *Forecasting Elections*, *The Austrian Voter* and *Applied Regression*. He may be reached at michael-lewis-beck@uiowa.edu.

 = pull quotes

There are various scientific approaches to election forecasting—poll aggregation, structural models, electronic markets, and citizen forecasting. With respect to the German case, the first two approaches, polls and models, have perhaps been the most popular. However, relatively little work has been done deploying citizen forecasting (CF), the approach we offer here. In principle, it differs considerably from other methods and appears, on its face, quite simple. Before an election, citizens are asked in a national survey who they think will win. As the percentage of expectations, for say party X, increases, the likelihood of an X win is judged to be higher. The method has been regularly applied with success in other established democracies, such as the United Kingdom and the United States.

We wish to extend the application to Germany, utilizing responses to the expectation question in *Politbarometer* surveys since 1980. The forecasting targets are the vote (or seat) shares of the following parties: CDU/CSU, SPD, FDP, Grüne, Die Linke, and Others.

Preliminary analysis of these data appear quite promising, as they generate little prediction error, in- or out-of-sample. This track record allows us a guarded optimism, in terms of the model's ability to accurately forecast the upcoming September 2021 election. Below, we review the literature, then data and measures, before turning to model building and relevant performance tests. Citizen forecasting, a hitherto neglected forecasting strategy for the German case, seems to shed considerable light on the sometimes dimly seen workings of this complex coalitional system.

Literature

Lewis-Beck and Skalaban (1989), who brought forth citizen forecasting as a distinct approach to the prediction of election outcomes, initially applied the method to the national probability samples of the American National Election Study (ANES). Dating from 1956, those surveys regularly asked respondents, in advance of the upcoming presidential election, “Who do you think will be elected President in November?” They found that, across these eight surveys, voters correctly foresaw the winner 69% of the time (Lewis-Beck and Skalaban, 1989, 148). Such a fundamental result, illustrating the potential of voter expectations as a forecasting device, has been replicated and elaborated upon in subsequent American studies (Graefe, 2014; Lewis-Beck and Tien, 1999; Murr, 2015). Further, the approach has spread to Europe, in particular the United Kingdom (Lewis-Beck and Stegmaier, 2011; Murr, 2011, 2016). Indeed, a current paper, drawing on 449 surveys from British elections (1950 to 2017), demonstrates that forecasting models based on vote expectations clearly outperform forecasting models based on the more common method—vote intention polls (Murr, Stegmaier, Lewis-Beck, 2021).

How much attention has been paid to CF in the case of German elections? Some, but not much. Ganser and Riordan (2015) looked at citizen forecasts of vote shares in an *ex post* study, wherein vote intention questions actually performed better. Lehrer, Juhl, and Gschwend (2019) carry out a “wisdom of the crowds” study of the 2017 national election, finding CF does a good job predicting the AfD vote, but again the effort was *ex post*. Graefe (2015) reviews four different forecasting methods applied to the 2013 German election—polls, prediction markets, expert judgement and quantitative models. He finds polls, on average, were more accurate than the other methods. Further, Graefe (2016) did a CF of the 2013 election, asking a non-representative sample of citizens to predict 14 election outcomes (e.g., which candidate will be the next Chancellor, which parties will exceed the 5% threshold, which coalition will form); he reports that they got 12 of the 14 right. However, this was also an *ex post* exercise. Thus, while CF seems to be gathering attention in the world of German election forecasting, the path forward remains open. We offer, for the first time, so far as we know, an *ex ante* German national election forecast, based on 2021 voter expectations. Below we discuss the database for that effort.

Data and Measures

The data we employ come from the established *Politbarometer*, going back to 1980 (Forschungsgruppe Wahlen 2020). Our replication archive is available on Harvard Dataverse (Murr and Lewis-Beck, 2021). The salient sampling characteristics of these surveys are fully discussed in the Online Appendix. For our purposes, the most relevant feature is that they are national probability samples. In forecasting, lead time is a *sine qua non*. That is, the prediction must be made before the contest, hopefully well before, in order that the forecast not be judged

trivial, e.g., a day-before-the-election forecast (Lewis-Beck, 2005). The lead time we focus on is two months. This gives a forecast that stands at a noteworthy distance from the election event. Moreover, it has the added advantage of being available for every election in the series. The median sample size of surveys with lead time of two months is large, at 1,518 (in 1980). (As a robustness check, we also experimented with one, three and five month leads, which had surveys for every election but one or two. See discussion in the Online Appendix).

Ideally, we would have multiple expectation questions, in order to accommodate different objects of the forecast. In addition to “who will lead the government?” we might ask “the vote share you expect for party X?” or “who will be chancellor?” or “will there be a single-party government or a coalition government?” Instead, we have available a generic “who will win?” question. **Certainly “winning” can mean different things to voters.** In an innovative paper, Stiers et al. (2018) explore voter beliefs about the meaning of their party winning. Voters of the party that received the most votes are almost unanimous in perceiving a win; however, even parties that got a smaller number of votes sometimes said their party won, perhaps because they made gains over the last contest (Stiers et al., 2018). In the same way, a German voter for a smaller party might declare a “win” for their party, if they expected it to be part of a ruling coalition. Further, even if a German major party such as the SPD joins the ruling coalition as a junior partner, this may count as the party “losing” the election.

Our question at hand, in addition to asking about winning, is also open-ended: "What do you think personally: who will win the general election?" (own translation). This said, we know from the cited UK and US studies that this question predicts well and can be coded to account for different meanings different people might have. We find the answers generally fall into three clear categories: single-party, candidate, or coalition responses. For example, few voters would

say a small party alone would win, but they might reasonably predict it would be part of a winning coalition. For the major parties CDU/CSU and SPD we counted the single-party mentions; for the minor parties FDP, Grüne, and Die Linke we counted both the single-party and the coalition mentions; (the remaining responses form the reference category.) Hence, we arrive at a forecasting model for a party. (Since AfD is a new party, emerging in 2013, we are unable to fit a regression equation to it alone; however, we include it in a generic vote share equation for Other Parties. We confess that the rise of the AfD could upset our predictions, especially since after 2017 the differentiation between major and minor parties has become less plausible; indeed, according to current polling it looks like there could be only one major party left.)

The Model

In terms of general theory, citizen forecasting holds that as collective voter expectations increase for a party, the party becomes more likely to win (i.e., govern in whole or in part). Of course, expectations are, by nature, adaptive, changing with the political wind. For the German case, a very strong wind concerns whether a “grand coalition” is made, or in the making. In fact, when the incumbent consists of a “grand coalition” such as CDU/CSU and SPD, that greatly changes the meaning of “winning” and hence the nature of expectations. The specification, then, for the prediction equation to be estimated (by ordinary least squares), reads Party Vote Share = f (Vote Expectation, Grand Coalition Status). Such a regression model for each of five parties, calculated from available data for the ten elections (1983 to 2017), appears in table 1. (We exclude the 1980 election as afterwards the FDP switches its electoral strategy away from the SPD and to the CDU/CSU.)

[Place table 1 here]

With the above vote share equation, then, we predict the vote share of each party. In terms of predicting the 2021 election, this means also reporting forecasts of which party will have the largest vote share, as well as which coalition scenarios are likely (e.g., which ones reach +50%). Overall, these regression results in table 1 encourage us in our task. First, examine the goodness-of-fit statistics. The R-squared are good to very good (with a median value of .74.) The within-sample error is low (with a median of .02). The out-of-sample error, based on a jackknife cross-validating procedure, excluding each case in turn, is surprisingly low, with a median RMSE again at .02. Note that the out-of-sample RMSE even slightly improves if we normalize the predictions; that is ensure that the sum of all predicted values, including that of Others, equals 100%. These estimates, especially the out-of-sample error, suggest the forecasting potential of the model. One realization of such potential appears in the prediction of the party of the chancellor. When we assume that the highest predicted party vote share will be the party of the chancellor, in a series of out-of-sample tests for each race, we manage to correctly predict 9 out of the ten chancellors (missing only in 2002).

The regression results themselves say a lot about the functioning of the German party system, and the campaign mechanics that allow for rather accurate forecasting. Note that the intercepts for the CDU/CSU and the SPD are the same (at .31), which suggests they have about the same, rather substantial, reservoir of votes. Also, note the intercepts for the Greens and the FDP are the same (at .06), emphasizing their underdog rivalry. Further, the big parties lose about equally (minus 9 or 10 points) when under a grand coalition, a fact that makes good sense and has value as confirmation of a bit of campaign lore. As well, the gradient of the expectations slope remains roughly equivalent (between .20 and .25) across CDU/CSU, SPD, and FDP. In fact, taking into account the standard errors of the coefficients, there is no evidence to suggest

that the gradient of expectations differs between CDU/CSU, SPD, FDP, and even Grüne. The only exception is the gradient for Linke, though this may be a result of the smaller set of elections.

Speaking broadly, the model's performance appears laudable, especially given its parsimony, with only two independent variables. Its contours can be further appreciated by examination of the scatterplots for vote share regressed on party expectation, in figure 1 below. Observe how closely the points hug the line. Indeed, the linearity of the fits generally seems exemplary, with no acceleration upward in a curve as the expectation of winning passes the 50% mark, a condition indicating the ongoing tight competition, at least between the two main parties.

[Place figure 1 here]

The Citizen Forecast Calculation for September, 2021

As of June 2021, one *Politbarometer* survey with vote expectations is available, enabling us to forecast with a lead time of about three-and-a-half months. We can plug the survey result into the estimated regression equations of table 1 to predict the vote shares and derive predictions for other outcomes, including which party will be the largest one and which coalitions would receive more than 50%. According to the *Politbarometer* newsletter (June 10, 2021, p.1), when 1,232 respondents were asked, "who will win the general election?" their answers distributed themselves as follows: CDU/CSU/Laschet = 64%, SPD/Scholz = 3%, Grüne/Baerbock = 9%, Other responses = 9%, Don't Know = 15%.

Overwhelmingly, respondents expect the CDU/CSU to "win." With the caveat that the reported survey results include candidate mentions and exclude coalition mentions, we translate them into the following vote share predictions:

$$\text{Vote Share (CDU/CSU)} = .31 + .20 * .64 - .09 = .34.$$

That prediction does not cross the 50% threshold needed to govern. What coalition would achieve that necessary majority? According to respondent expectations, SPD is predicted to receive the following vote share:

$$\text{Vote Share (SPD)} = .31 + .22 * .03 - .10 = .21.$$

Adding the SPD share to the CDU/CSU share yields a comfortable majority coalition, i.e., $.34 + .21 = .55$.

But, CDU/CSU may reject the possibility of joining with SPD yet again. If so, according to the currently available expectations data, a coalition with the Grüne would narrowly miss the threshold:

$$\text{Vote Share (Grüne)} = .06 + .60 * .09 + .04 = .15.$$

Adding the Grüne share to the CDU/CSU share yields no majority, i.e., $.34 + .15 = .49$.

What to conclude? Of course, these point estimates are not without error. With respect to coalition formation, the Union/SPD combination seems a far safer bet statistically (e.g., the absolute value of 55% clearly exceeds the alternative, 49%). However, the Union/Grüne combination seems not out of the question, either statistically (recall its small RMSE of .02) or politically (as an emergent strategy). Finally, what about the utility of the CF approach? We see that, in the German case, it enables us to identify clearly the leading party in the electorate, as well as identifying the most likely coalition combinations.

References

Forschungsgruppe Wahlen, Mannheim. 2020. "Politbarometer 1977-2019 (Partielle Kumulation)." GESIS Datenarchiv, Köln. ZA2391 Datenfile Version 12.0.0. <https://doi.org/10.4232/1.13631>

- Graefe Andreas. 2014. “Accuracy of Vote Expectation Surveys in Forecasting Elections.” *Public Opinion Quarterly* 78 (S1): 204–232.
- Graefe, Andreas. 2015. “German Election Forecasting: Comparing and Combining Methods for 2013.” *German Politics* 24 (2): 195–204.
- Graefe, Andreas. 2016. “Forecasting Proportional Representation Elections from Non-representative Expectation Surveys.” *Electoral Studies* 42: 222–288.
- Ganser, Christian, and Patrick Riordan. 2015. “Vote Expectations at the Next Level: Trying to Predict Vote Shares in the 2013 German Federal Election by Polling Expectations” *Electoral Studies* 40: 115–126.
- Lehrer, Roni, Sebastian Juhl, and Thomas Gschwend. 2019. “The Wisdom of Crowds Design for Sensitive Survey Questions” *Electoral Studies* 57: 99–109.
- Lewis-Beck, Michael S. 2005. “Election Forecasting: Principles and Practice” *British Journal of Politics and International Studies* 7:145–164.
- Lewis-Beck, Michael S., and Andrew Skalaban. 1989. “Citizen Forecasting: Can Voters See into the Future?” *British Journal of Political Science* 19 (1): 146–153.
- Lewis-Beck Michael S., and Mary Stegmaier. 2011. “Citizen Forecasting: Can UK Voters See the Future?” *Electoral Studies* 30 (2): 264–268.
- Lewis-Beck, Michael S., and Charles Tien. 1999. “Voters as Forecasters: a Micromodel of Election Prediction.” *International Journal of Forecasting* 15 (2): 175–184.
- Murr, Andreas E. 2011. ““Wisdom of Crowds”? A Decentralised Election Forecasting Model That Uses Citizens’ Local Expectations” *Electoral Studies* 30 (4): 771–783.
- Murr, Andreas E. 2015. “The Wisdom of Crowds: Applying Condorcet’s Jury Theorem to Forecasting US Presidential Elections.” *International Journal of Forecasting* 31 (3): 916–929.
- Murr, Andreas E. 2016. “The Wisdom of Crowds: What Do Citizens Forecast for the 2015 British General Election?” *Electoral Studies* 41 (1): 283–288.
- Murr, Andreas E., and Michael S Lewis-Beck. 2021. “Replication Archive for: Citizen Forecasts of the 2021 German Election” Harvard Dataverse. doi: 10.7910/DVN/WVTI2K
- Murr, Andreas E., Mary Stegmaier, and Michael S. Lewis-Beck. 2021. “Vote Expectations Versus Vote Intentions: Rival Forecasting Strategies.” *British Journal of Political Science* 51 (1): 60–67.
- Stiers, Dieter, Jean-François Daoust, and André Blais. 2018. “What Makes People Believe That Their Party Won the Election?” *Electoral Studies* 55:21–29.

Table 1. Regression Models of Vote Share on Expectations that Party will Win.

	CDU/CSU	SPD	FDP	Grüne	Linke	Others
Intercept	.31 (.05)	.31 (.02)	.06 (.01)	.06 (.01)	.04 (.01)	–
Expectations	.20 (.09)	.22 (.07)	.25 (.10)	.60 (.26)	8.09 (2.10)	–
Grand coalition	–.09 (.03)	–.10 (.03)	.05 (.01)	.04 (.01)	.04 (.01)	–
Number of observations	10	10	10	10	8	–
R-squared (in-sample)	.64	.86	.74	.64	.88	–
RMSE (in sample)	.04	.03	.02	.01	.01	–

RMSE (out-of-sample)	.04	.05	.02	.02	.02	–
RMSE (out-of-sample, normalised)	.03	.04	.02	.02	.02	.06
Chancellors correctly predicted (out-of-sample)	9 out of 10 (2002 is wrongly predicted)					

Note: To normalize predictions, we divided by the sum of predicted values across parties if their sum was larger than 1 (as in 1998 and 2005). The prediction for Others equals 1 minus the sum of the (normalized) predictions for the remaining parties.

Figure 1. Scatterplot of vote share on expectations. Grey dots and lines indicate elections preceded by grand coalitions; black dots and lines indicate elections without a preceding grand coalition. (As a diagnostic tool, we also examine jackknifed predicted versus actual values. See the discussion in the Appendix.)